Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Ab initio crystal structure determination of spherical viruses that exhibit a centrosymmetric location in the unit cell

Junichiro Taka, Hisashi Naitow, Masato Yoshimura, Naoyuki Miyazaki, Atsushi Nakagawa and Tomitake Tsukihara*

Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

Correspondence e-mail: tsuki@protein.osaka-u.ac.jp

Received 9 February 2005
Accepted 19 May 2005

Ab initio phasing by non-crystallographic symmetry averaging coupled with solvent flattening has previously been used to determine the structure of canine parvovirus [CPV; Tsao *et al.* (1992), *Acta Cryst.* A**48**, 293–301; Chapman *et al.* (1992), *Acta Cryst.* A**48**, 301–312]. As CPV particles were located at general positions, initial phases were generated with a spherical shell deviating from the centre of symmetry. In many virus crystals, the viral particles are located at positions with a centre of symmetry in the unit cell. Thus, the initial phases calculated with a spherical shell model have a centre of symmetry. The inherent difficulty in structural determination was breaking the centre of symmetry in the initial phase. The centric nature of the initial phases of the rice dwarf virus crystal, however, was successfully broken at low resolution by iteration of the density-modification method described by Tsao *et al.* (1992) [Naitow *et al.* (1999), *Acta Cryst.* D**55**, 77–84]. In this study, *ab initio* phasing was tested for seven viruses ranging from 82 to 344 Å in radius. Although the crystal structures of the initial spherical shell models for each virus had a centre of symmetry, the centric natures of the initial phases were successfully broken by non-crystallographic symmetry averaging coupled with solvent flattening at a low resolution; these phases were then successfully extended to high-resolution measurements. A novel procedure of *ab initio* phasing for spherical virus crystals is proposed.

1. Introduction

Crystal structure analyses of the tomato bushy stunt virus (TBSV; Harrison *et al.*, 1978) and the southern bean mosaic virus (SBMV; Abad-Zapatero *et al.*, 1981), pioneering works in virus crystallography, were performed using the multiple isomorphous replacement method followed by non-crystallographic symmetry (NCS) averaging. More than 100 viral structures have been deposited in the Protein Data Bank (PDB; Berman *et al.*, 2000) since these pioneering works. The majority of crystal structure analyses of viruses have been performed using the molecular-replacement (MR) method (Rossmann & Blow, 1963) followed by refinement using NCS averaging. As the NCS averaging refinement for virus crystals has a large radius of convergence (Rossmann, 1990), we could extend the initial phases determined at low resolution by MR to high resolution, even when using an unrelated model (Valegård *et al.*, 1990). The feasibility of *ab initio* phase determination has been demonstrated for spherical viruses (Tsao *et al.*, 1992). The initial phases of canine parvovirus (CPV), calculated at 20 Å resolution using a spherical hollow-shell model of the virus, were successfully refined and extended to 13.5 Å resolution (Chapman *et al.*, 1992). As CPV particles are located at general positions exhibiting non-

Table 1
Crystal data of viruses.

Virus	PhMV†	RDV	STMV‡	SBMV§	TBSV¶	DYMV††	TNV‡‡
Unit-cell parameters							
a (Å)	289.64	770.0	174.27	334.3	383.2	348.55	338.0
b (Å)	287.72	795.0	191.77	334.3	—	—	—
c (Å)	295.23	814.0	202.50	757.5	—	—	—
α (°)	63.59	—	—	90.0	—	—	—
β (°)	61.39	—	—	90.0	—	—	—
γ (°)	62.94	—	—	120.0	—	—	—
Space group	$P1$	$I222$	$I222$	$R32$	$I23$	$P4_232$	$P4_232$
Z§§	1	2	2	3	2	2	2
Centres of viruses	(0, 0, 0)	(0, 0, 0), (0.5, 0.5, 0.5)	(0, 0, 0), (0.5, 0.5, 0.5)	(0, 0, 0), (2/3, 1/3, 1/3), (1/3, 2/3, 2/3)	(0, 0, 0), (0.5, 0.5, 0.5)	(0, 0, 0), (0.5, 0.5, 0.5)	(0, 0, 0), (0.5, 0.5, 0.5)
No. of reflections used¶¶	231025	614051	26859	68556	23543	29132	8515
Resolution range (Å)	100.0–5.0	200.0–6.0	100.0–4.0	100.0–5.0	100.0–6.0	100.0–5.0	100.0–7.0
Completeness	0.7491	0.9995	0.9258	0.9754	0.9996	0.9133	0.7777
No. of observed reflections	224463	614055	26801	68403	23543	28984	8515
Resolution range (Å)	18.0–5.0	228.7–6.0	40.0–4.0	40.0–5.0	100.0–6.0	33.0–5.0	100.0–7.0
PDB code	1e57	1uf2	1a34	4sbv	2tbv	1ddl	1tnv

† Krishna *et al.* (2001). ‡ Larson *et al.* (1998). § Silva & Rossmann (1985). ¶ Olson *et al.* (1983). †† Larson *et al.* (2000). ‡‡ Bando *et al.* (1994). §§ Z is the number of viral particles in the unit cell. ¶¶ Reflections with a larger spacing than the observed lowest resolution were generated by calculation from the atomic parameters given in the PDB.

centrosymmetry, the initial phases deviate from the centre of symmetry. In contrast, in many virus crystals the viral particles are located at positions with centres of symmetry within the unit cell. In these cases, the initial phases calculated with a spherical shell have a centre of symmetry. We expected difficulties in breaking the centre of symmetry of the initial phase for virus crystals in which an icosahedral twofold axis coincides with a crystallographic even-fold axis (Chapman *et al.*, 1992). In $R32$ crystals of SBMV, in which the crystallographic twofold and threefold axes coincided with the icosahedral symmetry axes of the viral particles, NCS averaging was never able to break the centre of symmetry introduced by the initial model (Johnson *et al.*, 1976).

