

Supplementary Informations for

Near-atomic, non-icosahedrally averaged structure of giant virus *Paramecium bursaria* *chlorella virus 1*

Qianqian Shao, Irina V. Agarkova, Eric A. Noel, David D. Dunigan, Yunshu Liu, Aohan Wang, Mingcheng Guo, Linlin Xie, Xinyue Zhao, Michael G. Rossmann, James L. Van Etten, Thomas Klose, Qianglin Fang

This PDF file includes:

Supplementary Figs. 1 to 9

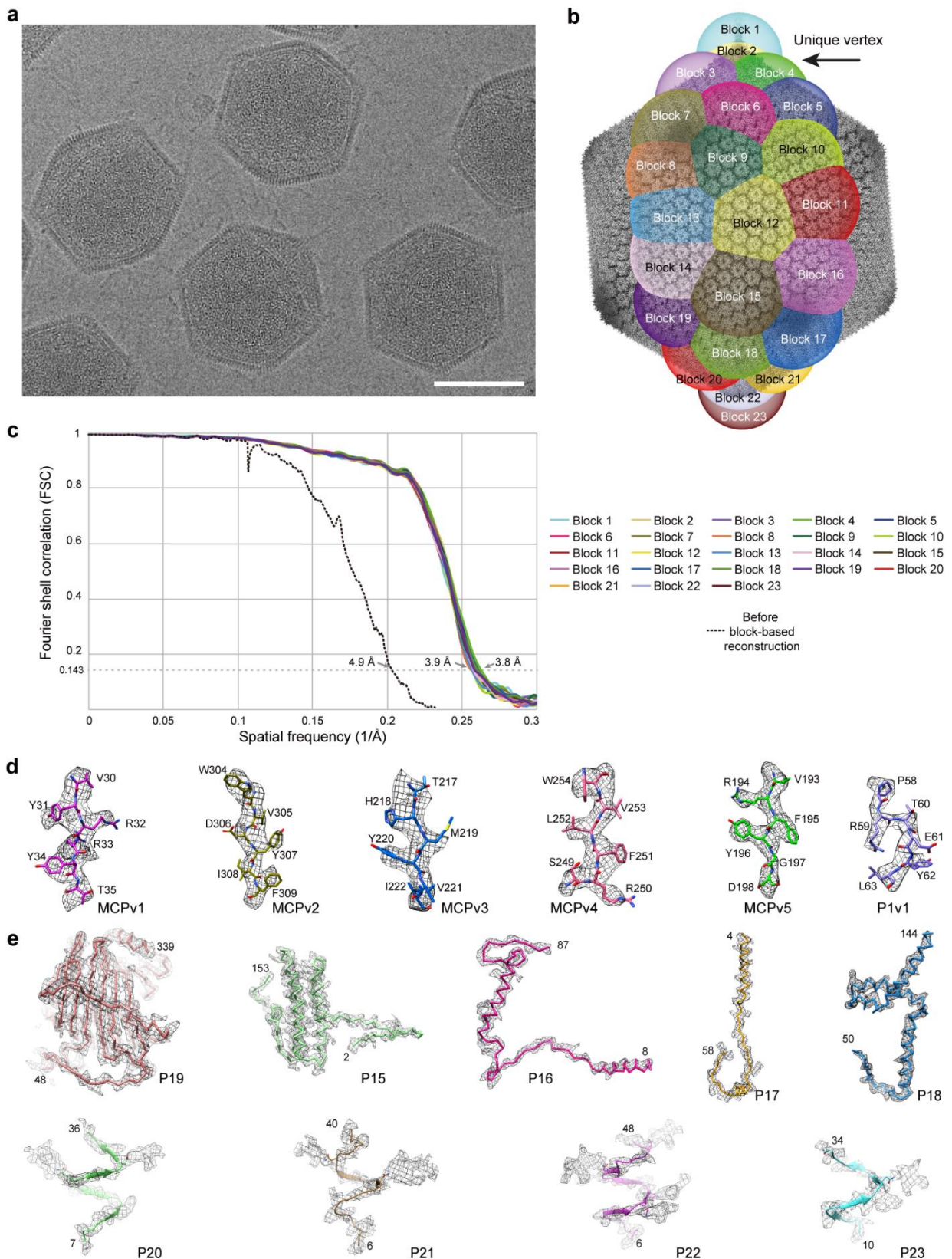
Supplementary Tables 1 to 6

Supplementary References

Other Supplementary Informations for this manuscript include the following:

