J. Mol. Biol. (1993) **231**, 375–391

Three-dimensional Structure of Satellite Tobacco Mosaic Virus at 2·9 Å Resolution

Steven B. Larson, Stanley Koszelak, John Day, Aaron Greenwood, J. Allan Dodds and Alexander McPherson

The Department of Biochemistry &
The Department of Plant Pathology
University of California, Riverside, Riverside, CA 92521, U.S.A.

(Received 1 October 1992; accepted 15 January 1993)

The crystal structure of satellite tobacco mosaic virus (STMV) has been solved by a combination of multiple isomorphous replacement and molecular replacement methods and refined at 2·9 Å resolution to a conventional R-factor of 0·215. STMV, a $T = 1$ icosahedral virus, is the smallest whose structure has been determined. The coat protein is an eight-stranded "Swiss roll" β-barrel with an amino-terminal strand that extends away from the β-barrel by more than 60 Å. This strand is primarily responsible for quaternary interactions within the capsid.

The most arresting feature of the virus structure is the intimate association of each capsid protein dimer with a Watson–Crick base-paired segment of RNA double helix on the interior of the virion. The icosahedral 2-fold axis of each dimer pair is coincident with that of the central base-pair of each helical RNA segment whose helical axis is along the edge of the icosahedron. The helical RNA segments are seven base-pairs in length with a stacked base at each 3′ end so that a total of 16 nucleotides is clearly visible. The character of the RNA helix is somewhat different than any of the canonical forms. Assuming full occupancy, then approximately 45% of the total RNA genome is present in the electron density map. The close association of capsid with highly structured nucleic acid suggests that assembly of STMV is likely to be a highly co-operative process involving both protein and RNA. The nucleic acid is distributed within the virion with a high degree of order. The capsid protein is a true double helical RNA binding protein and a number of prominent interactions between protein and RNA can be clearly seen.

Keywords: virus; RNA; double-helical; X-ray; crystallography

1. Introduction

Satellite tobacco mosaic virus (STMV†) is the naturally occurring satellite to the obligatory helper tobacco mosaic virus (TMV) without whose co-infection it cannot replicate. Although TMV is a rod shaped virus, STMV is a small spherical virus. This is the only case currently known of such a relationship. TMV has been known and studied for over 60 years (for reviews, see Stubbs, 1984; Holmes, 1984). STMV was only discovered in 1985 (Valverde & Dodds, 1986, 1987; Valverde *et al.*, 1991).

STMV is a $T = 1$ icosahedral virus (Caspar & Klug, 1962) whose 60 identical capsid proteins of $M_r = 17{,}500$ encapsidate a single-stranded RNA genome having 1059 bases of known sequence (Mirkov *et al.*, 1989). The RNA has two open reading frames that code for the coat protein plus one other small protein of $M_r = 6700$, whose function is not known. The RNA exhibits a high degree of sequence similarity with that of the helper TMV in the 3′ terminal 150 bases which, in TMV, form the recognition site for the TMV replicase enzyme (Joshi & Haenni, 1984), which the two viruses presumably share. There is otherwise no homology between STMV and TMV in terms of genome or coat-protein amino acid sequence, nor is there any immunological cross reactivity between STMV and TMV (Valverde & Dodds, 1987; Mirkov *et al.*, 1989; Dodds, 1991).

STMV has, from electron microscopy and X-ray diffraction measurements, a diameter of about 16 nm (Kim *et al.*, 1989; Koszelak *et al.*, 1989). It appears to be similar to another satellite virus, satellite tobacco necrosis virus (STNV: Liljas *et al.*,

† Abbreviations used: STMV, satellite tobacco mosaic virus; TMV, helper tobacco mosaic virus; STNV, satellite tobacco necrosis virus; TNV, tobacco necrosis virus; m.i.r., multiple isomorphous replacement; r.m.s., root-mean-square.

0022-2836/93/100375–17 $08.00/0

1982; Jones & Liljas, 1984), which has as its helper another spherical virus, tobacco necrosis virus (TNV). STMV is, however, about 20% smaller in all regards. The coat protein of STMV has an M_r of 17,500 as opposed to 21,000 for STNV, and the genome of STMV is 1059 bases *versus* 1239 bases for STNV. There is no amino acid sequence homology between the coat proteins of STMV and STNV, nor is there any immunological cross reactivity (Mirkov *et al.*, 1989; Dodds, 1991).

Because of the biological similarities between STMV and STNV, and in spite of the sequence and size differences, our initial assumption was that the structure of STMV would be similar to that of STNV. The structure of STNV was known from X-ray diffraction studies carried out some years ago (Liljas *et al.*, 1982; Jones & Liljas, 1984). As we will describe below, there are indeed similarities, but these are not so remarkable as the differences. Some of these results have been presented earlier (Larson *et al.*, 1993).

2. The Crystals and Their Diffraction Patterns

Crystals of STMV were observed by electron microscopy to form inside infected plant cells (Kim *et al.*, 1989), but the first crystals suitable for X-ray diffraction analysis were produced about five years ago (Koszelak *et al.*, 1989). They can be grown from ammonium sulfate solutions by vapor diffusion to sizes of about 0·7 mm in their longest dimension. The crystals have orthorhombic symmetry and a body-centered lattice with $a = 174{\cdot}3$ Å (1 Å = 0·1 nm), $b = 191{\cdot}8$ Å and $c = 202{\cdot}5$ Å. There are two virus particles in the unit cell. This requires that the virus be centered exactly on 222 symmetry points and hence, the space group is $I222$. Thus, one quarter of the virus, or 15 copies of the icosahedrally related coat proteins, comprises the asymmetric unit of the crystals.

The unit cell properties of the crystal presented two important opportunities with regard to structure determination. First, the 15-fold redundancy present within the asymmetric unit could be utilized as a tool in both phase determination and refinement. Second, the unit cell for STMV is small as virus structures go, and the X-ray diffraction data from the crystals could be collected accurately and in a straightforward manner on a multiwire area detector system with a rotating anode X-ray source.

3. Data Collection, Methods and Structure Solution

Crystals were mounted in quartz capillaries by conventional methods and all data were collected at 17 °C. Generally at least five days of data acquisition ensued before any perceptible deterioration of the diffraction pattern occurred. Data were collected using a San Diego Multiwire Systems two detector system (Xuong *et al.*, 1985) with 930 and 985 mm helium-filled crystal-to-detector paths. The X-ray source was a Rigaku RU-200 rotating anode generator fitted with a Supper graphite monochromator to produce CuKα radiation. The generator was operated at 45 kV and 175 mA. The crystal was rotated through 45 degree sectors about the omega axis in increments of 0·10 to 0·12 degree, at appropriate chi and phi settings. Each frame of data was measured for 120 to 200 seconds and the frames from all crystals for native, or for a particular heavy atom derivative, were merged as though they originated from the same crystal.

The structure solution and refinement to 2·9 Å resolution presented here was based on a data set that contained 69,351 reflections and was 92% complete to 2·9 Å. The merging R was 0·084 for the approximately one million observations. Data for heavy atom derivative crystals of STMV were collected to 3·5 Å resolution for ten different isomorphous candidates, of which eight proved useful.

There are two possible orientations of the virus particle in the unit cell consistent with 222 symmetry. The disposition of the icosahedral axes was immediately evident from a self-rotation function (Rossmann & Blow, 1962) calculated with the MERLOT program package (Fitzgerald, 1988) using 5 to 12 Å data. Hence, the correct orientation for the virus particle was unequivocally established.

Initially, the STNV structure (Liljas *et al.*, 1982; Jones & Liljas, 1984), whose co-ordinates were retrieved from the Brookhaven Data Bank (Bernstein *et al.*, 1977) was placed at the origin in the proper orientation followed by systematic rigid body shifting and refinement of the coat protein units using the program TNT (Tronrud *et al.*, 1987). This produced no significant decrease in the overall R-factor, which persisted around 0·48, and no improvement in the correlation coefficient to better than about 0·15. Variations on this approach were no more encouraging and we suspected at this time that STNV was simply an inadequate model for STMV.

From the rigid body refined STNV model, however, trial structure factors were calculated in the 6 to 15 Å resolution range. The phases were averaged through 20 cycles based on a protocol that utilized the programs of Bricogne (1974, 1976) for real space averaging in conjunction with TNT for map calculations. Subsequently, phases were extended to 2·95 Å resolution with the map for each cycle calculated with observed structure amplitudes scaled to the calculated amplitudes of the previous cycle. Scaling was done in shells of equal numbers of reflections (approx. 385 per shell) after sorting by resolution. An averaged map, calculated at 3·0 Å resolution, although different in features, showed no significant increase in resolution from a 5 Å map.

Using only the phases in the 5 to 15 Å range from this phase set, difference Fourier maps of the asymmetric unit were calculated for each of the potential heavy atom derivatives. The highest 300 peaks for each compound were then systematically compared to reveal icosahedrally related sites. Ideally, because the asymmetric unit contains 15 copies of the coat

Table 1

Heavy atom derivative data (3·5 Å to 15·0 Å) and refinement statistics

Derivative	Number of reflections	Percent complete	R_{sym}	R (nat-der)	R_c	Centric r.m.s.(F)/r.m.s.(ε) 3·5–15 Å	Centric r.m.s.(F)/r.m.s.(ε) 5·0–15 Å	Acentric r.m.s.(F)/r.m.s.(ε) 3·5–15 Å	Acentric r.m.s.(F)/r.m.s.(ε) 5·0–15 Å	X	Y	Z	B	Relative occupancy
K_2PtCl_4†	34,318	85	0·051	0·169	0·53	1·41	1·69	1·79	2·43	0·0502	0·2842	0·1479	55·3	1·93
										0·0334	0·2690	0·1549	48·8	1·08
										0·2486	0·0798	0·2109	46·9	1·42
										0·2474	0·0801	0·2198	21·1	0·78
K_2AuCl_4	33,082	82	0·063	0·145	0·50	1·43	1·93	1·50	2·20	0·0522	0·2873	0·1518	34·4	1·48
										0·0618	0·2863	0·1403	20·6	0·98
K_2PtCl_4‡	31,905	79	0·074	0·199	0·61	1·09	1·28	1·46	1·95	0·0507	0·2847	0·1479	79·8	2·69
										0·0323	0·2690	0·1558	60·7	1·33
										0·2486	0·0802	0·2128	51·3	1·95
										0·2507	0·0817	0·2269	30·2	0·83
K_2AuCl_4§	26,938	67	0·034	0·124	0·73	0·59	1·13	0·58	1·19	0·0570	0·2886	0·1467	121·8	1·52
o-Chloromercuriphenol	32,675	81	0·061	0·089	0·67	0·69	1·07	0·67	1·16	0·0550	0·2875	0·1466	48·1	0·74
$Hg(C_2H_4O_2)_2$	33,901	84	0·063	0·096	0·65	0·79	1·03	0·89	1·49	0·0525	0·2859	0·1496	25·7	0·55
										0·0402	0·2950	0·1300	18·3	0·46
$PtBr_2(NH_3)_2$	33,536	83	0·064	0·106	0·64	0·79	1·26	0·83	1·47	0·0506	0·2844	0·1501	42·7	0·87
										0·2455	0·0790	0·2119	118·3	0·86
$AgNO_3$	34,107	85	0·045	0·067	0·70	0·55	1·06	0·55	1·17	0·0537	0·2847	0·1511	66·0	0·53

