Acta Cryst. (1998). D**54**, 1261–1272

Structure Determination of Echovirus 1

D. J. Filman,[a] M. W. Wien,[b]† J. A. Cunningham,[c] J. M. Bergelson[c]‡ and J. M. Hogle[a,b]*

[a]*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA,* [b]*Committee on Higher Degrees in Biophysics, Harvard University, Cambridge, MA 02138, USA, and* [c]*Dana Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA. E-mail: hogle@hogles.med.harvard.edu*

(*Received 13 November 1997; accepted 13 February 1998*)

Abstract

The atomic structure of echovirus 1 (a member of the enterovirus genus of the picornavirus family) has been determined using cryo-crystallography and refined to 3.55 Å resolution. Echovirus 1 crystallizes in space group $P22_12_1$ with a = 352.45, b = 472.15 and c = 483.20 Å. The crystals contain one full virus particle in the asymmetric unit allowing for 60-fold noncrystallographic symmetry averaging. The diffraction pattern shows strong pseudo-B-centering with reflections with $h + l = 2n + 1$ being systematically weak or absent below about 6 Å resolution. The size of the unit cell and presence of pseudo-B-centering placed strong constraints on the allowed packing of the icosahedral particle in the crystal lattice. These constraints greatly facilitated the determination of the orientation and position of the virus by reducing the dimensionality of the search, but interactions between the crystallographic and noncrystallographic symmetries rendered the choice of space group ambiguous until very late in the structure determination. This structure determination provides a striking example of the power of packing analysis in molecular replacement and illustrates how subtle interactions between crystallographic and noncrystallographic symmetries can be resolved.

1. Introduction

Echovirus 1 (EV1) is a member of the well characterized picornavirus family. The virion consists of a single-stranded positive-sense RNA genome surrounded by a non-enveloped protein shell (Rueckert, 1996). A single open-reading-frame in the RNA genome is translated into a polyprotein which contains both structural and non-structural proteins. These proteins are cleaved from the polyprotein by virally encoded proteinases. The picornaviruses encode four structural proteins: three major capsid proteins, VP1, VP2 and VP3, and a small protein, VP4, which is located on the interior of the capsid protein shell. One copy of each of the proteins fold together to form the major building block of the capsid, termed the protomer. 60 copies of the protomer pack together with icosahedral symmetry $T = 1$ ($P = 3$) to form the viral shell.

Despite significant differences in the pathogenesis and amino-acid sequences of the picornaviruses, these viruses share a strikingly related architecture. The major capsid proteins, VP1, VP2 and VP3, share a similar eight-stranded β-barrel fold. Structural differences are primarily found in the loops that connect the strands, many of which contain antigenic sites, and in the extensions at the amino and carboxyl termini of each of the proteins. The amino-terminal extensions form an elaborate network on the interior of the capsid which is believed to contribute to the stability of the virion (Basavappa *et al.*, 1994; Hogle *et al.*, 1985).

EV1 is a member of the enterovirus genus. In addition to the echoviruses (32 serotypes), this genus includes the polioviruses (three serotypes), the group A (23 serotypes) and group B (six serotypes) coxsackieviruses and other human enteroviruses (at least four serotypes) (Melnick, 1990). This subgroup is distinguished by stability at low pH, which allows these viruses to infect the host *via* the acidic environment of the gut. The high-resolution structures of several enteroviruses are known, including the structures of all three serotypes of poliovirus (Filman *et al.*, 1989; Hogle *et al.*, 1985; Lentz *et al.*, 1997) and the structure of coxsackievirus B3 (Muckelbauer, Kremer, Minor, Diana *et al.*, 1995). EV1 shares about 75% sequence similarity with coxsackievirus B3 and about 50% sequence similarity with each of the poliovirus serotypes.

Clinically, echoviruses are frequently responsible for febrile illness and viral meningitis. Though many echovirus infections are apparently asymptomatic in adults, they can cause fatal disseminated infections in infants (Cherry, 1983; Melnick, 1996; Modlin, 1988). As several anti-echovirus drugs are currently under development, better diagnosis and understanding of this virus could improve the prognosis for infected infants.

In this manuscript we describe the methods used to determine the structure of EV1 at 3.55 Å resolution.

† Present address: Laboratoire d'Enzymologie et Biochimie Structurales, CNRS-UPR 9063, 1 Avenue de la Terrasse, Bâtiment 34, 91198 Gif-sur-Yvette CEDEX, France.

‡ Present address: Abramson 1202E, Children's Hospital of Philadelphia, 34th Street and Civic Center Boulevard, Philadelphia, USA.

ISSN 0907-4449

The virus was crystallized in the orthorhombic space group $P22_12_1$ with a = 352.45, b = 472.15 and c = 483.20 Å, and the crystals contain a complete particle in the asymmetric unit. The structure was solved by molecular replacement. Standard multidimensional molecular-replacement calculations for virus crystals containing a complete particle in the asymmetric unit can be very demanding computationally. The large size of the virus particles magnifies the effects of small rotations and thus requires an uncommonly fine sampling of the angular range in rotation searches. Moreover, interactions of crystallographic symmetry with the extensive noncrystallographic symmetry frequently result in the generation of pseudosymmetry in the diffraction data from virus crystals. This is the case with EV1 where the diffraction data exhibited approximate *B* centering extending beyond 6 Å resolution. The presence of pseudosymmetry results in extensive overlap between rotation-function peaks in standard intensity-based calculations. These closely spaced peaks arise from splitting of what would be single peaks in a truly centered lattice, owing to the small deviations in the angular orientation of the two particles related by the approximate centering operation. Resolution of the overlapped peaks, which is critical to determining the orientation of the virus particles, requires the use of highly accurate high-resolution data. The combination of high resolution and fine grids geometrically expand the computer memory and CPU requirements.

Paradoxically, the presence of pseudosymmetry also places significant constraints on the possible orientation and position of the particle. If appropriately exploited, these constraints can be used to reduce the dimensionality of the initial rotation and translation searches and thereby reduce the computational cost dramatically. Thus, in the EV1 structure determination, an analysis of the constraints imposed by the pseudo-*B*-centering made it clear that a short series of one-dimensional *R*-factor searches would yield a good approximate answer at low resolution. After that, highly localized multi-dimensional searches could be used to account for the deviations from perfect centering, as evident at higher resolution. Unfortunately, in the EV1 structure determination, the interactions between the noncrystallographic and crystallographic symmetry also precluded an identification of the space group until the final stages of the structure solution. Indeed, the structure was first solved in the 'incorrect' space group. The resulting maps were readily interpretable and consistent with the amino-acid sequence of EV1. The error was manifested only in nearly random agreement between calculated and observed amplitudes for one parity group of reflections ($h + l$ odd). This problem was easy to correct, as only one other choice of primitive orthorhombic space group was consistent with the observed packing, pseudo-centering and particle rotation and was capable of resolving the problem with the $h + l$ odd terms. Several of the virus crystal structures solved in recent years exhibit higher order pseudo-symmetries (Muckelbauer, Kremer, Minor, Tong *et al.*, 1995; Zlotnick *et al.*, 1997) suggesting that the use of pseudo-centering to limit the size of the molecular-replacement problem could be applied generally.

2. Methods and materials

2.1. *Virus propagation, purification and crystallization*

EV1 was propagated in HeLa cell monolayers. The virus seed stock used for this crystallographic study is one passage away from the virus used to determine the sequence of this strain (Genbank accession number AF029859). The virus was overproduced in expanded-surface (1700 cm^2) roller bottles infected at a multiplicity of infection of five plaque-forming units per cell (assuming 2×10^8 cells per roller bottle). The virus was allowed to attach in a low volume of medium (about 10 ml per bottle) after which additional medium (30–40 ml per bottle) was added and the infection allowed to proceed for 10 h (all at 310 K). The infected cells were harvested with 10 m*M* Tris, 10 m*M* NaCl, 1.5 m*M* $MgCl_2$, pH 6.3 (RSB), 1% Nonident P40 (NP40) and subsequently stored at 103 K.

