Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Structure determination of adeno-associated virus 2: three complete virus particles per asymmetric unit

**Qing Xie,[a,b] Thayumanasamy Somasundaram,[b] Smita Bhatia,[b] Weishu Bu[b] and Michael S. Chapman[a,b]***

[a]Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306-4380, USA, and [b]Kasha Laboratory of Biophysics, Florida State University, Tallahassee, FL 32306-4380, USA

Correspondence e-mail: chapman@sb.fsu.edu

Received 20 November 2002
Accepted 5 March 2003

**PDB Reference:** adeno-associated virus 2, 1lp3, r1lp3sf.

The atomic structure of adeno-associated virus 2 (AAV-2) has been determined to 3.0 Å resolution. AAV-2 crystallized in space group $P1$, with unit-cell parameters $a = 249.7$, $b = 249.7$, $c = 644.8$ Å, $\alpha = 90.0$, $\beta = 101.2$, $\gamma = 120.0^\circ$. The crystals contained three full virus particles in the asymmetric unit, allowing 180-fold non-crystallographic symmetry averaging. The particle orientations were determined using the self-rotation function and found to have similar but resolvably different orientations. Approximate alignment of icosahedral and interparticle threefold screw symmetry led to a native Patterson that was interpretable in terms of approximate particle positions. Accurate positions required a Patterson correlation search that was constrained to be consistent with non-crystallographic threefold projection symmetry evident in the diffraction intensities. Initial phases to 15.0 Å resolution were calculated by molecular replacement using the known structure of a distantly related homolog (23% sequence identity). Real-space averaging was performed and phases were extended from 15.0 to 3.0 Å. An atomic model was fitted and refined using a simulated-annealing real-space procedure.

## 1. Introduction

Adeno-associated virus 2 (AAV-2) is a member of the parvovirus family. Parvoviruses are small (~250 Å) unenveloped viruses in which a single-stranded (ss) DNA (about 5 kb) genome is surrounded by a $T = 1$ (Caspar & Klug, 1962; Luo *et al.*, 1988) icosahedral capsid containing 60 copies of the capsid protein (Muzyczka & Berns, 2001). Parvoviruses contain three variants of the capsid protein (VP1, VP2 and VP3), which differ at their N-termini, in a ratio of about 1:1:8. The function of the N-terminal extensions is not fully understood. In AAV-2, VP3 has 533 amino acids, while VP2 adds 65 residues to the N-terminus and VP1 adds another 137 (Rutledge *et al.*, 1998), the results of alternative start codons/splicings. The parvovirus family can be divided into autonomous parvoviruses, which are self-sufficient in their replication, and the dependoviruses, the replication of which depends on co-infection with another virus, usually 'adenovirus', hence the name adeno-associated virus (Moskalenko *et al.*, 2000). Apart from their role as a helper virus, adenoviruses are unrelated to adeno-associated viruses.

Adeno-associated virus is a leading candidate vector for human gene therapy (Carter, 1990), through which it is hoped that a wide variety of inherited diseases, autoimmune disorders and cancers will be treatable. In vector constructs, much of the wild-type genome is replaced by the therapeutic transducing DNA, which is carried to cells within the protein capsid shell of AAV. In one of several promising applications,

© 2003 International Union of Crystallography
Printed in Denmark – all rights reserved

AAV is being used in clinical trials as a potential cure for cystic fibrosis (Flotte *et al.*, 1996; Carter & Flotte, 1996), a disease that afflicts 30 000 people per year in the USA. For many proposed therapies, there is the challenge of reaching the target cell while avoiding vector neutralization by antibodies generated from prior exposure. An applied long-term objective is to develop the capability to engineer improved vectors; to generate, for example, a family of synthetic serotypes that could be used to evade immune interception on successive doses. It is also hoped that ligands for alternative cellular receptors can be inserted to change the cell specificity. For such modifications to be made rationally, the structure is required. The determination was very challenging because of the large crystal asymmetric unit containing three complete virus particles, totaling about $12 \times 10^6$ Da. This paper describes the crystallographic structure determination.

## 2. Methods and results

### 2.1. Sample preparation and diffraction data

Crystals were grown by hanging-drop vapor diffusion from AAV2 that had been purified by caesium chloride gradient ultracentrifugation. Aliquots of 2–6 µl of 8.5 mg ml$^{-1}$ AAV were mixed with equal volumes of reservoir solution and equilibrated against 100 m*M* HEPES pH 7.3–7.5, 50 m*M* $MgCl_2$, 25% glycerol and 0.03% sodium azide at room temperature. Crystals grew to average dimensions of 0.3 × 0.3 × 0.25 mm in about a week. Crystals were flash-frozen by immersion in liquid nitrogen (Rodgers, 1994). With 25% glycerol in the crystallization medium, no additional cryoprotection was required.

### 2.2. Data collection and processing

Diffraction data were collected at the biohazard-level 2 F1 beamline at the Cornell High Energy Synchrotron Source (CHESS). Crystals were transported pre-frozen and were maintained in a stream of nitrogen vapor at 100 K. Diffraction images were collected using a Quantum 4 CCD detector (ADSC, Poway, CA, USA) with an oscillation angle of 0.3°. About half of the frozen crystals showed useful diffraction; one was twinned and the rest diffracted poorly, often showing evidence of ice formation during freezing or transfer. The resolution limit of crystals varied between 2.6 and (more frequently) 3.5 Å (Fig. 1*a*). This is usually ample to determine a structure with high non-crystallographic symmetry. Cryo-freezing extended the crystal lifetime from 1–2 min to about 3 h (Fig. 1*b*).

Two crystal forms were observed (Table 1). This is common with virus crystals, even under identical crystallization conditions, presumably because there are several efficient ways to pack nearly spherical virus particles. A preliminary characterization indicated that the form 1 lattice was rhombohedral with a unit-cell length of 265 Å. However, form 2, grown under nominally identical conditions, was prevalent and became the basis for this structure determination. Form 2 had a lattice repeat of ∼650 Å, requiring a sample-to-detector distance of ∼500 mm to maintain the usual ∼10 pixel minimal spot separation. However, the unit cell was long and narrow, so the finest spot separation varied with crystal orientation; therefore, many images were collected at higher resolution with a closer detector (300 mm).

Indexing and processing of the finely spaced reflections required careful selection of parameters for the program *DENZO* (Otwinowski & Minor, 1997). About 400 peaks between 50 and 7 Å resolution were selected from single oscillation images in which the lunes were well separated. Therefore, it was easier to start the indexing of each crystal with a short unit-cell spacing approximately parallel to the X-ray beam. Other frames were bootstrapped from previously indexed frames. The data consistently indexed in a primitive rhombohedral Laue group ($R\bar{3}$), with unit-cell parameters $a = b = c = 644.8$ Å, $\alpha = \beta = \gamma = 22.3°$. The equivalent centered hexagonal cell would have $a = b = 249.7$, $c = 1885.3$ Å (Hahn, 1996). The diffraction symmetry revealed a threefold, but no six or twofolds.

**Table 1**
Two crystal forms and their lattice types.

| Crystal form and pseudo-lattice type | *a* (Å) | *b* (Å) | *c* (Å) | α (°) | β (°) | γ (°) | *Z*† |
|---|---|---|---|---|---|---|---|
| 1, primitive rhombohedral | 266.5 | 266.5 | 266.5 | 58.9 | 58.9 | 58.9 | 1 |
| 1, equivalent centered hexagonal | 262.5 | 262.5 | 658.2 | 90.0 | 90.0 | 120.0 | 3 |
| 2, primitive rhombohedral | 644.8 | 644.8 | 644.8 | 22.3 | 22.3 | 22.3 | 3 |
| 2, equivalent centered hexagonal | 249.7 | 249.7 | 1885.3 | 90.0 | 90.0 | 120.0 | 9 |

† No. of AAV in unit cell.

