Crystal Structural Analysis of Tobacco Necrosis Virus at 5 Å Resolution

By Masahiko Bando, Yukio Morimoto, Takao Sato and Tomitake Tsukihara

Department of Biological Science and Technology, Faculty of Engineering, The University of Tokushima, Tokushima 770, Japan

and Yukio Yokota, Keiichi Fukuyama and Hiroshi Matsubara

Department of Biological Science, Faculty of Science, Osaka University, Toyonaka 680, Japan

Abstract

X-ray diffraction intensities for tobacco necrosis virus crystals were collected at 5 Å resolution using a Weissenberg camera with a large cassette of radius 430 mm. The synchrotron radiation source at the Photon Factory was used. The crystal structure of the virus was obtained by 91 cycles of the non-crystallographic symmetry averaging. Secondary structures such as α-helices and β-structures were clearly identified in the electron-density map at 5 Å resolution. This virus resembles southern bean mosaic virus both in orientation of coat protein subunits and in their folding. Ordered and disordered parts of each subunit of tobacco necrosis virus are shorter and longer than the corresponding parts of the southern bean mosaic virus by 12 and 27 residues, respectively.

Introduction

Tobacco necrosis virus (TNV) can multiply by itself, and satellite tobacco necrosis virus (STNV) multiplies only in the presence of the TNV (Kassanis & Nixon, 1960). The extract from plants inoculated with a mixture of TNV and STNV gave two discrete light-scattering zones in a sucrose density gradient corresponding to TNV and STNV (Kassanis & Nixon, 1960). This suggests that their coat proteins have high specificity for the recognition of their cognate RNA's. One of the goals of the present study is to elucidate the inter-recognition mechanism between the TNV coat protein and its RNA.

Various molecular weights ranging from 22 600 to 33 000 have been reported for coat proteins of the different strains of TNV (Uyemoto & Grogan, 1969; Lesnaw & Reichmann, 1969). We have obtained a molecular weight of 30 000 for the coat protein of the present strain (Fukuyama, Hirota & Tsukihara, 1987). The subunit is composed of 275 amino acids (Saeki et al., unpublished). TNV crystals belong to the cubic space group $P4_232$ with cell dimension $a = 338$ Å. For the unit cell of 3.86×10^7 Å^3, containing two virus particles with a molecular weight of about 6.7×10^6, the V_m value (Matthews, 1968) is 2.9 Å^3 Da^{-1} (Tsukihara, Yokota, Koyama, Fukuyama & Matsubara, 1990). Symmetry ($T = 3$) of TNV capsid and subunit arrangement of 180

copies of TNV coat protein were determined in a previous X-ray analysis (Tsukihara et al., 1990). Assuming that the coat protein is a sphere, the virus structure was compared with that of southern bean mosaic virus (SBMV), which is known at atomic resolution (Abad-Zapatero et al., 1981). These two viruses resemble each other in the arrangement of spherical subunits on the icosahedral surface (Tsukihara et al., 1990). Utilizing the structural similarity between the two viruses, we determined the crystal structure of TNV by using the non-crystallographic symmetry (NCS) averaging procedure (Bricogne, 1974, 1976).

Intensity data collection

X-ray data collection was carried out on the BL6A2 station at the Photon Factory, using a screenless Weissenberg camera (Sakabe, 1983) with a cylindrical cassette of 430 mm radius. The X-ray wavelength was 1.488 Å. The diffraction intensities were recorded on a 400×200 mm imaging plate (Fuji Film Co. Ltd). A crystal measuring $0.7 \times 0.7 \times 0.7$ mm was mounted in a capillary in such a way that it could be rotated about the [111] axis. The total oscillation range of $31.625°$ was covered by 31 serial Weissenberg photographs, each with an oscillation angle of $1.125°$. Each individual oscillation range overlapped with adjacent ones by $0.125°$. The exposure period of each scan was 90.0 s. The temperature in the X-ray experimental hatch was kept at 281 K during the intensity-data collection. The diffraction patterns recorded on the plates were digitized at 100 μm intervals using a Fuji BA100 readout system. The intensity data were processed up to 5 Å resolution using the program system WEIS (Higashi, 1989). The mosaic spread of each plate was set to a constant value of $0.1°$. The details of data collection and scaling by the method of Hamilton, Rollett & Sparks (1965) are given in Table 1. The merging R factor was 0.115 for 90 251 reflections with $I > 5\sigma(I)$. The number of observed independent reflections and the theoretically possible reflections in each shell are listed in Table 2. The observed reflections with $F > 3\sigma(F)$ amounted to 83% of the theoretical number in the 100–5 Å resolution range,

