Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/5/10/eaaw7414/DC1

Supplementary Materials for

Structure and genome ejection mechanism of *Staphylococcus aureus* phage P68

Dominik Hrebík, Dana Štveráková, Karel Škubník, Tibor Füzik, Roman Pantůček, Pavel Plevka*

*Corresponding author. Email: pavel.plevka@ceitec.muni.cz

Published 16 October 2019, *Sci. Adv.* **5**, eaaw7414 (2019)
DOI: 10.1126/sciadv.aaw7414

This PDF file includes:

Fig. S1. Purified sample of P68 contains native virions, particles in process of genome release, and empty particles; attachment of P68 virions to *S. aureus* cell walls; and interactions of P68 virions with liposomes.
Fig. S2. Resolution and interpretability of cryo-EM reconstructions.
Fig. S3. Details of P68 head.
Fig. S4. Incorporation of P68 portal complex into capsid and changes in the structure of P68 portal complex upon genome release.
Fig. S5. Structures of P68 tail fiber and tail spike.
Fig. S6. Schemes of cryo-EM reconstruction strategies.
Table S1. Cryo-EM structure quality indicators.
Table S2. List of P68 proteins.
Table S3. HHpred searches for homologs of P68 tail fiber, head fiber, tail knob, and tail spike.
Table S4. X-ray structure quality indicators.

Supplementary figures:

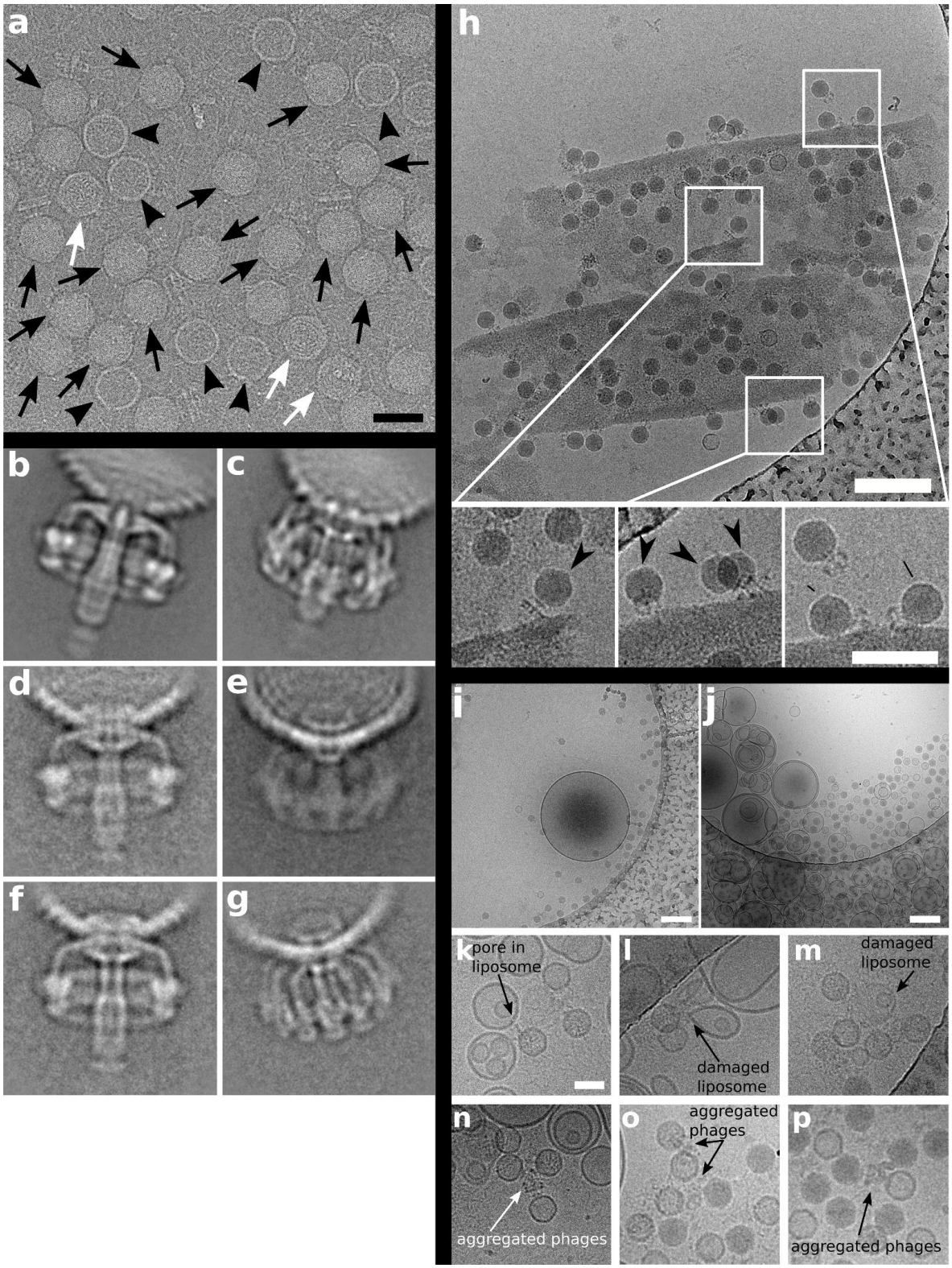

Fig. S1. Purified sample of P68 contains native virions, particles in process of genome release, and empty particles; attachment of P68 virions to *S. aureus* cell walls; and interactions of P68 virions with liposomes. (a) Electron micrograph of P68 showing native virions (black arrows), particles in process of genome release (white arrows), and empty particles (black arrowheads). Scale bar 50 nm. (b to g) Reference-free two-dimensional class averages of native virions with complete tails (b); genome release intermediates with complete (c) and shortened tails (d); and empty particles with complete (e) and shortened tails (fg). (h) Attachment of P68 to fragments of *S. aureus* cell wall. At the stage of initial attachment the tail axis of P68 is not perpendicular to the surface of the cell wall. Scale bar 200 nm, scale bar in inset 100 nm. (i to p) P68 particles do not bind to liposomes. P68 virions were mixed with liposomes with lipid composition that mimics that of *S. aureus* membrane at neutral (i) and acidic (j) pH. (k to m) Liposomes in mixture with P68 at acidic pH are distorted. (n-p) P68 particles aggregate with each other by interactions of their tails at acidic pH. Scale bars 200 nm in (ij) and 50 nm in (k-p).