Descriptions of the procedures of *ab initio* phasing of spherical viruses can be classified into three cases. (i) Viral particles are located at non-centric positions in the unit cell, as seen for CPV (Chapman *et al.*, 1992). (ii) The positions of viral particles exhibit a centre of symmetry within the unit cell; inversion of the crystal structure gives different NCSs from those of the original crystal structure. (iii) The positions of viral particles exhibit a centre of symmetry. Inversion of the crystal structure has the same NCSs as the original crystal structure, describing a crystal that contains a viral particle at the origin. A crystal in which the icosahedral twofold axis of the virus is parallel to the crystallographic even-fold axis is included in case (iii).

The initial phases of case (i) do not describe a centre of symmetry. Although those of case (ii) have a centre of symmetry, their centric nature is broken by one cycle of NCS averaging. In case (iii), however, NCS averaging could not break the centric nature of the initial phases, as was seen for SBMV (Johnson *et al.*, 1976). The *ab initio* phasing of rice dwarf virus (RDV), in which the crystallographic twofold axes coincided with the icosahedral twofold axes of the virus at the origin of the unit cell, belongs to case (iii). The centric nature of the initial phases, however, was successfully broken by iteration of the NCS averaging refinement (Naitow *et al.*, 1999). The centre of symmetry gradually disappeared after several cycles of refinement; the phases converged adequately after approximately 150 iterations at the initial low resolution. The non-centric phases of reflections that were affected by rounding-off errors during the computations served as triggers that break the centrosymmetric constraints. It remains to be investigated whether the centric nature of the initial phase can be broken for any other virus crystal. In this study, we investigated the *ab initio* structure-determination procedure for virus crystals in which the centres of the viral particles were located at the origin, as in case (iii).

Another difficulty with the *ab initio* method using the spherical shell model is the Babinet inversion (electron density is negative instead of positive), which leads to solutions in which the phases are incremented by π and to the production of negative electron density for the protein region (Tsao *et al.*, 1992). If the observed structure factors are absent at very low resolution, the Babinet inversion arises when the radius of the initial model includes a small error (Chapman *et al.*, 1992). In this study, the *ab initio* phasing procedures were evaluated based on the assumption that the low-resolution intensity data necessary for the phasing existed.

The *CCP*4 suite software for NCS averaging calculations (Collaborative Computational Project, Number 4, 1994) was used in place of the programs used in previous RDV phasing calculations, as *CCP*4 programs are free from space-group limitations.

2. Methods

2.1. Virus crystals for test calculation

The centres of viral particles within six independent crystals, determined from viral crystal structures previously deposited in the PDB, were uniquely located within their unit cells from

their unit-cell parameters, space groups and the approximate sizes of the viruses. In addition to RDV, these crystals were selected for test calculations. The *ab initio* phasing of these crystals was performed as described for case (iii) (crystal data are listed in Table 1). These examples can be categorized into four crystal systems: triclinic [physalis mottle virus (PhMV)], orthorhombic [(RDV and satellite tobacco mosaic virus (STMV)], trigonal (SBMV) and cubic [TBSV, TNV and desmodium yellow mottle virus (DYMV)]. With a radius of 344 Å, RDV was the largest virus, while STMV was the smallest, possessing a radius of 82 Å. TBSV possesses a protruding surface, which deviates significantly from a smooth spherical surface.

For this study, we acquired intensity data for RDV (Nakagawa *et al.*, 2003). The observed structure factors of PhMV, STMV, SBMV and DYMV were obtained from the PDB. The unobserved structure factors in low-resolution ranges were calculated from the atomic parameters provided in the PDB. Atoms within the capsids were not included in the calculations. Diffraction data with a low-resolution limit of approximately two-thirds of the radius of the viral particle were used for the calculation because low-resolution data prevent NCS refinement, effecting the Babinet opposite solution (Tsao *et al.*, 1992; Chapman *et al.*, 1992). As the observed structure factors of TBSV crystals had not been deposited in the PDB, we substituted the observed structure factors with those calculated from the provided atomic parameters.