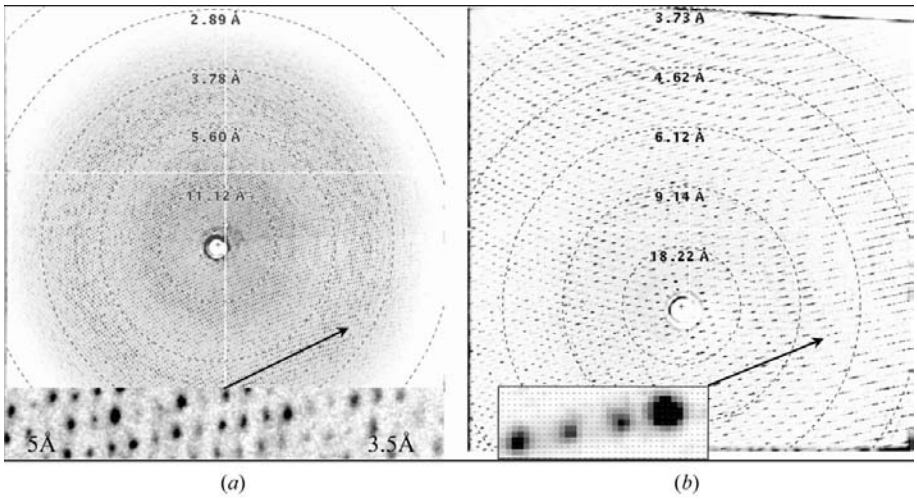


**Figure 1**
Diffraction from form 2 crystals in different orientations. (*a*) The sample-to-detector distance was 300 mm. (*b*) The distance was 500 mm. This is the 90th image collected from a crystal exposed to 2 h of beam, demonstrating the stability of samples at ∼100 K.

**Table 2**
Data-collection statistics.

(*a*) Data scaled in *R*3 and re-indexed as *P*1.

| | 100.0–3.0 Å | Outer shell, 3.56–3.43 Å |
|---|---|---|
| Crystal c7v3 (184 frames) | | |
| Reflections observed | 540067 | |
| Unique reflections | 467337 | 19941 |
| Completeness (%) | 17.6 | 11.3 |
| $R_{\rm merge}$† (%) | 9.0 | |
| Crystal c6v1 (87 frames) | | |
| Reflections observed | 108203 | |
| Unique reflections | 103821 | 3340 |
| Completeness (%) | 3.9 | 1.9 |
| $R_{\rm merge}$† (%) | 10.5 | |
| Merged form 2 crystals (271 frames in total) | | |
| Reflections observed | 648270 | |
| Unique reflections | 557228 | 23073 |
| Completeness (%) | 21.0 | 13.1 |
| $R_{\rm merge}$† (%) | 16.5 | 25.5 |

(*b*) Threefold expanded/merged data in space group *P*1.

| | 100.0–3.0 Å | Outer shell, 3.56–3.43 Å |
|---|---|---|
| Merged form 2 crystals (271 frames in total) | | |
| Reflections in data set | 756743‡ | |
| Unique reflections | 495356‡ | 25131 |
| Completeness (%) | 56.5 | 43.0 |
| $R_{\rm merge}$† (%) | 15.6 | 30.9 |

† $R_{\rm merge} = \sum_h \sum_n |I_{h,n} - \langle I_h \rangle| / \sum_h \sum_n I_{h,n}$, where $I_{h,n}$ is the $n$th observation of a symmetry equivalent of reflection $h$. The wavelength was 0.92 Å and the oscillation range was 0.3°. The expanded/merged data took advantage of the threefold observed in the diffraction pattern and were the data used to solve the ‡ ructure. † Many of the unobserved reflections have threefold-symmetry equivalents from which their intensities can be copied, increasing the number of reflections in the expanded data set. Some of the observations of the *P*1 cell are related by threefold symmetry and should not be considered unique in the expanded set. Likewise, the number of theoretically observable reflections is reduced by a factor of three, because each *P*1 reciprocal-lattice point has two equivalents with the added threefold symmetry. The effect is to increase the proportion of the unique data that has been measured.

With a mosaic spread that was larger than the oscillation range, essentially all reflections were partial. Following scaling with *SCALEPACK* (Otwinowski & Minor, 1997), the data were alternately post-refined (Rossmann *et al.*, 1979) and re-scaled. When possible, reflection intensities were obtained by summing partials from adjacent frames or, if not, estimated from the partiality. For scaling, it was important to use for the reference a (non-default) middle frame where summed partials were available. For the final post-refinement, the mosaicity was fixed at 0.5° and other parameters were refined for each frame. There were nearly $3 \times 10^6$ partial measurements of $\sim 7.5 \times 10^5$ independent observations of $\sim 5 \times 10^5$ unique reflections (Table 2); *i.e.* each observation spanned an average of four frames.

The current data set was merged from the best two crystals: 184 frames from the first crystal and 87 frames from the second crystal. Crystal c6v1 was transformed to an alternative indexing (with the same unit-cell parameters) for scaling compatibility to c7v3. The rhombohedrally processed data was 57% complete to 3.0 Å (80% complete to 3.5 Å). $R_{\rm merge}$ was 15.6% to 3.0 Å.

### 2.3. Phase determination

The strategy was to use molecular replacement followed by symmetry averaging. Initial phases would be calculated from a related virus structure at a low (15 Å) resolution (from which no molecular detail would be retained) then extended to high resolution using the icosahedral symmetry, so that the high-resolution phases would be model-independent (Rossmann, 2001). The prerequisites were the orientation of the icosahedral symmetry operators and the positions of the virus particles within the asymmetric unit.

### 2.4. Rotation function and implications for the space group

The particle orientations were determined by both ordinary and locked self-rotation functions using the program *GLRF* (Tong & Rossmann, 1997). The initial 14 Å ordinary RF showed all expected icosahedral rotation axes with one of the threefolds aligned approximately parallel to the putative rhombohedral threefold (Figs. 2*a*, 2*b* and 2*c*). The locked rotation function (Fig. 2*d*) gave only one solution consistent with the ordinary RF. Icosahedral twofolds were orthogonal to the diffraction threefold, so careful tests were made to check that the Laue diffraction symmetry was not 32.

At 3.0 Å resolution, locked and ordinary rotation functions resolved three slightly different orientations of the icosahedral symmetry (Fig. 3). Split rotation-function peaks have been seen in several previous virus structure determinations such as CPV (Tsao *et al.*, 1992), $\varphi$X174 (McKenna *et al.*, 1992), *Galleria mellonella* densovirus (Simpson *et al.*, 1998) and bateriophage HK97 (Wikoff *et al.*, 1999) and typically arises owing to approximate but inexact alignment of an icosahedral symmetry element with crystallographic symmetry. In this case, the icosahedral threefolds of the virus particles were tilted about 6° from the putative rhombohedral threefold axis.

The split rotation-function peaks and implied three distinct particle orientations were incompatible with packing in space group *R*3. The unit cell cross-section orthogonal to the rhombohedral axis could only accommodate one virus particle. A rhombohedral threefold would have to be internal to a particle and could not relate three separate particles within the cross-sectional plane. The diffraction threefold must therefore correspond to particles with three different tilts related by an aligned non-crystallographic threefold screw axis, not the 'pure' threefold (without a translational component) required of rhombodedral space groups. The alternative possibility that the corresponding hexagonal cell was not centered but contained $3_1$-related particles was eliminated by careful inspection of the diffraction images to verify that the 2/3 of the primitive lattice points predicted to be absent in the centered cell were truly absent. With only a threefold screw, the space group could at best be pseudo-rhombohedral and formally triclinic.

This unusual situation was cross-checked in several ways.

(i) Data processing was repeated in *P*1 with a single crystal to ensure that the threefold was not an artifact of improperly merging crystals.

(ii) The threefold symmetry was compared with Friedel symmetry using $R_{\rm merge}$: 10.3% compared with 9.1%. This represented an insignificant difference, indicating that the threefold was effectively exact. Although not constrained to be so, as far as we could tell the threefold was exactly along the body diagonal of the unit cell.

(iii) The possibility that twinning accounted for the apparent symmetry of the diffraction was eliminated by tests on the distribution of intensities (Yeates, 1997).

### 2.5. Reprocessing of the diffraction data

Laue diffraction symmetry is insensitive to translational components of symmetry elements. Thus, the impact upon diffraction of an exactly aligned threefold NCS screw axis is indistinguishable from that of a crystallographic threefold rotation. This was taken advantage of during scaling. The data were processed as if they were rhombohedral $R3$, imposing additional symmetry constraints. Observations that were related by the threefold were mapped to the $R3$ asymmetric unit and averaged. Following scaling, the data were threefold-expanded and reindexed to a triclinic $P1$ cell, with unit-cell parameters $a = 249.7$, $b = 249.7$, $c = 644.8$ Å, $\alpha = 90.0$, $\beta = 101.2$, $\gamma = 120.0^\circ$. Such expansion made the coverage of reciprocal space with diffraction observations more uniform and complete, reducing the deleterious effects of incomplete data on direct-space operations. (Note that the same amplitude estimates would have been generated later in the structure determination by the common process of 'filling' missing observations with those calculated by back-transformation of an NCS-averaged map.)

