PNAS
www.pnas.org

Supplementary Information for

Cryo-electron microscopy structure of the filamentous bacteriophage IKe

Jingwei Xu[1, 3], Nir Dayan[2], Amir Goldbourt[2*], Ye Xiang[1*]

[1]Center for Infectious Disease Research, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Beijing Advanced Innovation Center for Structural Biology, Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing 100084, China

[2]School of Chemistry, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 6997801, Israel

[3]Present address: Institute of Molecular Biology and Biophysics, Eidgenössische Technische Hochschule Zürich, CH-8093 Zürich, Switzerland

[*] To whom correspondence should be addressed: Y.X.: yxiang@mail.tsinghua.edu.cn, tel:+86-10-62772587. A.G.: amirgo@tauex.tau.ac.il, tel:+972-36408437

www.pnas.org/cgi/doi/10.1073/pnas.1811929116

This PDF file includes:

Supplementary text
Figs. S1 to S7
Tables S1
References for SI reference citations

Supplementary Information Text

Materials and Methods

Sample Preparation.

We obtained from Prof. Marjorie Russel (The Rockefeller University, NY) initial stocks of a hybrid phage, fIKe, in which wild-type p2 was replaced by p2 originating from f1 phage, thus allowing for significantly higher yields (1). The genome has 7337 nucleotides and the particle mass is ~21 MDa. The capsid envelope is thus identical to IKe and the final structure does not depend on small changes in the ssDNA length. Initial phage stocks were used to infect the *Escherichia coli* strain A527 that bears a tetracycline-resistant IncN conjugative plasmid encoding N pili (obtained from Prof. James Bull, University of Texas). The preparation protocol is based on previous reports (2). In short, single colonies on agar plates that contain 12.5 mg/ml tetracycline were collected and grown in LB media to log phase. Then, the speed of shaking was reduced to 60 rpm for an hour to allow efficient growth of pili, followed by infection for another hour at 60 rpm prior to resuming growth at 225 rpm for another 24 hrs. Purification was performed by separation of cells and virions by repeated centrifugations and virus precipitation (5% PEG 8000, 0.5M NaCl) followed by separation through CsCl gradient ultracentrifugation. The final concentration of the purified IKe virion was typically ~10^{13} particles/ml.

Cryo-EM Grid Preparation

Series dilutions of the IKe sample were made with TMS buffer (50 mM Tris at pH 7.5, 10 mM $MgCl_2$, 100 mM NaCl). The diluted samples were vitrified over 200

mesh Quantifoil grids (1.2 μm hole size) using a Vitrobot Mark III (FEI company). The vitrified samples were checked with an FEI Arctica microscope operating at 200 kV. The grids prepared from samples with a concentration of ~2.5×10^{12} particles/ml (4 times volume/volume dilution) were the best for data collection.

Cryo-EM data collection

CryoEM data of the bacteriophage IKe were collected manually as movie frame stacks at a nominal magnification of 22,500 (an effective pixel size of 1.32 Å) on an FEI Titan Krios electron microscope operating at 300 kV and equipped with a K2 Summit camera (Gatan Inc.). Each stack has 32 image frames and the accumulated electron dose rate for the entire stack is ~ 40 e/Å^2. The eTas software was used for all data collection (developed by Xueming Li's group at Tsinghua University). The movie frames of each stack were aligned initially at the micrograph level using MotionCor1. The frames were then dose weighted and summed up for subsequent processing into a single micrograph using the program MotionCor2 (3). The CTF parameters of the micrographs were calculated by using the program Gctf (4). A total of 573 micrographs with defocus values in the range of 1.0-2.2 μm were selected for further processing.

Image processing

To obtain the helical parameters of IKe, 80 micrographs were used for the initial reconstruction by following the routine helical reconstruction procedure of Relion2. Straight filaments were manually picked with Relion2 (5). The picked filaments were segmented into individual particles with an inter-box distance of approximately 52 Å (~3

times the helical rise). A total of 123895 segmented particles were picked. The segmented particles were subjected to 2D classification analysis (SI Appendix, Figure S1A & S1B). After the selection through 2D classifications, a total of 80047 particles were used for 3D auto-refinements with a five-fold symmetry imposed. The structure of the bacteriophage fd was low-passed to 60 Å and used as the initial model. The helical parameters of fd (PDB entry: 2HI5, twist=37.40º, rise=17.40 Å) were used for initial local helical symmetry searches. After the refinement, the particles were subjected to 3D classifications without performing orientation search. Helical parameters were searched and optimized after each iteration of 3D classification. The particles in good classes were combined and used for the next round of 3D refinement and classification. The helical parameter searches were converged after two rounds of iterative 3D auto-refinements and classifications. An additional round of 2D classification was performed to remove bad particles before the final refinement. A total of 10565 selected segments were used in the final 3D auto-refinement and helical parameter optimization. The final refined helical parameters were 38.52° for the twist and 16.77 Å for the rise. The resolution of the reconstruction was estimated by the Fourier shell correlation (FSC) curves calculated between the randomly split two halves of the data. The final resolution of the initial helical reconstruction was 6.9 Å. The local variation of the reconstruction was calculated with ResMap (6) and the protrusions in the helices suggested that the resolution of part of the capsid shell density could be better than 6.9 Å (SI Appendix, Figure S2A).

To obtain the high resolution structure of the IKe capsid shell, 493 micrographs were then used for further data processing by following the procedures described above (SI Appendix, Figure S3A). A total of 253156 segmented particles were picked and the

particles were selected through 2D and 3D classifications. An auto-refinement with 224416 selected particles and the optimized helical parameters yielded a ~4.2 Å reconstruction map. To further improve the resolution of the reconstruction, 3D classifications with multiple references were performed (7). A subset of data that consists of 112808 segmented particles was used for the final symmetric reconstruction, which resulted in a reconstruction at a resolution of 3.4 Å (SI Appendix, Figure S1D & S1E). The calculated ResMap of the reconstruction showed little variation in the capsid shell (SI Appendix, Figure S2B).

