

An optimal exposure strategy for cryoprotected virus crystals with lattice constants greater than 1000 Å

Lu Gan[‡] and John E. Johnson*

The Scripps Research Institute, USA. E-mail: jackj@scripps.edu

Received 3 August 2007

Accepted 30 November 2007

Studies of icosahedral virus capsids provide insights into the function of supramolecular machines. Virus capsid crystals have exceptionally large unit cells; as a result, they diffract weakly compared with protein crystals. HK97 is a dsDNA lambda-like bacteriophage whose 13 MDa capsid expands from 550 Å to 650 Å with large subunit conformational changes during virus maturation. The HK97 penultimate maturation intermediate was crystallized in a tetragonal unit cell that has lattice constants of 1010 Å × 1010 Å × 730 Å. The crystals could be cryoprotected, but diffracted to a modest resolution of 5 Å at a bending-magnet beamline. When these crystals were optimally exposed with two orders-of-magnitude more photons from a new insertion-device beamline, data extending to better than 3.8 Å resolution were obtained. Here, the strategies to collect and process such data are described. These strategies can be adapted for other crystals with large unit cells and for microcrystals.

© 2008 International Union of Crystallography
Printed in Singapore – all rights reserved

Keywords: virus crystals; bacteriophage HK97; insertion-device beamlines.

1. Optimized exposure: pushing diffraction to the max

The bacteriophage HK97 Head II structure was the first high-resolution structure of a mature dsDNA virus capsid (Wikoff, Liljas *et al.*, 2000; Helgstrand *et al.*, 2003). Cryo-EM and solution X-ray scattering studies of the procapsid and a number of maturation intermediate states have shown that HK97 capsids (and probably other dsDNA capsids) undergo a 'molecular ballet' during DNA-packaging-induced capsid maturation (Lata *et al.*, 2000; Lee *et al.*, 2004; Wikoff *et al.*, 2006). To fully understand the macromolecular mechanisms involved in capsid maturation, it is necessary to obtain high-resolution crystal structures of these intermediates. Biochemical and cryo-EM experiments revealed that the penultimate capsid state known as expansion intermediate IV (EI-IV) may be a candidate for crystallization (Gan *et al.*, 2004). Moreover, two capsid forms that are related to EI-IV, termed Head I and pepEI-IV, could also be crystallized in the same space group $P4_32_12$ (Gan *et al.*, 2006). Such crystals will now be referred to as 'HK97 tetragonal crystals'.

The structure of EI-IV was determined to 8 Å resolution from a single crystal, revealing important details of molecular rearrangement at the secondary structural level. However, this 8 Å resolution structure missed all of the important details at the amino-acid side-chain level that can be revealed at 3–4 Å resolution. We observed that HK97 tetragonal crystals either diffracted to worse than 30 Å resolution, where the diffraction image is dominated by the concentric small-angle solution scattering pattern of residual uncrystallized capsids, or to a moderate resolution of ~8 Å. The drastic difference probably resulted from the vagaries associated with flash freezing.

The crystals were exposed for 1 min at a bending-magnet (BM) beamline, which was optimized for the 3.44 Å resolution Head II dataset that was collected at room temperature (Wikoff, Schildkamp *et al.*, 2000). The best HK97 tetragonal crystals diffracted to 5.6 Å resolution for 200 one-minute exposures using a BM beamline. The crystals might be more ordered, but diffraction was limited by the small number of unit cells. Indeed, a typical protein crystal has at least 1000-fold more unit cells than an HK97 tetragonal crystal of comparable size. The HK97 crystals also have an unusually large solvent content because they were prepared from empty capsids. Not surprisingly, HK97 crystals produce almost undetectable diffraction when they are exposed to X-ray doses used for protein crystals.