The virus was purified using methods similar to the protocols previously described for poliovirus but with several modifications. Cell pellets were thawed and cell debris was pelleted by centrifugation for 10 min at 1000*g*. The supernatant was decanted and brought to a final concentration of 0.5% sodium dodecyl sulfate (SDS). The virus was pelleted by centrifugation for 3 h at 140000*g* through a 30% sucrose cushion. The pellet was overlaid with 0.2 ml 100 m*M* Tris–HCl, 100 m*M* NaCl, 100 m*M* EDTA, pH 7.4 (TNE), with 0.5% SDS and allowed to stand for 10 min. Pelleted virus was resuspended by pipetting and purified by CsCl density-gradient centrifugation (1.33 g ml^{-1} CsCl) in quickseal (Beckman) tubes. The virus band, located about one third from the top of the gradient, was removed and dialyzed against phosphate-buffered saline (PBS). Virus concentration was assessed by measuring the OD_{260} assuming an extinction coefficient of 7.7. The ratio of the OD_{260} to the OD_{280} was 1.85. This protocol yields about 0.1 mg of purified virus per roller bottle.

The virus was concentrated to about 10 mg ml^{-1} by pelleting through a cushion of 30% sucrose, 1 *M* NaCl in PMC7 buffer (10 m*M* PIPES, 5 m*M* $MgCl_2$, 1 m*M* $CaCl_2$, pH 7.0). EV1 was crystallized by microdialysis against 10 m*M* PIPES, 25 m*M* $CaCl_2$, 25 m*M* $MgCl_2$, 2.5% PEG 400 (Fluka) pH 6.0 at 277 K. Although these conditions most often yield two-dimensional plates, occasionally three-dimensional crystals also appear.

Table 1. *Data-collection and refinement statistics*

Data collection		
Crystals	1	
Temperature (K)	103	
Oscillation range (°)	0.2	
Frames	567	
Space group	$P22_12_1$	
Unit cell (Å)	a = 352.45, b = 472.15, c = 483.20	
Reflections observed	3522138	
Unique reflections	947283	
Multiplicity	3.7	
	Overall	Outer shell
Resolution range (Å)	50–3.55	3.61–3.55
Completeness	97.7%	81.3
R_{merge}	0.136	0.249

	Overall	Outer shell
Resolution range (Å)	50–3.55	3.66–3.55
R (model-based)	0.263	0.298
R (NCS)	0.258	0.293
Atomic model†		
Unique non-H atoms	6575	
Protein	6528	
Myristoyl (covalent)	15	
Palmitate (non-covalent)	18	
Solvent plus ions	14	
R.m.s. deviation from idealized standards		
Bond lengths (Å)	0.012	
Bond angles (°)	2.76°	

† Isotropic thermal motion in the model was taken into account by using 16 resolution-dependent bin scales. Individual temperature factors were not used. Individual occupancies were refined for all solvent molecules.

2.2. *Data collection and processing*

The structure was solved from data collected from a single frozen crystal of EV1. Prior to freezing, crystals were stabilized in 25% ethylene glycol in PMC7 at 277 K, transferred to a cryoprotectant of 30% ethylene glycol and 5% glycerol in PMC7 for about 1 min at 277 K and then flash frozen by plunging them into liquid nitrogen. X-ray diffraction data were collected using Cu $K\alpha$ radiation from an Enraf–Nonius GX-13 rotating-anode generator operated at 40 kV, 60 mA with a 100 μm focus and Franks mirror optics (Supper). Data were collected in 0.2° oscillations on a MAR Research 30 cm area detector using a crystal-to-detector distance of 320 mm. This distance, the minimum where reflections were separated by two or three pixels, limited the data collection to 3.55 Å resolution. Although under these conditions the crystals apparently diffracted to at least 3.0 Å, the data beyond 3.55 Å could not be obtained because the detector could not be shifted perpendicular to the beam. A total of 567 exposures were collected.

Diffraction data were integrated with *DENZO* (Otwinowski & Minor, 1997) and merged and scaled internally with *SCALEPACKenormous* (W. Minor, personal communication) with only minor modifications to the standard procedures. The data were readily indexed in a primitive orthorhombic Laue group with a = 352.4, b = 472.2 and c = 483.2 Å. In *SCALEPACK*, post-refinement of crystal mosaicity consistently yielded values too large to be physically reasonable. Consequently, the mosaicity was adjusted interactively in *DENZO* until the edges of the lunes were accurately predicted. The mosaicity of the crystal was estimated to be 0.30°, which is significantly higher than the mosaicity of frozen poliovirus crystals (Wien *et al.*, 1997). The final data set consists of 14182813 (partial) measurements of 947283 unique reflections and is 97.7% complete to 3.55 Å (81% complete in the highest resolution bin). R_{merge} for all measured pairs was 13.6, with R_{merge} = 24.1 in the highest resolution shell, where R_{merge} compares individual intensity measurements I_{hj} with their statistically weighted mean $\langle I_h \rangle$,

$$R_{merge} = (\sum_h \sum_j |I_{hj} - \langle I_h \rangle|)/ \sum_h \sum_j \langle I_h \rangle.$$

Data-collection statistics are provided in Table 1.

2.3. *Presence of pseudo-B-centering*

The space group was determined concurrently with obtaining a molecular-replacement solution (see below). Early in the indexing of the diffraction data it became apparent that the crystals exhibited pseudosymmetry

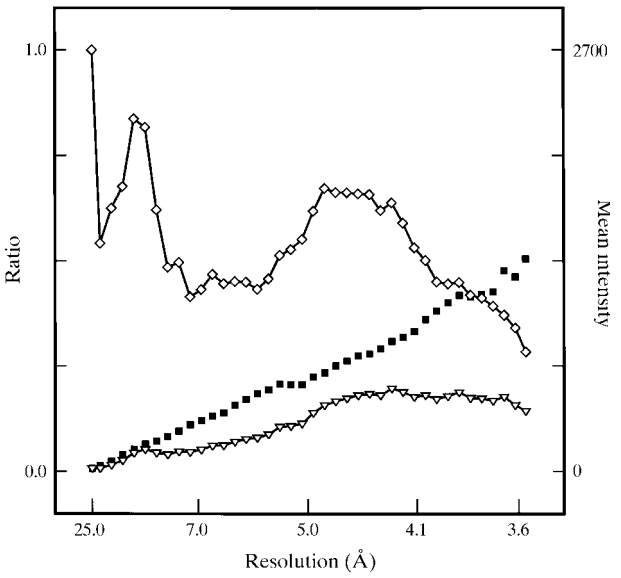


Fig. 1. Reflections with $h + l$ odd are systematically weak. The average intensity of the odd-parity $h + l$ reflections (open triangles), the average intensity of the even-parity $h + l$ terms (open diamonds) and the ratio between the odd-parity and even-parity averages (closed squares) are plotted as functions of resolution. Note that the ratio between the average odd and even order intensities for the $h + l$ reflections is near zero at low resolution and gradually increases. In contrast, the analogous ratios dependent on the parity of $h + k$, $k + l$ and $h + k + l$ uniformly approximate 1.0 and are independent of resolution.

which resulted in systematically weak zones. When the integrated data were separated into groups according to the parity of $h + k$, $h + l$, $k + l$ and $h + k + l$, pseudo extinctions were apparent for all $h + l$ odd reflections out to about 6 Å (Fig. 1), implying that the lattice was pseudo-*B*-centered. Correspondingly, a native Patterson map calculated with the 50–6.4 Å data was dominated by a large translation peak at ($u = 1/2$, $v = 0$, $w = 1/2$) that was five sixths of the height of the origin peak. Despite the presence of convincing axial extinctions extending to high resolution along all three principal axes, the possibility that the axial extinctions on the h and l axes were a consequence of the pseudo-centering precluded a definitive identification of the space group at this stage of the analysis.

