Supporting Information

Aksyuk et al. 10.1073/pnas.1207730109

SI Text

SI Materials and Methods. ***Virus purification.*** Group C *streptococcus* 26RP66 were propagated in the chemically defined medium (JHP Pharmaceuticals) (27.12 g/L), supplemented by sodium bicarbonate (2.5 g/L), cystein (0.5 g/L), and Todd Hewitt Broth (1% w/v). During the early log phase of bacterial growth, which corresponded to absorption of about 0.22 at 600 nm, the bacterial culture was infected with the C1 phage stock at a ratio of 1:10 (v/v). The virus was incubated with cells until the solution became clear, indicating complete lysis. The lysate was then filtered twice through 0.2 μm α-PES filter (Millipore). The lysate, intended for subsequent infection of the cells, was incubated at 50 °C for 10 min to inactivate phage lysine and then stored at 4 °C. The virus used for the structural work was not heated to 50 °C, because it was observed that some particles were damaged after heating. Virus was then pelleted by ultracentrifugation at 100,000 × *g* for 2 h and resuspended in buffer A, containing 20 mM Tris•HCl, 0.2 M sodium chloride, and 10 mM magnesium sulfate. The bacteriophage was further concentrated by an iodixanol (OptiPrep) cushion (50% iodixanol in buffer A described above). The virus band was then dialyzed against buffer A and purified by iodixanol gradient centrifugation (15%–35% iodixanol in buffer A). The virus was then buffer-exchanged using concentrators with 100 kDa cut off (Amicon).

Cryo-EM imaging and reconstruction. The virus was vitrified in liquid ethane on a holey grid (1). The images of C1 virus particles were collected on an FEI Titan Krios electron microscope at a magnification of 56000 × under low-dose conditions (approximately 25 electrons/Å^2). Micrographs were digitized using a Nikon scanner with a step size of 6.35 μm. The individual particles were selected using the E2BOXER program from the EMAN2 software package (2). The contrast transfer function parameters were determined by the program CTFIT, also from the EMAN software package. The initial icosahedral model of the C1 capsids was created from boxed particles with the best fivefold, threefold, and twofold rotational symmetry. About 17,000 particles were used for the icosahedral reconstruction, which was calculated to 8.3 Å resolution, judged by Fourier shell correlation (FSC) between two half-datasets using a threshold of 0.5. However, the tail portion has sixfold symmetry and, thus, would have been smeared out. Hence, to calculate a separate reconstruction of the tail, about 5,000 images that represented side views of the virus were selected. The 12-fold and sixfold reconstructions of the tail produced similar results, showing a tail with a terminal knob and twelve appendages at the head-tail junction. Using the orientations obtained in this reconstruction, a whole virus reconstruction was calculated, which included the head portion for these selected particles. Because of the symmetry mismatch the head did not have fivefold symmetry. This model was used for a reconstruction, assuming no symmetry using all the particles. After 20 cycles this reconstruction converged and the resulting capsid had approximate fivefold symmetry, whereas the tail retained 12 densities representing the appendages. Thus, this asymmetric reconstruction showed the relative orientation of the fivefold symmetric head and the sixfold symmetric tail. Using this reconstruction as a model, a fivefold reconstruction was performed. Due to the symmetry mismatch between the tail and the head, the fivefold reconstruction resulted in a smooth density for the tail and a continuous skirt of low density for the appendages. However, the capsid density had better features as a result of more accurate particle orientations obtained during the fivefold averaging. Subsequently, using the orientations obtained from the fivefold reconstruction, a better asymmetric reconstruction was determined based on the correlation coefficient between the five possible orientations around the tail axis of the asymmetric model and each image (3–5). This process was repeated iteratively. As a result a 24 Å resolution asymmetric reconstruction of the C1 phage was obtained. The steps described above are summarized in Fig. S1.

Cloning, purification, crystallization, and structure determination of gp12. ORF12 was cloned into the pET23d+ bacterial expression vector. The protein was expressed in *Escherichia coli* BL21 (DE3) at a temperature of 25 °C for 4 h. It was purified to homogeneity using affinity Ni-NTA chromatography followed by size-exclusion chromatography. As the native protein did not yield any crystals, an overnight trypsin digestion was carried out with a protein concentration of 1 mg/mL and a trypsin concentration of 4 mcg/mL at room temperature. The protein was then additionally purified on a Ni-NTA affinity column.