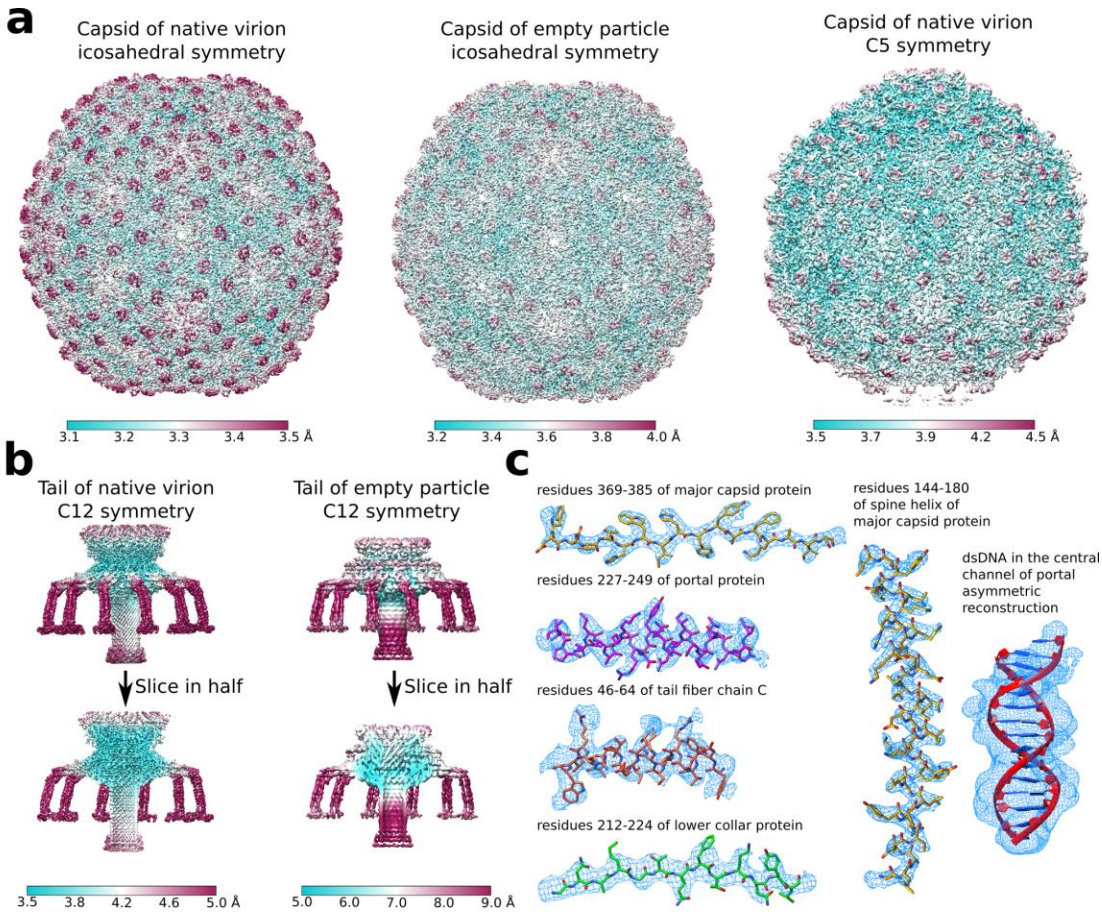

Fig. S2. – continues on next page.

d

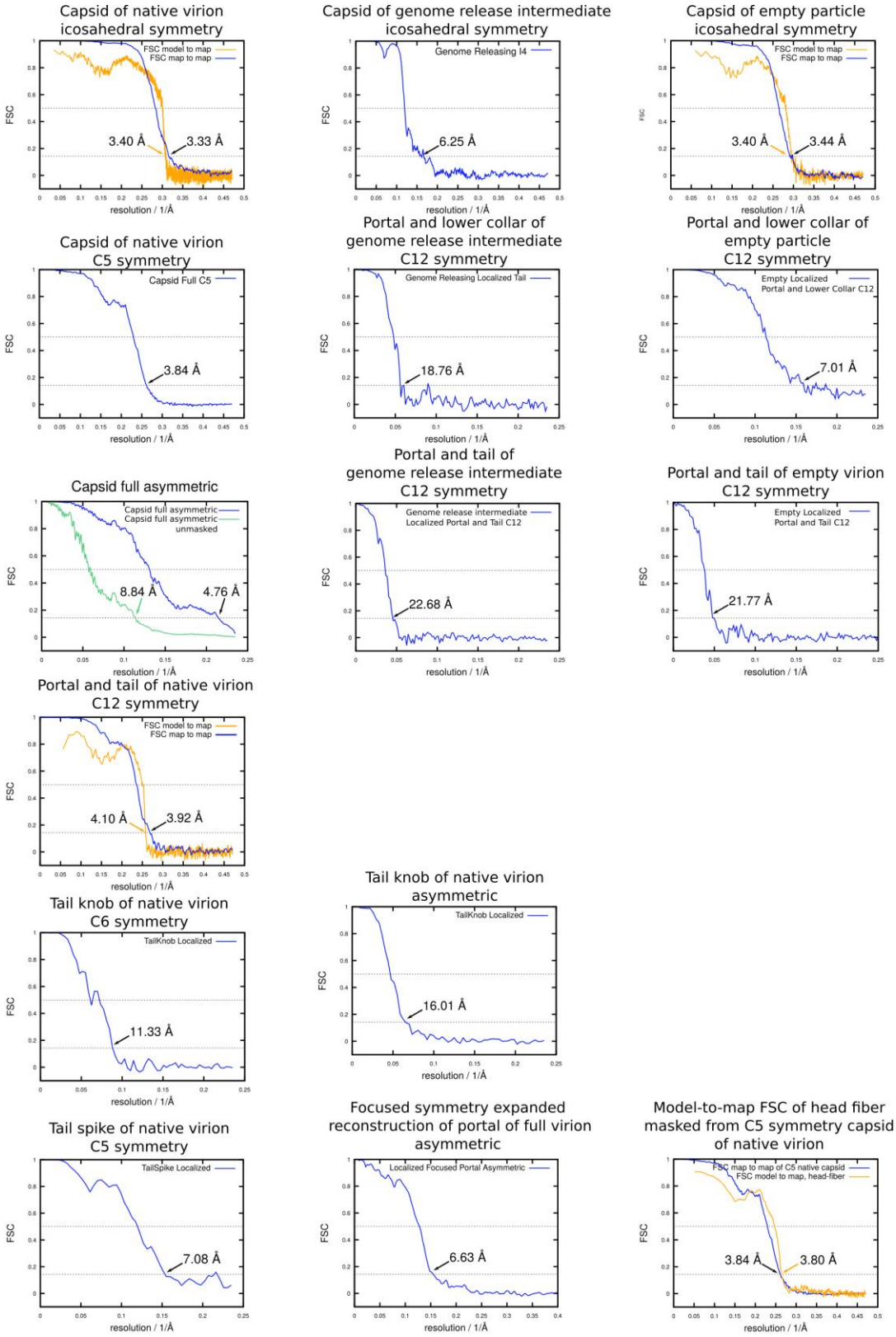

Fig. S2. Resolution and interpretability of cryo-EM reconstructions. (ab) Local resolution maps of cryo-EM reconstructions described in manuscript. (c) Representative cryo-EM density maps with the P68 virion proteins shown in stick representations. (d) FSC curves of cryo-EM reconstructions.