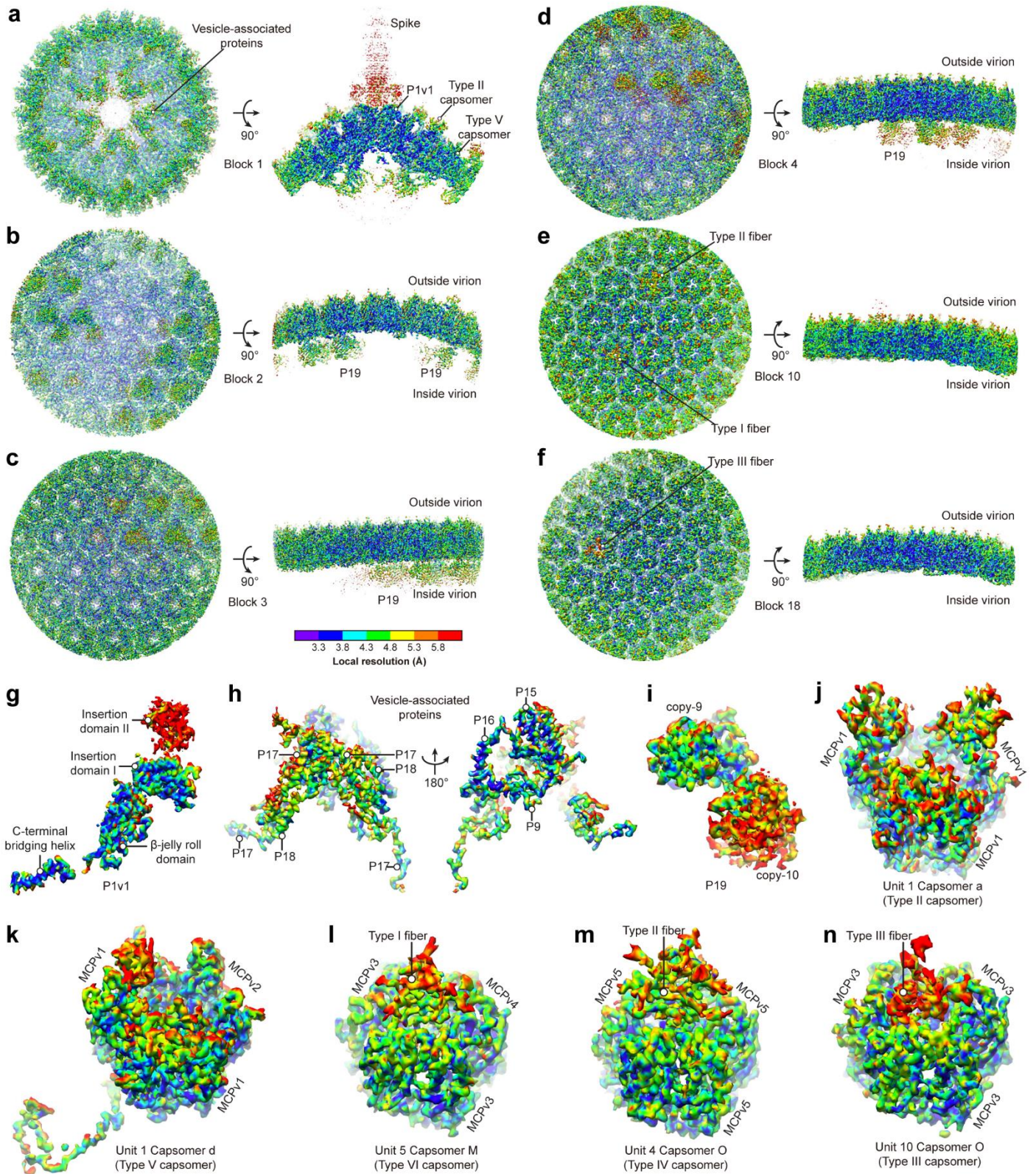
Supplementary Movies 1 to 2

Supplementary Figures



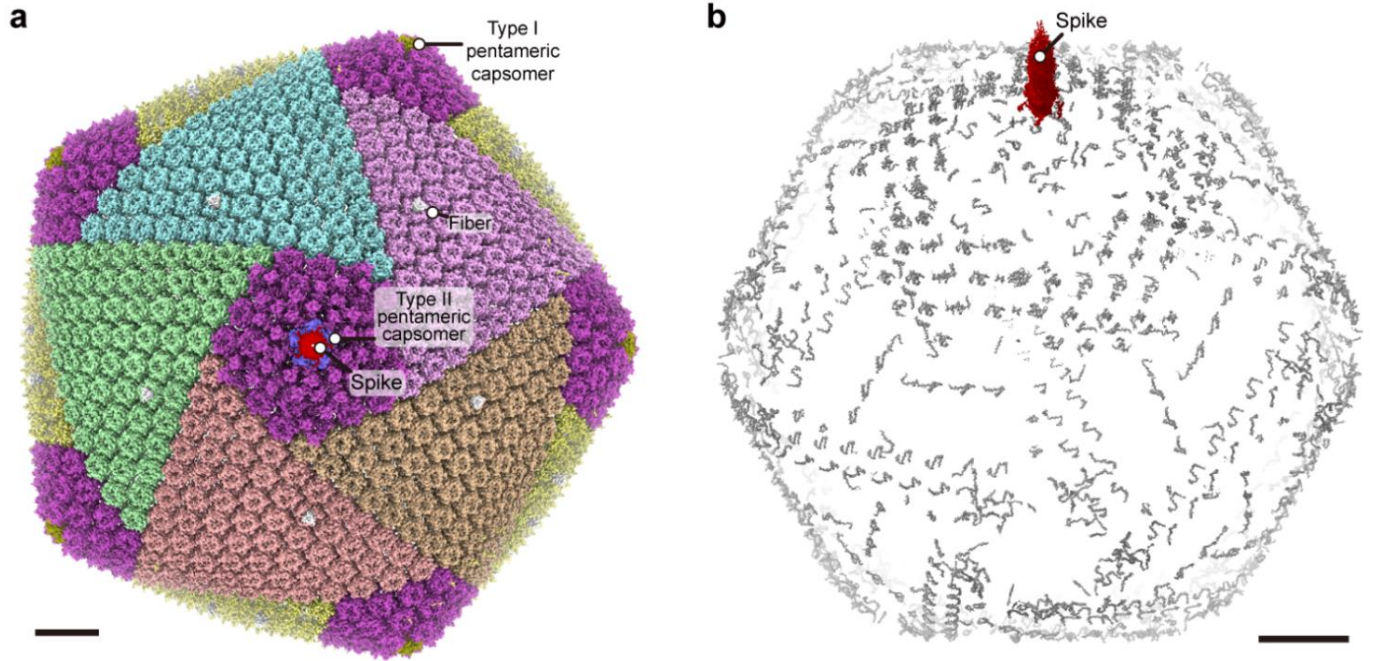
Supplementary Figure 1 | Resolution estimation and cryo-EM densities of the previously unresolved proteins.

a, Representative cryo-EM micrograph of PBCV-1. Scale bar = 100 nm. **b**, Block distribution within one five-fold asymmetric unit of the five-fold averaged, cryo-EM map of PBCV-1. **c**, FSC curves of each block and the cryo-EM map before applying the “block-based” reconstruction method calculated based on the FSC 0.143 cut-off. **d**, Cryo-EM densities of representative regions of the variants of the MCP and the penton protein, superposed with their atomic models. **e**, Cryo-EM densities of the previously unresolved minor capsid proteins, superposed with their atomic models.