Table 1. Intensity data collection

Film No.	Collected reflections Oscillation range (°)	No.*	Scaled reflections $I > 5\sigma(I)$ No.‡	R§
1	−6.5−−5.375	5221	3504	0.084
2	−5.5−−4.375	5512	3698	0.083
3	−4.5−−3.375	5419	3557	0.092
4	−3.5−−2.375	5254	3299	0.111
5	−2.4−−1.275	5285	3315	0.098
6	−1.6−−0.475	5288	3281	0.104
7	−0.7 0.425	4552	2514	0.130
8	0.3 1.425	4733	2566	0.113
9	1.3−2.425	4538	2396	0.121
10	2.9−4.025	5334	3275	0.117
11	3.9−5.025	5217	3127	0.122
12	4.9−6.025	5339	3273	0.130
13	5.9 7.025	5218	3079	0.121
14	7.0−8.125	5148	2977	0.114
15	8.0−9.125	5080	2904	0.116
16	9.0 10.125	5039	2802	0.111
17	10.0 11.125	5192	3040	0.117
18	11.0−12.125	5139	3070	0.110
19	12.0−13.125	5403	3252	0.112
20†	13.0 14.124	3918	543	0.209
21	14.0-15.125	5236	3102	0.116
22	15.0 16.125	5077	2886	0.121
23	16.0−17.125	5066	2882	0.125
24	17.0−18.125	5033	2790	0.125
25	18.0 19.125	4944	2716	0.129
26	19.0 20.125	4991	2849	0.119
27	20.0−21.125	5108	2948	0.120
28	21.0−22.125	4791	2626	0.120
29	22.0−23.125	4812	2663	0.122
30	23.0−24.125	4763	2634	0.123
31	24.0−25.125	4825	2678	0.128
Total		156475	90251	0.115

* Number of reflections accepted by WEIS.

† Reflections of this film are weaker than those of other films because of shorter exposure.

‡ Number of reflections accepted for scaling.

§ $R = \sum_h \sum_i |I_{hi} - \overline{I_h}| / \sum_h \sum_i I_{hi}$.

Table 2. Independent reflections in each shell

Resolution (Å)	Number of reflections Theoretical	Observed $F > 5\sigma(F)$
100–15	1244	1117
15–12	1090	958
12–10	1597	1501
10–8	3530	3218
8–7	3504	2901
7–6	6171	4917
6–5	12028	9515
Total	29161	24127

and uniformly covered the whole range of independent reflections in the resolution shell.