2.2. Outline of the *ab initio* phasing

The entire procedure of *ab initio* phasing is summarized in Fig. 1. Briefly, icosahedral NCS matrices were initially determined by a rotation function based on the observed structure factors. We determined the most probable inner and outer radii of the initial spherical shell model using an *R*-factor search. We then generated nine models, each with a different sized shell around the most probable one. For each spherical shell model, two unique masks with differing inner radii were prepared for use in NCS averaging and solvent flattening. We performed phase refinements for the nine initial models, using

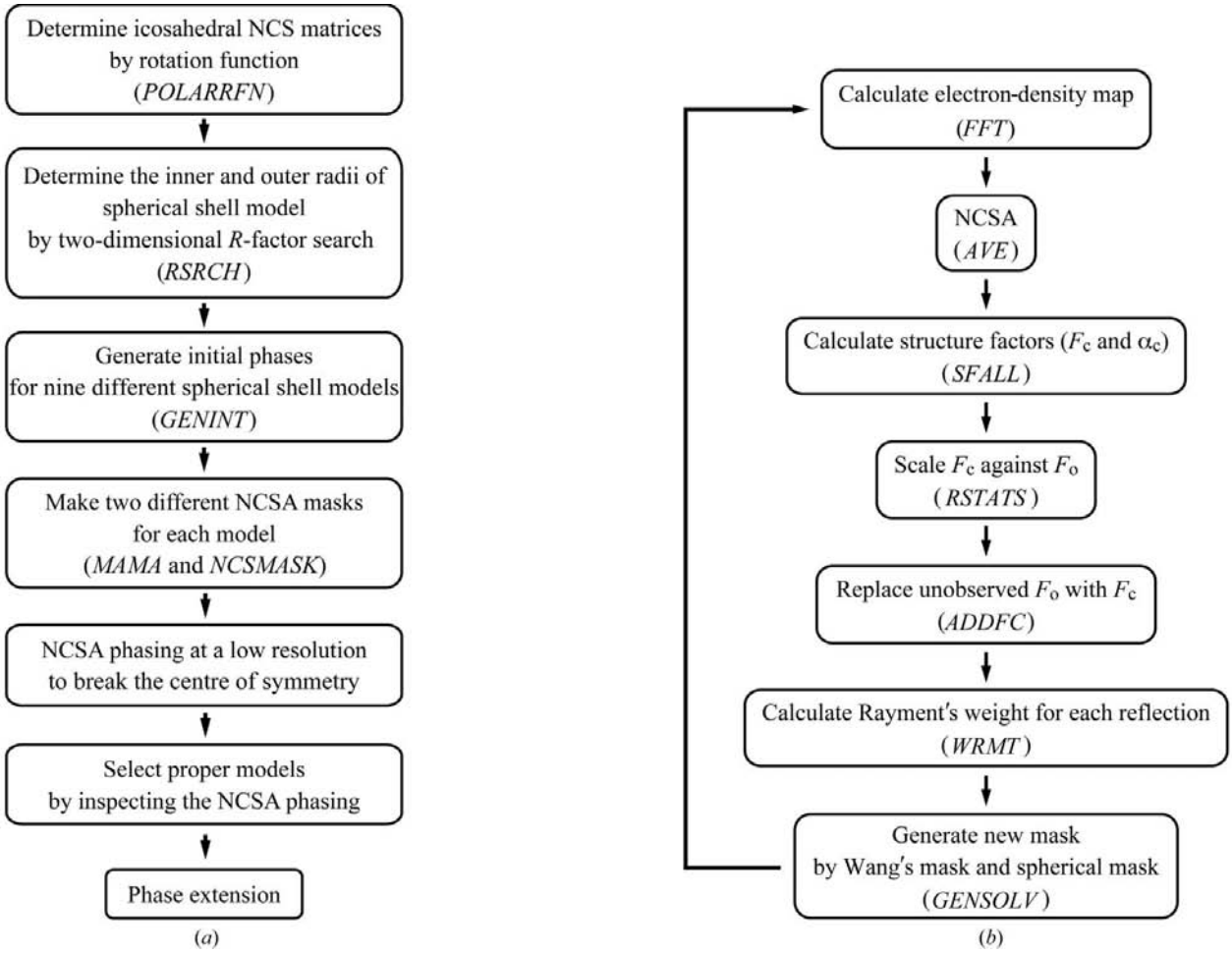

Figure 1
Flow charts of *ab initio* phasing by NCS averaging coupled with solvent flattening. The program names used for each calculation are given in parentheses. The programs *POLARRFN*, *NCSMASK*, *FFT*, *SFALL* and *RSTATS* are members of the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994), while the programs *MAMA* and *AVE* are part of the *RAVE* program package (Kleywegt *et al.*, 2001*a*,*b*). The *RSRCH*, *GENINT*, *ADDFC*, *WRMT* and *GENSOLV* programs were created in our laboratory. (*a*) Summary of the procedures from the determination of the viral orientation to phase expansion. (*b*) These procedures were applied to NCS averaging phasing at the initial low resolution and phase extension. The calculation using the *GENSOLV* program can be performed by running the *CCP*4 programs *SFALL*, *FFT* or *MAPMASK*.

Table 2
Resolution limits and number of reflections for the determination of the initial spherical shell parameters.

Virus	PhMV	RDV	STMV	SBMV	TBSV	DYMV	TNV
Radius† (Å)	144	344	82	159	166	151	147
Lower limit (Å)	100.0	200.0	100.0	100.0	100.0	100.0	100.0
Upper limit (Å)	18.1	50.0	12.7‡	23.7	24.0	21.8	21.1
Reflections							
Used	6436	1134	942	726	404	409	453
Observed	0	1134	884	573	0	261	453

† The radii of viruses were determined from the packing of spheres in the crystals. ‡ The diffraction data of STMV between 100.0 and 16.9 Å consisting of 409 reflections gave the same solution in the *R*-factor search as data between 100.0 and 12.7 Å.

two types of masks for each, by NCS averaging coupled with solvent flattening at the initial low resolutions. We then selected the refinements that successfully converged to break the centre of symmetry from the 18 sets of refinements. Phase extension was performed for the phase set selected by the initial refinement.