1.1. Experimental considerations

The highest-resolution data for the HK97 tetragonal crystals were obtained by what would be conventionally considered over-exposure of the crystals. Extended exposure times degraded the crystals but produced diffraction to much higher resolution than when a single crystal was used for a complete dataset. HK97 crystals were suitable for such a procedure because they are highly reproducible and can be flash-frozen. Multiple crystals were required to maximize the completeness of the dataset because they were not perfectly isomorphous, a problem often encountered when crystals are frozen. The diffraction images from many crystals were rejected owing to a small but significant difference in lattice dimensions. More than 20 crystals were screened for each capsid class studied, but data from up to eight crystals contributed to the final dataset. The EI-IV tetragonal crystals were not as reproducible and therefore a dataset could not be collected using extended exposures.

[‡] Present address: Division of Biology, California Institute of Technology, Pasadena, CA, USA.

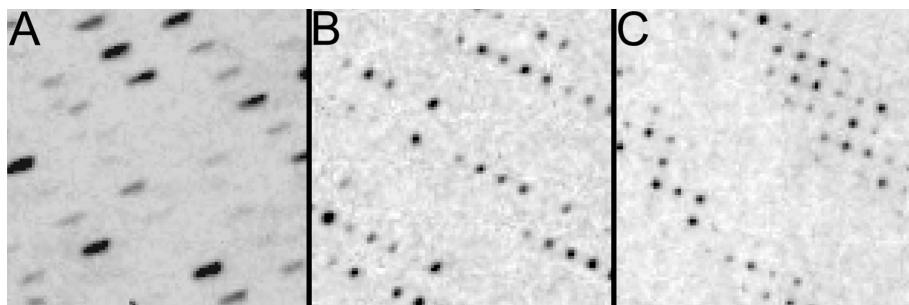


Figure 1
Comparison of HK97 tetragonal crystal reflections recorded (A) without and (B, C) with convergent focusing optics. The maximum resolutions of images A, B, and C were 4.5 Å, 5.6 Å and 3.8 Å, respectively.

Virus data have traditionally been collected at the brightest beamlines; for example, BioCARS 14-BMC was key to the Head II project (Wikoff, Schildkamp *et al.*, 2000). Since the new tetragonal crystals diffracted weakly compared with the room-temperature crystals, they were exposed to the unattenuated beam of a new GM/CA CAT insertion-device (ID) beamline. The flux (1.3×10^{15} photons s^{-1}) at the ID-D beamline allowed up to 100-fold more photons to be collected per reflection per unit time. By reducing the exposure time from 30 min to 20 s, the background radiation noise was minimized and hundreds of diffraction images could be collected in a day. These brilliance–dose relationships are not linear because the diffraction power depends on other factors such as wavelength and crystal size, even when holding crystal ‘quality’ constant.

GM/CA CAT’s ID-D beamline also features convergent optics, which enables the accurate integration of closely spaced reflections (Fig. 1). The convergent optics also increased the signal-to-noise ratio by imaging each reflection in an area smaller than 16 pixels, a value that was limited by the detector’s point spread function (Fig. 2). To maximize the resolution range covered, large multi-element MarMosaic CCD detectors were used that could resolve more than 120000 reflections in a single image. Future improvements in detectors might result in all photons being imaged in 1–4 pixels, which will significantly simplify intensity integration.

1.2. Outcome

HK97 tetragonal crystals diffracted to 5–6 Å resolution for up to 200 one-minute ‘standard exposures’ at a BM beamline. Comparable