The consequences of inappropriately enforcing threefold projection symmetry down the diagonal of the pseudo-$R3$ unit cell would have been catastrophic. Therefore, many of the subsequent calculations were repeated using data from a single crystal processed in $P1$ without threefold expansion. Locked rotation functions and translation functions with unexpanded data were entirely consistent with those of the expanded data.

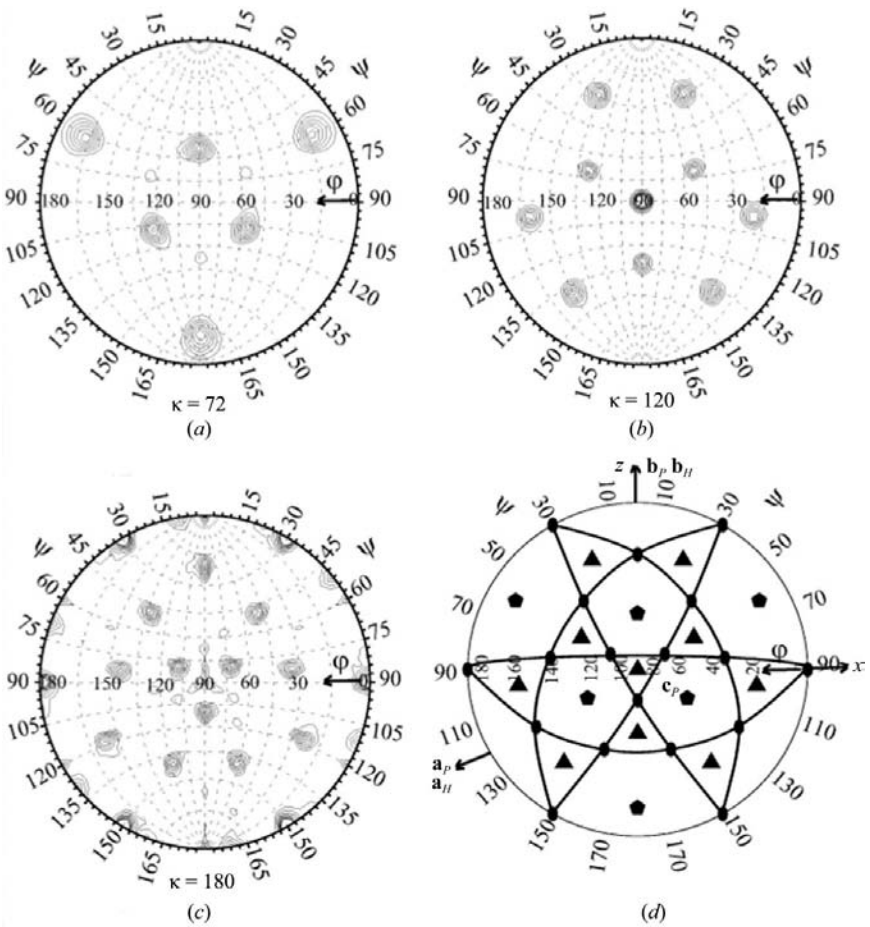


**Figure 2**
Spherical polar projection of AAV's ordinary (*a*)–(*c*) and locked (*d*) rotation functions at 14 Å. (*a*) shows the expected six icosahedral fivefolds for one hemisphere. (*b*) shows the ten icosahedral threefolds. The peak in the middle is stronger because it is coincident with the interparticle threefold screw axis. (*c*) shows the 15 icosahedral twofolds. (*d*) is the locked solution, searching for all symmetry elements at once. Pentagons represent fivefolds, triangles represent threefolds and ovals represent twofolds, which are connected with great circles. The directions of the crystallographic axes are as follows: $\mathbf{a}_P$, $\mathbf{b}_P$ and $\mathbf{c}_P$ for the $P1$ triclinic cell, and $\mathbf{a}_H$, $\mathbf{a}_H$, $\mathbf{b}_H$ for the hexagonal cell. The direction of the $\mathbf{c}_H$ axis is not shown as it is coincident with the central threefold.

### 2.6. Native Patterson and packing analysis

Approximate particle positions were determined with a native Patterson at 10 Å resolution (Fig. 4). As in prior virus structure determinations (Tsao *et al.*, 1992), this took advantage of the superposition of vectors between symmetry-equivalent atoms related by two operators: (i) an icosahedral rotation axis and (ii) a screw axis that is nearly parallel with approximately opposite rotational component. The result of applying both operations is nearly a pure translation, leading to the superposed Patterson vectors. Unlike past cases, the screw axis here is non-crystallographic, but the principle is the same.

For convenience in interpretation, the Patterson was calculated in the $H3$ hexagonal equivalent of the pseudo-$R3$ cell (Hahn, 1996, p78). Sections along the $H3$ $c$ axis were then parallel to planes of close-packed particles. The major peaks on sections $w = 1/3$ and $w = 2/3$ of about $100\sigma$ correspond to centering translations of the non-standard triply centered hexagonal cell and would be origin peaks in the real triclinic space group (Fig. 4). More interesting are the strong peaks in sections $w = 1/9$, $2/9$, $4/9\ldots$ of about $12\sigma$. These indicated (i) that there were nine layers in the $c = 1895$ Å $H3$ cell, (ii) that the particles were approximately spherically

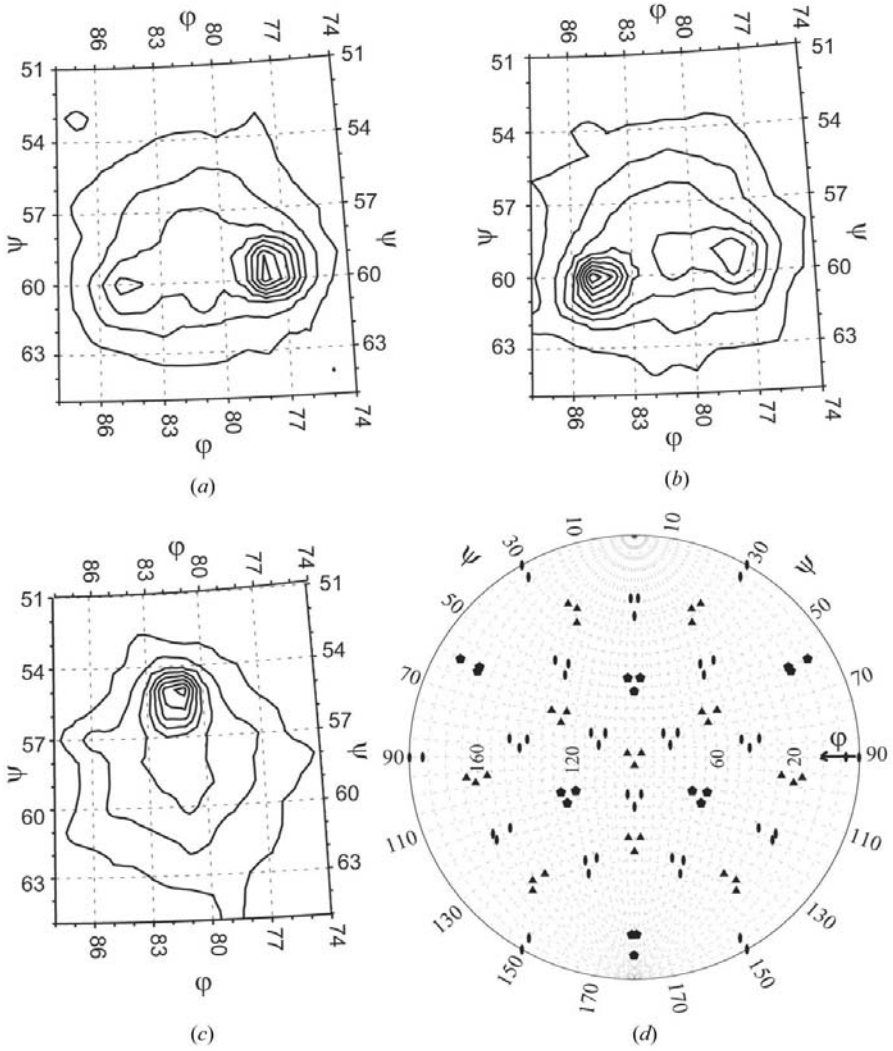


**Figure 3**
High-resolution (3.0 Å) locked self-rotation function search: (*a*)–(*c*) show the three solutions representing the spherical polar transformations required to rotate a standard set of icosahedral symmetry elements to the orientations of particles 1 to 3. (*a*), (*b*) and (*c*) show sections at $\kappa = 71$, 72 and 75°, respectively. (*d*) The full set of icosahedral symmetry elements for each of the solutions shown in (*a*)–(*c*) are shown in schematic form as in Fig. 2(*d*).

close-packed with particles of one layer sitting in one of two sets of gaps ('holes') between particles of the previous layer, (iii) that with height less than the origin peaks they corresponded to vectors between particles of only approximately the same orientation and (iv) that with Patterson vectors of (0, 0, 2/9), some particle centers separated by two layers were related by translation orthogonal to the close-packed planes (as can be seen in Fig. 5). The most important conclusions were that particles on neighboring layers had approximately (but not exactly) the same orientation and that particles three layers apart were of exactly the same orientation. The particles of this third layer sit not vertically above the particles of the zeroth layer (as required of a three-layer $P3_1$ cell), but above 'holes' in the zeroth layer, such that the interparticle vector is 79° from the hexagonal plan (not 90°), thereby forming a triclinic cell.