To determine the structure of the inner DNA core, asymmetric reconstructions of IKe were performed by assuming a fixed relationship between the capsid and the DNA with the particle orientations determined in the five-fold symmetric reconstruction (C5 expanded to C1). The capsid shell densities of 224416 particles were subtracted from raw images by following a previously described protocol (8). Five equivalent orientations of each particle were generated and were subjected to 3D classifications without orientation sampling. A featureless cylindrical volume was used as an initial model for the 3D classifications (SI Appendix, Figure S3B). The 3D classification results showed clear helical features for some of the classes. Further analysis showed that the helix structures of different classes were related to each other by a rotation along helical axis of a multiple of 72º (SI Appendix, Figure S3B). Simulations were performed by using the capsid shell and a randomly generated inner helical object that is either in a fixed or non-fixed orientation position relative to the outer capsid (SI Appendix, Figure S5A and S5B, respectively). The simulated three dimensional object was randomly projected. Then, reconstructions were calculated with the projected images that have either the correctly

assigned orientations or incorrectly assigned orientations with a constant n×72º (n is a fixed integrate value) aberration in the az angle or incorrectly assigned orientations with mixed n×72º (n is not a fixed integrate value) aberration in the az angle (SI Appendix, Figure S5A). The simulated reconstruction results showed that the reconstruction from the incorrectly assigned orientations with a constant n×72º (n is a fixed integrate value) aberration in the az angle show clear helix features and are related to the correct reconstruction by a n×72º rotation (SI Appendix, Figure S5A). However, the simulated reconstruction from the incorrectly assigned orientations with mixed n×72º (n is not a fixed integrate value) aberration in the az angle is featureless and similar to the initial cylindrical model. Therefore, the helical structures of the different classes are equivalent and related to each other through a rotation along the helical axis by a multiple of 72º. The particles in these classes were selected and subjected for the next round of 3D classification. A total of 202775 "particles" derived from 81600 individual segments were selected. The "particles" were further selected based on their *rlnLogLikeliContribution* values.

The final asymmetric reconstruction of IKe was calculated from a total of 53784 segmented particles and the resolution around inner DNA core was ~ 5 Å. The helical parameters of inner DNA core estimated in real space using *relion_helix_tools* package were rise = ~17 Å, twist = 180º.

Alternatively, the particles used in shell reconstruction were also subjected for a reconstruction without imposing any symmetry by using Relion2. The 5-fold symmetric reconstruction of IKe was low-pass filtered to 20 Å as the initial model. After a round of 3D auto-refinement, the particles were subjected to orientation-guided 2D classifications

without sampling and selected bad particles were discarded. A total of 74430 particles were used for the final refinement and 3D reconstruction, which resulted in a map of 7.5 Å resolution (SI Appendix, Figure S6B).

Structural modeling and validation.

The resolution around the capsid shell was ~ 3.4 Å and the map quality was good enough for *ab initio* modeling. Model of the major coat protein was built in *COOT* (9) and was refined against the density map using the *phenix.real_space_refine* package (10). The structural validation was performed by using *phenix.molprobity*. Some of the figures were prepared with the programs Chimera (11) and ESPript (12). The structure of IKe has been deposited to the PDB and EM databank (The PDB entry is 6A7F, and the EMDB entries are EMD-6993 and EMD-6994).

Quantification and Statistical Analysis

All reported resolutions are based on the gold-standard Fourier Shell Correlation (FSC) = 0.143 criteria (13).