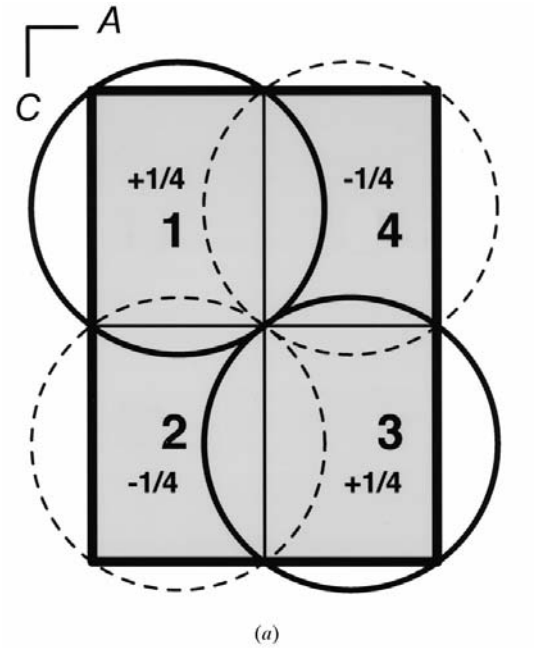
2.4. *Molecular-replacement models*

Starting phases [both for parameter trials and phase refinement based on non-crystallographic symmetry (NCS)] were obtained by molecular replacement. Two types of molecular-replacement models were tried. The first model was a homology model based on the structure of the P1/Mahoney strain of poliovirus, which has about 50% overall sequence homology to EV1 in the capsid region. In this model, loops on the outer surface of the virus were deleted and places where the sequence of the two viruses differed were changed to alanine (or glycine, if appropriate). In the second type of model, the 'spaghetti model' of the main chain, only the C, Cα and N of each residue were retained.

Somewhat surprisingly, the homology model with almost three times as many atoms had a model-based R factor ($R_{\rm model}$, as defined in Fig. 3) only about two points lower than the model with three atoms per residue for low-resolution trials. Therefore, to minimize both model bias and the size of the model-based calculations, the model with three atoms per residue was chosen as the most useful molecular-replacement model. Initially, the model with three atoms per residue was derived from P1/Mahoney. Later, after authentic EV1 electron density became interpretable, abbreviated search models were derived from the current echovirus model.

2.5. *Calculations*

Rotation searches, translation searches and electron-density averaging calculations were all carried out using a suite of programs developed in this laboratory (D. J. Filman, unpublished work). The R-factor trials (starting



(*a*)

Fig. 2. The choice of primitive orthorhombic space group is limited by packing, pseudo-centering and particle rotation. (*a*) Packing considerations require that the four spherical particles be distributed in the unit cell approximately as shown. (*b*) If the virus is rotated significantly, its axis of rotation must be nearly parallel to one of the three principal axes (shown in *A*, *B* and *C*). To preserve approximate *B* centering, viruses 1 and 3 must have similar orientations. The bow-tie shapes are convenient ways to indicate how the viruses have been rotated relative to an initial alignment with icosahedral twofold axes parallel to the crystallographic principals. The two following situations are consistent with a primitive orthorhombic space group. (*c*) If one intact virus constitutes the crystallographic asymmetric unit, then particle 3 is generated from 1 by a twofold or 2_1 (nearly) parallel to the particle's axis of rotation, and the choices for the axis of the larger rotation permits the choices of space group shown. Alternatively, (*d*), the asymmetric unit could contain two rotationally independent half-particles, each having an icosahedral twofold axis exactly coincident with a crystallographic symmetry axis. To emphasize the similarity in packing, some space groups are depicted with a non-standard choice of origin. Observe how knowledge of the rotation axis of the particle restricts the possible choices of space group.

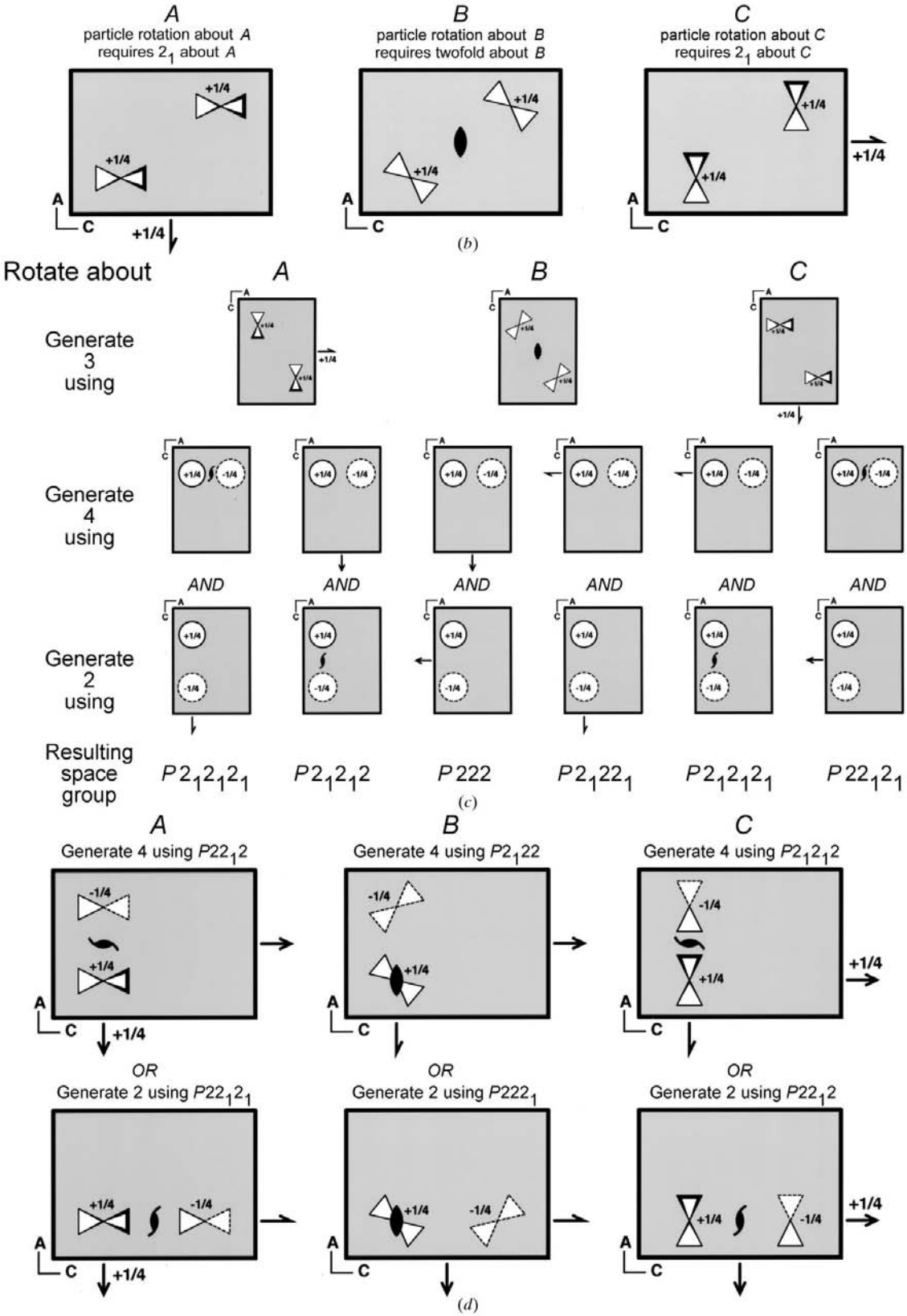

A
particle rotation about A
requires 2$_1$ about A
B
particle rotation about B
requires twofold about B
C
particle rotation about C
requires 2$_1$ about C
(b)
Rotate about
Generate 3 using
Generate 4 using
Generate 2 using
AND
Resulting space group
$P2_12_12_1$
$P2_12_12$
$P222$
$P2_122_1$
$P2_12_12_1$
$P22_12_1$
(c)
Generate 4 using $P22_12$
Generate 4 using $P2_122$
Generate 4 using $P2_12_12$
OR
Generate 2 using $P22_12_1$
Generate 2 using $P222_1$
Generate 2 using $P22_12$
(d)