Figure of merit: 0·60 at 5·0 Å resolution: 0·51 at 3·5 Å resolution.
Average phase difference from oriented STNV to rigid-body refined STNV: $\langle\Delta\phi\rangle = 83·4°$ (5 to 15 Å resolution).
Average phase difference from rigid body refined STNV to refined m.i.r. phases: $\langle\Delta\phi\rangle = 92·0°$ (5 to 15 Å resolution).
Average phase difference from refined m.i.r. phases to phase extended values to 2·95 Å: $\langle\Delta\phi\rangle = 48·2°$ (3·5 to 15 Å resolution).
† 4 day exposure to heavy atom compound.
‡ 2 week exposure to heavy atom compound.
§ 8 month exposure to heavy atom compound.

protein, one might have expected to find 15 strong peaks for a good single site substitution.

Difference Fourier syntheses of five of the seven non-platinum containing derivatives yielded a single outstanding set of 15 icosahedrally related peaks that were virtually identical between derivatives. Difference Fourier maps of the platinum derivatives showed two sets of 15 icosahedrally related peaks. These sets were identical among the platinum derivatives and one set of 15 corresponded to the site in the non-platinum derivatives. Thus, of eight substituted derivatives, there were only two substitution sites, both about 62 Å from the origin.

The major heavy atom substitution sites for all eight derivatives were refined against the origin removed Patterson as implemented in the program HEAVY (Terwilliger & Eisenberg, 1984). Difference Fourier syntheses based on progressively improved phases extended to 3·5 Å resolution yielded minor substitution sites that were then included in the heavy atom refinement process. The overall figure of merit for the m.i.r. phases was 0·51. Heavy atom refinement statistics are found in Table 1.

An m.i.r. map of the asymmetric unit of the crystal was calculated at 3·5 Å resolution. It immediately confirmed the correctness of the heavy atom derivatives. The region of the map within a radius of about 55 Å of the origin contained weak and basically disorganized density, while a strong band of electron density appeared between 55 Å and 90 Å.

The m.i.r. map was icosahedrally averaged, again using the programs of Bricogne (1974, 1976), yielding a map of outstanding clarity. The envelope of the coat protein was well defined and the course of the entire polypeptide could be traced by inspection. It was clear at this point, however, that the twist of the β-sheets of the β-barrel was opposite to that observed in other viruses and other proteins (Branden & Tooze, 1991). Therefore, the $-z$ coordinates of the heavy atoms were used to determine initial phases for phase extension (Arnold *et al.*, 1987; Arnold & Rossmann, 1988) to 2·95 Å resolution. The envelope was a 50 to 91 Å shell with an icosahedrally shaped exterior boundary. Rayment weights (Rayment, 1983) were applied to the observed structure amplitudes and F_{calc} values were used to fill in missing data. Observed structure amplitudes were not scaled to calculated amplitudes because high resolution shells showed a greater increase in the correlation coefficient when $|F_{obs}|$ were not scaled. After phase extension to 2·95 Å resolution (statistics shown in Table 2), an averaged electron density map was computed.

Residues 17 to 159 were easily identified on a minimap by virtue of their side-chain density. A model of the capsid protein and its associated RNA was built into the electron density using the program FRODO (Jones, 1985). All subsequent Fourier, difference Fourier, and refinement was based on phases calculated from successively improved models. Maps were, in general, icosahedrally averaged, and icosahedral symmetry was maintained throughout refinement as well. At this point crystallographic refinement to 2·9 Å resolution has been completed using conjugate gradient procedures in the program XPLOR running on the

Table 2

Phase refinement statistics for 2·95–15·0 Å resolution data

Max res. (Å)	Corr. coeff.	R	Scale factor	Number of reflections
7·90	0·872	0·095	1·69	3137
6·39	0·828	0·094	1·73	3123
5·61	0·818	0·089	1·74	3123
5·12	0·848	0·090	1·72	3123
4·77	0·855	0·085	1·78	3123
4·49	0·862	0·074	1·79	3123
4·27	0·857	0·090	1·83	3123
4·08	0·829	0·103	1·87	3123
3·92	0·832	0·100	1·89	3123
3·78	0·821	0·101	1·91	3123
3·66	0·782	0·114	1·94	3123
3·55	0·780	0·112	1·96	3123
3·45	0·780	0·123	1·99	3123
3·36	0·772	0·136	2·03	3123
3·28	0·750	0·143	2·06	3123
3·21	0·722	0·167	2·09	3123
3·13	0·669	0·193	2·17	3123
3·07	0·618	0·215	2·22	3123
3·01	0·581	0·226	2·32	3123
2·95	0·510	0·292	2·95	3123
2·95–15	0·953	0·152	—	62,474

$$\text{Corr. coeff.} = \frac{\Sigma(F_o - \langle F_o\rangle)(F_c - \langle F_c\rangle)}{[\Sigma(F_o - \langle F_o\rangle)^2 \Sigma(F_c - \langle F_c\rangle)^2]^{1/2}}$$

$R = (\Sigma||F_o| - |F_c||)/\Sigma|F_o|$.

Table 3

Refinement statistics for current STMV model

Resolution range	Number of reflections	R-factor shell	R-factor sphere
4·78–6·00	7895	0·233	0·233
4·18–4·78	7777	0·177	0·203
3·80–4·18	7534	0·194	0·200
3·53–3·80	7386	0·195	0·199
3·32–3·53	7204	0·216	0·202
3·15–3·32	6875	0·226	0·205
3·02–3·15	6643	0·250	0·210
2·90–3·02	6550	0·262	0·215

	Geometry†	
r.m.s. deviations for:	Protein	RNA
Bond lengths (Å)	0·013	0·010
Bond angles (deg.)	3·26	3·58
Dihedral angles (deg.)	15·2	37·3
Improper dihedral angles (deg.)	0·8	0·6
ω (from 180°) (deg.)	1·3	—

Resolution: 2·9 Å to 6·0 Å.

Number of reflections: ($F > 5\sigma$) 57,864.

XPLOR weight W_A: 80,213.

Phase angle change (3·5 Å to 6·0 Å) for m.i.r. to model: $\langle\Delta\phi\rangle = 55·0°$.

Phase angle change (3·0 Å to 6·0 Å) for extended phases to model: $\langle\Delta\phi\rangle = 26·6°$.

† Protein geometry based on bond and angle parameters found in Engh & Huber (1991) and RNA geometry defined in XPLOR parameter file parnah1e.dna.

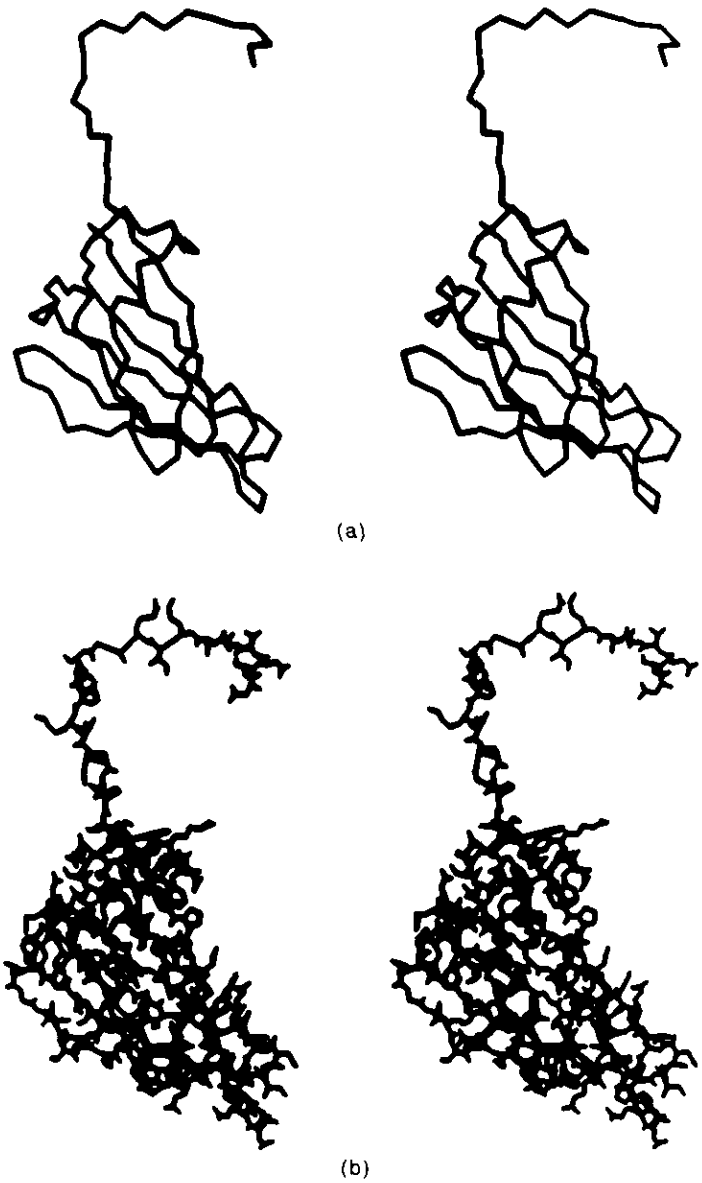


Figure 1. The C^{α} backbone of the STMV capsid protein from amino acid residues 16 through 159 is shown in (a). (b) All non-hydrogen atoms of the corresponding residues are shown. Amino acid residues 1 to 15 were not visible in the electron density map. The capsid protein is composed of an 8-stranded Swiss roll β-barrel and an extended amino-terminal strand.