Supplementary Figures

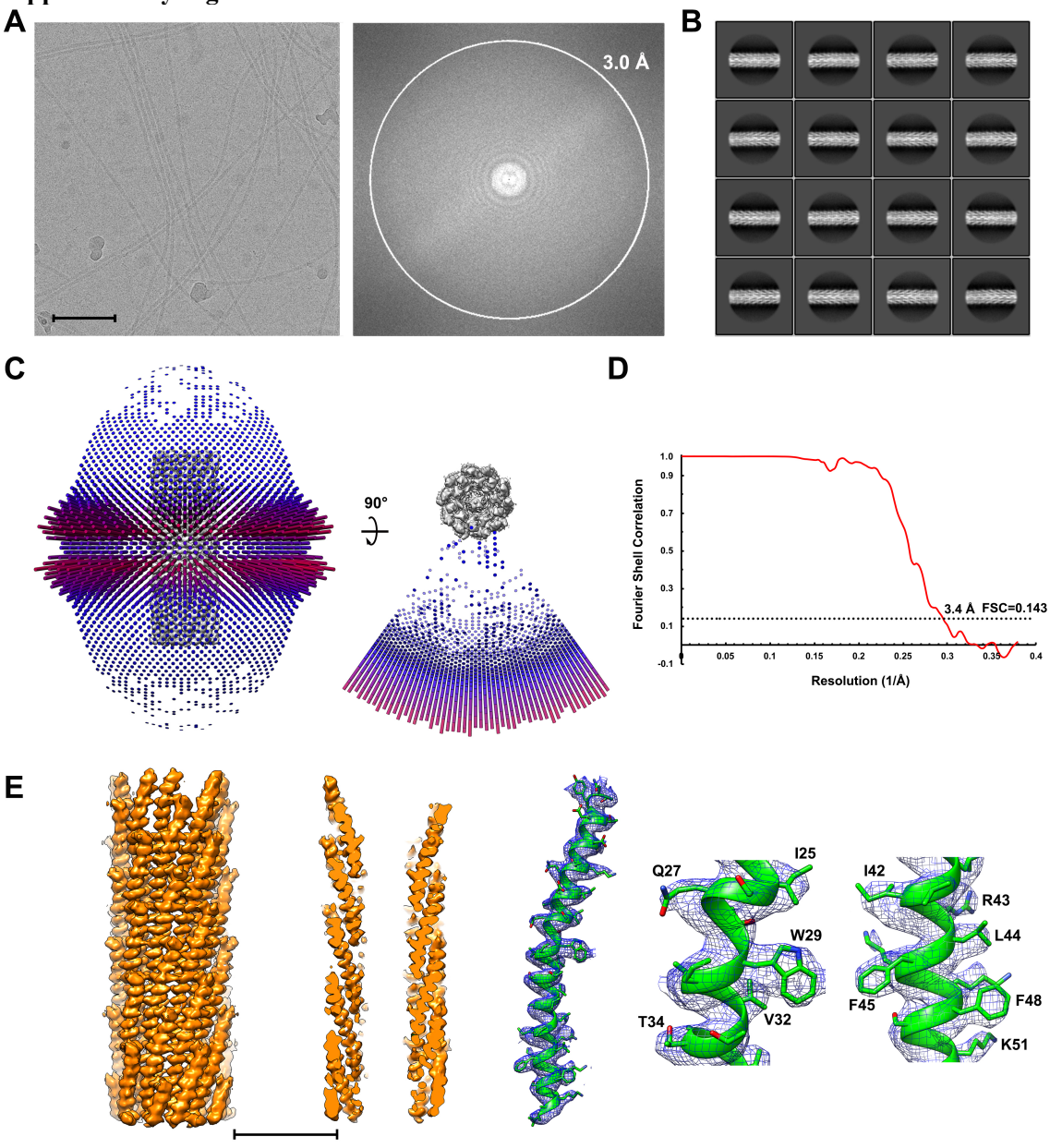


Fig. S1. Cryo-EM analysis of the filamentous phage IKe.

(**A**) A representative raw cryo-EM image of the filamentous phage IKe (left) and its Fourier transform (right). The scale bar is 100 nm.

(**B**) Two-dimensional class averages of the segmented IKe particles. Features of the helical shell are visible.

(**C**) Angular distribution of the particles used in the final reconstruction. Each cylinder represents one orientation and its height is proportional to the number of particles from that orientation used in reconstruction.

(**D**) Gold standard Fourier shell correlation (FSC) curves of the reconstruction.

(**E**) Left: surface shadowed diagrams showing the outer surface and the central section of the capsid. Right: representative diagrams showing EM densities around parts of the major coat protein. The scale bar is 50 Å.

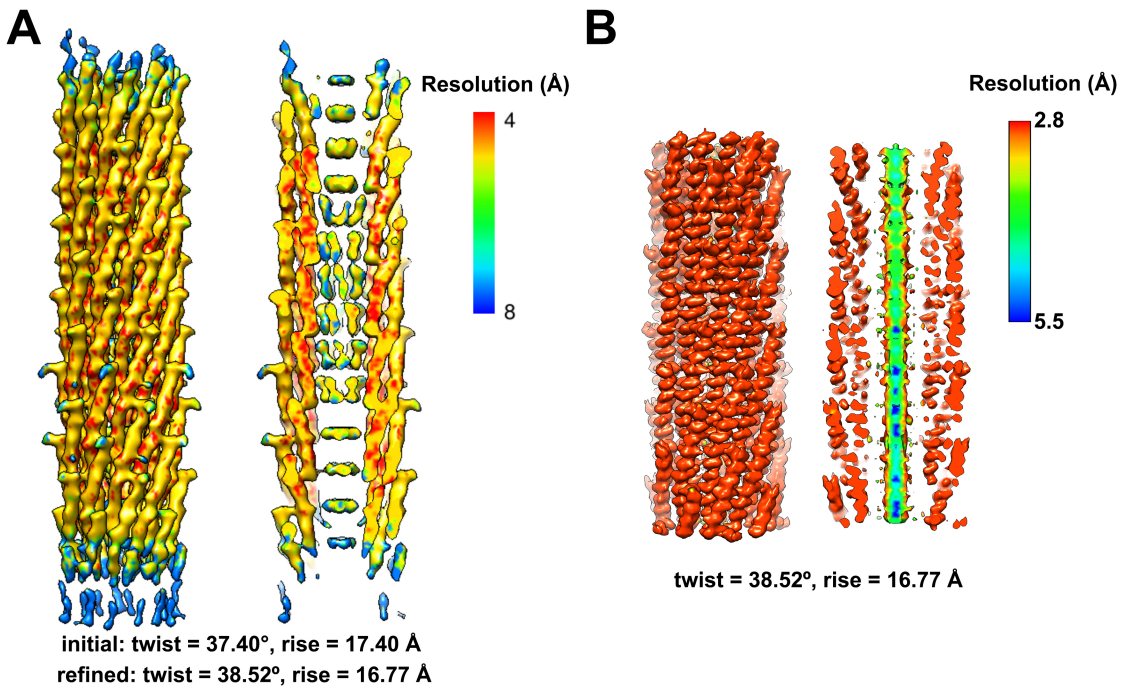

A
Resolution (Å)
4
8
initial: twist = 37.40°, rise = 17.40 Å
refined: twist = 38.52°, rise = 16.77 Å
B
Resolution (Å)
2.8
5.5
twist = 38.52°, rise = 16.77 Å

Fig. S2. Local variation of the reconstructions.

(A) Surface shadowed diagram calculated with ResMap (6) showing the local variation of the filamentous phage IKe reconstruction after the helical parameter optimization. **(B)** Surface shadowed diagram calculated with ResMap (6) showing the local variation of the final 3.4 Å reconstruction.