Fig. 2 (*cont.*)

either with an atomic model or with averaged electron density) were carried out using Unix C-shell scripts to link the programs together. This method of organization facilitated coarse-grained parallelization of the trials process over a network of non-identical work stations.

3. Results

3.1. *Pseudo-B-centering and packing analysis*

Once the NCS-unique volume is specified, a complete trial structure for the crystal can be specified by knowing the following: the number of independent particles, the space group, the particle location (T_x, T_y, T_z), the particle orientation (R_a, R_b, R_c) and the unit-cell parameters (a, b, c). The size of the unit cell (the *ac* and *ab* face diagonals) are almost exactly twice the expected particle diameter and the presence of pseudo-*B*-centering place severe restrictions on the values of these parameters. The rotations and translations can be divided into two groups: those (potentially large ones) that are consistent with exact *B* centering, and those (necessarily small ones) that disrupt *B* centering only at high resolution. The absence of significant odd-parity $h + l$ terms at resolutions below about 6 Å puts an upper limit on the small translations and rotations. It would take about a 1 Å shift (one sixth of the resolution) to disrupt the centering at 6 Å. A rotation of the virus by only 0.4°, operating over the 150 Å radius of the virus particle would shift the equatorial atoms by this distance.

3.2. *Rotation search*

At least one of the viral twofold axes must remain approximately parallel to one of the unit-cell edges, in order for pseudo-*B*-centering to occur (Fig. 2*b*). This allows the initial molecular-replacement orientation trials to be carried out inexpensively as three one-dimensional trial series at low (50–6 Å) resolution (Figs. 3*a*, 3*b*, 3*c*). Starting with a perfectly aligned particle, a single copy of the 'spaghetti' model, with three atoms per residue, was placed in the unit cell and rotated at 0.33° intervals about each principal axis. A highly significant sharp *R*-factor minimum was obtained when the virus model was rotated by 31.5° about *c*, the 483 Å cell edge (Fig. 3*c*).

Beyond about 6 Å resolution, pseudo-*B*-centering begins to break down (Fig. 1). The simple one-parameter rotation model does not predict the presence of odd-order $h + l$ terms with significant intensities. To generate these intensities, the model might possibly be rotated by a small amount around the *a* and/or *b* axes. Using model-based *R*-factor trials, the single-particle rotation search was expanded to include the possibility of additional small rotations about the other two axes. As these trials progressed, the resolution of the data included was extended to the 3.55 Å limit of the available data. At this stage, an *R*-factor minimum was located at 0.35° around *a*, 0.07° around *b* and 31.47° around *c*. The small rotations about *a* and *b* were well within the permissible limits. The unit-cell parameters were also refined over the course of these rotation searches.

3.3. *Space-group determination*

Once the rotation of the particle was known, the translation and space group could be addressed. Logical analysis (Fig. 2) indicated that the possible choices of space group were coupled, *a priori*, with the possible choices of rotation axis. The presence of a large rotation about *c* (Fig. 3*c*) limited the alternatives to four distinct choices of space group, setting and origin (Figs. 2*c* and 2*d*). (These choices were $P2_12_12_1$ or $P22_12_1$, each with one complete virus per asymmetric unit, or $P22_12$ or $P2_12_12_1$, with two independent half-viruses per asymmetric unit, each located along a crystallographic twofold axis.)

The presence of non-zero rotations about *a* and *b* then excluded the possibility that the virus lies along a crystallographic twofold axis and eliminated two of the four possible choices of space group. Of the two remaining possibilities, $P22_12_1$ and $P2_12_12_1$, the latter was provisionally chosen, based on the presence of convincing extinctions of odd-order axial reflections on all three axes (and particularly on the *a* axis) at resolutions well beyond the limit where the general $h + l$ odd extinctions broke down. Late in the structure analysis it became apparent that the correct space group was $P22_12_1$ rather than $P2_12_12_1$. Because the entire packing analysis was conducted in the initial choice of space group with axes arbitrarily placed in the order of ascending unit-cell lengths, the nonstandard setting of the $P22_12_1$ is maintained in this manuscript. Matrices for generating coordinates in both the nonstandard and standard ($P2_12_12$) setting are provided in Table 2.

3.4. *One-dimensional translation search*

Significant translation of the particles along the *c* axis is permissible. Observe (in Fig. 2*b*) that particles 1 and 3, a pair linked by *B* centering, are related to one another by a twofold screw axis parallel to *c*. Clearly, the relative orientations and disposition of these two particles would not be affected by any translation parallel to *c* (in either possible space group). In contrast, significant (>1 Å) translations in the *ab* plane would be inconsistent with *B* centering. This argument explains why the translation problem can be addressed as a one-dimensional trial search at low resolution (50–6 Å).

To assess the magnitude of the translation along *c*, a series of model-based parameter trials was carried out, using a simplified atomic model (Fig. 4). Previously refined values for the orientation and unit cell were

used. Although this trial was carried out in the incorrect space group ($P2_12_12_1$), it nevertheless provided an unambiguous indication of the correct translation, some 26 Å from the initial position at $Z = 0.5$.

3.5. *Multi-parameter refinements*

To obtain more accurate values for all nine molecular-replacement parameters, an extensive series of multi-dimensional model-based trial searches was carried out, in the neighborhood of the current rotation and translation solution. When a multi-dimensional minimum was found, the resolution of the calculation was increased gradually as the accuracy of the solution improved, first to 4.0 Å, and then to 3.55 Å which was the limit of the available data.

NCS-based phase refinements were begun at the same time, using the best available model-based trials as sources of starting phases. Whenever phase refinement converged and an improved version of the 'averaged' wedge map became available, electron-density-based *R*-factor trials were carried out to further refine the molecular-replacement parameters. In each such trial, the 'wedge' map was expanded to fill the unit cell in a way that depended on the values of the trial parameters; then the resulting map was Fourier transformed and compared with the data. Parameter refinement and averaging were alternated until no further improvement was seen in the agreement statistics.

Table 2. *Final values of molecular-replacement parameters*

Final values of the refined parameters (in $P22_12_1$). The number of significant digits shown for each parameter was determined empirically: in each case, a variation in the most significant decimal place made a detectable difference in the *R* factor trials. The ratio between unit cell edges is known much more precisely than their absolute lengths. Rotation matrix applied to the model = $R_aR_bR_c$.