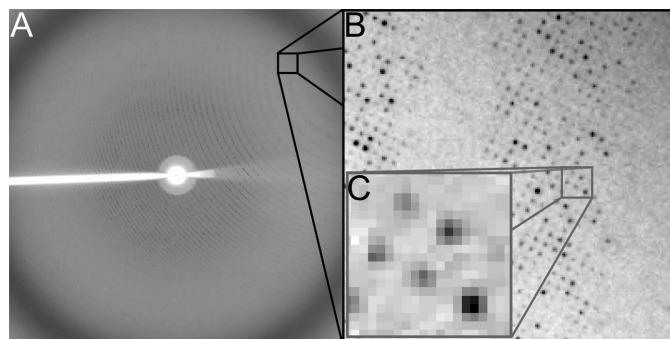


Figure 2
(A) Diffraction image (0.25° oscillation) of a cryoprotected Head II tetragonal crystal. (B) Zoomed view of an area corresponding to ~ 4.2 Å resolution. (C) Zoomed view of five reflections. This crystal measured $0.4 \text{ mm} \times 0.4 \text{ mm} \times 0.15 \text{ mm}$ and was exposed for 20 s to the unattenuated convergently focused beam of GM/CA CAT’s 23-ID-D beamline. The image was recorded on a Mar225 during the 2005 beamline commissioning phase.

crystals (same mother liquor, different preparation) diffracted to 3.8–4.2 Å resolution for up to 20 twenty-second exposures at the GM/CA CAT ID-D beamline. Such ‘extended exposures’ were optimal for answering key questions about HK97 capsids, but dramatically reduced the number of diffraction images that could be collected per crystal. Fortunately, many crystals were large enough that data could be collected by rastering with the small beam ($40 \mu\text{m} \times 60 \mu\text{m}$). Owing to the random orientation of each crystal, the resulting dataset was not complete. Systematic absences were difficult to detect

in the non-contiguous images, so an almost complete dataset was collected at low resolution with conventional exposure times. However, no $(0, 0, 4n)$ systematic absences were observed, perhaps owing to the unfortunate orientation of the crystal’s four-fold axis parallel to the spindle axis. A fortuitous $(0, 0, 4n)$ systematic absence was eventually confirmed from a single test-crystal diffraction image, cutting the molecular replacement search time by half during the structure determination.

1.3. Pitfalls

There are a few important disadvantages to extended-exposure data collection. First, it is time-consuming at both the beamline and the data-processing stages. Second, it is not guaranteed to produce an interpretable structure because crystals that are apparently isomorphous at the integration stage may not scale well. Furthermore, a scaling cut-off must be chosen to reject images from non-isomorphous crystals (see below). However, these time costs were justified because the diffraction resolution was sufficient to answer the key biological questions, and therefore saved many months of potentially fruitless crystal optimization. These time costs may be impractical for determination of structures by isomorphous replacement or anomalous scattering. Moreover, crystals that do not have high non-crystallographic symmetry (NCS) may yield noisier electron density maps because frozen crystals are never perfectly isomorphous with each other.

2. Data processing strategies

Data from the optimally exposed crystals had many properties in common with data collected from crystals at room temperature. Mosaicity increases and diffraction falls off rapidly as the crystals are exposed to the unattenuated ID X-rays; many crystals are non-isomorphous; resolution limits are heterogeneous. All of these properties reduce the resolution to significantly lower than that of the best crystal. To obtain the best possible structure from optimally exposed datasets, the data must be integrated and scaled with great care. HK97 tetragonal crystal diffraction images were processed and scaled using the methods implemented for Head II and Prohead II (Wikoff, Liljas *et al.*, 2000; Wikoff *et al.*, 2003), but with significant modifications.

2.1. Indexing and intensity integration

Individual HK97 diffraction images used to be indexed independently using the program *DENZO* (Otwinowski & Minor, 1997), even if they came from the same crystal (Wikoff *et al.*, 2003). This strategy allowed the resolution cut-off to be optimized for each

individual frame. Sequential images that are indexed independently are often assigned different symmetry-related orientation matrices, so the partially recorded reflections that account for the majority of reflections cannot be summed conveniently. Therefore, the images were auto-indexed once per crystal. Indeed, it was found that the post-scaling statistics, especially completeness and $I/\sigma(I)$ of the highest-resolution bins, could be improved if all images from the same crystal were integrated using the starting orientation matrix of the first image (Gertsman *et al.*, 2008).