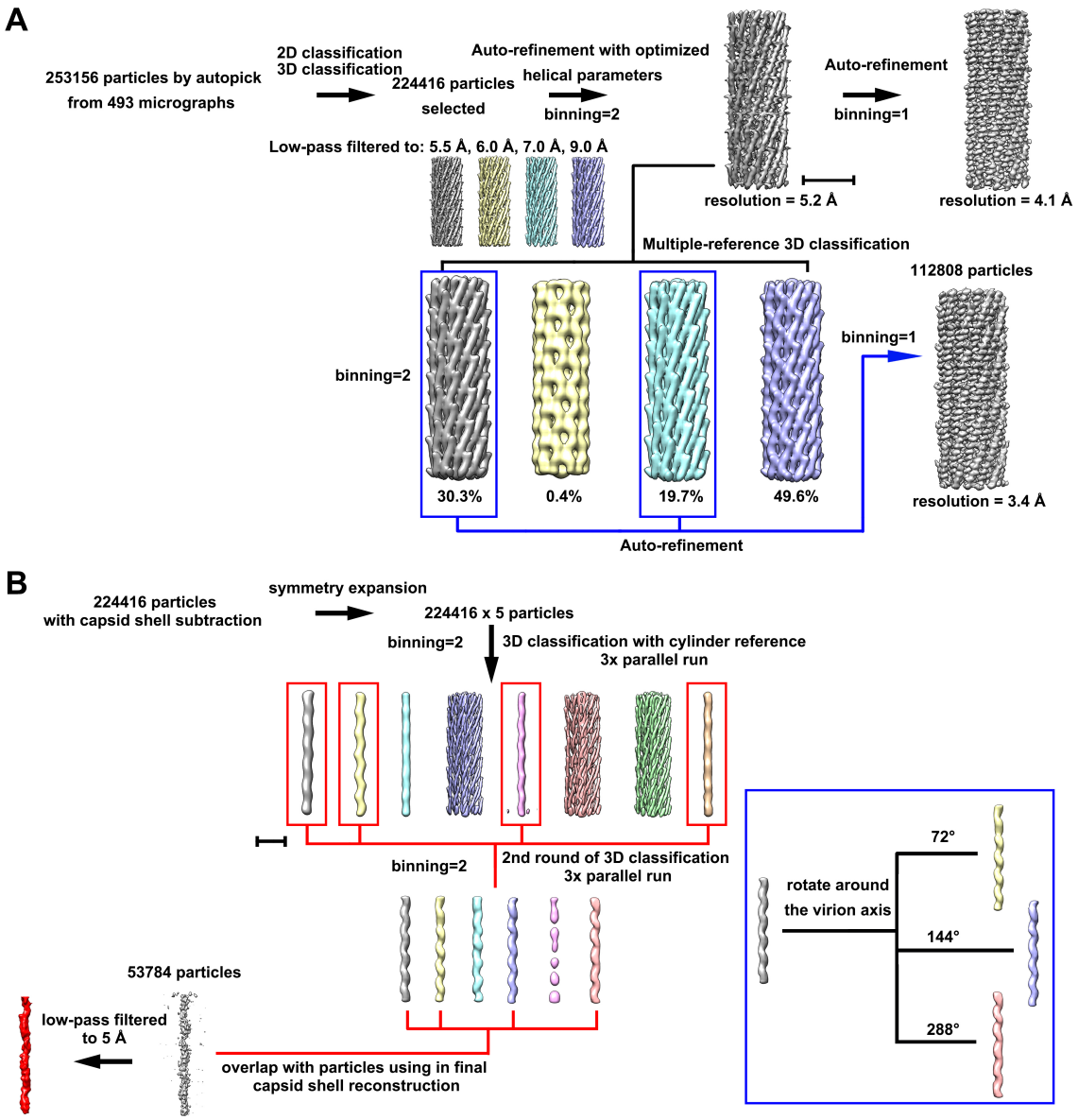
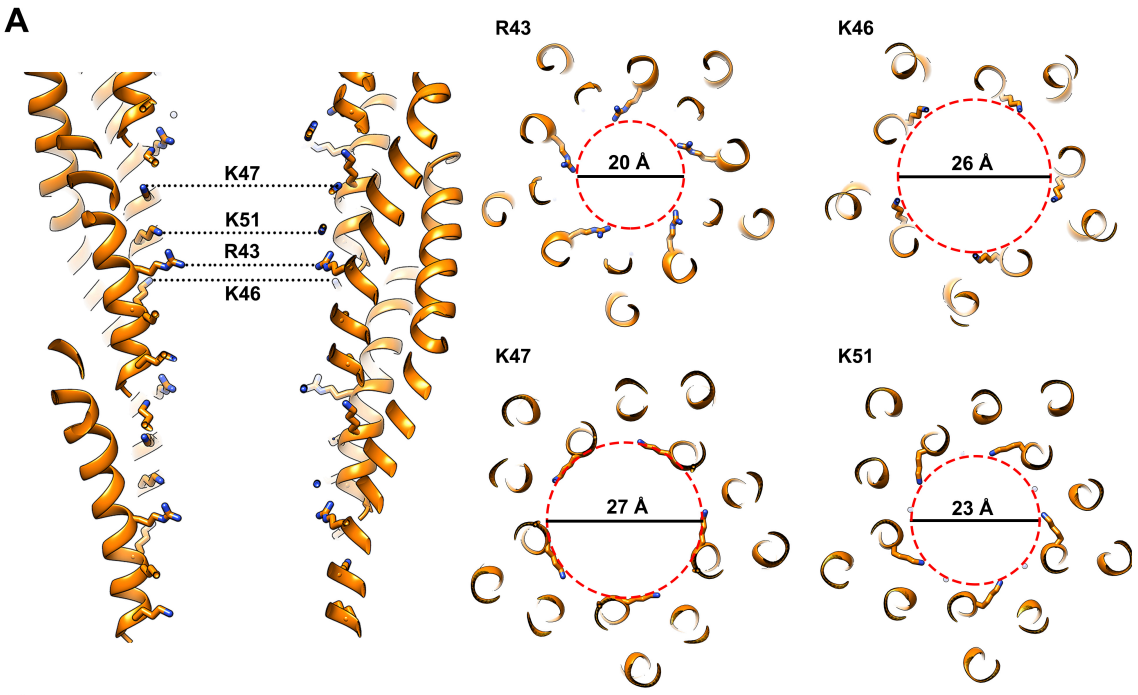