Peaks in the native Patterson became weaker beyond 10 Å resolution (owing to the imperfect alignment), limiting the precision of the particle positions. Other ambiguities remained. It was neither possible to distinguish between packing schemes that were enantiomers of each other nor to assign one of the three particle orientations (from the rotation function) to each of the three particle positions. Patterson analysis reduced the space that would have to be searched, but higher resolution methods were needed.

### 2.7. Translation function

Translation functions were initially calculated in three ways: using *CNS* (Brünger, 1997*a*, 1998) and *GLRF* (Tong, 1996) with a standard translation function and an *R*-factor search. In all three procedures, the search model was CPV, reoriented according to the AAV rotation function. The first particle was fixed at the origin. The second particle was located by searches with a partial model before a third particle could be added. Searches for particles 2 and 3 were then alternated. Searches were initially at 10 Å resolution, limited to regions close to those indicated by the Patterson analysis, and gave clear solutions. Recalculation at 3.0 Å resolution gave similar solutions, but the solutions for three methods differed from each other and from threefold projection symmetry by ∼2 Å, a precision inadequate for phase extension.

### 2.8. Constrained Patterson correlation search and refinement of particle positions

The standard methods lacked appropriate constraints to enforce the threefold projection symmetry expected from the diffraction symmetry. The searches were six-dimensional [($x$, $y$, $z$) coordinates for each of the two non-origin particles] with additional discrete ambiguities in assigning particle orientations for each position. An appropriately constrainable translation function would be four-dimensional. Two of the parameters would describe the ($x$, $y$) location of the threefold screw which would determine the ($x$, $y$) positions of the two non-origin particles. The other two parameters would describe the translations of the non-origin particles parallel to the threefold screw axis. Packing dictates that the spacing between layers be similar, but nothing constrains it to be exactly even, as in a crystallographic $3_1$ axis. Even inter-layer spacing was an optional constraint applied only at the start of the search, reducing it to two dimensions. Aside from the AAV-specific

**Table 3**
Rigid-body refinement of a CPV atomic model against AAV diffraction amplitudes.

| | | Change in orientation (°) | | | Change in position (Å) | | | | *R* factors | |
|---|---|---|---|---|---|---|---|---|---|---|
| Refinement type | Batch No. | $\theta_x$ | $\theta_y$ | $\theta_z$ | $x$ | $y$ | $z$ | Resolution (Å) | Start | End |
| Subunit | 1 | −0.77 | 0.94 | 0.0 | −0.62 | −0.01 | −0.10 | 15.0–4.0 | 0.4677 | 0.4631 |
| | 3 | −1.09 | 1.05 | −0.03 | 0.02 | 0.02 | −0.03 | 15.0–3.0 | 0.4682 | 0.4656 |
| | Overall | −1.87 | 1.99 | −0.03 | −0.60 | 0.01 | −0.13 | | | |
| Entire particle | 2 | −0.03 | 0.09 | −0.08 | 0.13 | −0.08 | | 15.0–4.0 | 0.4596 | 0.4593 |
| | 4 | 0.06 | 0.04 | −0.01 | 0.18 | 0.28 | | 15.0–3.0 | 0.4657 | 0.4653 |
| | Overall | 0.03 | 0.13 | −0.09 | 0.31 | 0.20 | | | | |

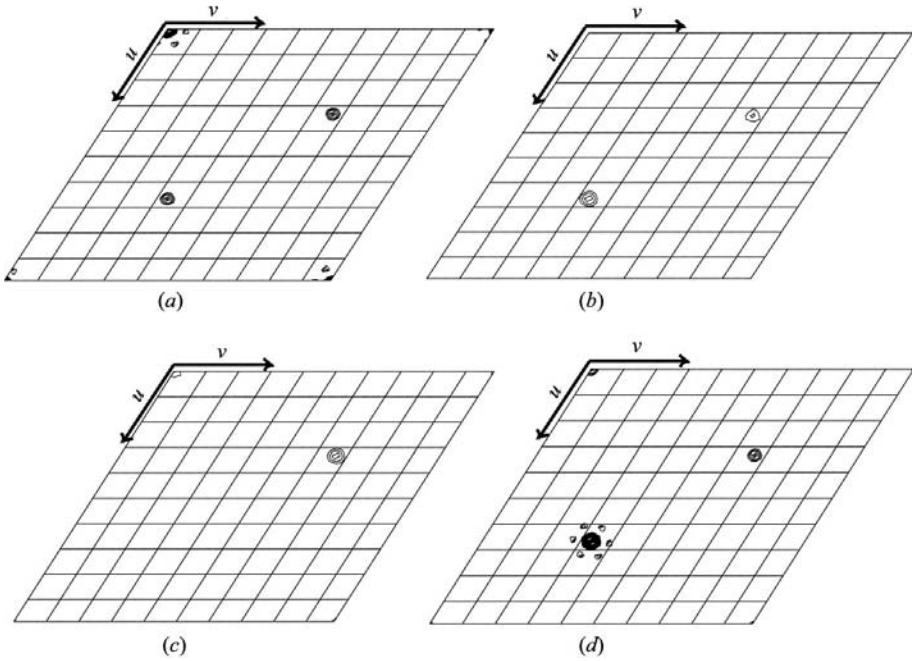


**Figure 4**
Native Patterson, calculated in the hexagonal *H*3 setting, equivalent to the pseudo-rhombohedral cell. (*a*), (*b*), (*c*) and (*d*) show sections at $w$ = 0, 1/9, 2/9 and 1/3, respectively. All are contoured from 6$\sigma$ in 2$\sigma$ steps. The 92$\sigma$ peak at (2/3, 1/3, 1/3) (which corresponds to an origin peak in the real unit cell) is surrounded by ripple peaks resulting from truncation at 10 Å resolution. The other 12$\sigma$ peaks are in positions compatible with approximate spherical close-packing.

constraints, the new program encoded a standard Patterson correlation search and refinement (Brünger, 1990, 1997*a*; Tong, 1996).

Even without constraints, the inter-layer spacing refined to within 0.1 Å of even spacing at 5 Å resolution. Constraints for a screw axis with even spacing moved the *xy* positions by ∼2 Å, with no significant change in correlation (0.196 compared with 0.197). Constrained steepest descent optimization moved the particle positions by 0.22 Å, improving the correlation for data to 5.0 Å (3.0 Å) to 0.201 (0.154) and crudely scaled *R* factors to 0.485 (0.494). Optimization close to the positions of an oppositely handed screw axis at 5 Å resolution led to inferior statistics (CC = 0.159), eliminating this possibility from consideration.

Rigid-body least-squares refinement was then applied to fit CPV atomic coordinates to the AAV diffraction amplitudes. The program *CNS* (Brünger *et al.*, 1998) was used with different constraints for alternate batches: (i) the position and orientation of an individual subunit was refined, with fixed 180-fold non-crystallographic symmetry between subunits, and (ii) *xy* positions and orientations of entire 60-fold symmetric particles were refined while threefold screw symmetry was constrained between the particles. Overall, the subunit was rotated by 2.7° and translated 0.6 Å and the particle was rotated by 0.16° and translated by 0.35 Å for a total decrease in *R* factor of 0.8% (Table 3). The second iteration moved the particle about 0.2 Å with a marginal change in *R* factor, indicating convergence and giving a sense of the final precision. One of the limitations of this refinement was the low sequence identity (23%; Chapman & Rossmann, 1993) and presumed poor structural homology between the CPV atomic model and AAV diffraction. However, this was the best that could be performed prior to the start of symmetry averaging.