The ability to record high-resolution reflections from a 1000 Å primitive-unit-cell crystal presented a new challenge: there were too many reflections to integrate. After extensive testing, it was found that up to 100000 reflections could be integrated, so each image was integrated using a low- and a high-resolution pass. The resolution cut-off was chosen so that equal numbers of reflections were in each resolution band. Scaling of the resulting data revealed an artificially high $I/\sigma(I)$ in the overlapping resolution band, but this did not produce any noticeable artefacts in the electron density maps.

2.2. Scaling, image rejection and reindexing

Optimally exposed data were scaled in three passes to eliminate the bad crystals and images that showed severe radiation damage. First, the reflection files from all crystals were scaled together. The non-isomorphous and ‘bad’ crystals, with much higher R_{merge} and χ^2 values than most other frames, were detected in the log files and excluded from further processing. While decreases in resolution can readily be detected by looking at the images, increases in mosaicity are detected by scaling. HK97 data were scaled in *Scalepack* (Otwinowski & Minor, 1997) using the option ‘fit batch mosaicity’, which refines independently the mosaicity of each image. The log file revealed that the mosaicity increased slowly at first, but then rapidly. The mosaicity of each crystal was classified as low (early exposures) and high (late exposures). All images were re-integrated using the highest mosaicity value from each mosaicity class to ensure that all reflections were integrated; each crystal was therefore treated as two

Table 1

Statistics from optimally exposed crystals.

Additional details can be found by Gan *et al.* (2006). Note that the Head II dataset was collected using the smallest MarMosaic detector.

Dataset	No. of crystals	Refs _{measured}	Refs _{unique}	R_{merge}	R_{cryst}	$B_{\text{Wilson}} (-B_{\text{sharp}})$	d_{min} (Å)
Head II	2	7863609	1737016	20.3	30	6	3.9
pepEI-IV	5	9116091	2144154	17.0	34	39	3.8
Head I (trip 1)	5	14857665	1678427	16.2	41	55	4.2
Head I (trip 2)	8	6446051	1240774	16.3	50	75	4.0

crystals, one with low mosaicity and one with high mosaicity. Spurious reflections, such as those with negative $\sigma(I)$ or an $I/\sigma(I)$ value less than -1.5 , were removed from the reflection files using a locally written Perl script. These trimmed image files were scaled one last time, but with the option ‘fit crystal mosaicity’. Although the completeness could have been increased by scaling partially recorded reflections to full reflections, this procedure was not carried out because there were enough data to calculate interpretable maps.

2.3. Results and final maps

Four datasets were collected from three capsid forms (Table 1). The maps were sharpened by $-B_{\text{Wilson}}$ and the structures were refined extensively using real-space refinement methods (Korostelev *et al.*, 2002). There is a clearly positive correlation between the number of crystals, R_{cryst} , and B_{Wilson} . Since the maps were of comparable quality, it is likely that by using more (slightly non-isomorphous) crystals the errors in the averaged amplitudes increased significantly. The effects of such errors in electron density maps were mitigated by 30-fold NCS real-space averaging.

An optimally exposed dataset was collected from two cryoprotected Head II crystals to control for crystallization artefacts. The electron density map calculated from this dataset was compared with the room-temperature Head II dataset, revealing similar quality (Fig. 3). Owing to the smaller number of crystals used for the cryo dataset (2 *versus* 60), R_{cryst} was significantly better (30% *versus* 36%).