2.3. Positions and NCS matrices of viral particles used for the test calculations

As the unit cell with a *P*1 space group contained one particle in an asymmetric unit, the centre of the PhMV particle was fixed at the origin. The self-rotation function calculated for a PhMV crystal gave NCS parameters consistent with those given in the PDB. Thus, the NCS matrices given in the PDB were used for further calculations. By considering the number of particles in the unit cell, we determined that the centres of the viral particles in the six remaining crystals were uniquely located at the origin or equivalent positions in the unit cell; the NCS matrices for these crystals were determined by coinciding symmetries of the virus with crystallographic symmetries.

2.4. Determination of the outer and inner radii of a spherical shell for the initial model of a virus

The outer radius of a virus, referred to as r_p, was determined approximately from the packing of viral particles within the crystal. For a series of values for the inner (r_i) and outer (r_o) radii of the spherical shell model of a virus, we calculated the *R* factors according to the equation: $R = \sum(F_o - F_c)/\sum F_o$, in which F_o is an observed structure factor and F_c is a structure factor calculated from the electron-density distribution in the unit cell. Using a step size of 2 Å, the outer radii (r_o) that were examined in the *R*-factor search ranged from r_p + 20 Å to r_p − 30 Å, while the inner values (r_i) ranged from r_o − 20 Å to $r_p/3$. The inner sphere of radius r_i, which was filled with nucleic acid, had an electron density of ρ_{NA}. According to Chapman *et al.* (1992), the relative density of the nucleic acid, $\rho'_{NA} = \rho_{NA}/\rho_P$, where ρ_P is the electron density of the protein region, was fixed at 1.4 for *R*-factor calculations. ρ'_{NA} for TBSV, however, was set to 0.0 because the structure factors calculated without nucleic acids were substituted for those observed for the TBSV crystal. The ρ'_{NA} values for PhMV and DYMV were also set to 0.0, as most of the structure factors of the crystals in the low-resolution range were estimated from the atomic parameters of the protein capsids. The lower and upper resolution limits of the reflections for determination of the spherical shell parameters were approximately two-thirds and one-seventh of the viral radius, respectively, approximately determined from the packing of the viral particles within the crystal. By setting the higher resolution for each virus to one-seventh of the viral radius, the data-to-parameter ratio of any virus was sufficiently larger than the necessary value of 200 given by Chapman *et al.* (1992).

2.5. Assessment of the correctness of refined phases

Two criteria were calculated to monitor the progress of the phase refinement. The first was the conventional crystallographic *R* factor described above. The second was a correlation coefficient, C1, describing the relationship between the electron density obtained by refinement and the theoretical electron density calculated from the atomic parameters. This value was estimated by the equation

$$\mathrm{C1} = \frac{\sum(\rho_o - \langle\rho_o\rangle)(\rho_T - \langle\rho_T\rangle)}{[\sum(\rho_o - \langle\rho_o\rangle)^2]^{1/2}[\sum(\rho_T - \langle\rho_T\rangle)^2]^{1/2}},$$

in which ρ_o and ρ_T are the refined and theoretical electron densities, respectively, and $\langle\rho_o\rangle$ and $\langle\rho_T\rangle$ are the average ρ values. As it is impossible to determine ρ_T in advance, this value can only be estimated for the test calculation.

2.6. Initial phase refinement to break the centre of symmetry

Initial phase refinements were performed for the nine spherical shell models using the same resolution limits as used for the *R*-factor search (Table 2). Nine models were generated from all combinations of three values each for both the inner and outer radii. The three values chosen comprised the result of the *R*-factor search and two additional values plus and minus 2 Å. Two classes of masks were prepared for each model; the outer radius of both masks was set to 150% of the outer radius determined by the *R*-factor search. The inner radius of the first mask was the same as that described in the initial spherical shell model, while the second was prepared using an inner radius 10 Å less than that of the first one to account for the unevenness of the internal surface of the viral particle. The masks were then truncated to remove sections overlapping with those regions related by either crystallographic symmetries or NCSs. In each refinement cycle, the electron density outside of the mask was set to zero. The structure factors calculated from the electron-density distributions were substituted for unobserved structure factors. The observed structure factors used in the calculation of electron density were weighted according to the following equation from Rayment (1983),

$$w = \exp[-(\big||F_o| - |F_c|\big|/|F_o|)],$$

where F_o is the observed structure factor and F_c is the calculated structure factor from the refined electron-density distribution.

Table 3
The most probable inner and outer radii of viruses determined from the *R*-factor maps and the statistics of the *R*-factor search.

Virus	PhMV	RDV	STMV†	SBMV	TBSV	DYMV	TNV
Inner radius (Å)	112	254	40	88	116	100	106
Outer radius (Å)	140	340	84	144	160	154	142
Inner atom‡ (Å)	104	251	53	105	109	104	108
Range of *R*-factor regions searched							
r_i (Å)	90–120	240–270	20–50	70–100	100–130	90–120	90–120
r_o (Å)	130–160	320–350	70–100	130–160	140–170	130–160	120–150
R factor§							
Minimum	0.649	0.543	0.584	0.507	0.628	0.645	0.665
Second	—	—	—	—	0.649	0.664	—
Average	0.691	0.582	0.705	0.631	0.685	0.722	0.745
E.s.d.	0.017	0.028	0.059	0.057	0.031	0.040	0.036

† When the upper limit of resolution for STMV was set to 16.9 Å, the minimum, second and e.s.d. values of the *R*-factor search were 0.487, 0.665 and 0.071, respectively; no second minimum existed. ‡ The shortest distance between the protein atom and the centre of the virus particle was calculated from the atomic coordinates given in the PDB. § Four different *R* factors are given. These are the minimum and the second minimum values and the averaged values over the whole range and the estimated standard deviation (e.s.d.).