Once symmetry averaging was under way (see §2.9), it was possible to further refine some of these parameters. Searches were made for the particle orientation and position that optimized the correlation coefficients between $F_{obs}$ and $F_{calc}$ from the back-transformed averaged map. (This converged more quickly than map correlation between $\rho_{obs}$ and $\rho_{calc}$.) The methods were similar to those of Rossmann *et al.* (1992), with the addition of optional constrains on the screw symmetry between particles and the inter-layer spacing. The process was attempted at resolutions of 15.0, 9.0, 6.0 and 3.75 Å as phase refinement proceeded (Table 4). Map and reciprocal-space correlations at 15 Å gave different optimal positions and were ignored. The correlation coefficients were calculated after eight cycles of averaging, which were sufficient to remove bias from previous parameters used in averaging. Each parameter was refined in turn, with full convergence in a single round. The interlayer spacing stayed within 0.1 Å of exact $3_1$ symmetry, which was considered insignificant evidence of deviation. By 3.75 Å resolution, refinement had converged. Overall, the shift was about 0.4 Å in position and 0.1° in orientation.

**Table 4**
Refinement of symmetry operators during phase refinement.

| | Change relative to rigid refinement of CPV model *versus* AAV data | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Orientation (°) | | | Position of screw axis (Å) | | | Patterson Correlation coefficient (Brünger, 1997) | |
| Resolution (Å) | $\theta_x$ | $\theta_y$ | $\theta_z$ | $x$ | $y$ | Correlation coefficient maximized | Start | End |
| 40–15 | +0.40 | +0.27 | −0.30 | −0.80 | +1.00 | Real-space | 0.574 | 0.586 |
| 40–15 | −0.05 | +0.27 | +0.20 | +0.50 | +0.10 | Reciprocal space | 0.545 | 0.547 |
| 40–9.0 | +0.01 | −0.05 | +0.05 | +0.32 | +0.42 | Real space | 0.682 | 0.692 |
| 40–9.0 | −0.01 | −0.05 | +0.05 | +0.45 | +0.25 | Reciprocal space | 0.621 | 0.623 |
| 40–6.0 | −0.10 | −0.03 | +0.08 | +0.25 | +0.35 | Real space | 0.713 | 0.716 |
| 40–6.0 | −0.01 | −0.07 | +0.05 | +0.35 | +0.30 | Reciprocal space | 0.672 | 0.673 |
| 40–3.75 | −0.01 | −0.09 | +0.05 | +0.20 | +0.30 | Real space | 0.650 | 0.650 |
| 40–3.75 | −0.01 | −0.09 | +0.05 | +0.21 | +0.30 | Reciprocal space | 0.617 | 0.617 |

**Table 5**
Statistics from selected points in phase refinement by symmetry averaging.

The reciprocal-space correlation coefficient is between $F_o$ and $F_{\langle map\rangle}$. The real-space correlation is between the densities of NCS equivalent points in a map with coefficients $2F_o - F_{\langle map\rangle}$, $\varphi_{\langle map\rangle}$. $F_{\langle map\rangle}$ and $\varphi_{\langle map\rangle}$ are Fourier terms of the back-transformed averaged map.

| | | | | Correlation coefficients | |
|---|---|---|---|---|---|
| Batch No. | Comments | Resolution limit (Å) | *R* factor for the back-transformed map | Reciprocal space | Real-space (map) |
| 1 | | 15.0 | 0.389 | 0.559 | 0.575 |
| 12 | | 9.09 | 0.329 | 0.621 | 0.680 |
| 13 | After refinement of averaging parameters | 8.77 | 0.314 | 0.655 | 0.711 |
| 26 | | 6.02 | 0.315 | 0.666 | 0.712 |
| 27 | After refinement of averaging parameters | 5.88 | 0.303 | 0.690 | 0.720 |
| 51 | | 3.76 | 0.326 | 0.601 | 0.641 |
| 52 | After refinement of averaging parameters | 3.70 | 0.316 | 0.627 | 0.652 |
| 67 | | 3.03 | 0.332 | 0.641 | 0.613 |
| 68 | Filling missing observations | 3.70 | 0.314 | 0.630 | 0.759 |
| 83 | Filling missing observations | 3.03 | 0.321 | 0.634 | 0.775 |
| 116 | After three rounds of atomic mask improvement | 3.03 | 0.315 | 0.647 | 0.781 |
| 117 | Starting with atomic model phases | 3.70 | 0.307 | 0.648 | 0.765 |
| 132 | Extended from model phases by averaging | 3.03 | 0.314 | 0.651 | 0.782 |

After building an AAV atomic model, it was possible to return to rigid-body refinement with a more appropriate atomic model. An *R*-factor search, constrained by screw symmetry, translated the particles 0.1 Å and rotated them 0.01°, with a marginal change in *R* factor (0.349 to 0.346). This indicated that the refinement for optimal averaging parameters had converged upon the correct position and orientation.

### 2.9. Phase improvements and extension

Phases calculated from the homologous model would be refined at low resolution and then extended to high resolution using the 180-fold non-crystallographic symmetry. Initial attempts starting at 6.0 Å resolution using the *DM* program (Cowtan & Main, 1993) and CPV or DNV phasing models resulted in model-biased maps, the reasons for which remain unclear. Subsequent attempts with *RAVE*/*CCP*4 (Jones, 1992; Collaborative Computational Project, Number 4, 1994) started conservatively at 15 Å resolution with a CPV backbone phasing model that lacked molecular detail. The appearance of side chains with the AAV sequence therefore provided strong validation of the process.

Averaging and phase extension to 3.0 Å were performed in 67 steps (of at least six cycles per step), in which the increment corresponded to approximately half the longest reciprocal-lattice unit. In the first extension, missing observations were omitted. After reaching 3.0 Å, missing observations between 40 and 3.75 Å resolution were 'filled' (Chapman *et al.*, 1992) with estimates calculated from the back-transformed averaged map. These were incorporated in a new extension to 3.0 Å resolution.

Most prior capsid structure determinations have involved only closed point group non-crystallographic symmetry, so masks could define the entire assembly (Vellieux & Read, 1997). With several particles per asymmetric unit, it is a subunit mask that must be defined. The unaveraged map was not of sufficient quality to demarcate the subunit boundary, so the mask was based on the homologous CPV and DNV subunit structures. With their low sequence identity (∼20%), a generous 3.5–4.0 Å margin was allowed around all atoms to allow for structural differences. There was extensive overlap between the masks of subunits from neighboring particles that is now seen in retrospect to be partly a consequence of the entirely different surface-loop structures of AAV, CPV and DNV. The presence of high-order NCS and absence of crystallographic symmetry led to a large number of different inter-particle contact surfaces with which the mask should be as consistent as possible. Overlap was removed in ten progressively stringent iterations of the program *MAMA* (Kleywegt & Jones, 1994), in which pixels whose count of symmetry equivalents most exceeded the number of NCS operators were removed first. The masks were a compromise between overlap removal and coverage of the CPV/DNV models. The mask used until 9 Å resolution had 10%

remaining overlap, while that used at 3 Å had only 1.3% overlap; both covered 94% of the CPV/DNV atoms.

Our first atomic model was missing several loops whose density protruded outside this mask. New masks were made from the most recent AAV model. Spheres (10 Å in radius) were added to the mask where missing residues were predicted to be. Removal of the worst overlap left a mask with 2.1% overlap that covered 96% of the AAV model. With three iterations of mask improvement, phase refinement and atomic modeling, all of the protruding loop density was recovered. The correlation coefficient between densities of equivalent regions in the final averaged map was 0.781 and the correlation between $F_{\rm obs}$ and structure amplitudes calculated by back-transformation from the averaged map was 0.647 (Table 5). Lest phases in some regions of reciprocal space had converged upon Babinet or enantiomeric opposites (Valegård *et al.*, 1991; Chapman *et al.*, 1998; Munshi *et al.*, 1998), phase refinement was restarted from the AAV model phases. The lack of improvement (Table 5) indicated that the phases were already self-consistent. Maps and phases from the restarted phase refinement were discarded to eliminate any question of model bias.