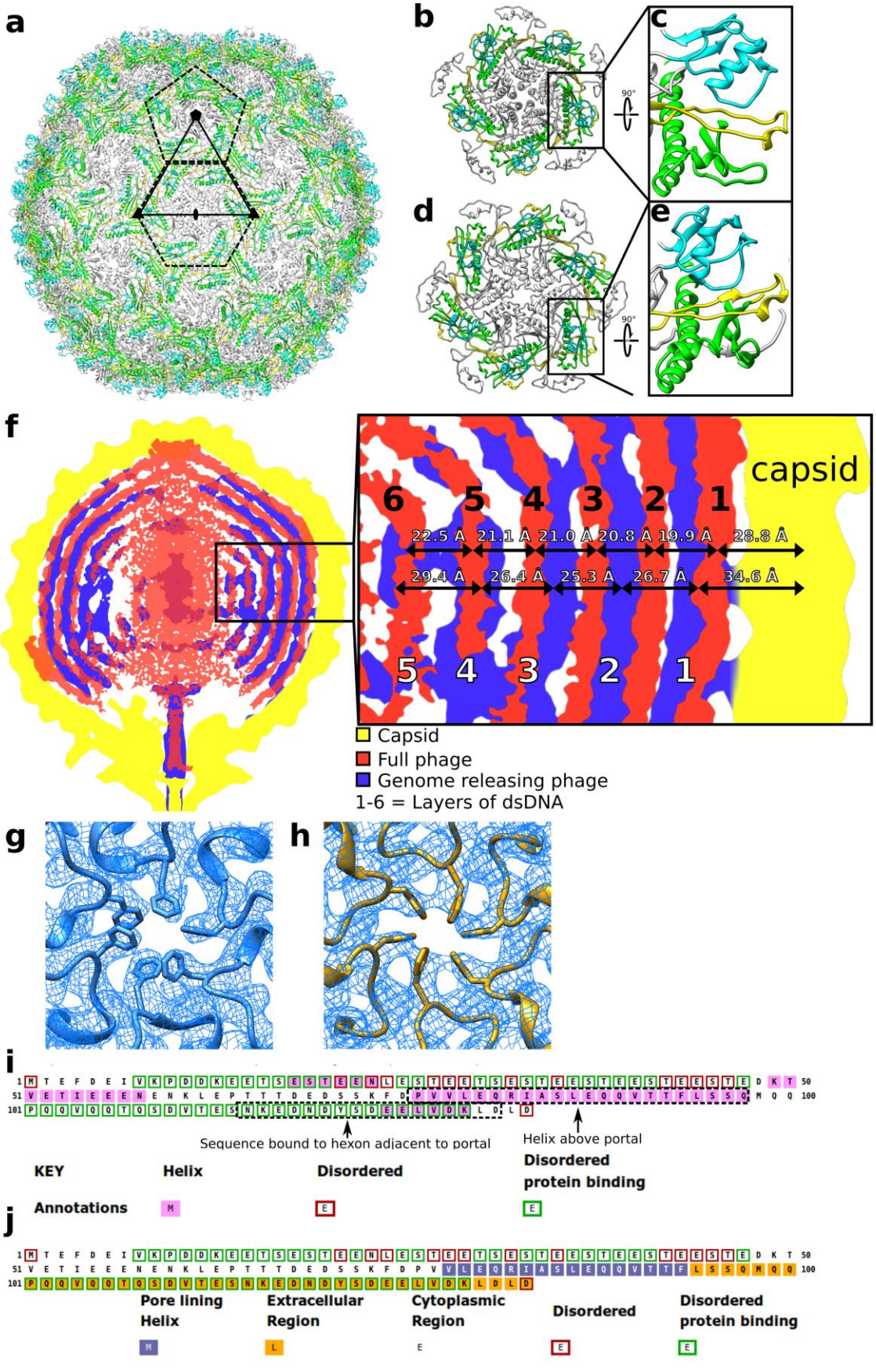
a
b
c
d
e
90°
90°
f
capsid
6 5 4 3 2 1
22.5 Å 21.1 Å 21.0 Å 20.8 Å 19.9 Å 28.8 Å
29.4 Å 26.4 Å 25.3 Å 26.7 Å 34.6 Å
5 4 3 2 1
Capsid
Full phage
Genome releasing phage
1-6 = Layers of dsDNA
g
h
i
Sequence bound to hexon adjacent to portal
Helix above portal
KEY
Helix
Disordered
Disordered protein binding
Annotations
j
Pore lining Helix
Extracellular Region
Cytoplasmic Region
Disordered
Disordered protein binding

Fig. S3. Details of P68 head. (a to e) Insertion and peripheral domains of major capsid protein form cleft that binds extended loop of neighboring capsid protein within pentamer or hexamer. (a) Overview of capsid protein interactions in P68 head. The major capsid proteins are shown in cartoon representation in grey, with insertion domains highlighted in blue, peripheral domains in green, and extended loops in yellow. Interactions of major capsid proteins in pentamers (b) and hexamers (c). Details of binding of extended loops by insertion and peripheral domains in pentamers (d) and hexamers (e). (f) Spacing of DNA layers in native P68 virions and genome release intermediate. The DNA of native virions is shown in red and that of genome release intermediates in blue. Phage proteins are shown in yellow. (gh) Comparison of structures of side chains of phe 259 of major capsid proteins in hexamers that form attachment sites for head fibers (a) and in hexamers that do not bind head fibers (b). Cryo-EM maps are shown as blue mesh. (ij) Inner core protein of P68 contains predicted pore lining helix. Sequence analysis of inner core protein of P68 (gp22). The program PSIPRED predicts propensity of several of its sequences to form α-helices (a). Longest α-helix of inner core protein has properties of pore-lining helix (b), indicating that it may participate in forming the pore in the cytoplasmic membrane for genome delivery.