18 sets of phase refinements were performed at the initial low resolution until we observed convergence of the phases. More than 100 iteration cycles were required before phase convergence was achieved. As the C1 values of five viruses were low despite good *R* values, the electron-density distribution was converted to its inverse, resulting in a higher C1 value. Although C1 is not available in a real structure determination, the correctness of the hands can be decided from the electron-density distribution of an α-helix at high resolution. The behaviour of the NCS averaging at initial low resolutions was inspected for each virus by comparing the individual phase at each cycle with the corresponding phase refined at the highest resolution. The phase refinement of RDV proceeded as described previously (Naitow *et al.*, 1999). NCS averaging of the phase improvement of other viruses exhibited similar behaviour to that of RDV. Briefly, keeping the centre of symmetry, the phases approached the target values until approximately the seventh cycle of NCS averaging. After approximately 20 cycles of NCS averaging, the phases deviated from the centre of symmetry and began to approach their target angles. The phase improvement became very small after about the 60th cycle and converged at approximately the 135th cycle.

2.7. Phase extension

The models with the lowest *R* values were selected from the nine trials of the initial phasing with the larger and smaller inner radii for further phase extension by repeating the NCS averaging cycles. The mask used in the phase extension was derived from combination of the spherical shell mask used in the initial phase refinement and a mask generated by the method of Wang (1985). The *DM* program (Cowtan, 1994) from the *CCP*4 suite was modified to generate Wang's mask. NCS averaging was repeated in the same resolution range until the phases converged. After achieving phase convergence, the upper limit of resolution was extended by one unit of the minimum reciprocal cell constant in a^*, b^* and c^*. In this test calculation, the highest resolution limits of the phase extension for the majority of virus crystals were set to 6 Å, while those for RDV and STMV were set to 10 and 4.3 Å, respectively.

3. Results

3.1. The inner and outer radii of the spherical shell model

The spherical shell parameters for seven viruses were determined from the *R*-factor maps. The outer radii of the viral particles were restricted from the unit-cell parameters and packing conditions of the crystals. For example, the RDV crystal, which possesses unit-cell parameters a = 770.0, b = 795.0, c = 814.0 Å, contained two viral particles at the origin and the body-centred position in the unit cell. As the closest distance between two virus centres was 687 Å, the outer radius (r_o) was estimated to be approximately 344 Å. *R* factors were calculated for a range of outer radii between 315 and 364 Å and inner radii between 121 and 344 Å. In the map, the most probable values for the outer and inner radii were determined to be 340 and 254 Å, respectively. The *R*-factor search yielded a global minimum. Starting from the lowest peak, the NCS refinement converged well, while the second lowest peak did not result in a successful refinement. The statistics of the *R*-factor search are detailed in Table 3.

3.2. Phase refinement at the initial low resolution

As all crystals used in these test calculations contained viral particles at special positions within the unit cell, the initial phases calculated using a spherical shell model exhibited a centre of symmetry. When the inner volume of the capsid was set to a constant value of ρ_{NA}, the iterative NCS averaging provided abnormally thin electron-density distributions in the protein shell in comparison to cases calculated with $\rho_{NA} = 0$. After more than 100 iterations of NCS averaging using both masks, at least one of the nine trials converged to a reasonable electron-density distribution, with the exception of STMV. Convergence was judged on the criteria of C1 values and successful breakage of the centric nature for each virus crystal. Refinement calculations for STMV using the mask with the small inner radius did not successfully converge, as determined by both *R* and C1 values. The statistics of the phase refinements at the initial resolution are summarized in Table 4. The refinement conditions resulting in the smallest *R* value among the nine refinements gave the largest or second largest C1 value, indicating that the *R* factor is a useful criterion upon

Table 4
Phase refinements at the initial resolution.

(*a*) Results of calculations using a mask with the same inner radius as that of the spherical shell model. (*b*) Results of calculations using a mask with an inner radius that is smaller than the most probable value by 10 Å.