Diffracting crystals of gp12 were obtained in 40% 2-methyl-2,4-pentanediol (MPD) and 0.1 M sodium citrate, pH 6.0, as well as in 30% PEG400 and 0.1 M CHES, pH 9.5. Both conditions did not require additional soaking in cryo-protectants. The MPD grown crystals were reproducible and diffracted to about 3 Å resolution. The crystals belonged to the space group $P2_12_12$ and contained six molecules per asymmetric unit. Only one dataset was collected using a PEG grown crystal that had a C2 space group with 12 molecules per asymmetric unit. Although crystals of the selenomethionine derivative of gp12 were obtained, a heavy atom search was unsuccessful due to the presence of many methionines (96 per asymmetric unit). The MPD grown crystals were used for soaking with platinum and mercury compounds. Several datasets were collected at the Advanced Photon Source, beamline 23 ID-D GM/CA. The diffraction images were processed using HKL2000 (6). A multiple anomalous dispersion dataset collected using the Hg-derivative of gp12 was used for phasing. The Hg data was truncated at 6 Å, as in the next resolution shell the correlation coefficient between anomalous differences between the wavelengths dropped to less than 30%. The SHELXD program was used to locate the Hg atoms (7, 8). From the list of heavy atom sites, twelve heavy atoms formed a ring-like structure and were selected to calculate initial phases. The quality of the initial electron density map was very poor with no recognizable secondary structural elements. However, the map contained a ring-like density with an empty hole. Subsequently, a combination of solvent flattening and NCS averaging with phase extension was carried out using the program RESOLVE (9). The IMP program from the Uppsala Software Factory was used to refine the NCS operators (10, 11). As a result, about half of the structure had electron density of sufficient quality to be automatically traced by the program RESOLVE (12), the rest of the structure was completed manually using the program COOT (13, 14). Refinement was carried out using PHENIX.REFINE (15) from the PHENIX software package (16). NCS was enforced during refinement and the map was NCS averaged during manual rebuilding between the cycles of refinement. If the map was not NCS averaged, the N-terminal domain (120 residues) was visible only for two out of six NCS-related molecules. These two copies make crystal contacts to the neighboring hexamers, explaining the more ordered density. The structure of gp12 in the C2 space group with two hexamers in the asymmetric unit was determined with the molecular replacement method using the program PHA-

SER (17). To avoid model bias, the N-terminal domain was truncated from the model used in the molecular replacement search. Nevertheless, the density of the N-terminal domain after 12-fold averaging showed the same fold as the one found in the hexamer from the $P2_12_12$ space group. Data collection and refinement statistics are summarized in Table S1.

Fitting of crystal structures into EM maps. The fitting statistics are summarized in Table S2. The best fit of gp12 corresponds to the position of N-terminal D1 at the proximal part of the knob, suitable for interaction with the lower collar and the position of the missing D4 at the distal end of the knob. However, at the current resolution it is not possible to completely exclude the 180° related fitting solution. Additionally, because the cryo-EM density of the tail knob is smooth, several roughly equivalent fits of the gp12 hexamer related by the rotational symmetry around the vertical axis are possible.

1. Baker TS, Olson NH, Fuller SD (1999) Adding the third dimension to virus life cycles: Three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiol Mol Biol Rev* 63:862–922.
2. Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128:82–97.
3. Jiang W, et al. (2006) Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature* 439:612–616.
4. Morais MC, et al. (2001) Cryoelectron-microscopy image reconsruction of symmetry mismatches in bacteriophage φ29. *J Struct Biol* 135:38–46.
5. Tang J, Sinkovits RS, Baker TS (2010) Three-dimensional asymmetric reconstruction of tailed bacteriophage. *Methods Enzymol* 482:185–210.
6. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, eds Charles W, Carter J (Academic Press, New York), Vol 276, pp 307–326.
7. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64:112–122.
8. Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* 58:1772–1779.
9. Terwilliger TC (2002) Statistical density modification with non-crystallographic symmetry. *Acta Crystallogr D Biol Crystallogr* 58:2082–2086.
10. Kleywegt GJ, Zou JY, Kjeldgaard M, Jones TA (2001) Around O. *International Tables for Crystallograhy, Vol F, Crystallography of Biological Macromolecules*, eds Rossmann MG, Arnold E (Kluwer Academic Publishers, Dordrecht/Boston/London), pp 353–356, 366–367.
11. Kleywegt GJ, Read RJ (1997) Not your average density. *Structure* 5:1557–1569.
12. Terwilliger TC (2003) Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr D Biol Crystallogr* 59:38–44.
13. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501.
14. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
15. Afonine PV, Grosse-Kunstleve RW, Adams PD (2005) The Phenix refinement framework. *CCP4 Newsletter July 2005*:Contribution 8.
16. Adams PD, et al. (2002) *PHENIX*: Building new software for automated crystallographic structure determination. *Acta Crystallogr sect D* 58:1948–1954.
17. McCoy AJ, et al. (2007) *Phaser* crystallographic software. *J Appl Cryst* 40:658–674.