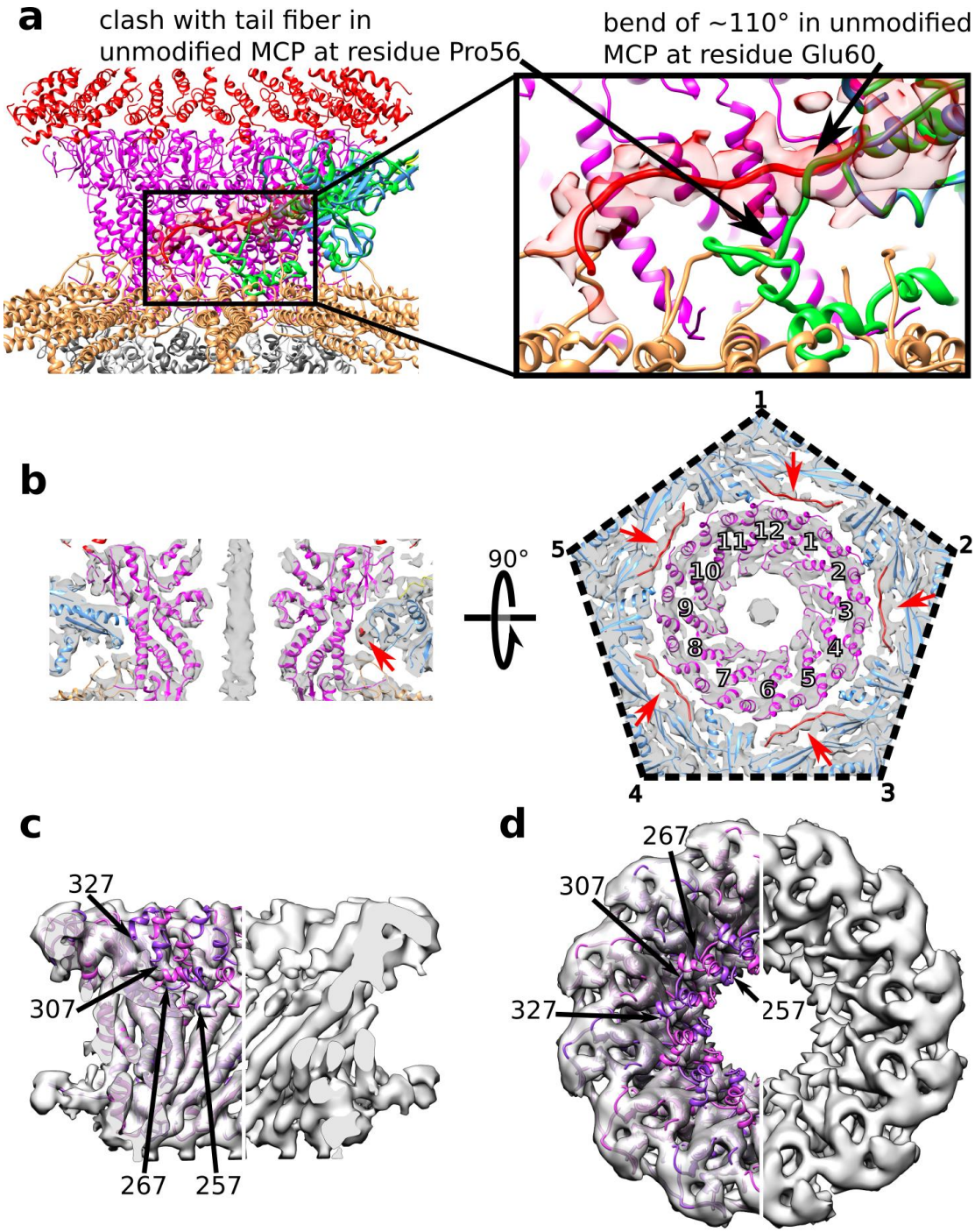
a
clash with tail fiber in
unmodified MCP at residue Pro56
bend of ~110° in unmodified
MCP at residue Glu60
b
90°
1
2
3
4
5
6
7
8
9
10
11
12
c
327
307
267
257
d
267
307
327
257

Fig. S4. Incorporation of P68 portal complex into capsid and changes in the structure of P68 portal complex upon genome release. (ab) Comparison of structures of N-terminal arms of major capsid proteins that interact with pentamers of capsid proteins and portal complex. Portal proteins are shown in magenta, major capsid proteins in blue, N-terminal arms of capsid proteins interacting with pentamers of capsid proteins in green, and N-terminal arms of capsid proteins interacting with portal complex in red. (a) If the N-terminal arm of major capsid protein binding to the portal complex had the same structure as in the vertices occupied by capsid proteins, the N-terminal arm would clash with the portal and tail fibers. (b) Side and top views of capsid-portal interactions. (cd) Differences between structures of portal complexes of native P68 virion (left halves of images and cartoon structures) and empty particle (right halves of images). Side-view of portal complex cut in half (a) and top view of portal complex (b). Cryo-EM densities of the portal complexes are shown as grey semi-transparent surfaces.

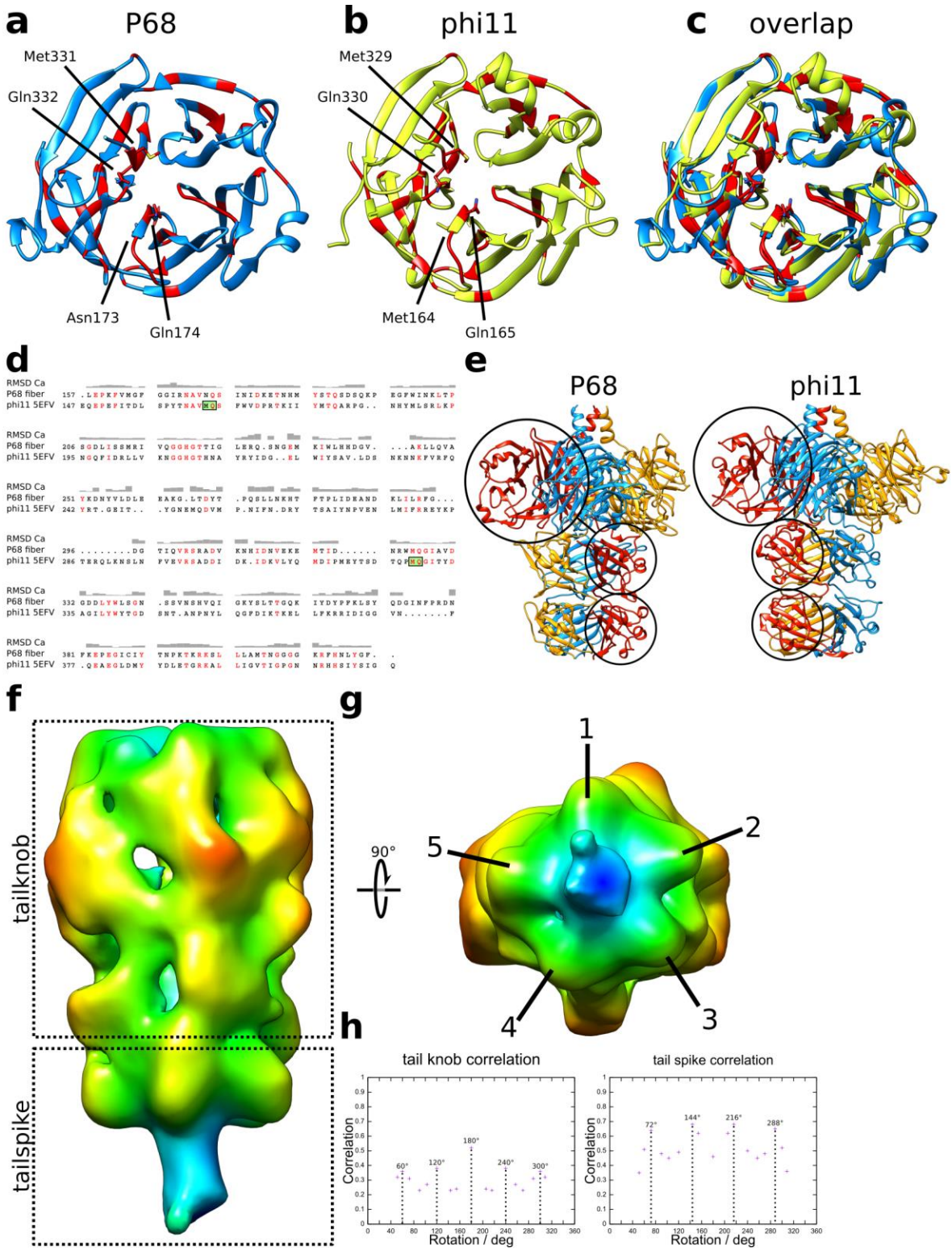
a
P68
Met331
Gln332
Asn173
Gln174
b
phi11
Met329
Gln330
Met164
Gln165
c
overlap
d
RMSD Ca
P68 fiber
phi11 5EFV
e
P68
phi11
f
tailknob
tailspike
g
1
2
3
4
5
90°
h
tail knob correlation
tail spike correlation
Correlation
Rotation / deg
60°
120°
180°
240°
300°
72°
144°
216°
288°