Orientation and location of coat protein subunits in the icosahedral capsid

In the previous study icosahedral particles were uniquely located at the Wyckoff positions *a* (0,0,0;1/2,1/2,1/2) (*International Tables for Crystallography*, 1983, Vol. A, pp. 626–628) in the unit cell (Tsukihara *et al.*, 1990). A cross-rotation function between a TNV crystal and a model crystal composed of an icosahedral asymmetric unit of SBMV was calculated using the fast-rotation function program written by Crowther (1972). The atomic parameters of SBMV were taken from the Protein Data Bank (Abad-Zapatero *et al.*, 1981; Bernstein *et al.*, 1977). The reflection data from 10 to 12 Å spacing were included and integration was carried out within the region between 5 and 58 Å in the calculation. Significant peaks appeared close to peak positions of self-rotation function of TNV. Consequently, TNV is similar to SBMV in molecular orientation of the subunits in the icosahedral capsid. Thus, a poly-alanine model of
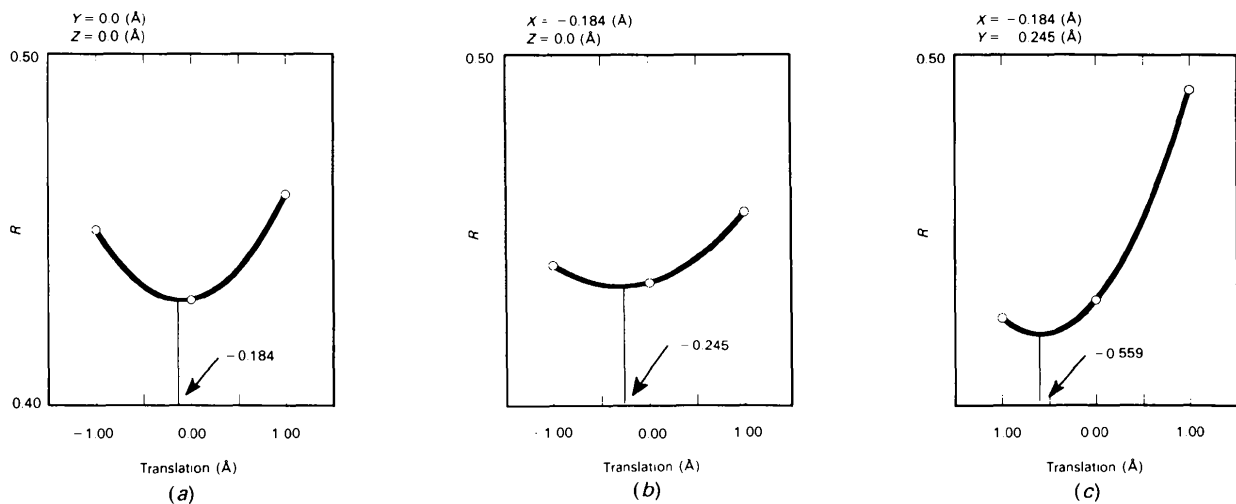

Fig. 1. Translational search along the *a*, *b* and *c* axes for an icosahedral asymmetric unit of the initial model of TNV. (*a*) R factors at three points of −1.0, 0.0 and 1.0 Å from the initial position along the *a* axis. (*b*) Those along the *b* axis, where translations along *a* and *c* axes were fixed at −0.184 and 0.0 Å, respectively. (*c*) The third step of the calculation, the translation was estimated as (−0.184,−0.245,−0.559) Å.

SBMV placed on each of the Wyckoff positions a was applied for an initial model of TNV for the structure analysis. The position of the icosahedral asymmetric unit of the initial model was refined by an R-factor ($R = \Sigma|F_o - F_c|/\Sigma F_o$) search. The center of the virus was fixed at the Wyckoff positions a during the R-factor search. The most probable position of the icosahedral asymmetric unit was obtained as $(0.184,-0.245,-0.559 \text{ Å})$ from that of the initial TNV model (Fig. 1). The refined TNV model with the R factor of 0.419 was used for further refinement.

Phase extension from 8 to 5 Å resolution by NCS averaging

The course of phase refinement by NCS averaging is shown in Fig. 2. An electron-density map was calculated by the *FFT* program originally written by Ten Eyck (1973). The second and the third steps were performed by Wang's method (Wang, 1985). In the second step, radius R in equation (2) of Wang (1985) was set at 15 Å for 8 Å resolution and 11 Å for 5 Å resolution. A threshold level of the electron density to define the molecular boundary was chosen so that the number of grid points within the molecular region was half of the total number of grid points. In addition, the molecular region was restricted to a shell whose inner and outer radii were 50 and 170 Å, respectively.

In the third step, the constant, S, in equation (1) of Wang (1985) was set at 0.30 throughout the iterations, and a constant ρ^c was estimated by equation (1) of Wang (1985). This ρ^c was added to the electron density in the molecular region of the map and any remaining negative density in the region was set to zero to reduce noise. The electron density of the solvent region was fixed to a constant value of $\rho^c + \langle\rho_{sol}\rangle$, where $\langle\rho_{sol}\rangle$ is the average electron density of the solvent region.

A new electron density was generated at each grid point in the region defining the molecule by averaging five electron densities at icosahedral equivalent points. These five electron densities were evaluated by a linear interpolation in a fine-grid cell with grid-point spacings of about 1/6 of the current resolution.