Cray Y-MP system at the San Diego Supercomputer facility (Brunger *et al.*, 1987; Brunger, 1991; Engh & Huber, 1991). The current refinement results are shown in Table 3.

4. Results

The electron density map obtained by 15-fold averaging and phase extension of the m.i.r. phased Fourier about the non-crystallographic icosahedral symmetry elements was virtually free of noise, showed high contrast and excellent connectivity. Antiparallel β structure was clearly evident and distinctive side-chains could, in general, be readily identified. The envelope of the capsid protein was clear and there were no ambiguities in that regard. Heavy atom derivative site 1 was found 1·73 Å from Cys157 and 2·24 Å from Met76 near the carboxyl terminus and, somewhat unexpectedly, on the interior surface of the capsid. Heavy atom site 2 was found associated with Met22, also on the interior of the protein shell. The interior of the virus, with the exceptions discussed below, was characterized by a substantially lower electron density than that of the protein shell.

The course of the capsid polypeptide chain was traced without interruption from the carboxy terminus at amino acid residue 159 back to amino acid residue 17. Residue 16 was later placed from a difference Fourier map. The polypeptide segment from 1 to 15 at the amino terminus, which contains five lysine and arginine residues, could not be unambiguously identified and is apparently associated with nucleic acid on the interior of the capsid.

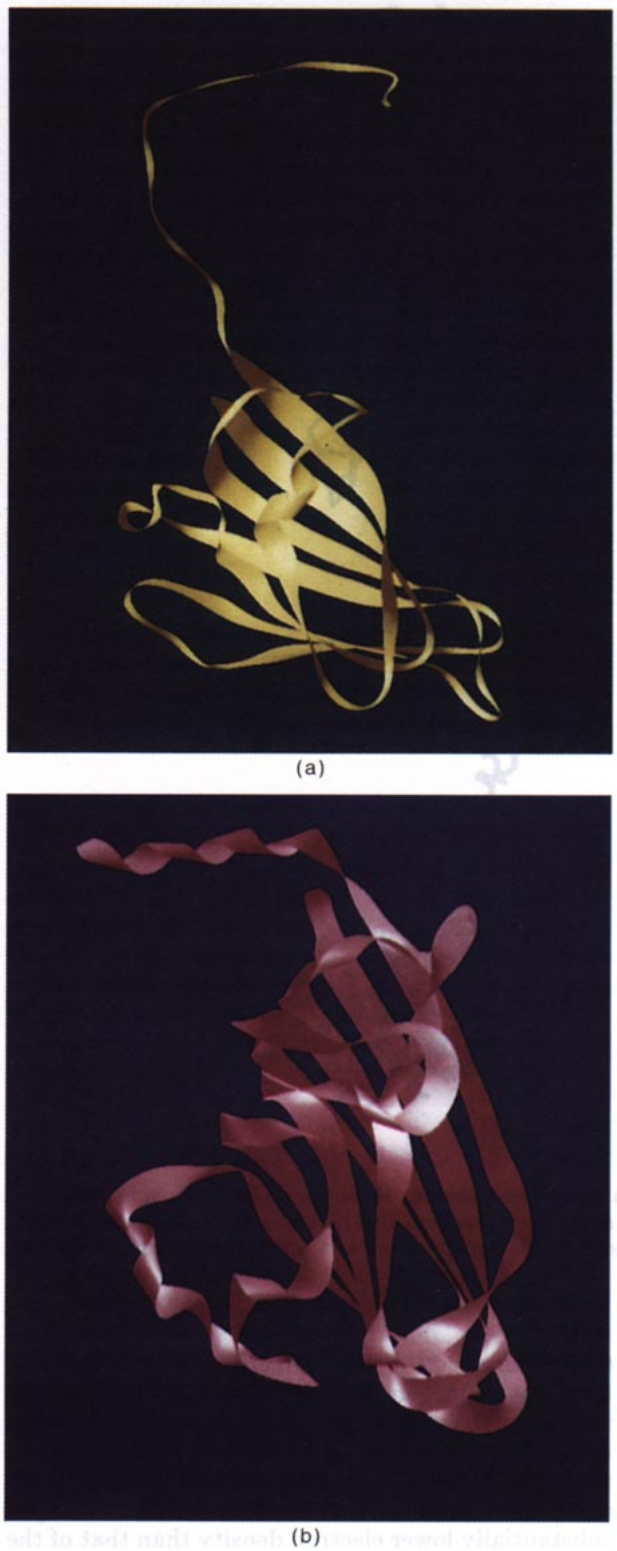


Figure 2. (a) The polypeptide chain of the capsid protein of STMV is presented. (b) The capsid protein from STNV is shown in a corresponding orientation based on approximate superposition of the respective β-barrels. The scale of the 2 coat proteins is the same.

There, it presumably assumes an icosahedrally inconsistent or otherwise disordered conformation. A similar lack of structure in the amino-terminal portion of the polypeptide has been observed for other virus structures as well (Rossmann & Johnson, 1989; Rossmann & Erickson, 1985; Harrison, 1991; Liljas, 1986).

The alpha carbon trace of the capsid protein, excepting amino acid residues 1 to 15, is shown in Figure 1(a) and the entire structure, main and side-chain atoms, is seen in Figure 1(b). Amino acid residues 37 to 159 form a compact, eight-stranded, antiparallel β-barrel having the frequently observed Swiss roll topology. The β-barrel portion of the molecule is similar to the carboxy-terminal β-barrel that is found in the capsid protein of STNV. The capsid proteins of STMV and STNV are shown for comparison in similar orientations in Figure 2. In terms of this β-barrel, the capsid proteins of STMV and STNV are homologous in a structural sense in spite of the absence of any discernible amino acid sequence homology. Swiss roll β-barrels similar to those seen in these two satellite viruses have also been found in the coat proteins of other more complex viruses whose structures have been solved by X-ray crystallography (Rossmann & Johnson, 1989; Rossmann & Erickson, 1985; Liljas, 1986; Harrison, 1991).

Aside from the ubiquitous β-barrel, the capsid protein contains only one other secondary structural element, which encompasses amino acid residues 1 through 36. This portion of the protein consists of a long polypeptide chain that extends well away from the core of the monomer. Given that residues 1 to 15 are not visible and are not shown in Figures 1 and 2, the reach of this extended chain is exceptional. The contour distance of the polypeptide between residues 37 and 15 is about 60 Å.

In the intact virus, as can be seen in Figure 3(a), a considerable length of the amino-terminal polypeptide is closely associated, and is hydrogen bonded through main-chain atoms, with a strand forming the far edge of the β-barrel of a protein subunit related to it by a viral 2-fold axis of symmetry. Hydrogen bonding and other tertiary interactions between dyad-related subunits are so extensive that one would almost certainly conclude that it is dimers of coat protein that must undergo assembly to form the viral capsid. There is abundant evidence that dimers are the aggregating species for numerous other viruses as well. The remaining, visible part of the amino-terminal polypeptide is, in addition, the basis for the majority of interactions with a 3-fold related protein subunit as seen in Figure 3(b). These contacts are responsible for the formation of the trimer facet of the icosahedral shell. Again, the mode of interaction is through hydrogen bonding and extension of an antiparallel β-sheet. In Figure 3(c) are five capsid proteins organized around an icosahedral pentamer axis. The contacts here are the least involved. The structure of the amino-terminal portion of the capsid protein, the interactions it makes with adjacent subunits and its contribution to the formation of the viral capsid are significantly different in STMV than was seen in STNV.

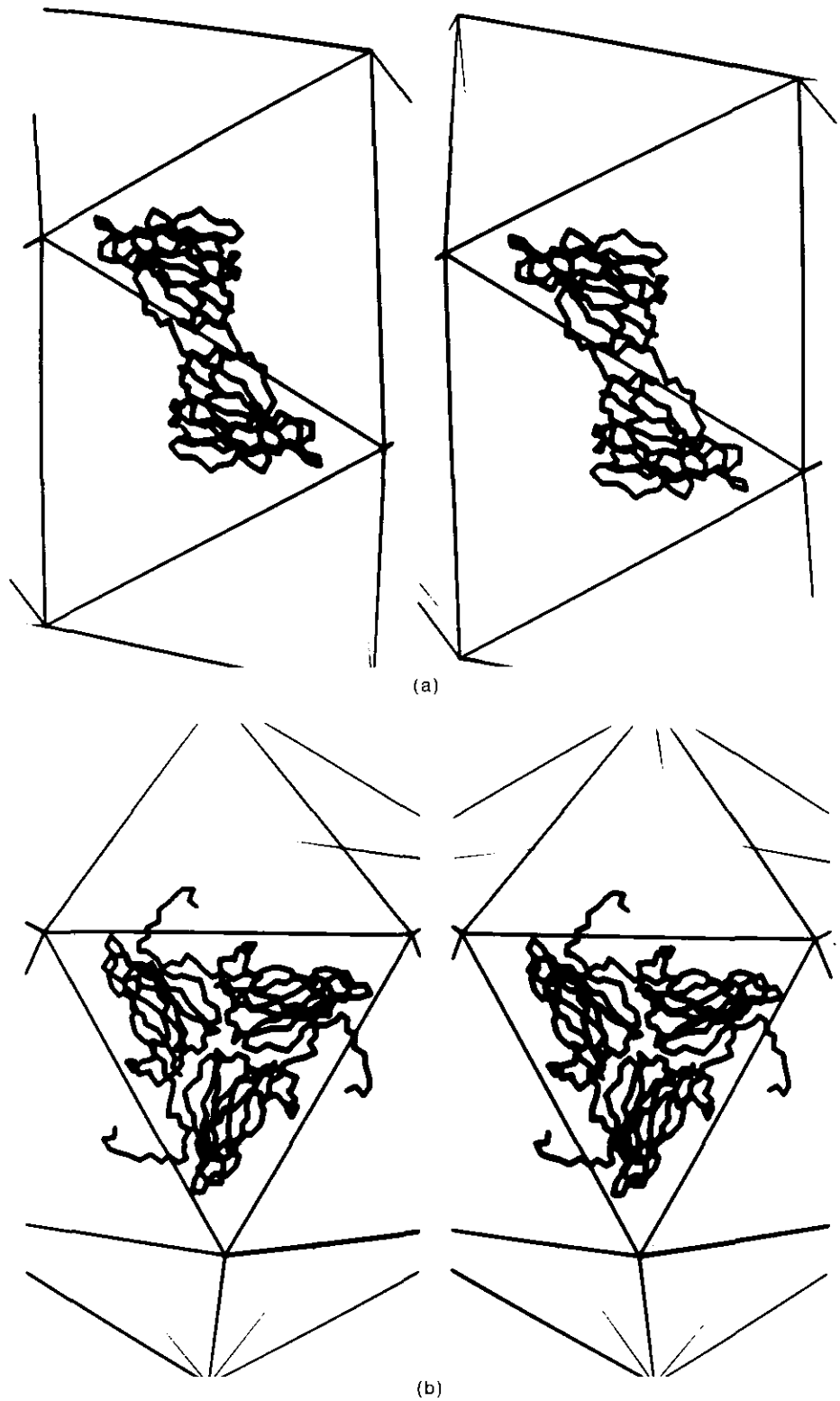
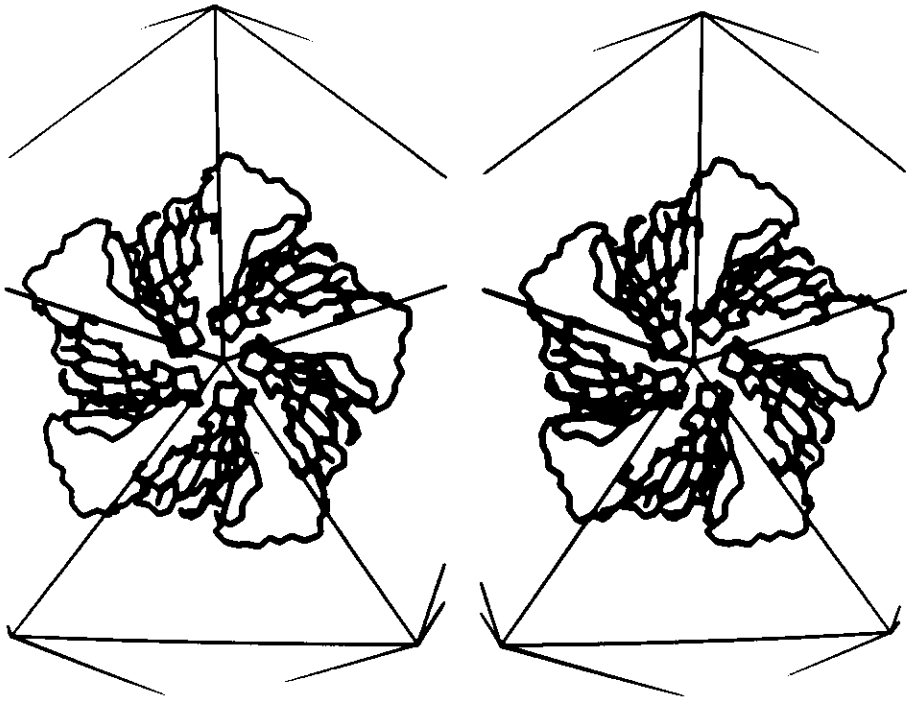