Fig. S5. Structures of P68 tail fiber and tail spike. (a to d) Comparison of structures of platform domains of P68 solved by X-ray crystallography at 2.0 Å in blue (a) and phi11 in green (b). Residues conserved among the two viruses are shown in red in both of the structures. Residues that are involved in the receptor binding of phi11 and the corresponding residues of P68 are shown as sticks and indicated with labels. (c) Superposition of the two structures. (d) Structure-based sequence alignment of platform domains of P68 and phi11. Relative distances of the corresponding Cα atoms are indicated. Residues of phi11 involved in receptor binding are highlighted with a green background. Residues conserved among the two viruses are shown in red. (e) X-ray crystallography structure solved at 2.0 Å resolution of the tail-fiber shows swapping between platform and tower domains of tail fiber proteins of P68 and phi11. (f to h) Asymmetric reconstruction of tail knob and tail spike complexes. (fg) The surface of the map is rainbow colored based on its distance from the tail axis. (h) Correlation coefficients of tail knob and tail spike parts of the reconstructions with rotated copies of themselves. The results indicate that the tail knob has sixfold symmetry and the tail spike has fivefold symmetry.

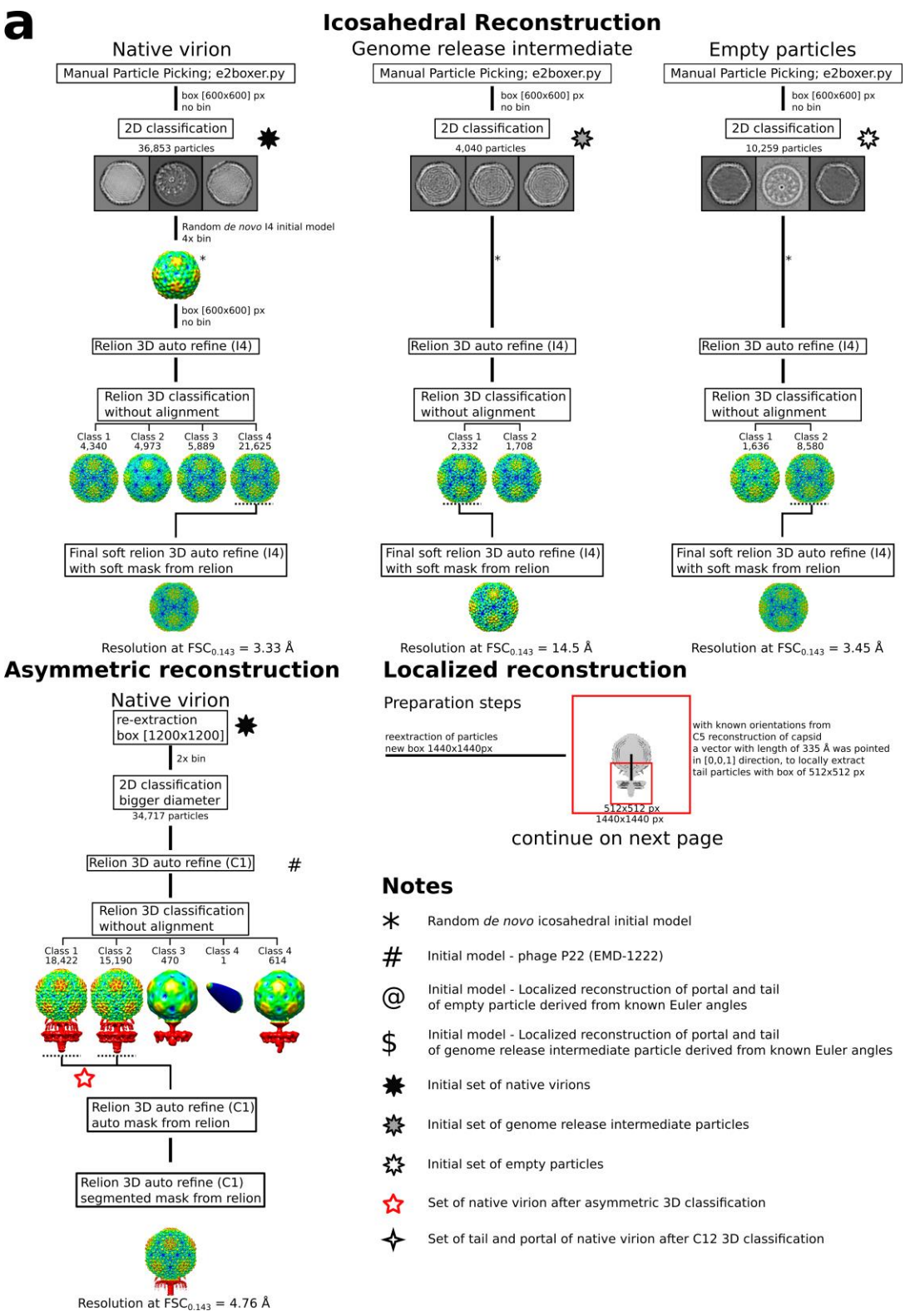

Fig. S6. Schemes of cryo-EM reconstruction strategies.