Cell (Å)	$a = 352.45$	$b = 472.15$	$c = 483.20$
Translation (fractional)	$T_x = 0.25000$	$T_y = 0.24950$	$T_z = 0.19623$
Rotation (°)	$R_a = 0.370$	$R_b = 0.035$	$R_c = 31.42$
Rotation matrix†	0.85337	−0.52129	0.00389
	0.52131	0.85335	−0.00519
	−0.00061	0.00646	0.99998

Transformation of refined parameters (into $P2_12_12$)

Cell (Å)	$a = 472.15$	$b = 483.20$	$c = 352.45$
Translation (fractional)	$T_x = 0.24950$	$T_y = 0.19623$	$T_z = 0.25000$
Rotation matrix‡	0.52131	0.85335	−0.00519
	−0.00061	0.00646	0.99998
	0.85337	−0.52129	0.00389

† Applied to the model to place it into the $P22_12_1$ cell. ‡ Applied to the model to place it into the $P2_12_12$ cell.

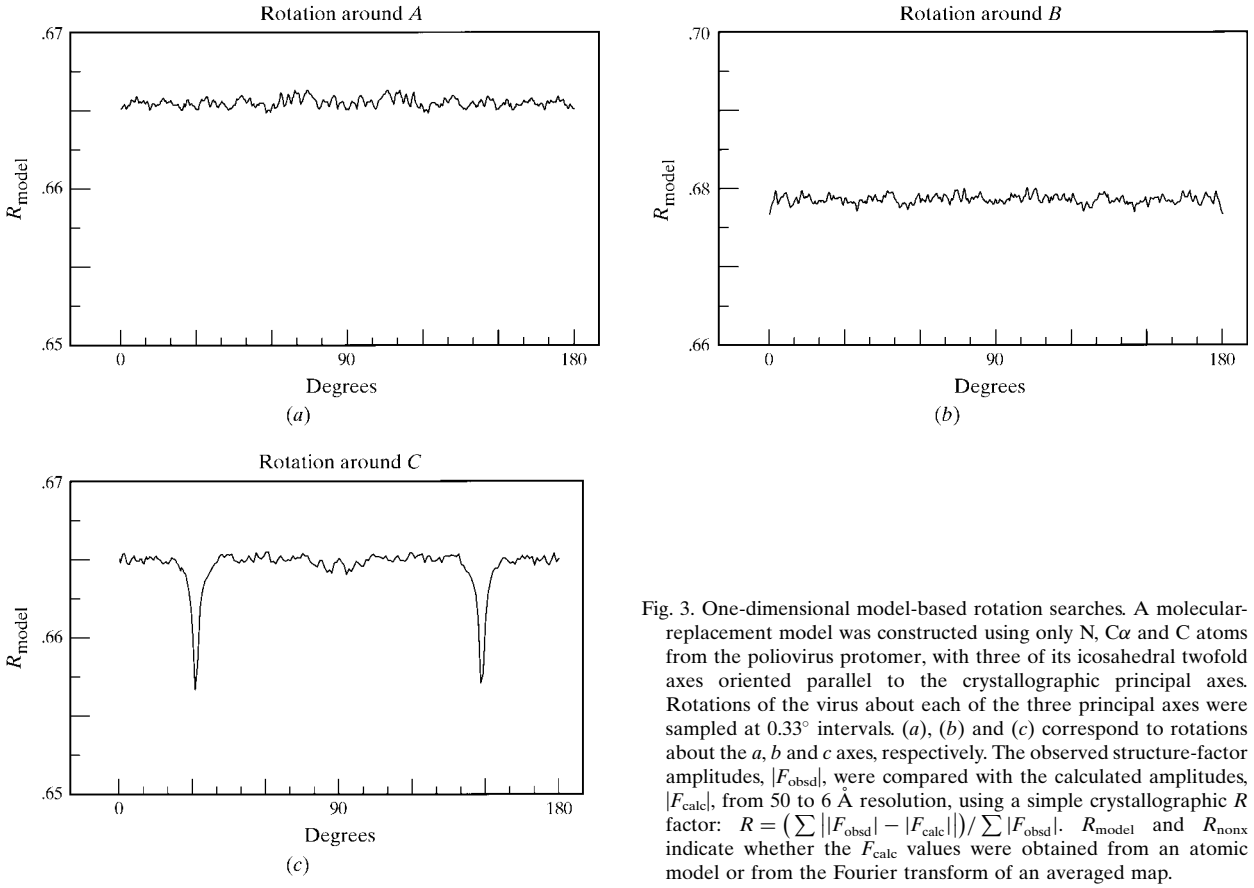


Fig. 3. One-dimensional model-based rotation searches. A molecular-replacement model was constructed using only N, Cα and C atoms from the poliovirus protomer, with three of its icosahedral twofold axes oriented parallel to the crystallographic principal axes. Rotations of the virus about each of the three principal axes were sampled at 0.33° intervals. (*a*), (*b*) and (*c*) correspond to rotations about the *a*, *b* and *c* axes, respectively. The observed structure-factor amplitudes, $|F_{obsd}|$, were compared with the calculated amplitudes, $|F_{calc}|$, from 50 to 6 Å resolution, using a simple crystallographic *R* factor: $R = \left(\sum \left||F_{obsd}| - |F_{calc}|\right|\right) / \sum |F_{obsd}|$. R_{model} and R_{nonx} indicate whether the F_{calc} values were obtained from an atomic model or from the Fourier transform of an averaged map.

3.6. *Map interpretation and model building*

When all available data to 3.55 Å had been included, and the nine parameters had been refined as best as possible, two seemingly contradictory results were obtained: although the averaging statistics were quite poor (R_{nonx}, as defined in Fig. 3, was 0.35 overall) the averaged map was readily interpretable (see Fig. 5) and contained new chemical information that was not present in the phasing model. Specifically, side chains corresponding to the sequence of EV1 were clearly evident in the averaged electron density even though no side chains were present in the molecular-replacement model; continuous density was seen for all loops on the surface of the virus and this density was consistent with the appropriate insertions and deletions relative to the phasing model; and portions of the EV1 structure were visible which corresponded to disordered portions of the P1/Mahoney structure and were, therefore, not included in the model.

3.7. *Correcting the space group*

After phase refinement had converged, the reflections were separated, as before, into even- and odd-parity groups in various ways, depending on the parity of h, k, l, $h + k$, $k + l$, $l + h$ or $h + k + l$. Model-based and NCS-based R factors were evaluated separately for each group and tabulated as functions of resolution. All of the parity classes had similar statistics (data not shown), with R_{nonx} equal to 0.35 overall, except for the $h + l$ even and odd groups (Fig. 6). Using the $h + l$ even-parity reflections, R_{nonx} was 0.26 overall and varied with resolution. For the $h + l$ odd group it was 0.52 and essentially independent of resolution. This behavior was diagnostic of an incorrect choice of space group.

The only acceptable alternative choice of space group was $P22_12_1$ (see above). By factoring the structure-factor expression for both choices of space group, it can be shown that (with an appropriate choice of particle translation) the structure factors for the $h + l$ even reflections are identical in both cases, but the structure factors for the $h + l$ odd reflections differ. In order to

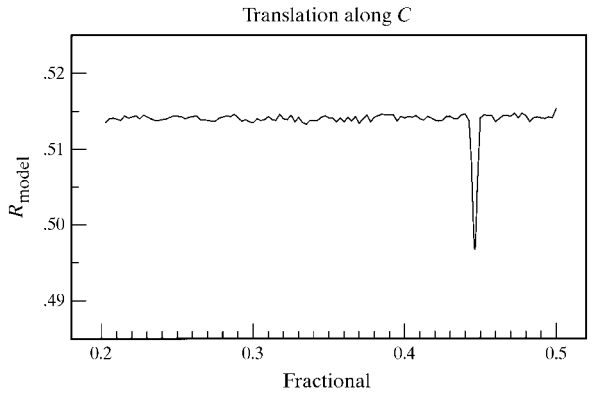


Fig. 4. One-dimensional translation trials. A single virus particle, oriented nearly correctly, was placed at the fractional coordinates (1/2, 1/2, Z) for all positions between $Z = 0.20$ and $Z = 0.50$, sampled at about 1 Å intervals, and the three symmetry-related particles were generated with the assumed space-group operators. The sharp peak identifies a 26 Å shift away from $Z = 0.5$.