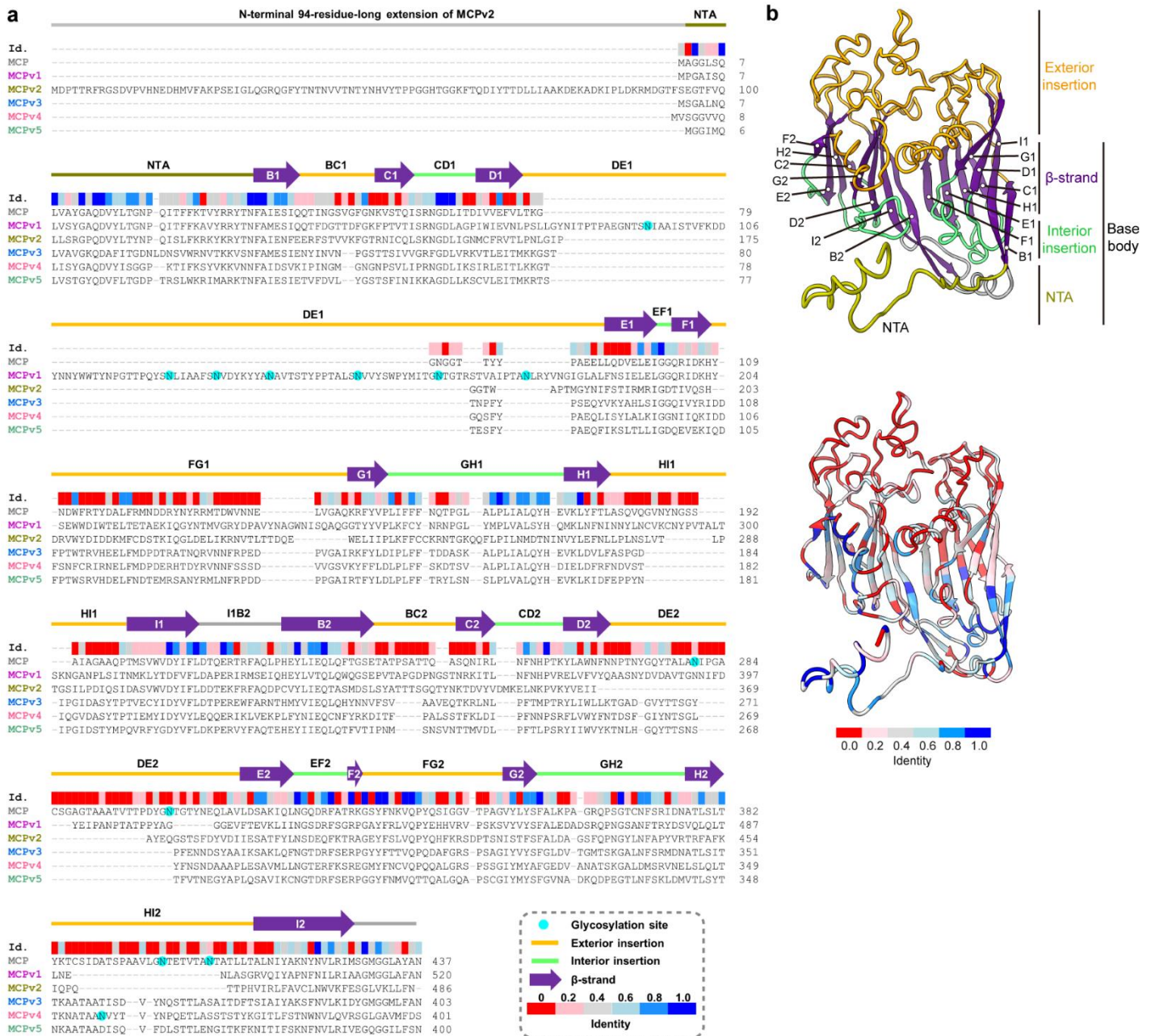


Supplementary Figure 2 | Local resolution estimation of the cryo-EM maps.

Local resolution maps of block 1 (**a**), block 2 (**b**), block 3 (**c**), block 4 (**d**), block 10 (**e**) and block 18 (**f**), P1v1 (**g**), vesicle-associated proteins (**h**), two copies of P19 (**i**), type II pseudo-hexameric capsomer (**j**), type V pseudo-hexameric capsomer (**k**), type I fiber and type VI pseudo-hexameric capsomer (**l**), type II fiber and type IV pseudo-hexameric capsomer (**m**), and type III fiber and type III pseudo-hexameric capsomer (**n**), as estimated using RESMAP¹.