Structure factors were calculated by Fourier transformations of the new electron densities in a grid cell with grid-point spacings of about 1/3 of the current resolution. A correlation coefficient,

$$C_{av} = \Sigma(F_o - \langle F_o\rangle)(F_{MR} - \langle F_{MR}\rangle)/$$
$$[\Sigma(F_o - \langle F_o\rangle)^2(F_{MR} - \langle F_{MR}\rangle)^2]^{1/2},$$

where F_o and F_{MR} are observed and calculated structure factors, and an R factor, $R_{av} = \Sigma|F_o - F_{MR}|/\Sigma F_o$, were calculated to monitor the progress of the refinement. Following Sim (1959, 1960), weighted Fourier coefficients for the electron-density calculation of the next cycle were generated using the calculated and the observed structure factors. These procedures were repeated at the same resolution until the correlation coefficient and the R factor converged to constant values. The step size of each phase extension was kept within 1.5 reciprocal cell dimensions. The criteria used at the final stage of refinement are shown in Fig. 3. After 91 cycles of the iteration the C_{av} and the R_{av} were converged to 0.92 and 0.16, respectively. The final electron-density map was constructed with the observed structure factors and the calculated phases in the resolution shell between 100 and 5 Å. Fig. 4 shows the distribution of electron densities in the final map. The maximum and minimum electron densities of the map were 6.70×10^4 and $-5.49 \times 10^4 \text{ e Å}^{-3}$, respectively, on an arbitrary scale.

Step 1.   Electron density calculation
Step 2.   Definition of the molecular boundary
Step 3.   Filtering noise in real space
Step 4.   Averaging electron densities among NCS equivalents
Step 5.   Calculation of structure factors
Step 6.   Generation of Fourier coefficients

Fig. 2. The course of NCS averaging. After the phase refinement is converged by several cycles of the iteration, the refinement was extended at a resolution step to a higher resolution.
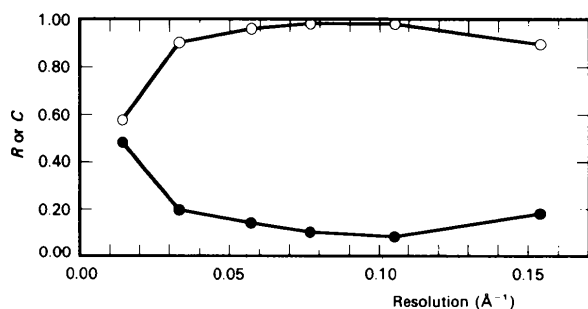

Fig. 3. R factor and correlation coefficient for each resolution shell at the final stage of the refinement.
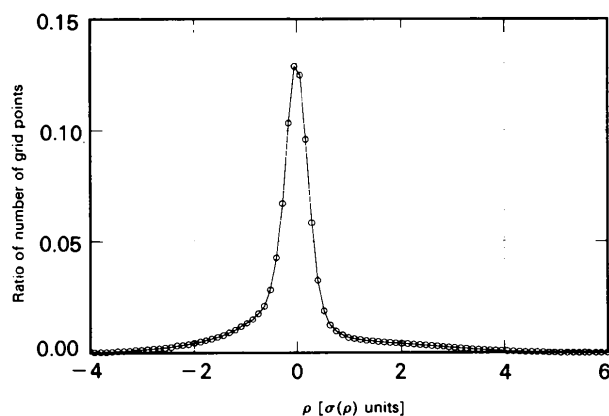

Fig. 4. Statistics of electron-density distribution calculated on an arbitrary scale using reflections between 100 and 5 Å spacings. Since $F(000)$ is not included in the calculation, the averaged electron density is 0. The vertical axis represents the ratio of the number of grid points to the total number of grid points in the unit cell. The horizontal axis represents the electron-density range expressed in units of electron-density variance.

The estimated standard deviation (e.s.d.) of the electron densities, σ, is $1.08 \times 10^4 \, \mathrm{e} \, \text{Å}^{-3}$.