Fig. 3.

A very strong peak appeared in the electron density map exactly on the 5-fold axis of each pentamer at a radius of 73 Å. Two asparagine side-chains extend from each 5-fold related monomer and straddle this peak so that ten ligands are provided. In addition, this peak occludes the only obvious channel to the interior of the virus. Our initial assumption was that these peaks represented calcium ions as was found in a similar location in STNV (Liljas *et al.*, 1982; Jones & Liljas, 1984), and in several other viruses (Rossmann & Johnson, 1989; Liljas, 1986; Harrison, 1991). Other consider-



(c)

Figure 3. (a) The dimer interaction of the STMV capsid proteins is seen as it appears on the surface of the virus. Only the C^{α} backbone is shown. (b) The trimer aggregate, and (c) the organization of the proteins around an icosahedral 5-fold axis of STMV are shown.

ations, however, suggest that this is not the case. While asparagine residues might be suitable ligands for Ca^{2+}, we note that the epsilon oxygen atom of Asn115 is hydrogen bonded to a main-chain amino group of its own subunit. It is the NH_2 group that most closely approaches the axial peak. The distance from the center of the peak on the 5-fold axis to ND2 of asparagine 115 is about 4 Å and ND2 of asparagine 117 to the axial peak is 4·3 Å. These distances would be quite improbable if the peak were Ca^{2+}. In addition, the virus is quite stable in the presence of EDTA and/or EGTA and crystals can be grown equally well, and of the same form, whether there is an excess of Ca^{2+} or a large excess of both EDTA and EGTA. The crystals can also be readily grown with 25% saturated citrate as the precipitant. Furthermore, we see no evidence of the binding of metal ions elsewhere in the structure in contrast to what was observed for STNV (Liljas *et al.*, 1982; Jones & Liljas, 1984). We believe that the strong electron density peak we observe on the 5-fold axis is, in fact, an anion. Because of the crystal growth conditions, we suspect that it is probably a sulfate ion.

The entire viral capsid of STMV is shown in two different orientations in Figure 4(a) and (b) along with the capsid of STNV in Figure 4(c) and (d) for comparison. When viewed in their entirety, the complexity of structure obscures detailed differences. Nevertheless, it is clear that the capsids are quite different. The diameter of STMV is somewhat smaller than that for STNV, reflecting its 20% smaller capsid protein and its smaller genome. At the 3-fold axes, STMV and STNV have radii of 80 Å and 86 Å, respectively while at the 5-fold axes they have radii of 86 Å and 96 Å. The prominent protrusions, or cusps, at the 5-fold axes of STNV are considerably reduced in STMV.

The most compelling aspect of the virus structure was what could be clearly observed inside the virion. Once electron density belonging to the capsid protein was assigned, and the amino acids (with the exception of 1 to 15) accounted for, a substantial region of well-defined electron density remained. This density, seen in Figure 5, was centered on an icosahedral dyad between radii of 52 Å and 64 Å and took the form of two antiparallel arcs bridged by bands of density. Each arc consisted of a sequential array of strong density peaks separated by a more or less constant distance of 5·8 Å with prominent but weaker areas connecting them.

The electron density pattern was unmistakably that of Watson–Crick base-paired, double helical nucleic acid. The appearance of double helical nucleic acid within the viral capsid was unanticipated because the electron density map had been icosahedrally averaged. The expectation was that only features consistent with symmetry of the virus capsid should be seen. The clear implication was that the double helical nucleic acid segments were indeed distributed with icosahedral symmetry (Larson *et al.*, 1993).

A model of the RNA composed of guanosine and cytosine nucleotides was constructed into the density using computer graphics. The result is seen in Figure 6. The electron density exactly accommodated two antiparallel strands, each composed of

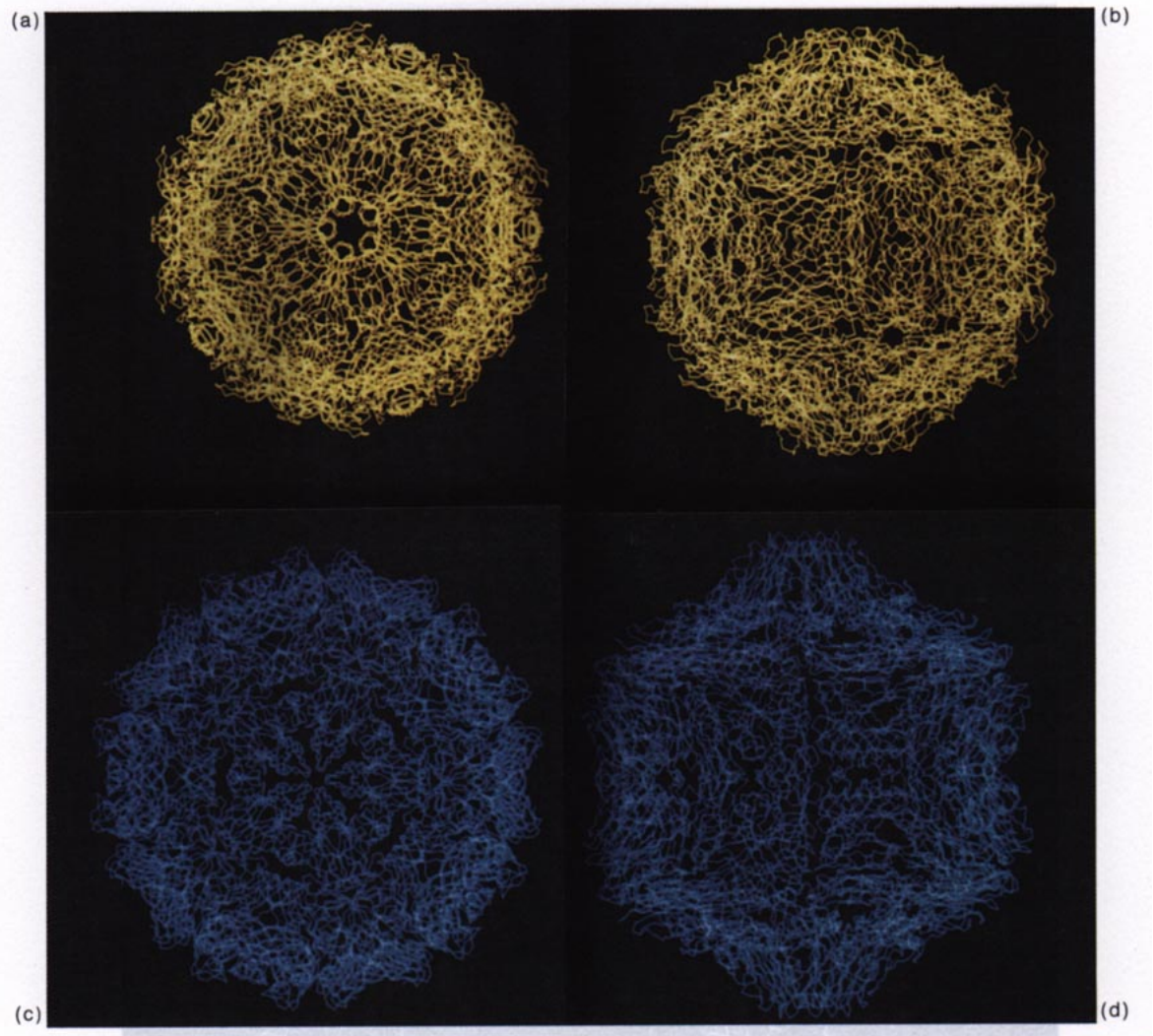


Figure 4. The STMV capsid is shown (a) viewed along an icosahedral 5-fold axis and (b) along a 2-fold axis. The STNV capsid is shown (c) and (d) in corresponding orientations. The scale for both viruses is the same. Apparent in these comparisons is the smaller more spherical shape of the STMV and a significant reduction of the cusps at the 5-fold vertices that are prominent features of STNV.

eight nucleotides. Seven base-pairs were clearly evident with one final, apparently unpaired but stacked, nucleotide at the 3′ ends of each double helical segment. The segments of RNA lay precisely on viral icosahedral 2-fold axes of symmetry, that is, the 2-fold axes of the nucleic acid helical segments and the viral icosahedral 2-fold axes were coincident. The dyads lay precisely in the planes of the central base-pairs, perpendicular to the helical axes of the RNA segments, thus relating the two base-paired polynucleotide strands that make up the double helices.