A
253156 particles by autopick
from 493 micrographs
2D classification
3D classification
224416 particles
selected
Auto-refinement with optimized
helical parameters
binning=2
Auto-refinement
binning=1
Low-pass filtered to: 5.5 Å, 6.0 Å, 7.0 Å, 9.0 Å
resolution = 5.2 Å
resolution = 4.1 Å
Multiple-reference 3D classification
112808 particles
binning=2
binning=1
30.3%
0.4%
19.7%
49.6%
resolution = 3.4 Å
Auto-refinement
B
224416 particles
with capsid shell subtraction
symmetry expansion
224416 x 5 particles
binning=2
3D classification with cylinder reference
3x parallel run
binning=2
2nd round of 3D classification
3x parallel run
72°
rotate around
the virion axis
144°
288°
53784 particles
low-pass filtered
to 5 Å
overlap with particles using in final
capsid shell reconstruction

Fig. S3. Structure determination workflow.

(**A**) Flowcharts showing the particle selection and cryo-EM 3D reconstruction procedure. Please refer to SI Appendix, Image Processing for details.

(**B**) Flowcharts showing the orientation determination and cryo-EM 3D reconstruction of the inner DNA core. Please refer to SI Appendix, Image Processing for details. The scale bar is 5 nm. The DNA core classes with helical features in 3D classification were related to each other by an n×72º rotation along virion axis (n is a fixed value for each class).


A
K47
K51
R43
K46
R43
20 Å
K46
26 Å
K47
27 Å
K51
23 Å

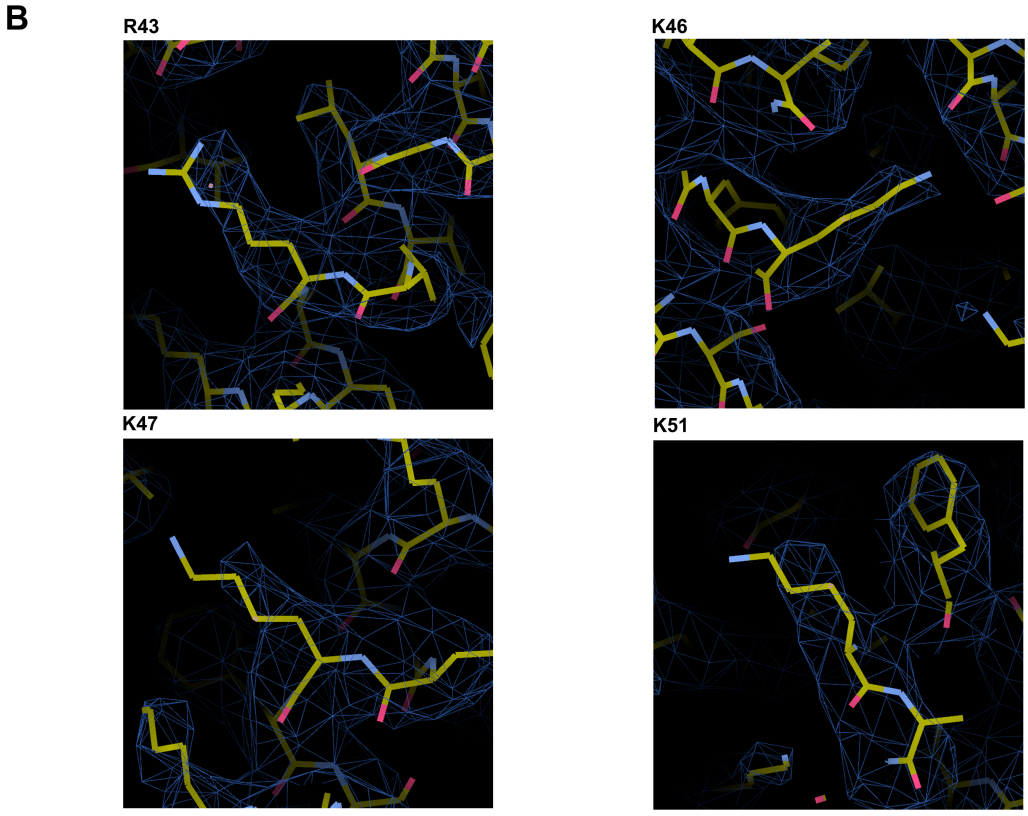

B
R43
K46
K47
K51

Fig. S4. Structural analysis of the positively charged residues in H_C.

(**A**) Left: central section ribbon diagrams showing the lumen of the IKe shell assembly. The side chains of four positively charged residues (R43, K46, K47 and K51) are shown in sticks. Right: cutoff views showing the four positively charged residues. Distances between the positively charged residues are measured and shown.

(**B**) Diagrams showing the side chains of the four positively charged resides and the relative electron density maps.