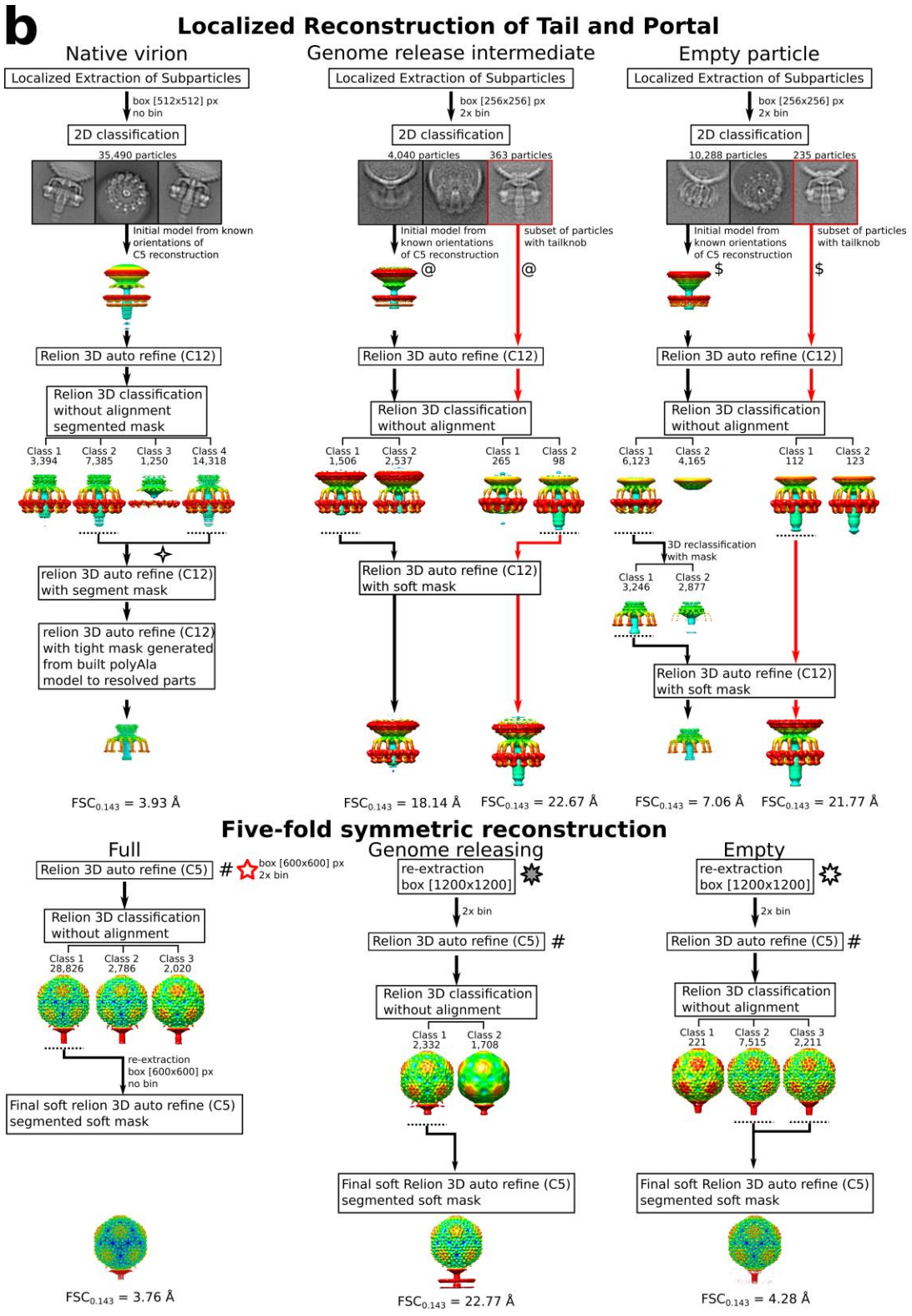

Fig. S6. Schemes of cryo-EM reconstruction strategies - continued.

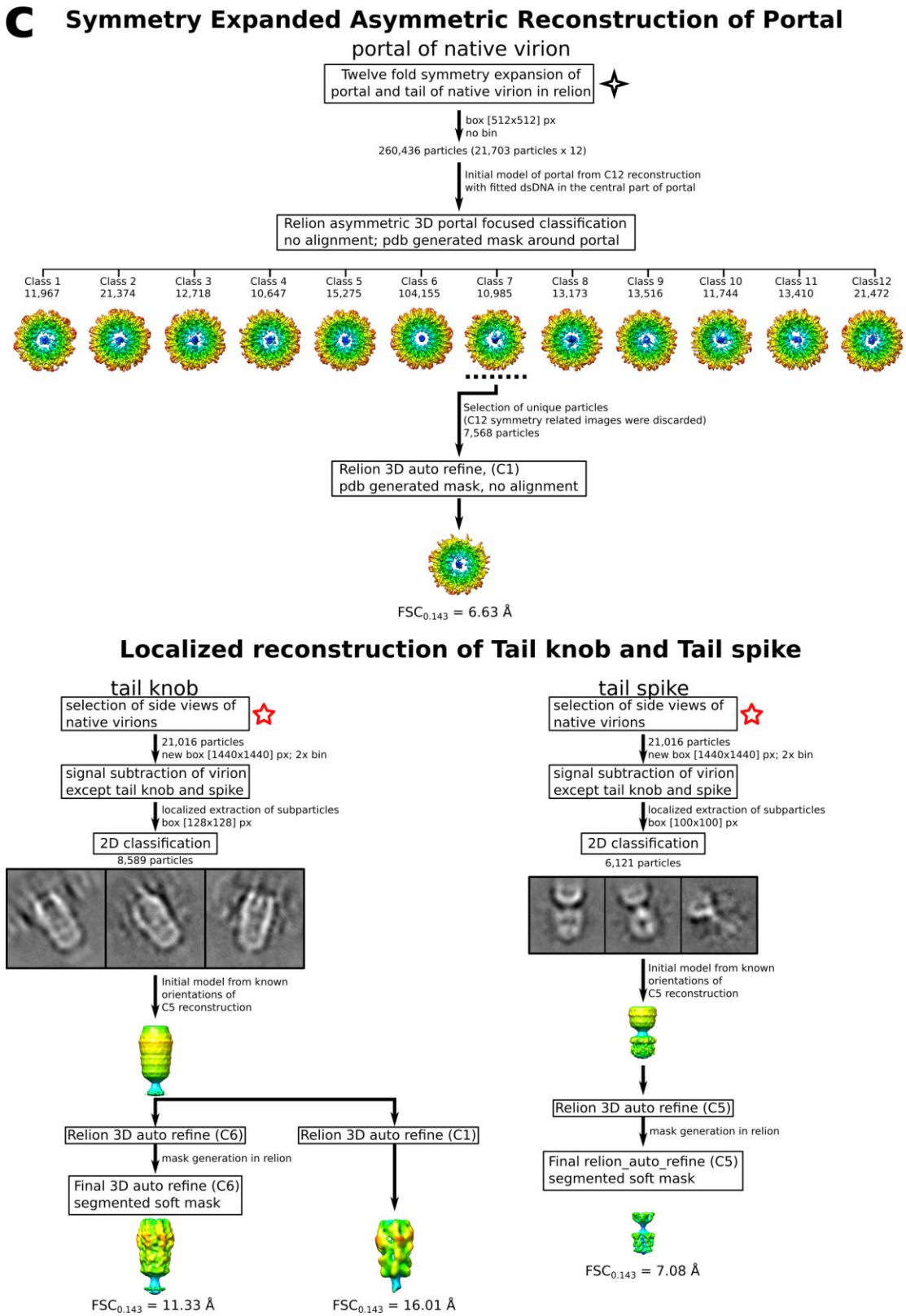

Fig. S6. Schemes of cryo-EM reconstruction strategies - continued.

Supplementary tables:

Table S1. Cryo-EM structure quality indicators.