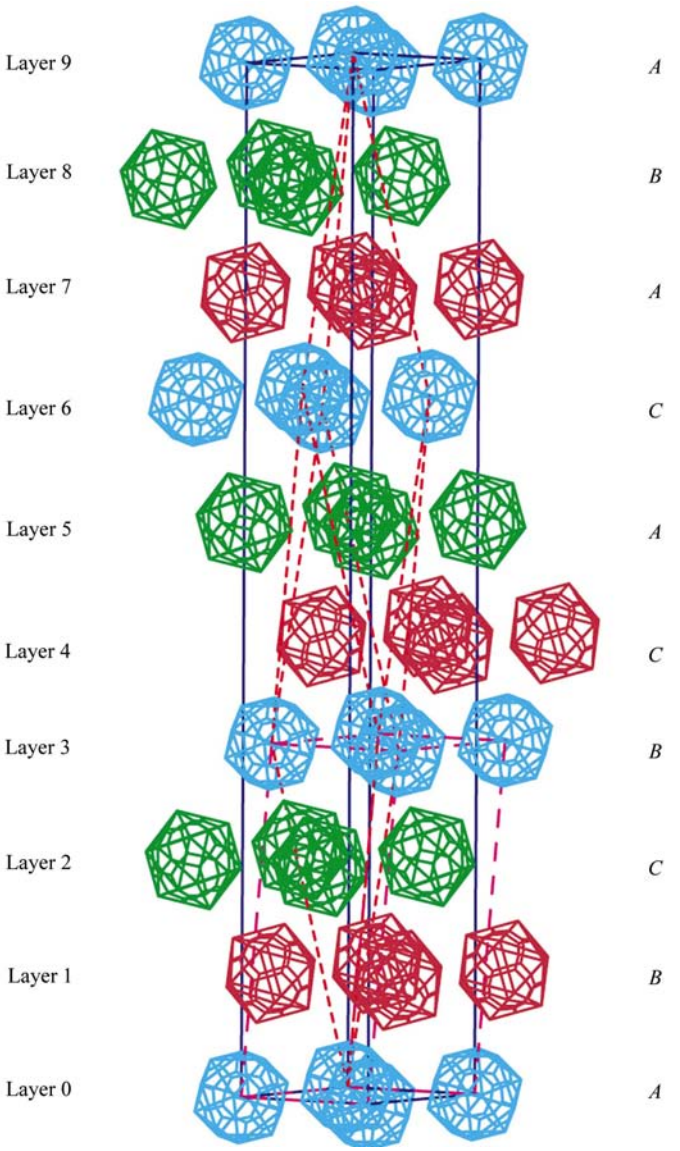


**Figure 5**
Packing diagram for AAV form 2 crystals. The *P*1 unit cell is outlined by a dashed purple line, the pseudo-rhombohedral cell by a dotted red line and the equivalent hexagonal setting by a solid blue line. Icosahedra representing AAV particles are reduced in size for clarity and colored to represent different orientations. Equivalent unit cells can be interconverted with the transformations $\mathbf{a}_H = \mathbf{c}_R - \mathbf{a}_R$; $\mathbf{b}_H = \mathbf{a}_R - \mathbf{b}_R$; $\mathbf{c}_H = \mathbf{a}_R + \mathbf{b}_R + \mathbf{c}_R$; $\mathbf{a}_T = \mathbf{a}_H$; $\mathbf{b}_T = \mathbf{b}_H$; $\mathbf{c}_T = \mathbf{c}_R$, where the subscripts *T*, *R* and *H* refer to triclinic, pseudo-rhombohedral and hexagonal triple cells, respectively. The positions within the planes orthogonal to the pseudo-rhombohedral axis are designated *A*, *B* or *C*. Particles in layers labeled with the same letter are 'vertically' above each other. Sometimes the positions are alternating as in hexagonal close packing, but this is interrupted with layer dislocations that lead to three consecutive layers with different positions, characteristic of cubic close packing. Thus, it is a hybrid between cubic and hexagonal close packing in which the combination of position and orientation repeat with a periodicity of nine layers.

### 2.10. Model building and refinement

An approximate model was fitted into the density using *O* (Jones & Kjeldgaard, 1997). Clear electron density then followed from residue 15 to the C-terminus (a total of 519 residues) without a break or ambiguity as to how the chemical sequence should be aligned. No attempt was made to model the 14 N-terminal residues owing to the diffuse density. Although only at 3.0 Å resolution, the density was of excellent quality owing to the phase refinement/averaging with high-order NCS (Fig. 6).

A preliminary refinement has been completed, mostly using our real-space extension of the *CNS* refinement program (Brünger *et al.*, 1998; Korostelev *et al.*, 2002). This protocol supports fully restrained torsion-angle simulated-annealing refinement of local regions. Slow-cooling annealing protocols were applied to three or four residues at a time, selecting annealing temperatures of 2000–35 000 K and weights of 50.0 to 80.0 on the experimental term, based on trial-and-error and inspection in *O*. Upon completion, discontinuities were smoothed by real-space refining the whole model by slow-cooling from 6000 to 500 K with a weight of 80.0. The combined 'manual' rebuilding and local real-space annealing refinement was a single step, reducing *R* from 0.465 to 0.349. Subsequent rigid-body refinement against amplitudes in *CNS* only decreased *R* to 0.346 (see above). The final model was obtained after highly restrained *B*-factor maximum-likelihood refinement against diffraction amplitudes in *CNS* (Pannu & Read, 1996; Adams *et al.*, 1997) while constraining exact 180-fold non-crystallographic symmetry.

The final $R/R^{\rm free}$ was 0.338/0.342. The root-mean-square (r.m.s.) deviation from ideal bond lengths was 0.020 Å and that from ideal bond angles was 2.2°; only 3% of $\varphi$, $\psi$ angles were in generously allowed regions (none were in disallowed regions; Laskowski *et al.*, 1993). Threefold relatives of test reflections were omitted from phase extension and refinement. However, with such high NCS redundancy it was impractical to exclude all neighbors of off-lattice points related to test reflections by NCS. (The alternative of omitting resolution

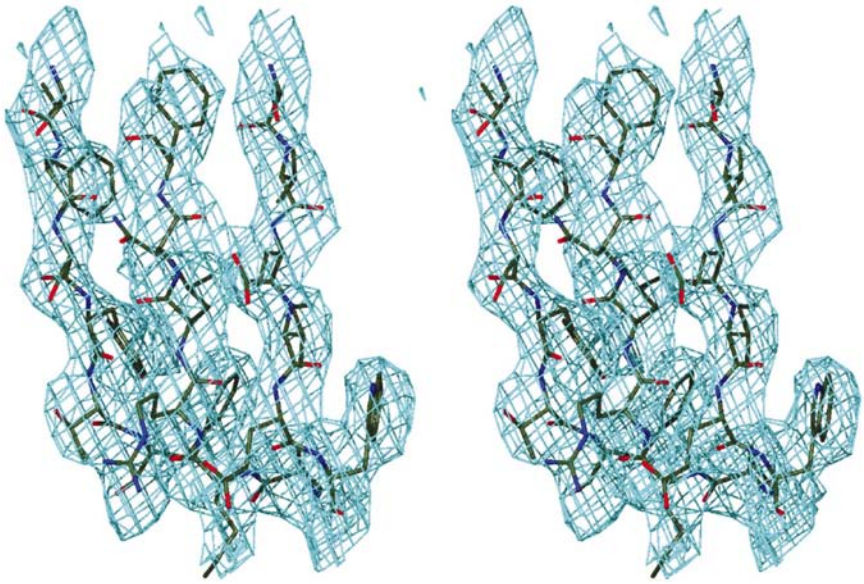

**Figure 6**
Electron density for a typical part of the $\beta$-barrel. The figure is shown in stereo with density contoured at 0.8$\sigma$ using the programs *O* and *MOLRAY* (Jones & Kjeldgaard, 1997; Harris & Jones, 2001) and clearly shows the fit of aromatic side chains.

shells would, it was feared, have frustrated phase determination, which was entirely dependent on slow extension in steps of approximately half the reciprocal-lattice unit.) The exclusion of just threefold relatives was not sufficient to eliminate bias from $R^{\rm free}$. However, the high data-to-parameter ratio achieved with high NCS redundancy eliminates the possibility of significant over-fitting with tight stereochemical restraints (Brünger, 1997*b*).

## 3. Discussion

The structure determination of AAV presented unusual challenges requiring non-standard responses, the success of which can now be examined with hindsight.

Work with cryofrozen samples of potentially biohazardous materials requires appropriate containment to avoid the risks of aerosols should failure of cooling lead to sample thawing. With increased BL2 availability, synchrotron data collection is not the problem that it was. For preliminary laboratory characterization, the choice is installation of diffraction equipment in a biohazard laboratory or of containing just the sample space on top of the X-ray generator. An apparatus developed for this work that was consistent with usual methods of cryocooling is described in the supplementary material[1].

The strategies for dealing with the large and anisotropically shaped unit cell met with mixed success. Imaging plates (IPs) might have achieved greater coverage of resolution than CCDs, but the recycling of IP cassettes, scanned off-line, would have limited the number of exposures collectable from each frozen crystal. The fine spot separation of some crystal orientations was resolved with a detector distance of 500 mm (off-centering the detector to cover part of the high-resolution diffraction). For other crystal orientations, the 300 mm distance captured 2.6 Å resolution data with well resolved spots. With a long cell axis roughly parallel to the beam, lunes overlapped and autoindexing of many of these images with several programs failed (Leslie, 1993; Powell, 1999; Rossmann & van Beek, 1999; Otwinowski & Minor, 2001). The problem was not immediately appreciated, as on-line processing was not available at the time, so the opportunity to index these through 'setting images' at other crystal orientations was lost.