Model building

Electron-density cages of the final map of 5 Å resolution were constructed. The total volume of the cages was 10% of that of the unit cell. A poly-glycine model of a C subunit of SBMV was superposed on the electron-density map on a Silicon Graphics graphics system using *TURBO-FRODO* (Jones, 1978). The residues 179, 180, 210–214, 234 and 235 of TNV C subunit in

the SBMV numbering system were located outside the electron-density cages as shown in Fig. 5. Most of the remaining parts fitted the electron-density map well. 23 residues of an N-terminal arm (βA arm) of the C subunit exhibit ordered structure. It is shorter than that of SBMV by three residues. The corresponding main-chain segments of the subunits A and B were disordered like those in SBMV (Abad-Zapatero et al., 1981). TNV C subunit was rebuilt in the electron-density distribution by deleting residues outside the electron-density cages. The C subunit model except for the βA arm from residues 66 to 88 was superposed well after pseudo-threefold rotation on the electron-density cages corresponding to the A and B subunits. The structures were idealized using a program by Hendrickson & Konnert (1980). The backbone structures around the gaps in the primary structure are superposed well onto the electron-density cages as shown in Fig. 5. The average electron density at
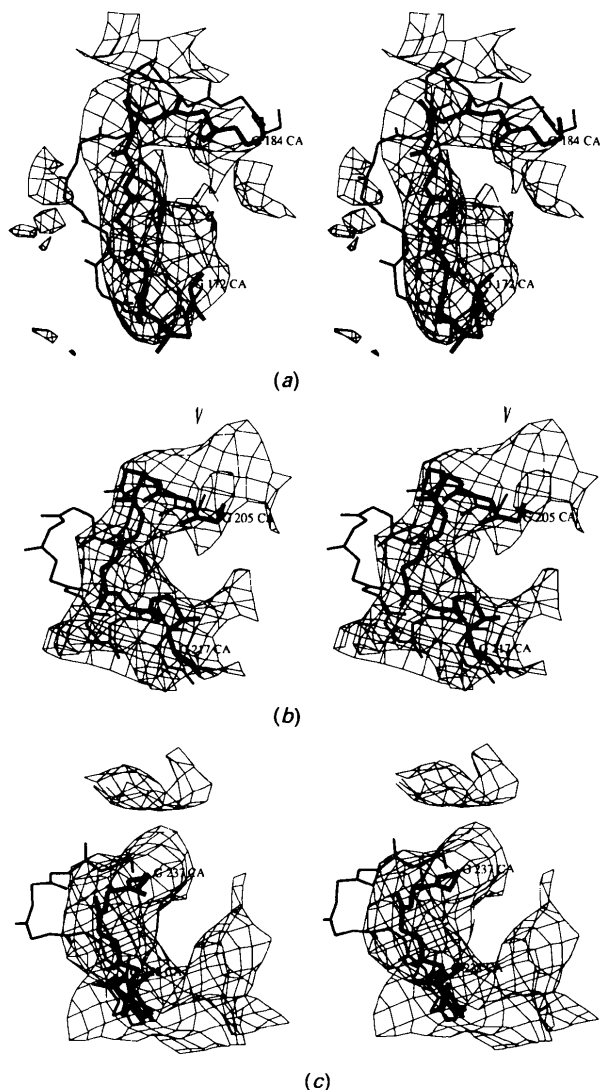

(a)

(b)

(c)