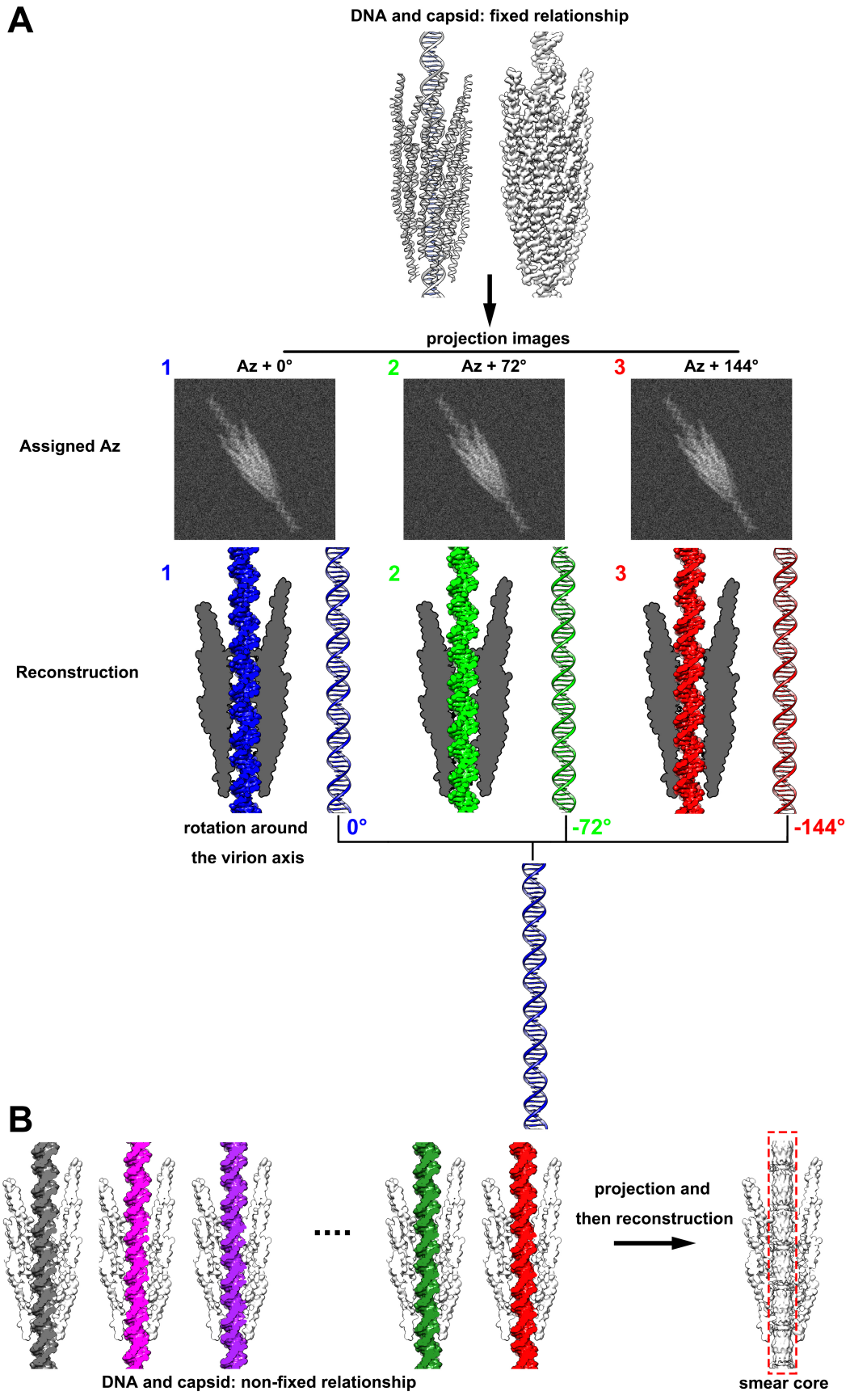

A
DNA and capsid: fixed relationship
projection images
1
Az + 0°
2
Az + 72°
3
Az + 144°
Assigned Az
Reconstruction
rotation around
the virion axis
0°
-72°
-144°
B
projection and
then reconstruction
DNA and capsid: non-fixed relationship
smear core

Fig. S5. Helical reconstruction simulations suggested a fixed relationship between the inner DNA core and the capsid shell.

(**A**) Top: The initial structure model was generated from the capsid model and a double stranded DNA helix model. The relative position of the helix to the capsid shell was fixed. Middle: Random projections generated from the model. Three sets of the projections were generated, in which the az angles of the projections were set to be the correct az, az+72° and az+144°, respectively. Three dimensional reconstructions were calculated from the datasets with the arbitrarily assigned az angles. Differences were only observed for the helix core in the simulated reconstructions.

Bottom: The helix core is the same as the original model with the dataset that has the correct az angle, whereas the helix core calculated from the dataset with the assigned wrong angle of az+72° or az+144° is identical to the original model after being rotated around the long axis by 72° or 144°.

(**B**) Left: structure model generated from the capsid model and a double stranded helix model, in which the relative position of the helix to the capsid shell was not fixed. Right: a 3D reconstruction calculated from the random projections of the non-fixed models showing smeared featureless densities inside.