Because the electron density was icosahedrally averaged, it might be concluded that nucleic acid segments were positioned on each of the icosahedral dyad axes of the virus, and that, therefore, 30 such segments were distributed with icosahedral symmetry about the interior of the virion. Some caution must be exercised, however, in quantitatively applying that criterion. Because of the symmetry averaging of the electron density map, the double helical segments would appear even if only a portion of the icosahedrally equivalent sites contained the nucleic acid. An averaged map would appear to have RNA density elements at all sites, representing the statistical average of the helices actually present with otherwise presumably random density. The resultant electron density, however, would appear with a magnitude in direct proportion to the effective occupancy. Estimates of occupancy based on relative density is imprecise but our evaluation of the icosahedrally averaged electron density map is that there is little, if any, difference in density strength between capsid protein and nucleic acid. Thus, we would say that if the 30 icosahedrally equivalent sites are not all occupied, then certainly at least a very high percentage of them are filled.

A second consequence of the icosahedral averaging is that base-pairs observed in the helical

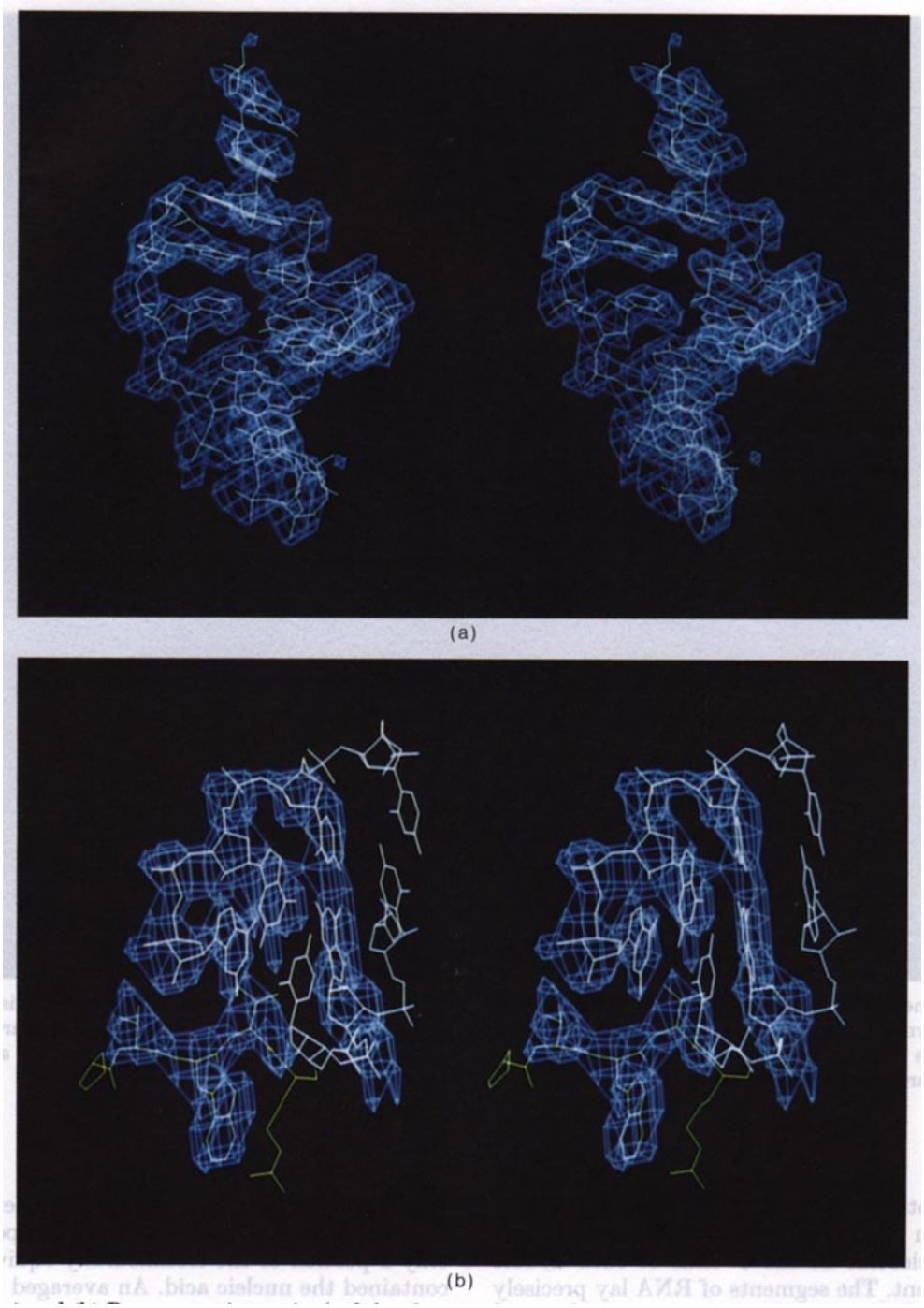


Figure 5. (a) and (b) Representative regions of the electron density from the icosahedrally averaged, phase extended m.i.r. Fourier map that arise from nucleic acid on the interior of the capsid. (a) The density belonging to the 7 base-pairs to which the RNA model was fit is shown. (b) Density belonging to both capsid protein and double helical RNA is shown. Some density has been omitted in (b) to provide clarity. The refined model for both protein and RNA is seen superimposed on the density.

segments of RNA arise from density resulting from the superposition of all possible base-pairs. In addition, the RNA segments must have the form and geometry of the average of all helical segments in the virus. While all helices may be virtually identical, except for the order of base-pairs, they could also vary significantly from the mean structure. Again, however, this would produce an effective reduction in occupancy or an apparent dispersion of the electron density.

The 1059 bases of the viral RNA were analyzed for their tendency to form secondary structure arising from Watson–Crick base-pairing. This was done using the program FOLD (Devereux *et al.*, 1984) employing the algorithm of Zucker & Stiegler (1981). The result is shown in Figure 7. The single strand of 1059 RNA bases displays an exceptional capacity to base-pair and to form extensive secondary structure consistent with double helical RNA. From the 65% of the bases predicted to be

Figure 6. The 16 nucleotide, Watson–Crick base-paired, double helical RNA segment found associated with capsid protein dimers. The segments, modeled to the icosahedrally averaged phase extended m.i.r. electron density map and refined using XPLOR, consist of 7 consecutive base-pairs with a stacked base at the 3′ ends of each helical strand. The axis of each helical segment is along an edge of the icosahedron while the dyad axis of the central base-pair is coincident with an icosahedral 2-fold axis.

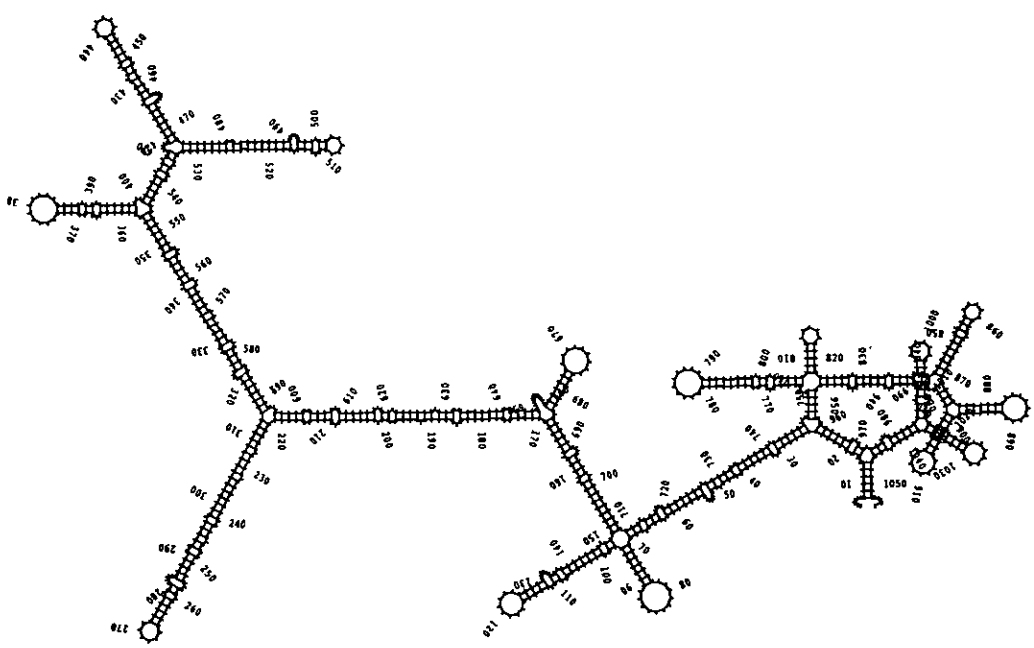


Figure 7. The predicted secondary structure of the RNA genome of satellite tobacco mosaic virus showing the high degree of Watson–Crick base-pairing (65% of total bases). The base-paired regions might be expected to form double helical 3-dimensional structure as seen bound to the interior of the capsid proteins in the STMV. The overlapped segments on the right, composed of bases 1 to 30 and 750 to 1059 are believed to comprise the recognition region for the TMV provided RNA transcriptase.

Watson–Crick base-paired, one can, with little difficulty, find more than 30 candidate segments of the required secondary structure to provide those observed in the virus. While the secondary structure shown in Figure 7 is only a model and is probably inaccurate in its details, it must be fundamentally correct. The secondary structure prediction, along with the X-ray results, indicate that the RNA must fold back upon itself in a manner similar to that seen in Figure 7 to form extensively base-paired helical stems.

The RNA segments seen in Figure 6 possess rather good helical geometry. This may be due in part to the averaging process, but it might also be an accurate indication of their uniformity. The model RNA double helix was constructed with 3′ *endo* sugars, although confirmation of the ribose conformation must await high resolution refinement.