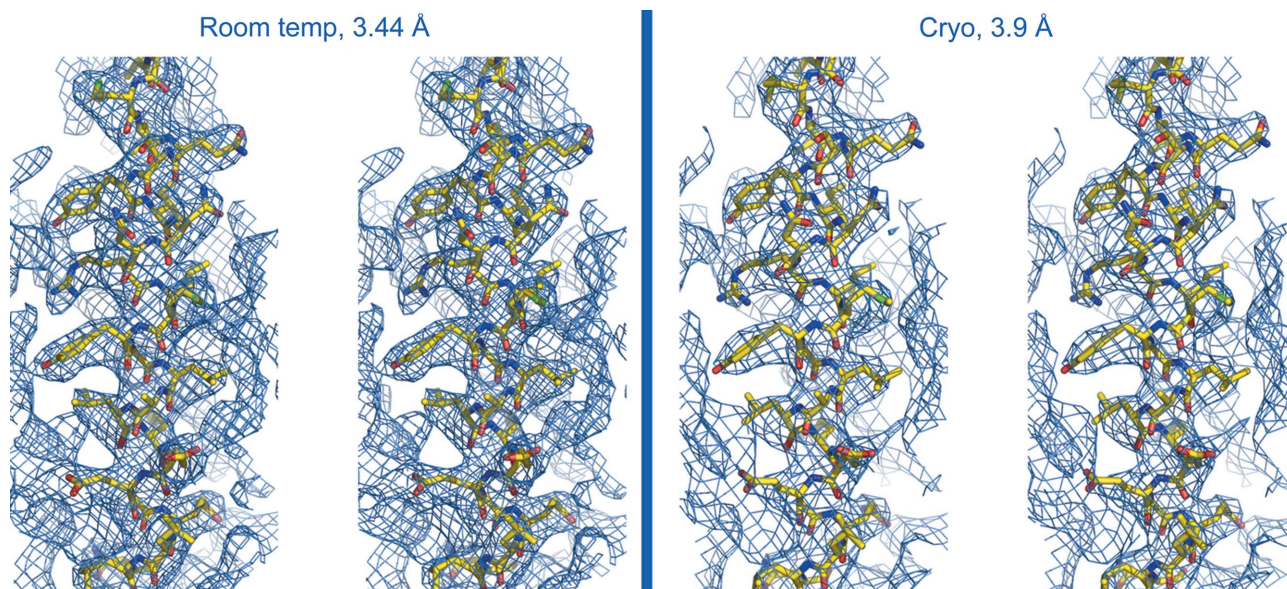


Figure 3

Comparison of Head II electron density, in stereo. The densities were real-space averaged using (left) 60-fold or (right) 30-fold NCS and contoured at 1.5σ RMS. Note that the maps are comparable in terms of noise and side-chain definition despite the differences in resolution and crystal exposure method.

3. Summary and prospectus

The lessons learned by determining structures from optimally exposed data are useful for other challenging crystals and can be summarized as follows. First, a complete low-resolution dataset should be collected to determine the space group and molecular replacement parameters. If a second dataset can be collected from the same crystal, the crystal should be manually rotated 45° from the spindle axis to increase the chances of recording a line of systematic absences. Next, a crystal should be exposed with 10- to 100-fold more dose than used for the low-resolution dataset. If the resolution increases significantly, the crystals are considered to be 'weakly' diffracting instead of 'poorly' diffracting and should be exposed in a dose series to determine how many images can be collected from one crystal. To maximize completeness, a large number of crystals should be shot. The data should be processed carefully, paying special attention to individual mosaicity and R_{merge} values. Extended exposures are feasible using the brightest beamlines, which have convergent beam fluxes of $\geq 1 \times 10^{13}$ photons $\text{s}^{-1} \text{mm}^{-2}$ at the specimen. To minimize air scatter, the beamline should have high-energy (>20 kV) X-ray capability or a helium/vacuum path between the sample and detector.

A survey of the VIPERdb revealed that virus crystals with a unit-cell parameter greater than 500 Å rarely diffract beyond 3.4 Å resolution (Shepherd *et al.*, 2006). Most of these crystals were exposed at room temperature, leading one to suspect that radiation damage is the limiting factor. Another resolution-limiting factor is the large background from the water solution-scattering ring. If suitable cryoprotection conditions can be found for these crystals, diffraction may be extended beyond the 3.4 Å resolution limit by optimizing the exposure. We suspect that breaking this resolution barrier may be especially challenging for crystals that have enormous Matthews coefficients (V_m). An HK97 tetragonal crystal has a V_m of 14 and produces up to three times more solvent scattering than a typical protein crystal, which has a V_m of 2–3 (Kantardjieff & Rupp, 2003).