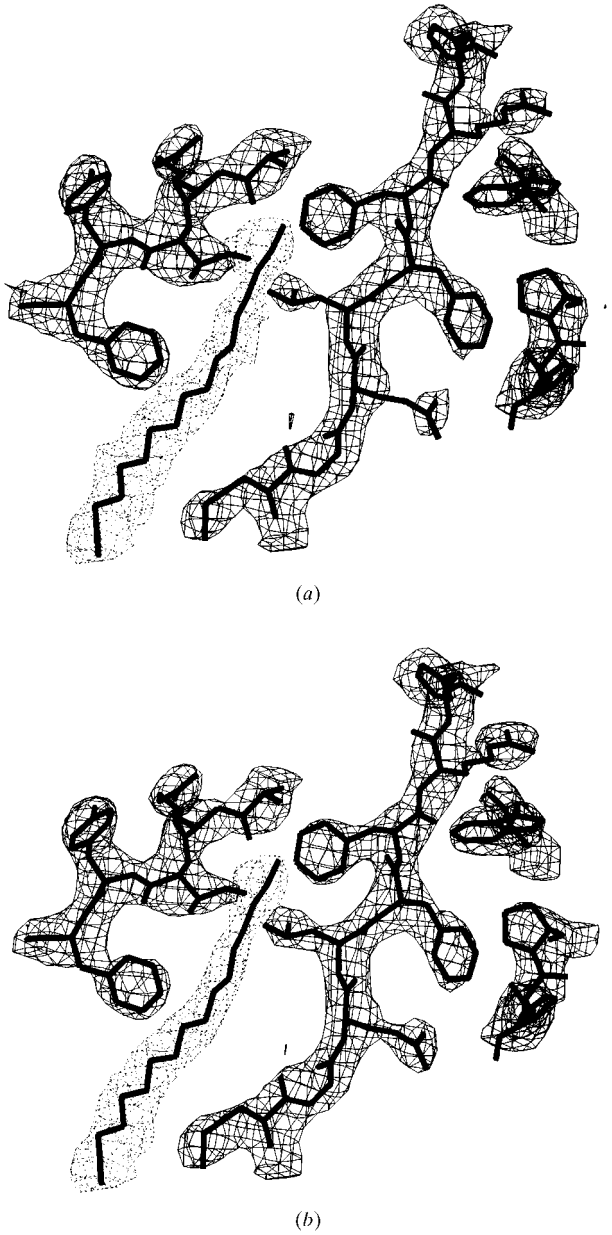


Fig. 5. The initial and final 3.55 Å maps are very similar. (*a*) An early map, phased with the three-atom model, and averaged in the incorrect space group $P2_12_12_1$. (*b*) The final map, averaged in the correct space group $P22_12_1$. Portions of the final model are included as landmarks. Electron density associated with other portions of the structure is omitted for clarity. A lowered contour level (dashed lines) was used for the putative palmitate ligand. An intermediate map phased with a complete atomic model and averaged in $P2_12_12_1$ (not shown) is almost indistinguishable from (*b*).

maintain pseudo-*B*-centering in the standard setting of the new space group, the particle was shifted by (1/4, 1/4, 1/4). Then, using model- and density-based *R*-factor trials in the correct space group, the nine parameters were re-refined, altering the sign of a rotation or translation, if necessary. Refined values for the parameters changed only slightly.

The question was posed whether refinement against the odd- or even-parity terms would give more reliable results. The rationale for using the odd terms alone was that their *R* values might be more critically sensitive to the parameters and choice of space group. On the other hand, a possible reason to exclude the odd terms was the tendency of the odd-parity terms toward zero, whenever perfect centering was approached, resulting in a singularity that made refinement problematic.

To address this question, two series of model-based parameter trials were carried out in parallel: one *versus* the data with $h + l$ odd and one *versus* the data with $h + l$ even. Surprisingly, the refined rotation about the *a* axis differed significantly between the two, being 0.535° (odd) and 0.370° (even). To resolve which, if either, of these orientations was preferable, NCS-based phase refinements were carried out in parallel, starting with each of these possible orientations or with their average. The orientation specified by the odd-parity data alone was clearly incorrect, as it led to a $R_{\rm nonx}$ value which was worse by 0.05.

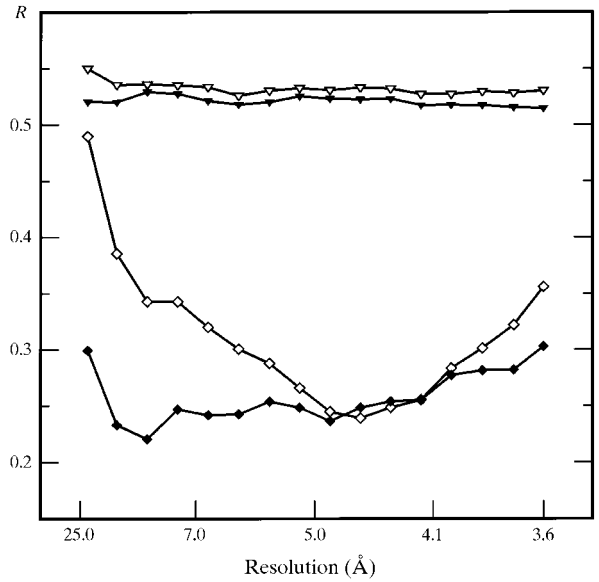


Fig. 6. In the incorrect space group, the *R* values for the $h + l$ odd-parity reflections are high and independent of resolution. Triangles indicate the odd-parity data and diamonds indicate the even-parity data. $R_{\rm model}$ (open symbols, 0.3751 overall) was obtained from a complete atomic model that was partly refined in the wrong space group. $R_{\rm nonx}$ (closed symbols, 0.3476 overall) at convergence of an NCS-based phase refinement that was initially phased by the same model.

3.8. *Phase refinement*

The final NCS-based phase refinement was initialized using the orientation specified by the even-parity reflections (Table 2). The atomic model at this stage was essentially complete and not substantially different from the final refined version. Over the course of five cycles, the standard crystallographic *R* dropped gradually from 0.2815 (based on the atomic model) to 0.2582 (based on the averaged map at convergence). Model-based and NCS-constrained phases differed by less than 20° (by *F*-weighted average). At convergence, $R_{\rm nonx}$ for the even-parity $h + l$ reflections was fairly similar in all resolution ranges (Fig. 7*a*). $R_{\rm nonx}$ for the odd-parity $h + l$ reflections was higher in all resolution ranges, consistent with the relative weakness of these data, and increased markedly at low resolution, where their average intensities approach zero.

3.9. *The atomic model and its refinement*

When averaging yielded an interpretable image (see below) the atomic model for EV1 was built using *FRODO* (Jones, 1985), and refined using the *XX12* procedure (Jacobson *et al.*, 1996). This refinement focuses on a single copy of the NCS-unique structural motif (the protomer consisting of one copy each of VP1, VP2, VP3 and VP4). The method achieves computational efficiency by treating a parallelepiped enclosing the protomer as if that neighborhood was a periodic object. Using the averaged map of this same region as a refinement standard, this pseudo-real space procedure refines the atomic coordinates so as to cause the resolution-bin-scaled Fourier transform of the model to approach the transform of the standard, both in amplitudes and phases.