The atomic co-ordinates from the model in Figure 6 were analyzed using the program NEWHELIX (Dickerson *et al.*, 1985) to determine detailed helical

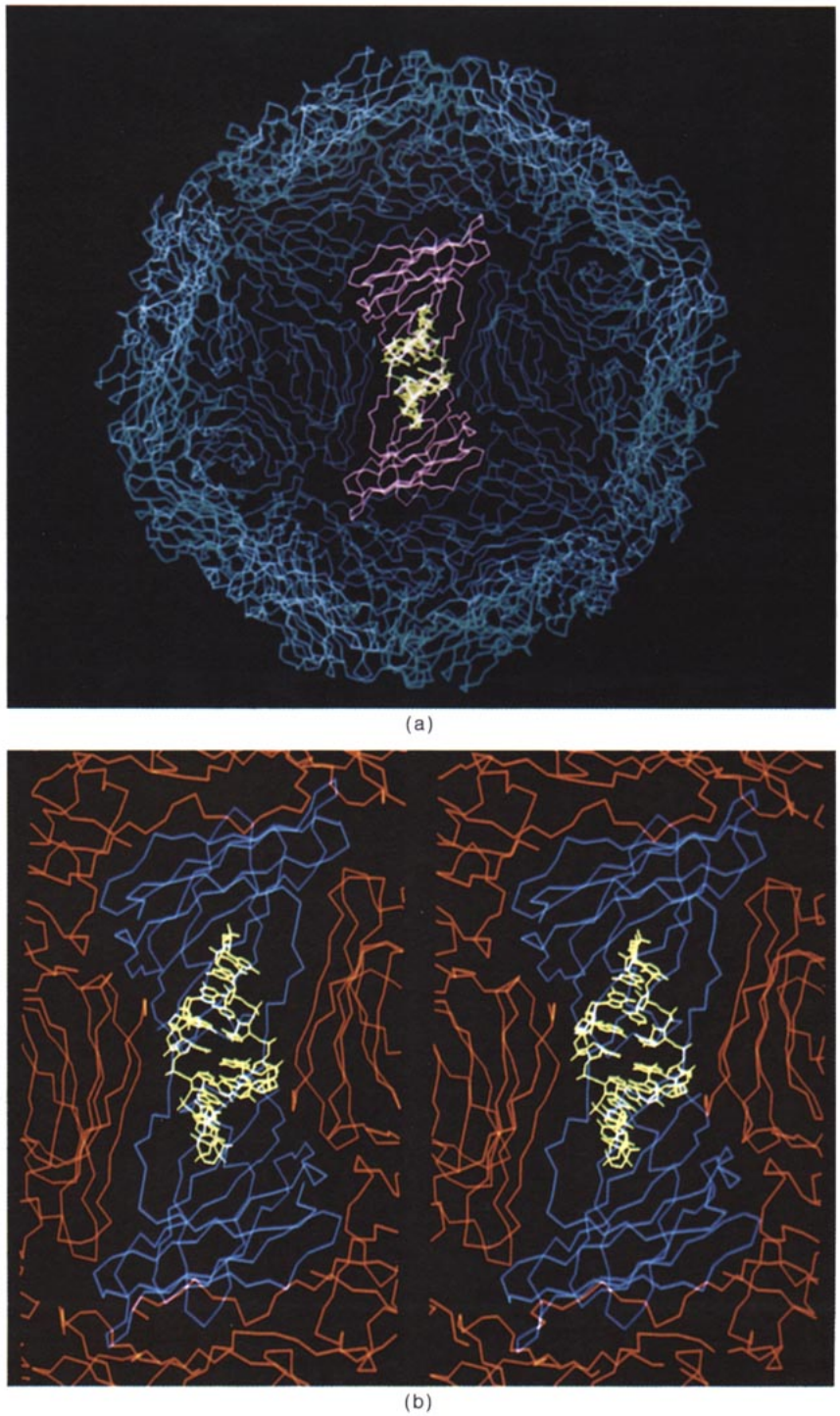


Figure 8. The STMV virion along a 2-fold axis with the front half cut away to show a dimer of capsid protein associated with a segment of RNA double helix is shown. (b) The protein dimer is seen in greater detail as it is embedded in the virion shell while maintaining association with the segment of helical RNA.

parameters. The base tilt with respect to the helical axis is about 70°, which is characteristic of *A*-form DNA. This was observed in helical portions of tRNA as well (Kim *et al.*, 1974; Robertus *et al.*, 1974). The overall shape of the RNA helix is, otherwise, more similar to *B*-form DNA, with a pitch of 30 to 32 Å, 10 base-pairs per turn and a diameter of 17 Å. The width and depth of the major and minor grooves also more closely resemble *B*-form DNA. Thus the RNA double helix seen in STMV is something of a hybrid structure that incorporates some of the properties of both *A* and *B*-form DNA.

Nucleic acid helices have dyad axes lying in each base plane, perpendicular to and intersecting the helical axis. The 2-fold axes are exact for the two antiparallel polyphosphate backbones of the base-paired strands, for the sugar moieties and for the glycosidic bonds. In the virus, the dyad axes of the

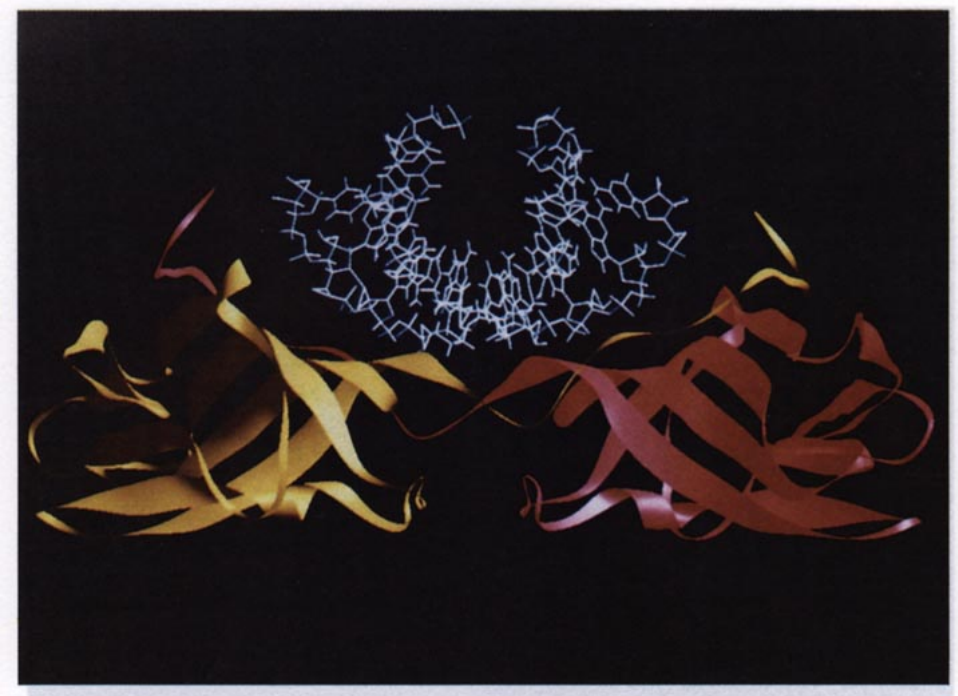


Figure 9. The disposition and orientation of the double helical RNA segment is shown with respect to a dimer of the capsid protein. Such an interaction occurs at the 30 icosahedral dyad axes. In this illustration the 2-fold axis is vertical and coincides with the central dyad of the RNA. The illustration was made using the program RIBBONS (Carson & Bugg, 1986).

central base-pairs of each helical segment are precisely coincident with the icosahedral 2-fold symmetry elements of the capsid. The helical axes of the RNA segments, on the other hand, lie in the plane defined by sequential 5-fold axes, that is, parallel to the edges of the icosahedron.

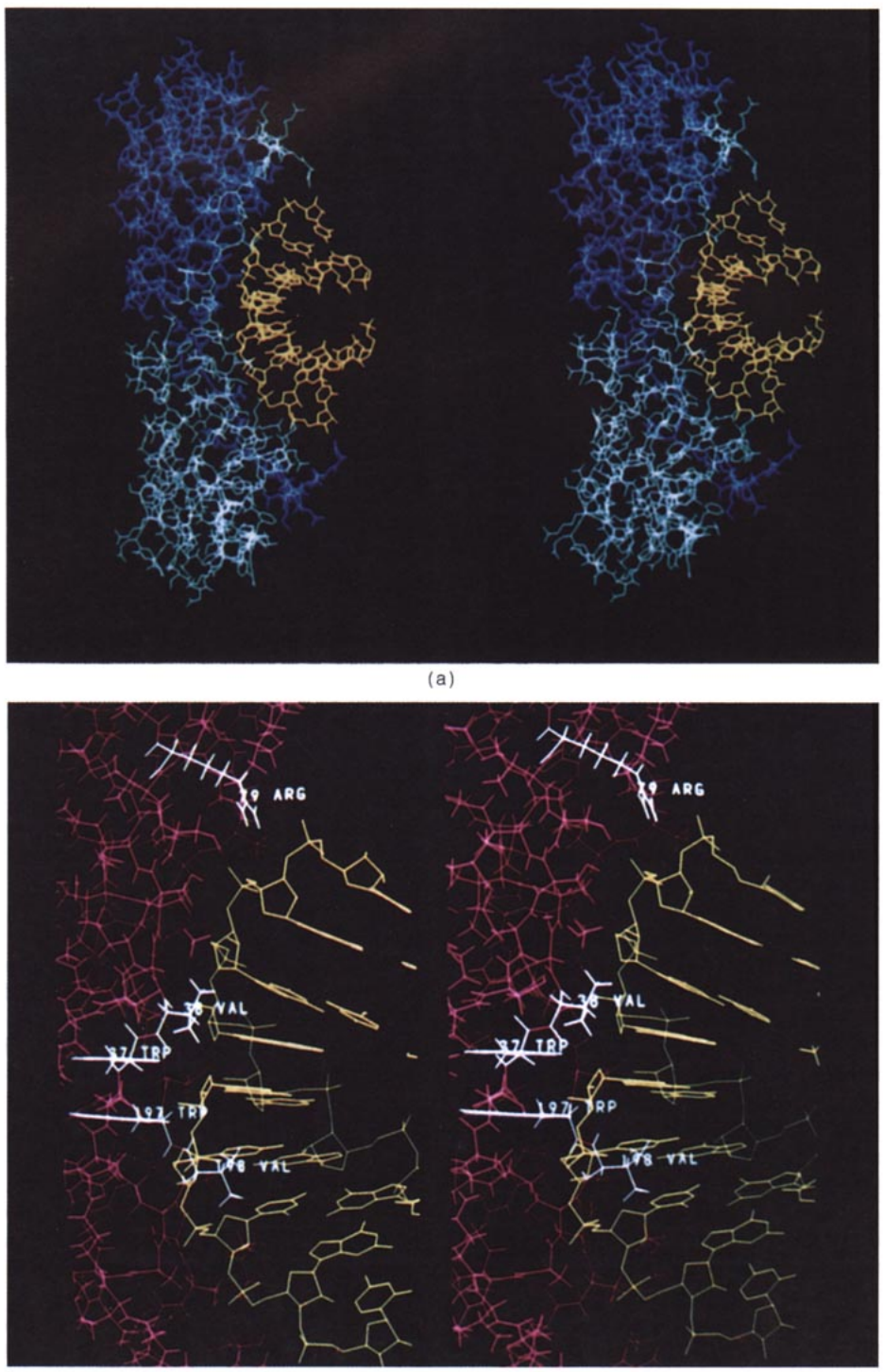
As seen in Figure 8, the double helical RNA segments are closely associated with the inside surface of the STMV capsid. Each RNA segment is in intimate contact with a dyad related pair of capsid proteins on the viral surface with which they share a common 2-fold axis. The dimer of the capsid protein thus constitutes a non-base-pair specific (so far as we know), double helical RNA binding protein. The broad features of the association between the RNA segments and the capsid proteins that bind them are shown in Figure 9. The protein dimers form cradles at their centers on the inside surface of the virus. The RNA segments lie crosswise in the cradles. The van der Waals surfaces of the RNA helices and the protein dimers are in direct contact, although water molecules could be a part of the interface. Contacts between protein and nucleic acid arise primarily from main and side-chain atoms contributed by several strands of the β-barrels, and from residues contained in the visible portions of the amino-terminal polypeptide.