Parameters	Whole virion (combination of maps)	Capsid	Native virion	Capsid	Native virion	Capsid	Native virion	Head Fiber	Gen. release intermediate	Capsid	Gen. release intermediate	Capsid	Empty particle	Capsid	Empty particle	Localized Tail and Portal	Native virion	Localized Tail and Portal	Gen. release intermediate	Localized Tail and Portal	Empty particle	Localized Tail and Portal	Native virion	Localized Knob	Native virion	Localized Spike	Gen. release intermediate	Localized Tail with knob	Empty particle	Localized Tail with knob
Data collection and processing																														
Magnification	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x	75,000x
Voltage (kV)	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300
Exposure ($e^- \cdot Å^{-2}$)	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2	21.2
Pixel size (Å)	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063	1.063
Symmetry	-	I4	C5	C1	C5 (masked from C5 map)	I4	C5	I4	I4	C5	I4	C5	I4	C5	C12	C12	C12	C6	C5	C12	C5	C12	C12	C12	C12	C5	C12	C12	C12	C12
Init. No of particles	-	36,853	34,717	34,717	34,717	4,040	4,040	10,259	10,259	9,726	10,259	9,726	10,288	6,123	8,598	6,129	7,08	235												
Final No of particles	-	21,625	28,826	33,512	28,826	2,330	2,332	8,580	3,44	4,28	21,703	1,506	18,76	7,06	11,33	7,08	22,68	21,77												
Map resolution (Å)	-	3,33	3,82	4,76	3,82	8,37	22,77	3,44	4,28	3,92	1,506	18,76	7,06	11,33	7,08	22,68	21,77	0,143												
FSC threshold	-	0,143	0,143	0,143	0,143	0,143	0,143	0,143	0,143	0,143	0,143	0,143	0,143	0,143	0,143	0,143	0,143	0,143												
Refinement																														
Initial model	-	*de novo*	code (EMD-1222)	code (EMD-1222)	code (EMD-1222)	*de novo*	code (EMD-1222)	*de novo*	code (EMD-1222)	code (EMD-1222)	*de novo*	code (EMD-1222)	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*												
Accession codes																														
EMDB	EMD-4459	-	EMD-4436	EMD-4437	EMD-4440	-	EMD-4453	EMD-4442	EMD-4456	EMD-4435	EMD-4454	EMD-4457	EMD-4449	EMD-4450	EMD-4455	EMD-4458														
PDB	6Q3G	6IAT	-	-	6IAW	-	-	6IB1	-	6IAC	-	-	-	-	-	-														
Model composition																														
Atoms (except hydrogens)	1,074,183	14,138[a]	-	-	21,915	-	14,138[a]	-	8,218[a]	-	-	-	-	-	-	-														
Residues	133,777	1,728[a]	-	-	2,790	-	1,728[a]	-	1,029[a]	-	-	-	-	-	-	-														
B-factors																														
Protein	-	40,1	-	-	153,2	-	40,8	-	107,1	-	-	-	-	-	-	-														
R.M.S. deviations																														
Bond Lengths (Å)	-	0,009	-	-	0,009	-	0,009	-	1,17	-	-	-	-	-	-	-														
Bond Angles (°)	-	1,17	-	-	1,45	-	1,17	-	1,30	-	-	-	-	-	-	-														
Validation																														
MolProbity score	-	2,03 (100[th])	-	-	2,30 (99[th])	-	1,77 (100[th])	-	1,87 (100[th])	-	-	-	-	-	-	-														
(percentile)																														
ClashScore (percentile)	-	7,95 (97[th])	-	-	10,0 (97[th])	-	4,48 (100[th])	-	4,62 (100[th])	-	-	-	-	-	-	-														
Poor Rotamers (%)	-	1,21	-	-	1,77	-	0,57	-	1,09	-	-	-	-	-	-	-														
Ramachandran plot																														
Outliers (%)	-	0,35	-	-	0,22	-	0,23	-	0,19	-	-	-	-	-	-	-														
Favored (%)	-	90,68	-	-	88,53	-	90,10	-	87,85	-	-	-	-	-	-	-														
R factors																														
R work	-	0,3193	-	-	-	0,3459	-	0,3559	-	0,3597	-	-	-	-	-	-														
R free	-	0,3181	-	-	-	0,3461	-	0,3390	-	0,3803	-	-	-	-	-	-														

[a] - in asymmetric unit of icosahedron
[b] - in asymmetric unit of dodecameric tail and portal

Table S2. List of P68 proteins.

ORF No.	Protein name/ Function	Structural?/ Found in MS	Amino acid sequence length	Symmetry	No. of chains in ASU	No. of modelled residues	Not modelled residues	Model Completness	Homologous proteins PDB codes
1	hypothetical protein/-	No/No	65	-	-	-	-	-	-
2	hypothetical protein/-	No/No	100	-	-	-	-	-	-
3	hypothetical protein/-	No/No	78	-	-	-	-	-	-
4	hypothetical protein/ Putative ssDNA binding protein	No/Yes	122	-	-	-	-	-	-
5	hypothetical protein/-	No/No	58	-	-	-	-	-	-
6	hypothetical protein/-	No/No	136	-	-	-	-	-	-
7	hypothetical protein/-	No/Yes	55	-	-	-	-	-	-
8	hypothetical protein/-	No/No	160	-	-	-	-	-	-
9	Putative Encapsidation Protein	No/No	378	-	-	-	-	-	-
10	DNA polymerase	No/No	755	-	-	-	-	-	-
11	Tailspike/ Cell Wall Degrading Enzyme	Yes/Yes	479	C5	-	-	-	-	4F88, 5T1Q, 2K3A, 2LRJ,4CT3 5D74, 5UDN, 4CGK, 4HZ9
12	Holin	No/No	140	-	-	-	-	-	-
13	Tailknob	Yes/Yes	587	C6	-	-	-	-	5FB5, 4EP0
14	Head Fiber/ Receptor Binding	Yes/Yes	481	C3 (trimer) C5 (in whole phage, pentamer of trimers)	1	A: 54	A: 1; 56-481	A: 11.4%	(C-terminal part) 4HEM, 2F0C, 3DA0, 3D8M, 2DSD, 1ZRU
15	Holin	No/No	92	-	-	-	-	-	-
16	Endolysin	No/No	250	-	-	-	-	-	5UDN, 5D74, 4CT3, 2K3A, 5O1Q, 2MK5
17	Tail Fiber/ Receptor Binding	Yes/Yes	647	C3 (trimer) C12 (dodecamer of trimers, in native phage structure)	3	A: 633 B: 636 C: 617	A: 1-7; 322-327; 646-647 B: 1-4; 322-327; 646-647 C: 1-23; 322-327; 646-647	A: 97.8% B: 98.2% C: 95.3%	5EFV
18	Lower Collar	Yes/Yes	251	C12	1	A: 221	A: 1-2; 155-183	A: 88.0%	-
19	Portal	Yes/Yes	327	C12	1	A: 301	A: 1-5; 82-103	A: 92.0%	1FOU, 2JES, 3JL4, 3JA7
20	Major Capsid Protein	Yes/Yes	408	Quasi-icosahedral	4	A: 393 B: 386 C: 386 D: 393	A: 1-3; 396-408 B: 1-10; 396-408 C: 1-10; 396-408 D: 1-3; 396-408	A: 96.3% B: 94.6% C: 94.6% D: 96.3%	all HK97 folds
21	Arstotzka Protein	Yes/Yes	60	Quasi-icosahedral	4	A: 54 B: 55 C: 27 D: 55	A: 1-4; 60 B: 1-4; 60 C: 1-4; 32-60 D: 1-4; 60	A: 91.6% B: 91.6% C: 46.6% D: 91.6%	-
22	Inner Core Protein	Yes/Yes	133	C3, C2 C12 (dodecamer of hexamers in native phage)	6	A: 17 B: 17 C: 23 D: 22 E: 21 F: 21	A: 1-75; 94-133 B: 1-75; 94-133 C: 1-73; 98-133 D: 1-73; 97-133 E: 1-73; 96-133 F: 1-73; 96-133	A: 12.7% B: 12.7% C: 17.2% D: 16.5% E: 15.7% F: 15.7%	-