Fig. 5. Stereo diagrams of three regions of the polypeptide chain of the C subunit of TNV (heavy lines) and SBMV (thin lines) superposed on the electron-density cages drawn at the 0.97σ level. (a) 11 residues from 196 to 206 of TNV and 13 residues from 172 to 184 of SBMV. (b) Eight residues from 227 to 234 of TNV and 13 residues from 205 to 217 of SBMV. (c) Eight residues from 245 to 252 of TNV and ten residues from 228 to 237 of SBMV.
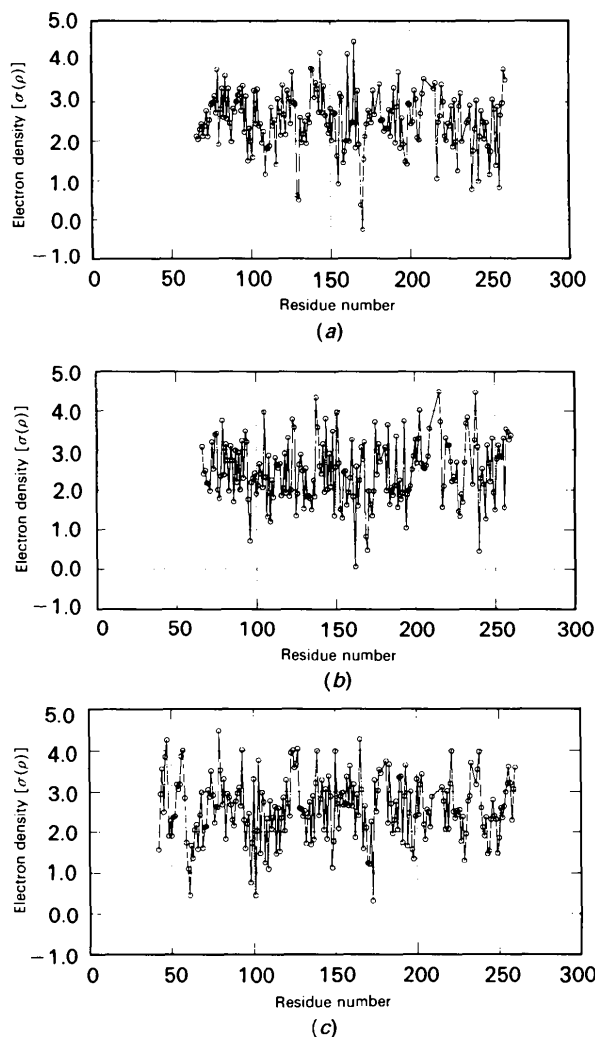

Fig. 6. Average electron density for each residue expressed in $\sigma(\rho)$ units. (a) A subunit, (b) B subunit and (c) C subunit.

atomic positions in the whole molecule is 2.31σ, and the e.s.d. is 1.27σ. 97% of the backbone atoms have electron densities higher than the average density of the whole cell. The average electron density for each amino-acid residue is shown in Fig. 6. The statistics of electron-density distributions show that the structural models are consistent with the electron-density distribution.

Concluding remarks

Secondary structures of TNV coat protein and those of SBMV are schematically depicted in Fig. 7. They were named βA, βB, βC, αA, βD, βE, αB, βF, αC, βG, αD, αE, βH and βI along the polypeptide chain, following the nomenclature for SBMV (Hermodson et al., 1982). Each of A and B subunits has 186 ordered residues in the icosahedral surface, and C subunit has 24 additional ordered residues at βA. Tertiary structure comparison between TNV and SBMV subunits shows that TNV subunits have two gaps in a loop region between αC and βG, five gaps in continuous segments of αD and αE, and two gaps in a loop between βH and βI. The gaps between αC and βG are inside the virus capsid, and the other gaps are on the surface of the virus particle