AAV joins a growing list of virus crystals with pseudosymmetry (Tsao *et al.*, 1992; Muckelbauer *et al.*, 1996; Zlotnick *et al.*, 1997; Filman *et al.*, 1998; Natarajan & Johnson, 1998). The underlying reasons are likely to be the same as give rise to multiple crystal forms for several viruses (Tsao *et al.*, 1992; Wu & Rossmann, 1993). The near-spherical shape leads to high-symmetry packing schemes that can be slightly perturbed by non-spherically symmetric features. Adoption of the same or similar contact interfaces with several neighboring particles can cause the icosahedral symmetry of the particles to be aligned exactly or approximately with the packing symmetry (Natarajan & Johnson, 1998). Prior examples have included near-centering and near-alignment of icosahedral axes with the lattice. Here, we see exact alignment of a non-crystallographic packing symmetry with a lattice diagonal. As with echovirus 1, the true space group was not obvious until relatively late in the structure determination and became apparent only by integrating the results of molecular replacement with packing considerations (Filman *et al.*, 1998).

Averaging and expansion of the threefold diffraction symmetry resulting from the non-crystallographic projection symmetry within the triclinic space group was unorthodox. There was much to be gained and nothing to be lost by this approach. The threefold screw axis could never have contributed to the phasing power, which came from just the 60-fold internal viral symmetry. The reason is apparent by considering an alternate choice of *P*1 cell with principal axes along the edges of a rhombohedron (Fig. 5). The phasing power of non-crystallographic symmetry comes from the over-sampling created by mapping amplitudes from one equivalent to off-lattice points in a symmetry-equivalent region of reciprocal space (Tong & Rossmann, 1995; Chapman, 1998). For a non-crystallographic screw axis across the diagonal of the real-space cell, there would be a corresponding threefold rotation axis in reciprocal space that would map lattice point to lattice point, not generating any over-sampling or phasing power. The only effects would be to average out errors from related

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: SX0053). Services for accessing this material are described at the back of the journal.

diffraction amplitudes and to fill in for missing observations. The benefits of this are greater if applied in data processing so that the increased number of symmetry equivalents improves scaling and post-refinement.

Although differences from threefold diffraction symmetry are not detectable statistically, there is a question as to whether small differences might result from subtle deviations between the particles in the crystal that could in principle be refined against unaveraged structure amplitudes. There are several indications that this would not be possible. NCS-constrained refinement of individual atomic positions gave essentially the same results whether expanded or unexpanded data were used. NCS-unconstrained refinement of individual atomic positions against medium-resolution data would lead to gross over-fitting. Rigid-group refinement was possible, allowing icosahedral particles to deviate in position and orientation from exact threefold screw symmetry, but no improvement was seen. Refinement against unexpanded data also restricts the available data to a single crystal. There are three orientations of AAV's triclinic lattice that can be chosen for each crystal. Without any statistically significant difference between the threefold-related intensities, there is no way to determine if the same choice of lattice has been made for each crystal. Thus, the merging of data from different crystals implicitly assumes threefold symmetry. Once assumed, there are then advantages to explicit expansion to fill missing reflections in terms of map quality. Imposition of strict NCS constraints also allows real-space refinement (Korostelev *et al.*, 2002) of an icosahedral asymmetric unit, using wide-convergence torsion-angle simulated-annealing methods (Rice & Brünger, 1994) that would not have been tractable for a full million-atom crystallographic asymmetric unit on the computers that were available.

Considering the fine spot separation, reduced background areas to limit overlap and the difficulties of processing relatively weak diffraction images, the quality of integrated data compares favorably with other medium-resolution virus structure determinations (Tate *et al.*, 1999; Grimes *et al.*, 1998; Llamas-Saiz *et al.*, 1997). However, compared with protein data, the values of $R_{\rm merge}$ of this and other virus crystals probably overstate the quality. Virus structures are often solved with data that is even more incomplete than that of AAV (80% at 3.5 Å); for example, canine parvovirus (61%; Tsao *et al.*, 1991), *Galleria* densovirus (58%; Simpson *et al.*, 1998) and bluetongue virus (54%; Grimes *et al.*, 1998). Redundancy often drops precipitously with resolution (Fig. 7), so that there are few redundant high-resolution reflections, leaving $R_{\rm merge}$ disproportionately reflecting the higher quality low-resolution data.

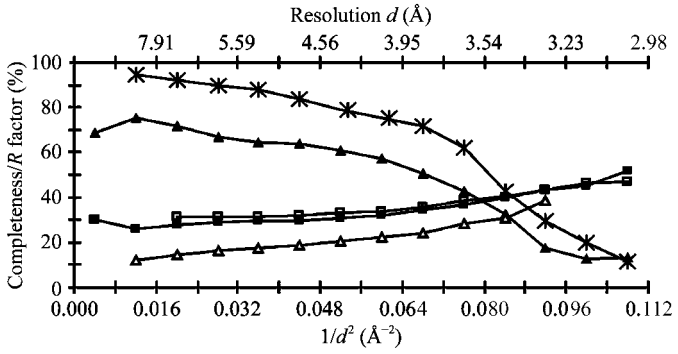


**Figure 7**
Summary statistics for symmetry averaging, data collection and atomic model. Stars, percentage of reflections; filled triangles, correlation coefficient for symmetry averaging; filled squares, *R* factor for symmetry averaging; squares, *R* factor for the final atomic model; triangles, $R_{\rm merge}$ (see Table 2).

The presence of three particles complicated molecular replacement. It did not affect high-resolution determination of the particle orientations: the rotation-function values were within 0.2° of the final values. This corresponds to a shift of about 0.2 Å averaged over the capsid, a precision achieved by locking the 532 icosahedral symmetry with *GLRF* (Tong & Rossmann, 1997). Approximate particle positions were, as in a few previous cases (Tsao *et al.*, 1992; Tars *et al.*, 2000), apparent from native Patterson functions owing to near-alignment of icosahedral and inter-particle symmetry axes. The 2 Å discrepancy between various standard translation functions is likely to have arisen from the low homology of the search model (23% sequence identity) and the need to search independently for >1 particle position. The dimensionality of the search was reduced using the constraints of the non-crystallographic threefold screw axis, but this required the programming of a tailor-made translation function.

Earlier versions of the real-space refinement used here were developed to make the refinement of large symmetric assemblies more efficient and computationally tractable (Chapman, 1995; Chapman & Rossmann, 1996). They have the advantage that only one of the icosahedrally symmetric units of a virus need be considered and an analogous approach has been adopted by Hogle and colleagues (Jacobson *et al.*, 1996). While real-space refinement was an efficient approach for the bulk of refinement, a final batch of reciprocal-space refinement sometimes lowers the final *R* factor slightly (Blanc *et al.*, 2001). The latest version of our real-space refinement supports a more powerful local simulated-annealing refinement (Korostelev *et al.*, 2002) that, even with improved computers, would still be intractable in reciprocal space for such a large complex. In the case of AAV, essentially no further improvement upon real-space simulated-annealing refinement was possible by reciprocal-space conjugate-gradient refinement.

Overall quality appears to be limited by phase extension/averaging or some upstream steps. The *R* factors for the atomic model (0.338) and for the back-transformed averaged map (0.325) are similar, indicating little scope for further improvement of the protein model with the current data and confirming that there has been little over-fitting. Averaging statistics (Fig. 7) have been better, especially for viruses diffracting more strongly at high resolution, but the statistics of AAV are not atypical and are similar to those for minute virus of mouse (MVM), echovirus 1 and bacteriophage PP7, for example (Llamas-Saiz *et al.*, 1997; Filman *et al.*, 1998; Tars *et al.*, 2000). Possible limitations include the quality of the

diffraction data, small deviations from the constrained $3_1$ non-crystallographic symmetry that are not yet detectable or the lack of a model for partially ordered nucleic acid inside the virus particles.

The structure, while only at medium resolution, has been sufficient to address some of the important issues in AAV biology and the design of improved gene-therapy vectors. AAV shares a $\beta$-barrel subunit topology, but differs from its homologs CPV and DNV (Tsao *et al.*, 1991; Simpson *et al.*, 1998) in the loop structures between $\beta$-strands and in the presence of peaks on the outer surface surrounding each threefold axis. Reliable mapping of the locations of amino acids and regions of the sequence previously associated with various phenotypes has been enough to determine that the likely sites of attachment to its cellular receptor are on the side of the peaks and to show that nearby regions are also antigenic (Xie *et al.*, 2002). This is prerequisite information for the design of vectors with altered cellular targeting and that can avoid immediate immune neutralization.