We cannot, as we have noted above, unambiguously identify from the electron density map the positions of residues 1 to 15. This segment of the polypeptide, which contains five lysine and arginine residues, must, we believe, penetrate into the interior of the virion where it is associated with other portions of the RNA. This interior nucleic acid may be either completely disordered, which we doubt, or spatially inconsistent with icosahedral symmetry.

Some details of the interactions between the protein and the double helical RNA segments are illustrated in Figure 10. Because dyad symmetry prevails, every interaction has a 2-fold related mate. Figure 10(a) shows the extent of the interface between protein and nucleic acid. In Figure 10(b) is seen the neutralization of an RNA phosphate group by arginine side-chain 79 and a valine side-chain lying squarely in the center of the minor groove of the RNA helix.

One of the most intriguing constellations of amino acid residues at the protein–RNA interface is seen in Figure 10(c). Tryptophan side-chain 37 and its dyad related mate are closely stacked with 3·6 Å separation, an icosahedral 2-fold axis passing between. The curious aspect of this formation is that the planes of the tryptophan rings are almost exactly parallel with the plane of the central base-pair of the RNA helical segment. Why the planes of all of these aromatic groups are parallel is not yet clear.

We are currently pursuing further analysis of the protein–RNA interface to delineate all of the interactions responsible for protein–RNA affinity. There are, for example, likely to be at least four pairs of 2-fold related arginine and lysine side-chains making salt bridges with the nucleic acid, including that seen in Figure 10(b). We can, in addition, see other probable contacts and hydrogen bonds. We prefer, however, to defer further discussion of these until higher resolution refinement is complete.



(a)

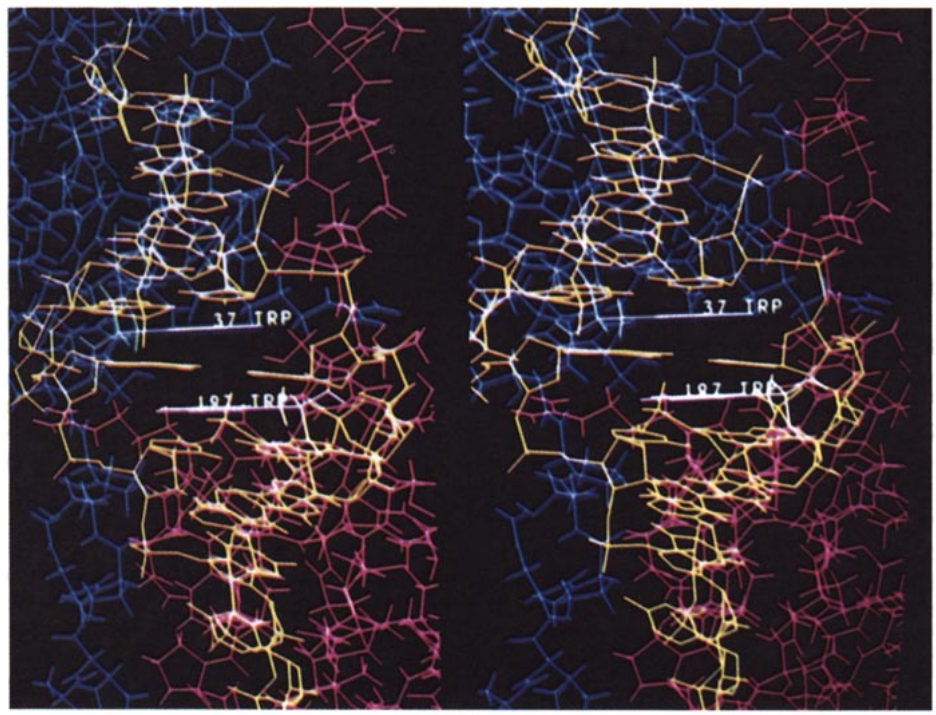
(b)

Fig. 10.

5. Discussion

The structure that we have found demonstrates that, for STMV at least, viral nucleic acid is not simply gathered into the capsid in a random or disorganized manner. It shows that the genomic RNA must possess a high degree of Watson–Crick base-paired secondary structure and that there is substantial order to the distribution of nucleic acid in the virion. It further shows that there are extensive and specific interactions between capsid protein, which is a nucleic acid binding protein, and the RNA. These likely serve not only to package the nucleic acid but to direct ordered assembly of the virus. Indeed, similar mechanisms have been proposed previously for a number of other viruses (for a review, see Casjens, 1985).

Full occupancy of the protein binding sites by RNA helices would alone account for about 45% of the nucleic acid in the virus. If these segments were



(c)

Figure 10. (a) The broad interface between a capsid protein dimer and a segment of helical RNA. (b) The guanidinium group of arginine 76 approaches a nucleic acid phosphate while a valine side-chain is positioned prominently in the minor groove of the RNA helix. Because of 2-fold symmetry, every interaction occurs in pairs. (c) One of the most intriguing dispositions of residues is shown here. Two symmetry related tryptophan side-chains (37 and 197) are stacked upon one another at the protein–RNA interface with their planes parallel with that of the central base-pair of the RNA double helix (view is along the dyad).

simply connected end to end in an expeditious but otherwise stereochemically sound manner, about 300 additional nucleic acid bases would be accounted for. Thus, full, or near complete occupancy of the RNA binding sites by continuous RNA would imply that something of the order of 75% of the RNA in the virus is found in the shell immediately interior to the protein coat. This is, in fact, consistent with neutron diffraction studies of crystalline STNV (Bentley *et al.*, 1987).

There are some conclusions that cannot be drawn with certainty, but which seem likely based on the results presented here. The capsid protein dimer appears to be an integral structural entity and is most probably the viral assembly unit. This follows not only from the structural complementarity and the close interactions between the two protein monomers within the dimer, but also because only a protein dimer would contain a composite binding site for the RNA segments. As with most proteins that bind double-stranded nucleic acid (Branden & Tooze, 1991), the protein dimer of the STMV capsid possesses an exact dyad axis of symmetry. In the virus this dyad serves the dual role of insuring compatibility with icosahedral packing symmetry and also providing a bi-directional binding site capable of accommodating the antiparallel strands of the nucleic acid helix.

Virus assembly is a highly co-operative process, involving the sequential ordered addition of protein units to form a capsid. Thus, co-operativity of RNA binding and co-operativity of viral capsid assembly are probably intimately interrelated if not part and parcel of one single process. That is, co-operative binding of protein dimers to double-stranded, helical RNA directs the ordered aggregation of protein dimers into an array that eventually closes around the nucleic acid to form an icosahedral capsid. There is good evidence from reassembly studies of other viruses, turnip crinkle virus for example (Sorger *et al.*, 1986), that this is true for other viruses as well. The secondary structure of the RNA serves the role of an organizational element essential to direct the assembly of RNA and protein dimers into a mature virus.

It seems that the nucleic acid must assume at least a unique secondary, if not tertiary, structure before protein binding initiates and viral assembly commences. This follows from the extraordinary propensity of the single-stranded RNA of STMV to form base-paired regions and, therefore, extensive helical structure. Although not essential, viral assembly directed by co-operative binding of protein to nucleic acid could require an initiation site on the RNA, as it does for TMV (Zimmern & Butler, 1977; Stubbs, 1984). This would further necessitate the prefolding or structuring of the RNA into some precursor format prior to viral assembly. A mechanism involving an ordered sequence of RNA replication, expression, folding and co-operative protein binding has the advantage that it would provide ordering in a temporal as well as a

spatial sense. The protein could not arrange itself into empty icosahedral shells (which have never been observed for STMV) before RNA were available for encapsidation, and improperly synthesized or folded RNA would be unable to direct virus assembly.

Co-operative protein–nucleic acid binding interactions are thought to proceed with an attendant conformational change that further enhances protein–protein interactions within the assembly (von Hippel *et al.*, 1977). Thus, it seems likely that the viral capsid protein may undergo some conformational change as it binds helical RNA and is incorporated into the growing array that becomes the capsid. The very extended and flexible amino-terminal polypeptide would seem ideal for mediating such changes.

The structure of STMV places constraints on viral disassembly and replication as well. There appear to be no metal ions integral to the virus structure, hence their removal would not provide an initiation mechanism. There are no open channels into the virus along which molecules of appreciable size could pass to the interior. Only the 5-fold axes provide passageways and these are blocked by what appear to be anions. Disassembly of virions would require disruption, not only of protein–protein interactions in the capsid, but protein–nucleic acid interactions as well. This might be accomplished by pH or ionic strength changes (Kaper, 1975), but we note that *in vitro* the virus particles are stable and infective over a broad range of both of these variables.