as shown in Fig. 8. The ordered βA arm of subunit C of TNV is shorter than that of SBMV by three residues. The primary structure of the present TNV subunit is not yet known. TNV strain A consisting of the same number of amino-acid residues as that of the present TNV was aligned with that of SBMV (Meulewaeter, Seurinck & Emmelo, 1990). The locations of gaps are consistent with the alignment by tertiary structure comparison and that by primary structural comparison. Out of the 275 residues in each subunit, 89, 89 and 65 residues of A, B and C subunits, respectively, are disordered and situated at the inner part of the virus particle. Consequently, a total of 14 580 disordered residues interact with RNA in the virus particle. Although TNV and SBMV have similar external radii, TNV has 4500 more disordered residues than SBMV. Most residues of His, Arg, Gln and Lys are located inside the virus and can interact with RNA as shown in Fig. 9. Interactions between these amino-acid residues and nucleic acid bases have been found by crystal structure analyses of protein–nucleic acid complexes (Pavletich & Pabo, 1991; Luisi, Xu, Otwinowski, Freedman, Yamamoto & Sigler, 1991; Koepke, Maslowska, Heinemann & Saenger, 1989; Nonaka, Mitsui, Irie & Nakamura, 1991)
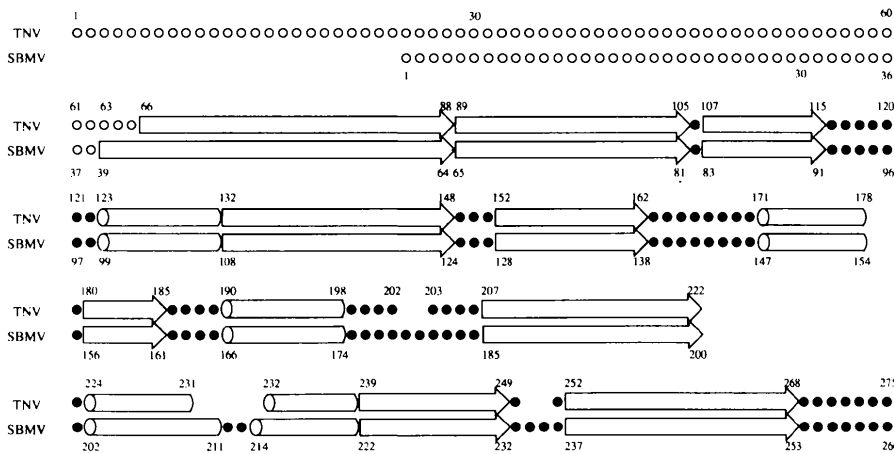

Fig. 7. Alignment of primary structures of TNV and SBMV coat proteins and the secondary-structure assignment. Closed and open circles are ordered and disordered residues, respectively. Open arrows and columns represent β-structures and α-helices, respectively.
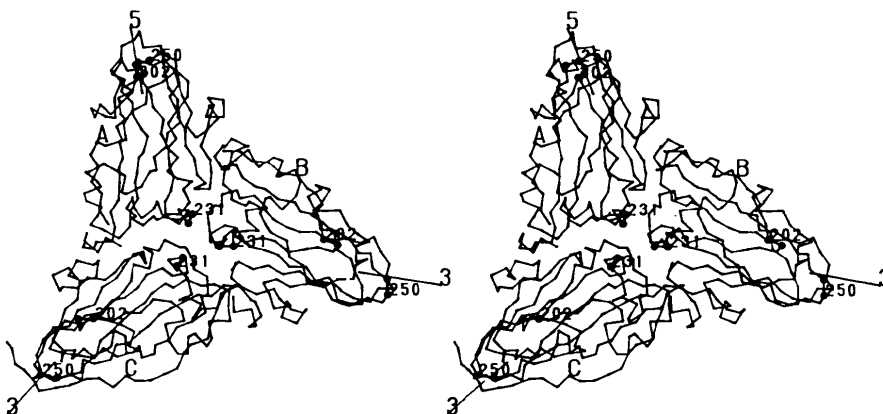

Fig. 8. Icosahedral asymmetric unit of TNV shown as a stereoscopic pair of Cα atoms. Icosahedral five- and threefold axes are indicated by numbers 5 and 3, respectively. Gaps are located between two successive residues which are shown by closed circles with a residue number.

or amino acid–nucleic acid complexes (Hata, Yoshikawa, Sato & Tamura, 1975; Takenaka, Ohki & Sasada, 1980). The high protein content in the RNA region and the interaction between RNA and amino-acid residues such as His, Arg, Gln and Lys, may be the main reason for the high RNA-recognition ability of TNV coat protein.
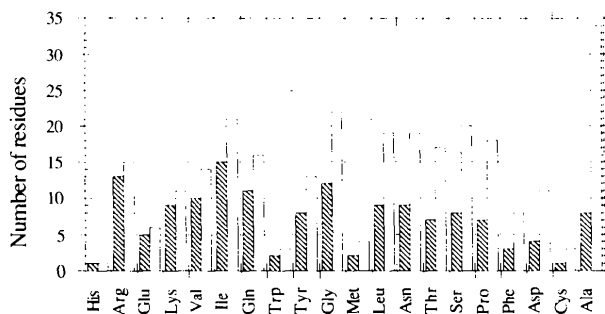
Fig. 9. The amino-acid composition of the C subunit, and the number of each amino-acid residue of a subunit that can interact with interior RNA are shown by open and hatched rods, respectively. Out of 275 amino acids in the subunit, 52% of the amino-acid residues are located inside the virus and can interact with RNA. Amino acids are arranged in the order of the ratio of amino-acid residues interacting with RNA to the respective total number. The ratios for His, Arg, Gln and Lys are higher than 81%.