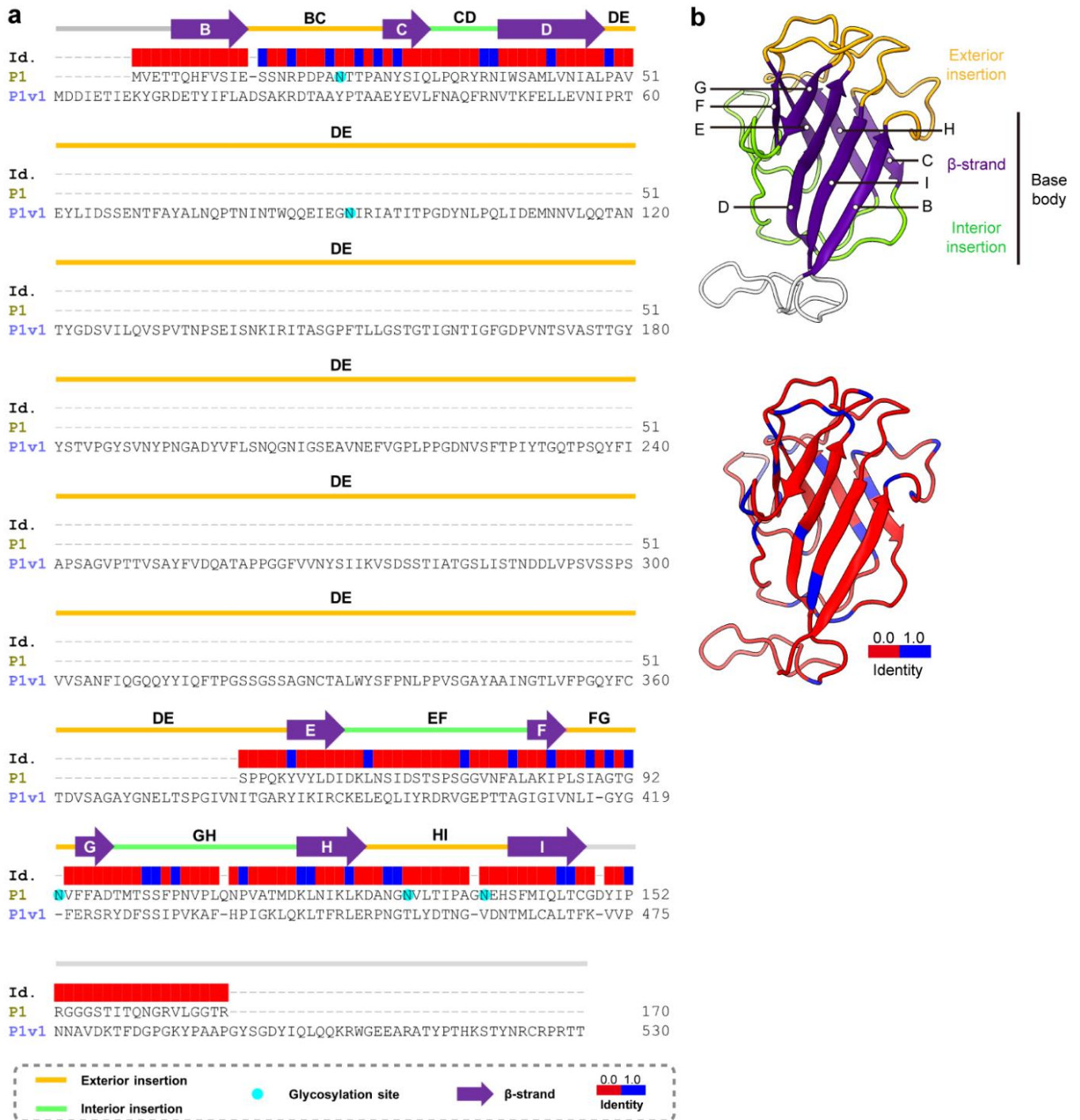


Supplementary Figure 3 | Symmetrons and uninterpreted densities of the five-fold averaged cryo-EM map of PBCV-1. a, Trisymmetrons and pentasymmetrons in the five-fold averaged cryo-EM map of PBCV-1, looking down the spike in the unique vertex. **b,** Uninterpreted densities of the five-fold averaged cryo-EM map of PBCV-1. Scale bar = 20 nm.