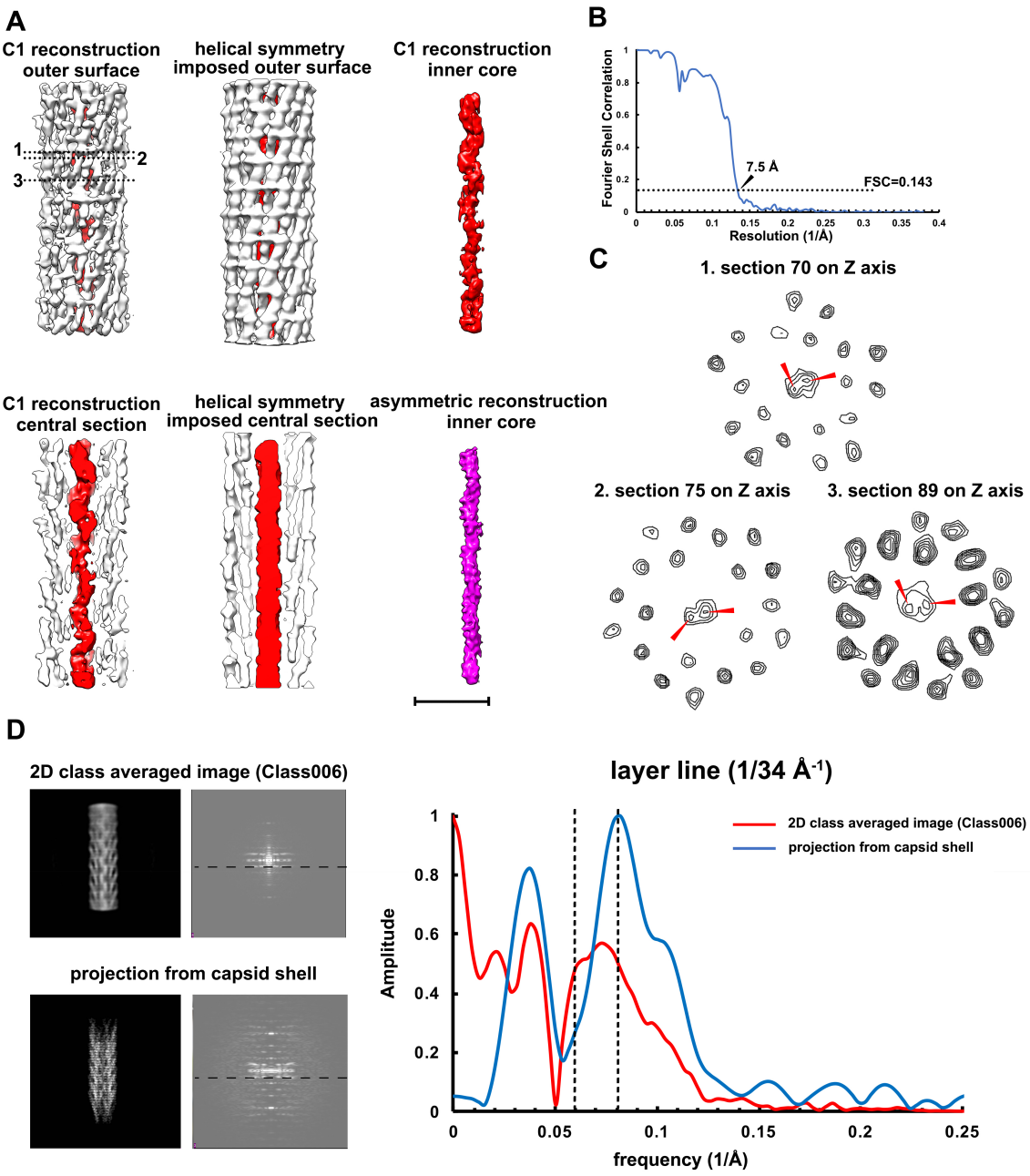

A
C1 reconstruction outer surface
helical symmetry imposed outer surface
C1 reconstruction inner core
1
2
3
C1 reconstruction central section
helical symmetry imposed central section
asymmetric reconstruction inner core
B
Fourier Shell Correlation
Resolution (1/Å)
7.5 Å
FSC=0.143
C
1. section 70 on Z axis
2. section 75 on Z axis
3. section 89 on Z axis
D
2D class averaged image (Class006)
projection from capsid shell
layer line (1/34 Å-1)
2D class averaged image (Class006)
projection from capsid shell
Amplitude
frequency (1/Å)

Fig. S6. Cryo-EM and Fourier analysis of the inner DNA core.

(A) Left: surface (top) and central section (bottom) diagrams of the reconstruction without imposing any symmetry (C1 reconstruction) showing the helical features of both the capsid and the inner ssDNA core. The capsid shell and the inner ssDNA core were colored white and red, respectively; middle: surface (top) and central section (bottom) diagrams of the C1 reconstruction with helical symmetry (rise = 16.77 Å, twist = 38.52°) applied in real space. Right: structural comparison of the inner ssDNA cores from the C1 (top, red) and the asymmetric (bottom, magenta) reconstructions. The scale bar is 5 nm.

(B) Gold standard Fourier shell correlation (FSC) curves of the C1 reconstruction.

(C) Sections along Z axis of the asymmetric reconstruction showing two peaks (indicated by arrows) of the inner DNA core, suggesting possible two strands of the circular ssDNA. The map is contoured from 10.0 σ to 20.0 σ with a step size of 1.0 σ.

(**D**) Left: Power spectra of a representative 2D class averaged image and a projection of the capsid shell. The layer line at 1/34 $Å^{-1}$ was highlighted with a dash line. Right: profile analysis of the layer lines at 1/34 $Å^{-1}$ showing there is a peak at R=0.06 $Å^{-1}$ in the 2D class averaged image.