	(*a*)		(*b*)	
(r_o, r_i)†	C1‡	*R*‡	C1‡	*R*‡
PhMV				
(−2.0, −2.0)	0.958	0.322	0.980	0.230
(−2.0, 0.0)	0.934	0.375	0.980	0.230
(−2.0, 2.0)	0.512	0.464	0.980	0.230
(0.0, −2.0)	0.958	0.322§	0.980	0.230
(0.0, 0.0)	0.935	0.375	0.980	0.230§
(0.0, 2.0)	0.511	0.464	0.980	0.230
(2.0, −2.0)	0.958	0.322	0.980	0.230
(2.0, 0.0)	0.934	0.375	0.980	0.230
(2.0, 2.0)	0.513	0.464	0.980	0.229
RDV				
(−2.0, −2.0)	0.763	0.233	0.959	0.195
(−2.0, 0.0)	0.768	0.235	0.969	0.178
(−2.0, 2.0)	0.776	0.237	0.969	0.178
(0.0, −2.0)	0.965	0.191§	0.963	0.192
(0.0, 0.0)	0.960	0.195	0.958	0.195
(0.0, 2.0)	0.775	0.236	0.968	0.177§
(2.0, −2.0)	0.965	0.191	0.952	0.193
(2.0, 0.0)	0.960	0.195	0.952	0.193
(2.0, 2.0)	0.959	0.198	0.949	0.190
STMV				
(−2.0, −2.0)	0.906	0.315	0.247	0.590
(−2.0, 0.0)	0.910	0.302	0.247	0.590
(−2.0, 2.0)	0.912	0.291	0.247	0.590
(0.0, −2.0)	0.906	0.315	0.146	0.580
(0.0, 0.0)	0.909	0.302	0.247	0.590
(0.0, 2.0)	0.911	0.289§	0.247	0.590
(2.0, −2.0)	0.906	0.315	0.147	0.580
(2.0, 0.0)	0.909	0.302	0.147	0.581
(2.0, 2.0)	0.911	0.289	0.146	0.580
SBMV				
(−2.0, −2.0)	0.861	0.214§	0.842	0.204§
(−2.0, 0.0)	0.862	0.215	0.839	0.207
(−2.0, 2.0)	0.858	0.217	0.848	0.205
(0.0, −2.0)	0.854	0.216	0.842	0.205
(0.0, 0.0)	0.856	0.216	0.834	0.208
(0.0, 2.0)	0.859	0.215	0.826	0.207
(2.0, −2.0)	0.844	0.222	0.835	0.204
(2.0, 0.0)	0.845	0.223	0.832	0.206
(2.0, 2.0)	0.845	0.224	0.827	0.211
TBSV				
(−2.0, −2.0)	0.822	0.306	0.437	0.350
(−2.0, 0.0)	0.897	0.278	0.554	0.321
(−2.0, 2.0)	0.780	0.339	0.587	0.327
(0.0, −2.0)	0.369	0.375	0.438	0.341
(0.0, 0.0)	0.791	0.339	0.426	0.349
(0.0, 2.0)	0.780	0.339	0.425	0.347
(2.0, −2.0)	0.455	0.373	0.545	0.330
(2.0, 0.0)	0.898	0.270	0.746	0.299
(2.0, 2.0)	0.933	0.233§	0.912	0.229§
DYMV				
(−2.0, −2.0)	0.964	0.160	0.963	0.164
(−2.0, 0.0)	0.970	0.144	0.963	0.164§
(−2.0, 2.0)	0.968	0.148	0.963	0.164
(0.0, −2.0)	0.964	0.158	0.629	0.542
(0.0, 0.0)	0.970	0.144§	0.629	0.542
(0.0, 2.0)	0.968	0.148	0.629	0.542
(2.0, −2.0)	0.964	0.160	0.641	0.552
(2.0, 0.0)	0.969	0.145	0.640	0.556
(2.0, 2.0)	0.968	0.156	0.640	0.556
TNV				
(−2.0, −2.0)	0.850	0.275	0.874	0.235
(−2.0, 0.0)	0.882	0.258§	0.874	0.235§
(−2.0, 2.0)	0.874	0.270	0.874	0.235
(0.0, −2.0)	0.866	0.262	0.816	0.272
(0.0, 0.0)	0.882	0.259	0.816	0.273
(0.0, 2.0)	0.845	0.287	0.816	0.273
(2.0, −2.0)	0.859	0.273	0.818	0.273
(2.0, 0.0)	0.854	0.278	0.818	0.273
(2.0, 2.0)	0.847	0.261	0.782	0.306

† r_i and r_o are the differences from the most probable inner and outer radii of the initial spherical shell model, respectively. ‡ C1 and *R* are defined in the text. § The model giving the smallest *R* factor among the nine trials was used to start further phase extension.

Table 5
The results of phase extension.

(*a*) shows the result of the calculation starting from the phase set giving the best result in (*a*) in Table 4, while (*b*) shows the result of the calculations starting from those values shown in (*b*) in Table 4.

	(*a*)		(*b*)		
Viruses	C1	*R*	C1	*R*	Resolution (Å)
PhMV	0.872	0.329	0.866	0.318	6.0
RDV	0.959	0.191	0.959	0.181	10.0
STMV	0.853	0.178	—	—	4.3
SBMV	0.731	0.146	0.830	0.128	5.9
TBSV	0.537	0.231	0.993	0.083	6.0
DYMV	0.904	0.089	0.905	0.087	5.5
TNV	0.846	0.231	0.842	0.196	7.0

which to evaluate the correctness of the refined electron density. Models with *R* factors within 0.01 of the minimum *R* factor among the nine models did not result in any differences in further refinement from the model with the minimum value. The respective best initial models for two different masks resulted in nearly identical convergence in further refinement, with the exception of STMV.