The model-dependent *R* values reported here ($R_{\rm model}$) were calculated in the authentic unit cell, using resolution-dependent bin scales (eight bins at 6 Å resolution and 12 or 16 bins at 3.55 Å resolution). These *R* values were calculated repeatedly whenever the refined model was used for model-dependent parameter trials and/or to re-initiate the NCS-based averaging calculation. Each time a round of averaging was completed, an improved image was available as a standard for further refinement of the model.

To compensate for the relatively low resolution of the three-dimensional image (3.55 Å), the default weights were adjusted to increase the penalty for poor variable torsions. Constants were set to 3.5 kcal mol^{-1} for most variable torsions and 100 kcal mol^{-1} for peptide bonds. Despite these adjustments, the r.m.s. deviations from idealized bond lengths and angles remained acceptable: 0.012 Å on bond lengths, 2.76° on bond angles, 24.66° on variable torsions and 1.29° on fixed dihedrals. Following refinement, the crystallographic *R* for this model was 26.3 overall (see Fig. 7*b*).

The refined model of EV1 contains 6575 non-H atoms: 6543 atoms for the protein and covalently linked myristate, 18 atoms for the non-covalent ligand in the hydrophobic core of capsid protein VP1, which is modeled convincingly as palmitate, and 14 bound solvent molecules with variable occupancies. These 14 bound solvents account for most of the significant features in the (bin-scaled) 'vector' difference map between 'averaged' and model-based electron density. Some of these solvent molecules are almost certainly bound metal ions, as they occupy negatively charged protein cages and have refined occupancies significantly larger than one. In light of the limited resolution of the structure, individual temperature factors were not refined. Instead, an arbitrary overall temperature factor was assumed when model-based electron density was calculated, and isotropic thermal motion in the model was taken into account by using resolution-dependent bin scaling in all structure-factor comparisons.

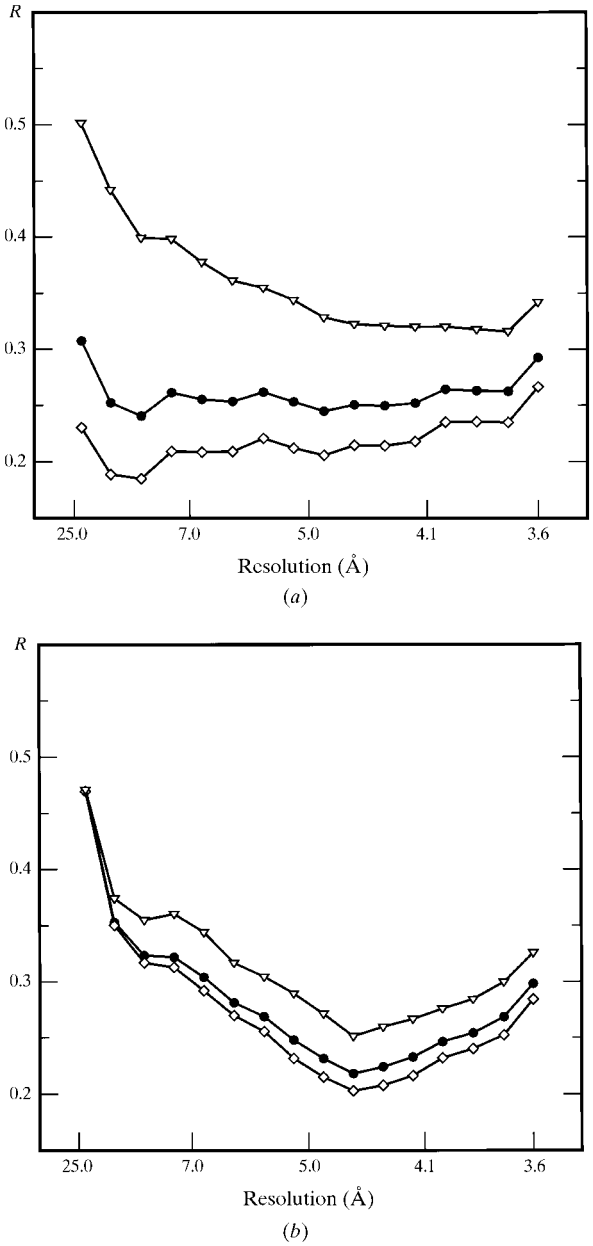


Fig. 7. In the proper space group, the R values are more similar for the $h + l$ even and the $h + l$ odd reflections. Open triangles indicate the odd-parity data and open diamonds indicate the even-parity data. Closed circles represent R values with all data included. (*a*) R_{nonx} compares the observations with the Fourier transform of the 'averaged' map after convergence of NCS-based phase refinement. Overall, R_{nonx} was 0.3359 for the odd-parity data, 0.2193 for the even-parity data, and 0.2581 for all data. (*b*) R_{model} compares the observations with the Fourier transform of the final atomic model. Overall, R_{model} was 0.2912 for the odd-parity data, 0.2518 for the even-parity data and 0.2629 for all data.

The atomic model includes all of VP1 (residues labeled 1001–1281), residues 2008–2261 of VP2, all of VP3 (residues 3001–3239) and most of VP4 (the myristate, labeled 4001, plus aminoacyl residues 4002–4015 and 4023–4069). The EV1 structure has few disordered polypeptide segments in comparison with most other picornavirus structures. Indeed, one of the most striking aspects of the EV1 structure is how little of the structure is disordered. Of 849 amino acids, the positions of only 14 are not obvious: residues 2001–2007 from the amino terminus of VP2, and residues 4016–4022 of VP4. A detailed description of the structure and its biological implications will be presented elsewhere (Wien *et al.*, in preparation).

4. Discussion

4.1. *The use of pseudosymmetry in molecular-replacement problems*

The presence of non-crystallographic symmetry in a crystal can provide a source of very powerful phase constraints. However, the success of the associated 'averaging' calculations hinges on accurate knowledge of orientations and positions in the unit cell. In completely general molecular-replacement problems, the relevant parameters must be obtained by a brute force three-dimensional calculation, at considerable expense. The costs increase with the unit-cell size, the fineness of angular sampling and the spacing between features to be resolved (Tong & Rossmann, 1990). Historically, the effort to make such calculations affordable has led to the invention of short cuts that, unfortunately, can affect the quality of the answers. These short cuts include carrying out the orientation calculations in Patterson space, so as to uncouple the rotation and translation searches; omitting much of the data from the calculation; sampling the angular range too coarsely; and re-casting the problem in terms of a limited set of basis functions.

In favorable cases, such as EV1, the molecular-replacement problem can be re-parameterized in a way

that reduces the number of parameters to be searched simultaneously or otherwise restricts the range of possible solutions. In EV1, these restrictions arise from the presence of pseudosymmetry in the crystals, and from an analysis of possible ways that the unit cell can be packed. Once the parameter space to be searched was limited, it became affordable to carry out closely spaced parameter trials focused on the relevant volume of parameter space. This computational approach lent itself easily to coarse-grained parallelization. Such calculations were shifted seamlessly between multiprocessor mainframe computers and a distributed network of desktop work stations, according to their availability.

4.2. *Solving the right structure in the wrong space group*

At first glance it seems impossible that the structure could be solved in the incorrect space group. Although the first step in the structure determination, the initial orientation search was by definition independent of space group, the success of subsequent steps (the initial translation search, refinement of position and orientation and NCS-based phase refinement) seems more problematic. The explanation lies in a very general relationship between amplitudes of certain even-parity zones of reflections in the two possible choices of space group and in the pseudo-*B*-centering of the echovirus crystals.