The close co-operation and interdependence between protein and RNA in the assembly and structure of the virion has evolutionary implications as well. For any mutation, diverse but essential functions would have to be maintained. Because the RNA specifies the amino acid sequence of the protein, any base change could alter (1) the structure of the coat protein, (2) the interaction among protein subunits in the capsid and therefore assembly and disassembly, (3) the secondary structure of the RNA, (4) the binding of protein to RNA, or (5) some of those numerous other biological functions, such as infectivity and replication, whose structural basis is still unknown. Thus, one would expect that mutation and evolution of the STMV must be a very delicate and stringently selective process. Despite this prediction, the genome of the type strain has been shown, like other replicating RNAs, to exist as a quasispecies population (Domingo & Holland, 1988) and that other strains show considerable divergences from the consensus sequence of the type strain (Kurath *et al.*, 1992).

STMV, though the smallest, and presumably the simplest, virus that has been crystallized and whose structure has been determined, is in some ways more subtle and sophisticated than viruses we might consider more complex. This is seen particularly in the extraordinary molecular economy implicit in the coat protein structure, the extensive self complementarity of the RNA, and the ingenious structural relationship that exists between protein and nucleic acid in the virion.

Crystals grown by a liquid-liquid diffusion method in a microgravity environment aboard International Microgravity Laboratory-I in January of 1992 provided us with crystals that were more than ten times the volume of any we had ever grown on earth and of enhanced quality (Day & McPherson, 1992). From eight of these crystals we were able to collect a highly redundant data set that was complete and statistically well above background to a resolution of 1·8 Å. We will, of course, continue refinement of the virus structure to the maximum resolution achievable. From that analysis we hope to extract more quantitative descriptions of the detailed atomic interactions that are fundamental to this virus structure.

This research was supported by grants from NIH, NSF and NASA. The authors wish to acknowledge the efforts of Mr J. Heick and M. Angel for the preparation of virus and to colleagues T.-P. Ko, N. Ban and M. Greene for their assistance and advice. We thank the San Diego Supercomputer Center for a grant of time on the Cray Y-MP.

References

Arnold, E. & Rossmann, M. G. (1988). The use of molecular replacement phases for the refinement of the human rhinovirus 14 structure. *Acta Crystallogr. sect. A*, **44**, 270–282.

Arnold, E., Vriend, G., Luo, M., Griffith, J. P., Kamer, G., Erickson, J. W., Johnson, J. E. & Rossmann, M. G. (1987). The structure of a common cold virus, human rhinovirus 14. *Acta Crystallogr. sect. A*, **43**, 346–361.

Bentley, G. A., Lewit-Bentley, A., Liljas, L., Skoglund, U., Roth, M. & Unge, T. (1987). Structure of RNA in satellite tobacco necrosis virus: a low resolution neutron diffraction study using $^1H_2O/^2H_2O$ solvent contrast variation. *J. Mol. Biol.* **194**, 129–141.

Bernstein, F. C., Koetzle, T. F., William, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanonchi, T. & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.

Branden, C. & Tooze, J. (1991). The structure of spherical viruses. In *Introduction to Protein Structure*, chapt. 11, pp. 161–177, Garland Publishing Inc., New York.

Bricogne, G. (1974). Geometric sources of redundancy in intensity data and their use for phase determination. *Acta Crystallogr. sect. A*, **30**, 395–405.

Bricogne, G. (1976). Methods and programs for direct-space exploitation of geometric redundancies. *Acta Crystallogr. sect. A*, **32**, 832–847.

Brunger, A. T. (1991). Simulated annealing in crystallography. *Annu. Rev. Phys. Chem.* **42**, 197–223.

Brunger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic *R* factor refinement by molecular dynamics. *Science*, **235**, 458–460.

Carson, M. & Bugg, C. E. (1986). An algorithm for ribbon models of proteins. *J. Mol. Graph.* **4**, 121–122.

Casjens, S. (1985). Nucleic acid packaging by viruses. In *Virus Structure and Assembly* (Casjens, S., ed.), chapt. 3, pp. 76–147, Jones & Bartlett Inc., Boston.

Casper, D. L. & Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.

Day, J. & McPherson, A. (1992). Macromolecular crystal growth experiments on International Microgravity Laboratory-1. *Protein Sci.* **1**, 1254–1268.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387–395.

Dickerson, R. E., Kopka, M. L. & Pjura, P. (1985). Base sequence, helix geometry, hydration, and helix stability in *B*-DNA. In *Biological Macromolecules and Assemblies. Nucleic Acids and Interactive Proteins* (Jurnak, F. A. & McPherson, A., eds), vol. 2, chapt. 2, pp. 37–126, John Wiley & Sons, New York.

Dodds, J. A. (1991). Structure and function of the genome of satellite tobacco mosaic virus. *Canad. J. Plant Pathol.* **13**, 192–195.

Domingo, E. & Holland, J. J. (1988). High error rates, population equilibrium and evolution of RNA replication systems. *RNA Genet.* **3**, 3–35.

Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. sect. A*, **47**, 392–400.

Fitzgerald, P. M. D. (1988). MERLOT, an integrated package of computer programs for the determination of crystal structures by molecular replacement. *J. Appl. Crystallogr.* **21**, 273–276.

Harrison, S. C. (1991). What do viruses look like? *Harvey Lectures*, **85**, 127–152.

Holmes, K. C. (1984). The structure determination of tobacco mosaic virus. In *Biological Macromolecules and Assemblies, Virus Structures* (Jurnak, F. A. & McPherson, A., eds), vol. 1, chapt. 4, pp. 121–148, John Wiley & Sons, New York.

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.

Jones, T. A. (1985). Interactive computer graphics: FRODO. In *Methods in Enzymology* (Wyckoff, H. W., Hirs, C. H. W. & Timasheff, S. N., eds), vol. 115, chapt. 12, pp. 157–171, Academic Press, New York.

Joshi, S. & Haenni, A.-L. (1984). Plant RNA viruses: strategies of expression and regulation of viral genes. *FEBS Letters*, **177**, 163–174.

Kaper, J. M. (1975). *The Chemical Basis of Virus Structure, Dissociation and Assembly*, North Holland, Amsterdam.

Kim, K. S., Valverde, R. A. & Dodds, J. A. (1989). Cytopathology of STMV and its helper virus in tobacco. *J. Ultrastruct. Mol. Struct. Res.* **102**, 196–204.

Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A., Seeman, N. C. & Rich, A. (1974). Three dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science*, **185**, 435–440.

Koszelak, S., Dodds, J. A. & McPherson, A. (1989). Preliminary analysis of crystals of satellite tobacco mosaic virus. *J. Mol. Biol.* **209**, 323–325.

Kurath, G., Rey, M. E. & Dodds, J. A. (1992). Analysis of genetic heterogeneity within the type strain of STMV reveals several variants and a strong bias for G to A substitution mutations. *Virology*, **189**, 233–244.

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

Liljas, L. (1986). The structure of spherical viruses. *Progr. Biophys. Mol. Biol.* **48**, 1–36.

Liljas, L., Unge, T., Jones, T. A., Fridborg, K., Lovgren, S., Skoglund, U. & Strandberg, B. (1982). Structure of satellite tobacco necrosis virus at 3·0 Å resolution. *J. Mol. Biol.* **159**, 93–108.

Mirkov, T. E., Mathews, D. M., DuPlessis, D. H. & Dodds, J. A. (1989). Nucleotide sequence and translation of STMV-RNA. *Virology*, **170**, 139–146.

Rayment, I. (1983). Molecular replacement method at low resolution: optimum strategy and intrinsic limitations as determined by calculations in icosahedral virus models. *Acta Crystallogr. sect. A*, **39**, 102–116.

Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. D., Clark, B. F. C. & Klug, A. (1974). Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature (London)*, **250**, 546–551.

Rossmann, M. G. & Blow, D. M. (1962). The detection of subunits within the crystallographic asymmetric unit. *Acta Crystallogr.* **15**, 24–31.

Rossmann, M. G. & Erickson, J. W. (1985). Structure and assembly of icosahedral shells. In *Virus Structure and Assembly* (Casjens, S., ed.), chapt. 2, pp. 30–73, Jones & Bartlett Inc., Boston.

Rossmann, M. G. & Johnson, J. E. (1989). Icosahedral RNA—Virus structure. *Annu. Rev. Biochem.* **58**, 533–573.

Sorger, P. K., Stockley, P. G. & Harrison, S. C. (1986). Structure and assembly of turnip crinkle virus II. Mechanism and reassembly *in vitro*. *J. Mol. Biol.* **191**, 639–658.

Stubbs, G. (1984). Macromolecular interactions in tobacco mosaic virus. In *Biological Macromolecules and Assemblies. Virus Structures* (Jurnak, F. A. & McPherson, A., eds), vol. 1, chapt. 5, pp. 149–202, John Wiley & Sons, N.Y.

Terwilliger, T. & Eisenberg, D. (1984). Unbiased three-dimensional refinement of heavy atom parameters by correlation of origin-removed Patterson functions. *Acta Crystallogr. sect. A*, **39**, 813–816.

Tronrud, D. E., TenEyck, L. F. & Matthews, B. W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Crystallogr. sect. A*, **43**, 489–501.

Valverde, R. A. & Dodds, J. A. (1986). Evidence for a satellite RNA associated naturally with the U5 strain and experimentally with the U1 strain of TMV. *J. Gen. Virol.* **67**, 1875–1884.

Valverde, R. A. & Dodds, J. A. (1987). Some properties of isometric virus particles which contain the satellite RNA of TMV. *J. Gen. Virol.* **68**, 965–972.

Valverde, R. A., Heick, J. A. & Dodds, J. A. (1991). Interactions between STMV, helper tobamovirus, and their hosts. *Phytopathology*, **81**, 99–104.

von Hippel, P. H., Jensen, D. E., Kelly, R. C. & McGhee, J. D. (1977). Molecular approaches to the interaction of nucleic acids with "melting" proteins. In *Nucleic Acid-Protein Recognition* (Vogel, H. J., ed.), pp. 66–89, Academic Press, New York.

Xuong, N.-H., Nielson, C., Hamlin, R. & Anderson, D. (1985). Strategy for data collection from protein crystals using a multi-wire counter area detector diffractometer. *J. Appl. Crystallogr.* **18**, 342–360.

Zimmern, D. & Butler, P. (1977). The isolation of TMV-RNA fragments containing the origin for viral assembly. *Cell*, **11**, 455–462.

Zucker, M. & Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucl. Acids Res.* **9**, 133–148.

Edited by A. Klug