3.3. Phase extension

The two phase sets that converged to the lowest *R* values in the initial refinement step using masks with the larger and smaller inner radii were used as the starting phase sets for further phase extension (Table 5). The phase extension of STMV, however, could only be performed with the mask using the larger inner radius. In contrast, the phase extension of TBSV using the mask with the smaller inner radius converged significantly better than that using the mask with the larger inner radius. The electron-density maps for PhMV and TBSV (Fig. 2) demonstrate that the C^{α} traces for both viruses superposed well with their electron-density maps. The C^{α} wire models for the remaining five viruses were adequately located in the electron-density map at a resolution of approximately 6 Å. The helical rods of RDV were identified in the 10 Å resolution electron-density map, while the β-barrel structure of STMV could be observed at 4.3 Å resolution. Refinements made using masks with both the larger and smaller inner radii converged well to similar structures for each virus, with the

exception of STMV, SBMV and TBSV. For SBMV and TBSV, the phase extension using the smaller inner radius mask successfully converged to a structure consistent with the ideal structure, while phase extension with the mask using the larger inner radius fell into a defective structure. Phase extension of TNV using the mask with a smaller inner radius resulted in a smaller R value than that generated using the larger inner radius mask by 0.025, despite similar C1 values for both calculations. All of the virus crystals gave C1 values greater than 0.830. With the exception of PhMV (R = 0.318), the R factors were lower than 0.200. To confirm the quality of the observed structure factors of PhMV, we performed test calculations using the structure factors derived from the PDB model to obtain values for C1 and R of 0.962 and 0.200, respectively. The high R value of 0.318 for PhMV may be caused by the low symmetry of the $P1$ space group and the low quality of the observed structure factors used for calculation.

4. Discussion

An R-factor search to determine the outer and the inner radii of each virus was successfully performed using diffraction data with resolutions ranging from two thirds (or 100.0 Å if the radius was less than 100.0 Å) and one seventh of the virus radius. The upper limit of resolution, which was as high as 12.7 Å for STMV, may be too high to obtain a reasonable R-factor map if the virus structure deviates significantly from the spherical shell model. The upper limit of resolution can be set to approximately 20 Å, as long as the data-to-parameter ratio remains larger than 200. R factors generated from reasonable outer and inner radii estimations in the initial density-modification refinements at a low resolution ranged from 0.144 to 0.289, while those generated from the proper phase extensions varied from 0.083 to 0.318. It was difficult to determine the threshold values for the R factor that would indicate adequate refinement for any virus, as the R factor is highly dependent on the accuracy of the intensity data. Regardless, as refinements with a smaller R factors between different trials with the same intensity data resulted in increasingly adequate structures, the R factor is a useful criterion to justify the progress of the density-modification refinement.

All trial conditions for STMV failed in initial refinement at low resolution when the mask with the smaller inner radius was used. The inner radius used was 30 Å, which is 75% of the inner radius of the spherical shell model and 57% of the distance between the innermost atom and the centre of virus particle. This inaccuracy resulted in the amplification of systematic errors in electron density arising within the shell during iterations of the phase extension. The most probable inner radius of TBSV was 116 Å, while the distance between the innermost atom and the centre of virus particle was 108 Å. When the mask with an inner radius of 116 Å was used, refinement at high resolution resulted in a high R factor, as the mask truncated the TBSV structure. These results indicated that the size of the inner radius of the mask is an important parameter that is required for the refinement to converge to a proper structure. While TBSV has large protrusions, these structures were successfully reproduced in the electron-density map obtained by phase extension at a high resolution by beginning with the spherical shell model at a low resolution. The large outer radius of the mask, set to 150% of the outer radius of the spherical shell model, significantly affected the entire structure.

The phase extensions for RDV, SBMV, TBSV, DYMV and TNV succeeded in approaching the ideal structures, regardless of the mask used. The mask with smaller inner radius gave a better convergence of the phase extension for PhMV than that using the larger inner radii. This is likely to result from the truncation of atoms by the shell with larger radius. For SMTV, the mask