We thank our many colleagues, especially James Conway, Robert Duda, Ilya Gertsman, Roger Hendrix, Gabriel Lander, Kelly Lee, Lars Liljas, Tianwei Lin, Vijay Reddy, Jeffrey Speir and Alasdair Steven for stimulating discussions. We thank the staff of beamlines 11-1 of SSRL, and 14-BMC and 23-ID-D of APS for assistance in data

collection. In particular, we thank Spencer Anderson, Keith Brister, Robert Fischetti, Nukri Sanishvili and Ward Smith for advising and helping with data collection. Research carried out at the SSRL was supported by the DOE, OBER and by the NIH, NCRR, Biomedical Technology Program, and the NIGMS. Research carried out at the APS was supported by the US DOE Contract No. W-31-109-Eng-38, research carried out at GM/CA CAT was supported by the NCI (Y1-CO-1020) and the NIGMS (Y1-GM-1104), and research carried out at BioCARS sector 14 was supported by the NIH, NCRR, under grant RR07707. LG was supported by a Fletcher Jones Foundation Fellowship. This work was supported by NIH grant R01 AI40101 to JEJ.

References

- Gan, L., Conway, J. F., Firek, B. A., Cheng, N., Hendrix, R. W., Steven, A. C., Johnson, J. E. & Duda, R. L. (2004). *Mol. Cell*, **14**, 559–569.
- Gan, L., Speir, J. A., Conway, J. F., Lander, G., Cheng, N. Q., Firek, B. A., Hendrix, R. W., Duda, R. L., Liljas, L. & Johnson, J. E. (2006). *Structure*, **14**, 1655–1665.
- Gertsman, I., Gan, L., Wikoff, W. R., Guttman, M., Lee, K. K., Duda, R. L., Hendrix, R. W., Komives, E. A. & Johnson, J. E. (2008). Submitted.
- Helgstrand, C., Wikoff, W. R., Duda, R. L., Hendrix, R. W., Johnson, J. E. & Liljas, L. (2003). *J. Mol. Biol.* **334**, 885–899.
- Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
- Korostelev, A., Bertram, R. & Chapman, M. S. (2002). *Acta Cryst. D* **58**, 761–767.
- Lata, R., Conway, J. F., Cheng, N., Duda, R. L., Hendrix, R. W., Wikoff, W. R., Johnson, J. E., Tsuruta, H. & Steven, A. C. (2000). *Cell*, **100**, 253–263.
- Lee, K. K., Gan, L., Tsuruta, H., Hendrix, R. W., Duda, R. L. & Johnson, J. E. (2004). *J. Mol. Biol.* **340**, 419–433.
- Otwinowski, Z. & Minor, W. (1997). *Macromol. Crystallogr.* **276**, 307–326.
- Shepherd, C. M., Borelli, I. A., Lander, G., Natarajan, P., Siddavanahalli, V., Bajaj, C., Johnson, J. E., Brooks, C. L. & Reddy, V. S. (2006). *Nucleic Acids Res.* **34**, D386–D389.
- Wikoff, W. R., Che, Z., Duda, R. L., Hendrix, R. W. & Johnson, J. E. (2003). *Acta Cryst. D* **59**, 2060–2064.
- Wikoff, W. R., Conway, J. F., Tang, J. H., Lee, K. K., Gan, L., Cheng, N. Q., Duda, R. L., Hendrix, R. W., Steven, A. C. & Johnson, J. E. (2006). *J. Struct. Biol.* **153**, 300–306.
- Wikoff, W. R., Liljas, L., Duda, R. L., Tsuruta, H., Hendrix, R. W. & Johnson, J. E. (2000). *Science*, **289**, 2129–2133.
- Wikoff, W. R., Schildkamp, W. & Johnson, J. E. (2000). *Acta Cryst. D* **56**, 890–893.