References

ABAD-ZAPATERO, C., ABDEL-MEGUID, S. S., JOHNSON, J. E., LESLIE, A. G. W., RAYMENT, I., ROSSMANN, M. G., SUCK, D. & TSUKIHARA, T. (1981). *Acta Cryst.* B37, 2002–2018.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* 112, 535–542.

BRICOGNE, G. (1974). *Acta. Cryst.* A30, 395–405.

BRICOGNE, G. (1976). *Acta. Cryst.* A32, 832–847.

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–178. New York: Gordon & Breach.

FUKUYAMA, K., HIROTA, S. & TSUKIHARA, T. (1987). *J. Mol. Biol.* 196, 961–962.

HAMILTON, W. C., ROLLETT, J. S. & SPARKS, R. A. (1965). *Acta Cryst.* 18, 129–130.

HATA, T., YOSHIKAWA, M., SATO, S. & TAMURA, C. (1975). *Acta Cryst.* B31, 312–314.

HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences/IUCr.

HERMODSON, M. A., ABAD-ZAPATERO, C., ABDEL-MEGUID, S. S., PUNDAK, S., ROSSMANN, M. G. & TREMAINE, J. H. (1982). *Virology*, 119, 133–149.

HIGASHI, T. (1989). *J. Appl. Cryst.* 22, 9–18.

International Tables for Crystallography (1983). Vol. A, edited by T. HAHN, pp. 626–628. Dordrecht: Kluwer Academic Publishers.

JONES, T. A. (1978). *J. Appl. Cryst.* 11, 268–272

KASSANIS, B. & NIXON, H. L. (1960). *J. Gen. Microbiol.* 25, 469–471.

KOEPKE, J., MASLOWSKA, M., HEINEMANN, U. & SAENGER, W. (1989). *J. Mol. Biol.* 206, 475–488.

LESNAW, J. A. & REICHMANN, M. E. (1969). *Virology*, 39, 729–737.

LUISI, F. B., XU, X. W., OTWINOWSKI, Z., FREEDMAN, P. L., YAMAMOTO, R. K. & SIGLER, B. P. (1991). *Nature (London)*, 352, 497–505.

MATTHEWS, B. W. (1968). *J. Mol. Biol.* 33, 491–497.

MEULEWAETER, F., SEURINCK, J. & VAN EMMELO, J. (1990). *Virology*, 177, 699–709.

NONAKA, T., MITSUI, Y., IRIE, M. & NAKAMURA, T. K. (1991). *FEBS Lett.* 283, 207–209.

PAVLETICH, P. N. & PABO, O. C. (1991). *Science*, 252, 809–817.

SAEKI, K., UMEOKA, T., TAKAHASHI, Y., KUNISHIMA, N., FUKUYAMA, K. & MATSUBARA, H. (1994). In preparation.

SAKABE, N. (1983). *J. Appl. Cryst.* 16, 542–547

SIM, G. A. (1959). *Acta Cryst.* 12, 813–815

SIM, G. A. (1960). *Acta Cryst.* 13, 511–512.

TAKENAKA, A., OHKI, M. & SASADA, Y. (1980). *Bull. Chem. Soc. Jpn*, 53, 2724–2730.

TEN EYCK, L. F. (1973). *Acta Cryst.* A29, 183–191

TSUKIHARA, T., YOKOTA, Y., KOYAMA, T., FUKUYAMA, K. & MATSUBARA, H. (1990). *Acta Cryst.* B46, 855–860.

UYEMOTO, J. K. & GROGAN, R. G. (1969). *Virology*, 39, 79–89.

WANG, B. C. (1985). *Methods Enzymol.* 115, 90–112.