4.2.1. *Translation search*. By factoring the structure-factor equations in the two possible space groups it can be shown that, depending on the choice of relative origin, either the amplitudes for the $h + k$ even reflections or the amplitudes for $h + l$ even reflections are identical in both choices of space group. Hence, when the space group is $P22_12_1$ but the search is conducted in $P2_12_12_1$, the three-dimensional translation function will have two distinct minima. In one of these minima, the amplitudes for the $h + k$ even reflections are calculated correctly but the $h + k$ odd are incorrect. In the other minimum, the amplitudes for the $h + l$ even reflections are calculated correctly but the $h + l$ odd reflections are incorrect. These minima are separated by the special translation (1/4, 1/4, 1/4). In a pseudo-*B*-centered structure the latter minimum will be favored by the statistics, since incorrectly calculated $h + l$ odd amplitudes are systematically weak.

In practice, the three-dimensional translation search was never actually calculated and the alternative minimum was never sampled. Packing considerations and the requirement for approximate *B* centering dictated the correct *x* and *y* translation in either space group. This choice of T_x and T_y caused the 2_1 axis along *c* to be superimposed in the two space groups, and assured that the strong $h + l$ even terms would be calculated identically in either space group during the one-dimensional translation search along *z*.

4.2.2. *Refinement of orientation and position*. The subsequent refinement of the orientation and position yielded substantially correct parameter values primarily because the strong $h + l$ even terms were predicted correctly regardless of the choice of space group. Collectively, the $h + l$ odd terms also tended to specify the correct parameters (as shown when they were used as the sole basis for refinement) though the residuals were higher, the refinements were noisier, and the refined parameters were less accurate.†

4.2.3. *NCS refinement*. Given that the correct orientation and position of the icosahedral point group can be determined despite the incorrect identification of the space group, the ability of the NCS refinement to produce high-quality, chemically reasonable electron-density maps is readily explained. NCS-based phase refinement seeks to converge on a set of unique electron densities that best accounts for the observed structure-factor amplitudes under circumstances where the assumed symmetries (crystallographic and noncrystallographic) and other molecular-replacement parameters are enforced (Bricogne, 1974). The NCS refinement can be thought of as a two-step procedure. In the first step, averaged electron densities are accumulated for the icosahedrally unique volume (the 'wedge map') by interpolation from a map of the crystallographic asymmetric unit that was calculated with the current phases. In the second step, these averaged densities are distributed into the asymmetric unit of the chosen space group and the resulting map is Fourier transformed, yielding an updated phase set.

Whenever the densities for the icosahedral asymmetric unit are accumulated, NCS averaging reinforces any information that is consistent with the assumed symmetry and dampens any information that is inconsistent with it. In either space group, the $h + l$ even reflections are consistent with the assumed arrangement of particles and contribute identically to the averaged map. However, in the incorrectly assumed space group, the $h + l$ odd terms are inconsistent with the assumed symmetry and, therefore, do not contribute significantly. In EV1, 60-fold averaging was sufficient to eliminate almost all of the noise, so that 'wedge' maps calculated in either choice of space group were chemically correct

† The behavior of the $h + l$ odd reflections is complicated. The four symmetry-related atomic contributions to each structure factor can be divided into two parts: one vector contribution that is unaffected by the error in the space group and a second vector whose proper contribution is negated when going from one space group to the other. Whenever these two vectors happen to be oriented at right angles to one another, the resulting structure-factor amplitudes are identical in the two space groups. However, whenever the two contributions are supposed to add in phase or to cancel one another, structure-factor amplitudes will be maximally different in the two groups. Thus, the strongest and weakest $h + l$ odd reflections have structure factors whose magnitudes are anti-correlated with their proper values. This curious effect becomes more severe as perfect centering is approached.

and virtually indistinguishable at this resolution. Presumably, the symmetry-consistent contribution of the $h + l$ odd terms in the correct space group would have had a greater impact if the terms were relatively stronger, or if the multiplicity of averaging was less.

The appearance in the electron-density map of chemically plausible information that was not specified by the original phasing model is ordinarily a good indication that a crystal structure has been solved correctly. The structure solution of EV1 provides an interesting counter-example, and illustrates two of the circumstances that make this kind of error possible: the presence of NCS and pseudo-centering. Evaluating NCS-based averaging statistics as a function of reflection parity is inexpensive, but has the potential to catch subtle errors in space-group assignment. Indeed, in the course of this structure determination, a second (non-virus) structure with an analogous ambiguity in the space group was encountered (Grant *et al.*, 1998). Our experience with the echovirus structure determination greatly facilitated the identification of the origins of the space-group ambiguity and its resolution in this latter structure.†

The authors thank Jason Sello for help with data collection and Robert Grant for useful discussion and for critical reading of the manuscript. This work was supported in part by AI20566 (to JMH), AI35667 (to JMB), by an NSF grant for High Performance Computing and Communication (MCB 9527181, Gerhard Wagner, PI). JMB is an Established Investigator of the American Heart Association. This is a contribution from the Harvard Center for Structural Biology, supported by the Giovanni Armenise-Harvard Foundation for Advanced Scientific Research.

† Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1EV1).

References

Basavappa, R., Syed, R., Flore, O., Icenogle, J. P., Filman, D. J. & Hogle, J. M. (1994). *Protein Sci.* **3**, 1651–1669.

Bricogne, G. (1974). *Acta Cryst.* A**30**, 395–405.

Cherry, J. (1983). In *Infectious Disease of the Fetus and Newborn*, edited by J. Remington & J. Klein. New York: Sanders.

Filman, D. J., Syed, R., Chow, M., Macadam, A. J., Minor, P. D. & Hogle, J. M. (1989). *EMBO J.* **8**, 1567–79.

Grant, R. A., Filman, D. J., Finkel, S., Kolter, R. & Hogle, J. M. (1998). *Nature Struct. Biol.* **5**(4), 294–303.

Hogle, J. M., Chow, M. & Filman, D. J. (1985). *Science*, **229**, 1358–65.

Jacobson, D. H., Hogle, J. M. & Filman, D. J. (1996). *Acta Cryst.* D**52**, 693–711.

Jones, T. A. (1985). *Methods Enzymol.* **115**, 157–171.

Lentz, K. N., Smith, A. D., Geisler, S. C., Cox, S., Buontempo, P., Skelton, A., DeMartino, J., Rozhon, E., Schwartz, J., Girijavallabhan, V., O'Connell, J. & Arnold, E. (1997). *Structure*, **5**, 961–78.

Melnick, J. L. (1990). In *Fields' Virology*, edited by B. N. Fields, D. M. Knipe, R. M. Chanock, J. Melnick, B. Roizman & R. Shope. Philadelphia: Lippincott-Raven.

Modlin, J. F. (1988). *Pediatr. Infect. Dis. J.* **7**, 311.

Muckelbauer, J. K., Kremer, M., Minor, I., Diana, G., Dutko, F. J., Groarke, J., Pevear, D. C. & Rossmann, M. G. (1995). *Structure*, **3**, 653–67.

Muckelbauer, J. K., Kremer, M., Minor, I., Tong, L., Zlotnick, A., Johnson, J. E. & Rossmann, M. G. (1995). *Acta Cryst.* D**51**, 871–887.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Rueckert, R. R. (1996). In *Fundamental Virology*, edited by B. N. Fields & M. K. Knipe. Philadelphia: Lippincott–Raven.

Tong, L. & Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 783–792.

Wien, M. W., Curry, S., Filman, D. J. & Hogle, J. M. (1997). *Nature Struct. Biol.* **4**, 666–74.

Zlotnick, A., Natarajan, P., Munshi, S. & Johnson, J. E. (1997). *Acta Cryst.* D**53**, 738–746.