Table S3. HHpred searches for homologs of P68 tail fiber, head fiber, tail knob, and tail spike.

Tower domain of tail fiber protein HHblits homolog search

Protein HHblits hit	UNIPROT code	Probability / %[#]	E-value[%]	P-value[*]	Score[$]	Aligned Cols[α]	Query HMM[β]	Template HMM[γ]	Sequence identity / %	Similarity[κ]
major teichoic acid biosynthesis protein C, *S. aureus*	A0A1D7UFH3	97.0	3.3×10^{-5}	1.1×10^{-10}	75.3	166	3-174	263-429	31	0.501
muramidase, *B. campestris*	W7D9P4	96.1	9.5×10^{-4}	3.1×10^{-9}	61.4	83	97-179	7-89	34	0.547
muramidase, *B. thermosphacta*	W7CPF7	96.0	1.5×10^{-3}	4.9×10^{-9}	64.9	161	4-174	290-452	29	0.492
bacteriophage related putative muramidase, *B. subtilis*	E0TX48	94.0	4.3×10^{-2}	1.4×10^{-7}	59.5	127	1-137	1007-1140	29	0.563
major teichoic acid biosynthesis protein C, *S. aureus* phage P240	A0A1X9IGV7	92.8	1.2×10^{-1}	4.1×10^{-7}	51.5	150	6-171	173-335	31	0.588

C-terminal domain of head fiber HHpred homolog search

Protein HHpred hit	pdb code	Probability / %[#]	E-value[%]	P-value[*]	Score[$]	Aligned Cols[α]	Query HMM[β]	Template HMM[γ]	Sequence identity / %	Similarity[κ]
receptor binding protein; phage TP901-1	4IOS	94.5	0.45	4.6×10^{-6}	33.4	77	23-101	7-87	3	-0.123
receptor binding protein; lactococcal phage p2	3DA0	92.4	2.2	2.2×10^{-5}	32.0	71	29-101	44-118	11	0.059
receptor binding domain; phage blL170	2FSD	90.5	3.4	3.5	33.2	87	3-99	32-124	15	0.211

Tail knob protein HHpred homolog search

Protein HHpred hit	pdb code	Probability / %[#]	E-value[%]	P-value[*]	Score[$]	Aligned Cols[α]	Query HMM[β]	Template HMM[γ]	Sequence identity / %	Similarity[κ]
bacteriophage C1 tail knob	4EO2	100.0	2×10^{-117}	2×10^{-112}	981.4	563	2-587	1-576	30	0.558
bacteriophage phi29 tail knob	5FB4	100.0	2×10^{-113}	2×10^{-118}	951.2	544	1-579	5-598	19	0.272

Tail spike protein HHpred homolog search

Protein HHpred hit	pdb code	Probability / %[#]	E-value[%]	P-value[*]	Score[$]	Aligned Cols[α]	Query HMM[β]	Template HMM[γ]	Sequence identity / %	Similarity[κ]
PlyCA, PlyCB; Lysin	4F88	100.0	2.0×10^{-75}	2.0×10^{-80}	609.36	418	4-476	7-463	19	0.215
N-acetylmuramoyl-L-alanine Amidase	5T1Q	99.3	1.3×10^{-14}	1.0×10^{-75}	118.45 6	122	350-477	243-358	17	0.201
CHAP domain protein	2K3A	99.3	8.1×10^{-15}	4.8×10^{-18}	106.56	93	342-476	39-153	20	0.149
staphyloxanthin biosynthesis protein	2LRJ	99.2	3.8×10^{-14}	2.3×10^{-17}	97.77	104	352-475	8-111	23	0.234
peptidoglycan protease, phage K	4CT3	98.9	1.5×10^{-10}	1.5×10^{-15}	104.46	106	350-463	44-154	19	0.143
putative phage lysin, phage phi7917	5D74	98.5	7.6×10^{-9}	7.8×10^{-14}	103.0	96	352-456	46-159	18	0.108
CHAP domain, S. pyogenes MGAS5005	5UDN	98.4	5.6×10^{-8}	5.7×10^{-13}	95.4	113	350-476	20-150	17	0.115
CHAP domain, S. pneumoniae	4CGK	98.0	9.0×10^{-7}	9.2×10^{-12}	87.6	107	352-476	284-390	17	0.046
endopeptidase	4HZ9	97.4	5.6×10^{-5}	5.7×10^{-10}	65.0	78	354-445	17-95	19	0.240
CHAP domain	4Q4G	92.5	0.82	8.4×10^{-6}	49.9	90	351-458	374-472	14	-0.030

[#] - estimated probability of the template to be (at least partly) homologous to the query sequence
[%] - the expected number of false positives per database search with a score at least as good as the score of this sequence match
[*] - equal to the E-value divided by the number of hidden markov models in the searched database
[$] - gives the total score that includes the score from the secondary structure comparison
[α] - the total number of matched columns in the querytemplate alignment
[β] - range of amino acid sequence from the query hidden markov model used for comparison
[γ] - range of amino acid sequence from the template hidden markov model used for comparison
[κ] - secondary structure similarity

Table S4. X-ray structure quality indicators.

X-ray Data Collection	Tail Fiber
Space group	R3
Cell dimensions	
a, *b*, *c* (Å)	100.73, 100.73, 125.46
$\alpha, \beta, \gamma(^\circ)$	90.0, 90.0, 120.0
Wavelength (Å)	0.97857
Resolution (Å)	2.0 (2.0502 - 2.0002)
R_{merge}	0.216
Completness (%)	99.9 (100.0)
Refinement	
Resolution (Å)	2.0
No. reflections	32065 (2146)
R_{work}/R_{free} (%)	16.4/20.2 (21.45/22.89)
No. of atoms (non-hydrogen)	
Protein	4021
Water	436
B-factors	
Protein	27.1
Water	30.5
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.57
Ramachandran outliers (%)	0.0
Rotamer outliers (%)	0.2