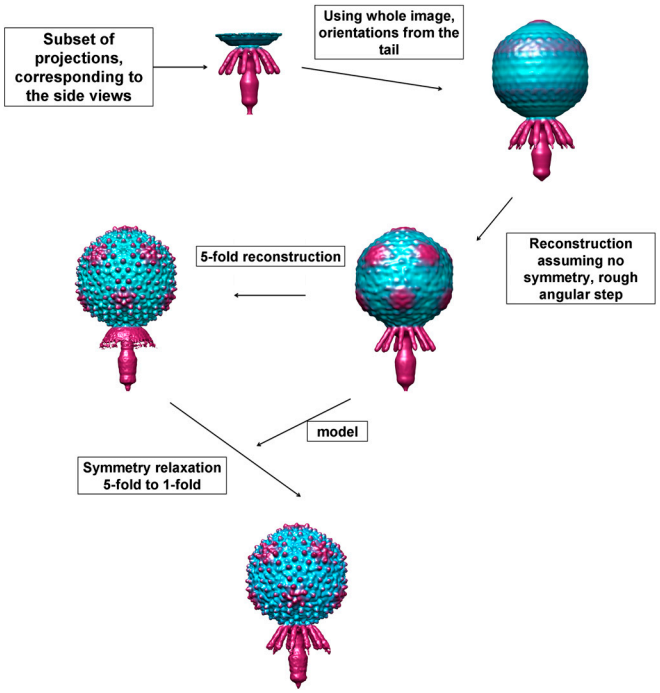

Fig. S1. Steps for acquiring asymmetric reconstruction of bacteriophage C1.

Table S1. Data collection and refinement statistics

	Crystal #1 gp12	Crystal #2 gp12Hg derivative (peak)	Crystal #3 gp12
Data Collection			
Space group	$P2_12_12$	$P2_12_12$	C2
Cell dimensions			
a, b, c (Å)	205.615, 209.640, 102.975	211.597, 214.475, 103.564	316.868, 211.380, 178.605
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 116.53, 90.0
Resolution (Å)	50 (3.11–3.0)*	50 (4.07–3.99)	50 (4.07–4.00)
R_{merge}	6.8 (42.5)	15.2 (47.5)	25.1 (53.8)
$I/\sigma I$	27.3 (3.7)	17.2 (5.3)	6.3 (3.1)
Completeness (%)	98.8 (91.0)	98.6 (99.1)	100 (99.9)
Redundancy	7.2 (5.8)	8.6 (8.0)	3.8 (3.8)
Refinement			
Resolution (Å)	50 (3.04–3.01)		50 (4.14–4.00)
No. reflections	87391 (2334)		88193 (2804)
R_{work}/R_{free}	0.2360/0.2814		19.49/23.32
No. atoms	23694		47376
R.m.s.d. deviations			
Bond length (Å)	0.018		0.010
Bond angles (°)	1.865		1.392

*Values in parentheses are for the highest resolution shell.

Table S2. Structure fitting statistics

Fitted structure	Cryo-EM map	*Sumf* *(%)	Negative density	Clash
HK97 capsid protein	Icosahedral reconstruction of the C1 capsid	55.88	13.41	0.32
Gp12	Asymmetric reconstruction of the tail	43.2	74.4	n/a

**Sumf* is the mean density of all atoms when fitted into the cryo-EM density normalized with respect to the highest density within the map.