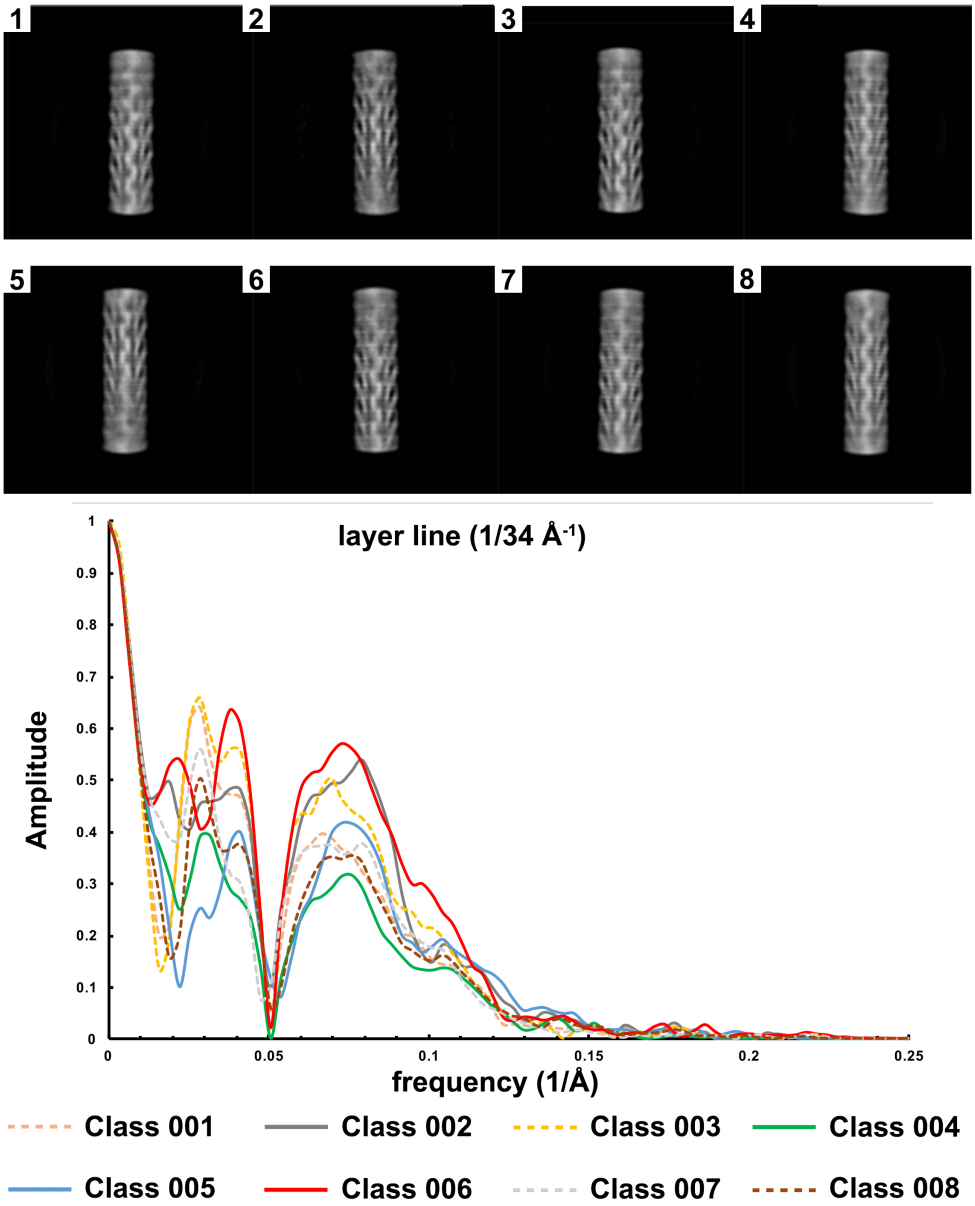

1
2
3
4
5
6
7
8
layer line (1/34 Å-1)
Amplitude
frequency (1/Å)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.05
0.1
0.15
0.2
0.25
Class 001
Class 002
Class 003
Class 004
Class 005
Class 006
Class 007
Class 008

Fig. S7. Power spectra of 2D class averaged images.

The 2D class averaged images are shown at the top. Bottom: Profiles of the layer lines at 1/34 $Å^{-1}$ in the power spectra from the 2D class averaged images.

Table S1. Statistics for structure determination and refinement

Data collection	Bacteriophage IKe
EM equipment	FEI Titan Krios
Voltage (kV)	300
Detector	K2 Summit
Electron dose ($e^-/Å^2$)	~40
Defocus range (μm)	1.0-2.2
No. of micrographs	573
No. of particles	253,156
Reconstruction	
Software	Relion2
No. of particles used	112,808
Final resolution (Å)	3.4
Map sharpening B-factor ($Å^2$)	-126.8
Model building	
Software	Coot
Refinement	
Software	*phenix.real_space_refine*
Model-map CC (mask)	0.8250
Model composition	
Protein residues	1410
No. of atoms	10830
No. of subunits	30
Validation	
R.M.S.Ds	
Bond length (Å)	0.0144
Bond angle (°)	1.41
Ramachandran statistics (%)	
Favored	100.00
Allowed	0.00
Outliers	0.00
MolProbity validation	
Rotamer outliers (%)	0.00
Clashscore	4.26
MolProbity score	1.21
C-beta outliers	0

References

1. Russel M (1992) Interchangeability of related proteins and autonomy of function. The morphogenetic proteins of filamentous phage f1 and IKe cannot replace one another. *J Mol Biol* 227(2):453-462.
2. Sambrook J & Russell DW (2001) *Molecular cloning : a laboratory manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.) 3rd Ed.
3. Zheng SQ*, et al.* (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* 14(4):331-332.
4. Zhang K (2016) Gctf: Real-time CTF determination and correction. *Journal of Structural Biology* 193(1):1-12.
5. He SD & Scheres SHW (2017) Helical reconstruction in RELION. *Journal of Structural Biology* 198(3):163-176.
6. Kucukelbir A, Sigworth FJ, & Tagare HD (2014) Quantifying the local resolution of cryo-EM density maps. *Nat Methods* 11(1):63-65.
7. Zhang XF*, et al.* (2017) An Atomic Structure of the Human Spliceosome. *Cell* 169(5):918-929.
8. Bai XC, Rajendra E, Yang G, Shi Y, & Scheres SH (2015) Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife* 4.
9. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography* 60:2126-2132.
10. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* 66:213-221.
11. Pettersen EF*, et al.* (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605-1612.
12. Gouet P, Courcelle E, Stuart DI, & Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15(4):305-308.
13. Scheres SHW & Chen SX (2012) Prevention of overfitting in cryo-EM structure determination. *Nature Methods* 9(9):853-854.