The assistance is gratefully acknowledged of Joan Hare in virus preparation and Genfa Zhou, Eric Blanc and Arezki Azzi in synchrotron data collection. Thanks are also due to Todd Yeates for running twinning tests and to Andrew Korostelev for assistance with the simulated-annealing real-space refinement. This work was funded by the American Cancer Society (RPG-99-356-01-GMC), the National Institutes of Health (GM 66875) and a pre-doctoral fellowship to Smita Bhatia from the Florida Division of the American Heart Association.

## References


Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.

Blanc, E., Zhou, G., Chen, Z., Xie, Q., Tang, J., Wang, J. & Chapman, M. S. (2001). *Crystallographic Computing 7: Proceedings of the IUCr Macromolecular Computing School, 1996*, edited by K. D. Watenpaugh & P. E. Bourne. Oxford University Press.

Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.

Brünger, A. T. (1997*a*). *Methods Enzymol.* **277**, 366–396.

Brünger, A. T. (1997*b*). *Methods Enzymol.* **276**, 558–580.

Brünger, A. T., Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Carter, B. (1990). *Handbook of Parvoviruses*, edited by P. Tijssen, pp. 247–284. Boca Raton, FL, USA: CRC Press.

Carter, B. & Flotte, T. (1996). *Curr. Top. Microbiol. Immunol.* **218**, 119–144.

Caspar, D. L. D. & Klug, A. (1962). *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.

Chapman, M. S. (1995). *Acta Cryst.* A**51**, 69–80.

Chapman, M. S. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 99–108. Dordrecht: Kluwer Academic Press.

Chapman, M. S., Blanc, E., Johnson, J. E., McKenna, R., Munshi, S., Rossmann, M. G. & Tsao, J. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 433–442. Dordrecht: Kluwer Academic Press.

Chapman, M. S. & Rossmann, M. G. (1993). *Virology*, **194**, 491–508.

Chapman, M. S. & Rossmann, M. G. (1996). *Acta Cryst.* D**52**, 129–142.

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. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.

Filman, D. J., Wien, M. W., Cunningham, J. A., Bergelson, J. M. & Hogle, J. M. (1998). *Acta Cryst.* D**54**, 1261–1272.

Flotte, T., Carter, B., Conrad, C., Guggino, W., Reynolds, T., Rosenstein, B., Taulor, G., Walden, S. & Wetzel, R. (1996). *Human Gene Therapy*, **7**, 1145–1159.

Grimes, J. M., Burroughs, J. N., Goued, P., Diprose, J. M., Malby, R., Zientara, S., Mertens, P. P. & Stuart, D. I. (1998). *Nature (London)*, **395**, 470–478.

Hahn, T. (1996). *International Tables for Crystallography*, Vol. A, 4th ed., p. 78. Dordrecht: Kluwer Academic Publishers.

Harris, M. & Jones, T. A. (2001). *Acta Cryst.* D**57**, 1201–1203.

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

Jones, T. A. (1992). *Proceedings of the CCP4 Study Weekend. Molecular Replacement*, edited by E. Dodson, S. Gover & W. Wolf, pp. 91–105. Warrington: Daresbury Laboratory.

Jones, T. A. & Kjeldgaard, M. (1997). *Methods Enzymol.* **277**, 173–208.

Kleywegt, G. J. & Jones, T. A. (1994). *Proceedings of the CCP4 Study Weekend. From First Map to Final Model*, edited by S. Bailey, R. Hubbard & D. Waller, pp. 59–66. Warrington: Daresbury Laboratory.

Korostelev, A., Bertram, R. & Chapman, M. S. (2002). *Acta Cryst.* D**58**, 761–767.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Leslie, A. G. W. (1993). *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey. Warrington: Daresbury Laboratory.

Llamas-Saiz, A. L., Agbandje-McKenna, M., Wikoff, W. R., Rossmann, M. G., Bratton, J. & Tattersall, P. (1997). *Acta Cryst.* D**53**, 93–102.

Luo, M., Tsao, J., Rossmann, M. G., Bassak, S. & Compans, R. W. (1988). *J. Mol. Biol.* **200**, 209–211.

McKenna, R., Xia, D., Willingham, P., Ilag, L. L., Krishnaswarmy, S., Rossmann, M. G., Olson, N. H., Baker, T. S. & Incardona, N. (1992). *Nature (London)*, **355**, 137–143.

Moskalenko, M., Chen, L., van Roey, M., Donahue, B. A., Snyder, R. O., McArthur, J. G. & Patel, S. D. (2000). *J. Virol.* **74**, 1761–1766.

Muckelbauer, J. K., Kremer, M., Minor, I., Tong, L., Zlotnick, A., Johnson, J. E. & Rossmann, M. G. (1996). *Acta Cryst.* D**51**, 871–887.

Munshi, S., Liljas, L. & Johnson, J. E. (1998). *Acta Cryst.* D**54**, 1295–1305.

Muzyczka, N. & Berns, K. I. (2001). *Virology*, 4th ed., edited by B. N. Fields, D. M. Knipe & P. M. Howley, pp. 2327–2360. Philadelphia: Lippincott, Williams & Wilkins.

Natarajan, P. & Johnson, J. E. (1998). *J. Struct. Biol.* **121**, 295–305.

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

Otwinowski, Z. & Minor, W. (2001). *International Tables for Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold. Dordrecht: Kluwer Academic Publishers.

Pannu, N. S. & Read, R. J. (1996). *Acta Cryst.* A**52**, 659–668.

Powell, H. R. (1999). *Acta Cryst.* D**55**, 1690–1695.

Rice, L. M. & Brünger, A. T. (1994). *Proteins*, **19**, 277–290.

Rodgers, D. W. (1994). *Structure*, **2**, 1135–1140.

Rossmann, M. G. (2001). *Acta Cryst.* D**57**, 1360–1366.

Rossmann, M. G., Leslie, A. G. W., Abdel-Meguid, S. S. & Tsukihara, T. (1979). *J. Appl. Cryst.* **12**, 570–581.

Rossmann, M. G., McKenna, R., Tong, L., Xia, D., Dai, J.-B., Wu, H., Choi, H.-K. & Lynch, R. E. (1992). *J. Appl. Cryst.* **25**, 166–180.

Rossmann, M. G. & van Beek, C. G. (1999). *Acta Cryst.* D**55**, 1631–1640.

Rutledge, E., Halbert, C. & Russell, D. (1998). *J. Virol.* **72**, 309–319.
Simpson, A. A., Chipman, P. R., Baker, T. S., Tijssen, P. & Rossmann, M. G. (1998). *Structure*, **6**, 1355–1367.
Tars, K., Fridborg, K., Bundule, M. & Liljas, L. (2000). *Acta Cryst.* D**56**, 398–405.
Tate, J., Liljas, J., Christian, P., Lin, T. & Johnson, J. E. (1999). *Nature Struct. Biol.* **6**, 767–773.
Tong, L. (1996). *Acta Cryst.* A**52**, 476–479.
Tong, L. & Rossmann, M. G. (1995). *Acta Cryst.* D**51**, 347–353.
Tong, L. & Rossmann, M. G. (1997). *Methods Enzymol.* **276**, 594–611.
Tsao, J., Chapman, M. S., Agbandje, M., Keller, W., Smith, K., Wu, H., Luo, M., Smith, T. J., Rossmann, M. G., Compans, R. W. & Parrish, C. (1991). *Science*, **251**, 1456–1464.
Tsao, J., Chapman, M. S., Wu, H., Agbandje, M., Keller, W. & Rossmann, M. G. (1992). *Acta Cryst.* B**48**, 75–88.
Valegård, K., Liljas, L., Fridborg, K. & Unge, T. (1991). *Acta Cryst.* B**47**, 949–960.
Vellieux, F. M. D. & Read, R. J. (1997). *Methods Enzymol.* **277**, 18–53.
Wikoff, W. R., Duda, R. L., Hendrix, R. W. & Johnson, J. E. (1999). *Acta Cryst.* D**55**, 763–771.
Wu, H. & Rossmann, M. G. (1993). *J. Mol. Biol.* **233**, 231–244.
Xie, Q., Bu, W., Bhatia, S., Hare, J., Somasundaram, T., Azzi, A. & Chapman, M. S. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 10405–10410.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.
Zlotnick, A., Natarajan, P., Munshi, S. & Johnson, J. (1997). *Acta Cryst.* D**53**, 738–746.