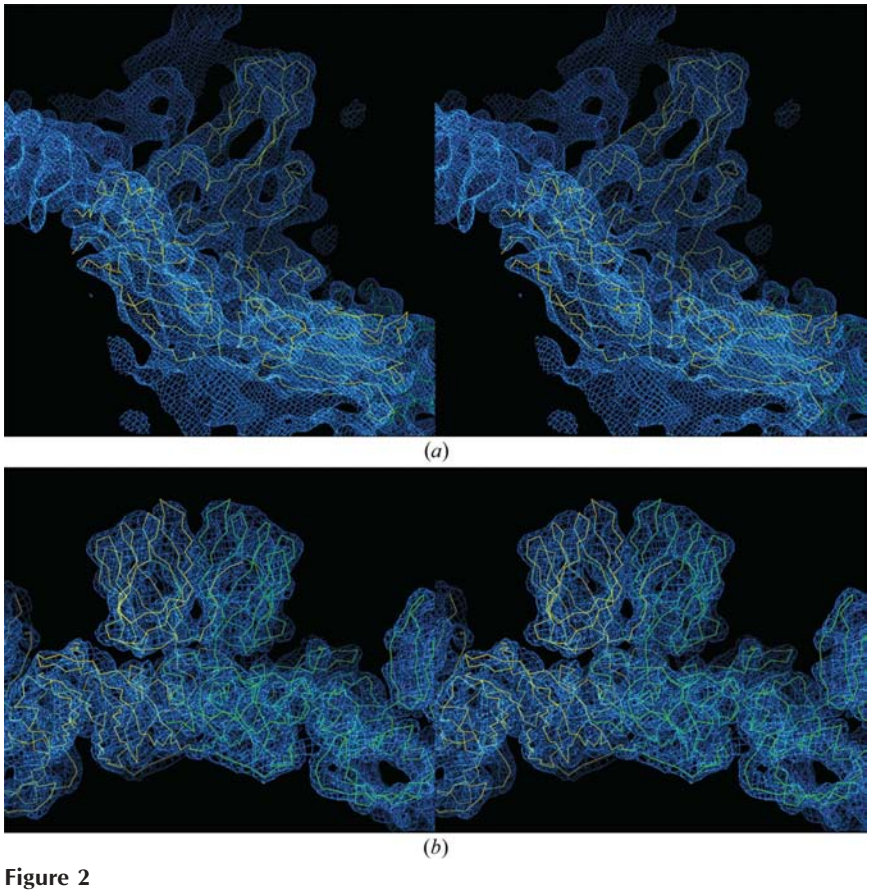

Figure 2
The electron-density maps calculated from the phases obtained by phase extension were superimposed on C^{α} trace models from the PDB. These values were calculated at a 6.0 Å resolution; their contours were drawn at the 1.8σ level using the program *O* (Jones *et al.*, 1991) for (*a*) PhMV and (*b*) TBSV.

with smaller inner radius had too large a vacant space to give a reasonable solution in the phase extension.

5. Conclusion

The crystal structure, consisting of spherical viral particles at the origin and equivalent positions, can be determined by the *ab initio* method. Intensity data with spacing from two thirds of the viral radius to a significantly higher resolution were required. The orientation of the viral particle was determined by a rotation function or by the packing in the crystal. The outer and inner radii of the virus were determined by an *R*-factor search using reflections whose resolutions ranged from two thirds to one seventh of the approximate outer radius of the virus. A mask using outer and inner radii of 1.5 and 1.0 times the spherical shell was prepared for NCS averaging coupled with solvent flattening. Nine spherical shell models using values surrounding the most probable model determined by the *R*-factor search were refined by NCS averaging using mask conditions at the same resolution as the *R*-factor search to break the centre of symmetry. Beginning with the phase set of the smallest *R* value for these nine trials, the phases were extended to the highest resolution. The same procedures as those used for the NCS averaging refinements for the nine models and the phase extensions were performed using another mask whose inner radius was smaller than the previous mask by 5–10 Å. The soundness of the obtained structures was evaluated by the *R* factors and the structural features of the electron-density map.

We greatly appreciate the precise discussions and suggestions of Professor Michael Rossmann of Purdue University. This work was supported in part by Grants-in-Aid for 'Research for the Future' Program (JSPS-RFTF96L00503 to TT) from the Japan Society of the Promotion of Science, a research grant from the Japan Biological Informatics Consortium entrusted from the New Energy and Industrial Technology Organization (to TT), Grants-in-Aid for 21st Century centre of Excellence Program from the Ministry of Education, Culture, Sports, Science and Technology, Japan (to TT) and Grants-in-Aids for Scientific Research in Priority Area (16087206 to TT) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

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.* B**37**, 2002–2018.
Bando, M., Morimoto, Y., Sato, T., Tsukihara, T., Yokota, Y., Fukuyama, K. & Matsubara, H. (1994). *Acta Cryst.* D**50**, 878–883.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Chapman, M. S., Tsao, J. & Rossmann, M. G. (1992). *Acta Cryst.* A**48**, 301–312.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. (1978). *Nature (London)*, **276**, 368–373.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Johnson, J. E., Akimoto, T., Suck, D., Rayment, I. & Rossmann, M. G. (1976). *Virology*, **75**, 394–400.
Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M. & Jones, T. A. (2001*a*). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 353–356. Dordrecht: Kluwer Academic Publishers.
Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M. & Jones, T. A. (2001*b*). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 366–367. Dordrecht: Kluwer Academic Publishers.
Krishna, S. S., Sastri, M., Savithri, H. S. & Murthy, M. R. N. (2001). *J. Mol. Biol.* **307**, 1035–1047.
Larson, S. B., Day, J., Canady, M. A., Greenwood, A. & McPherson, A. (2000). *J. Mol. Biol.* **301**, 625–642.
Larson, S. B., Day, J., Greenwood, A. & McPherson, A. (1998). *J. Mol. Biol.* **277**, 37–59.
Naitow, H., Morimoto, Y., Mizuno, H., Kano, H., Omura, T., Koizumi, M. & Tsukihara, T. (1999). *Acta Cryst.* D**55**, 77–84.
Nakagawa, A., Miyazaki, N., Taka, J., Naitow, H., Ogawa, A., Fujimoto, Z., Mizuno, H., Higashi, T., Watanabe, Y., Omura, T., Cheng, R. H. & Tsukihara, T. (2003). *Structure*, **11**, 1227–1238.
Olson, A. J., Bricogne, G. & Harrison, S. C. (1983). *J. Mol. Biol.* **171**, 61–93.
Rayment, I. (1983). *Acta Cryst.* A**39**, 102–116.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Silva, A. M. & Rossmann, M. G. (1985). *Acta Cryst.* B**41**, 147–157.
Tsao, J., Chapman, M. S. & Rossmann, M. G. (1992). *Acta Cryst.* A**48**, 293–301.
Valegard, K., Liljas, L., Fridborg, K. & Unge, T. (1990). *Nature (London)*, **345**, 36–41.
Wang, B.-C. (1985). *Methods Enzymol.* **114**, 90–112.