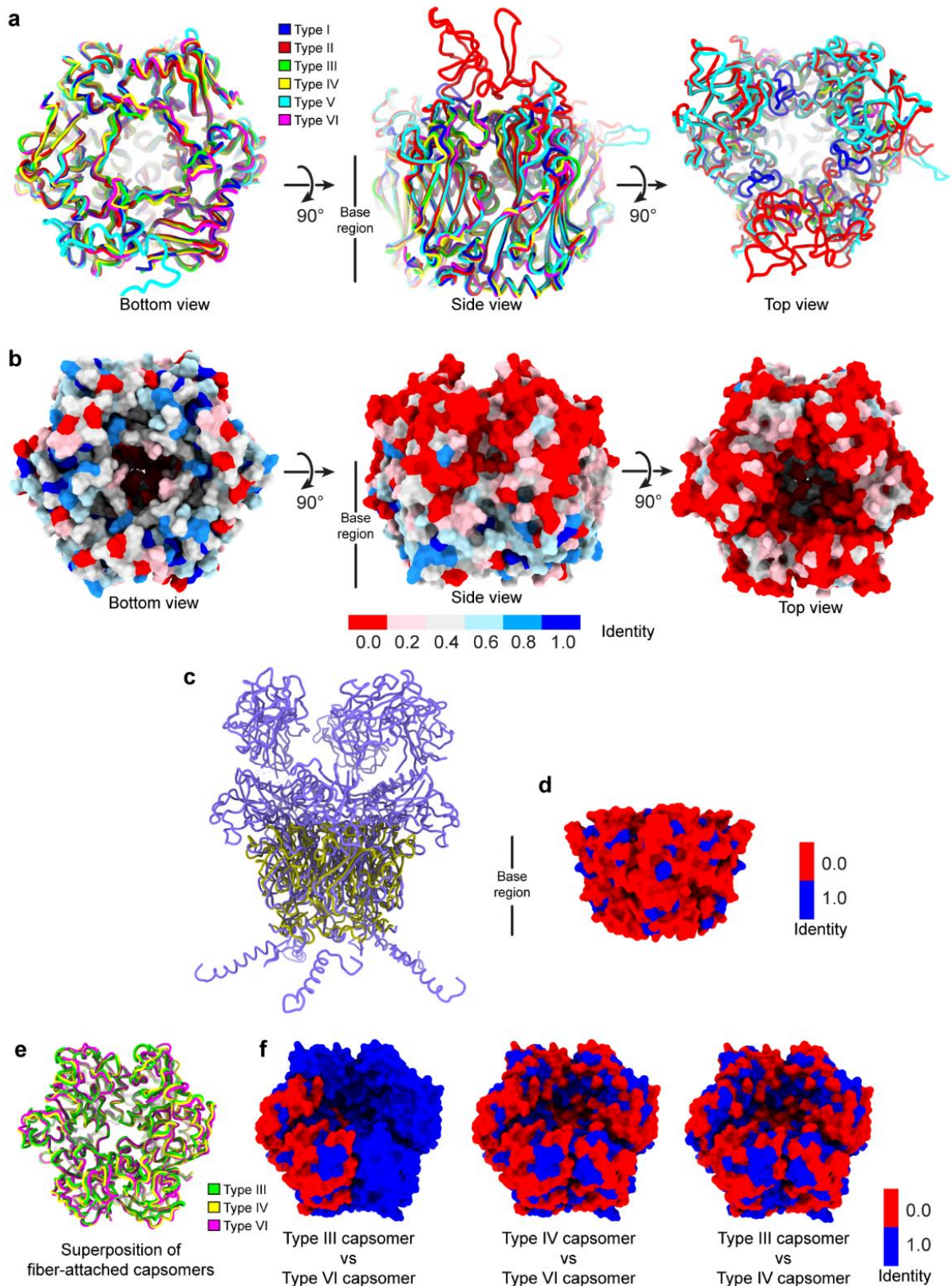
Supplementary Figure 4 | Sequence alignment of MCP and its variants.

a, sequence alignment of MCP and its variants. The identity values for each residue in the MCP sequence were calculated by comparing that residue of the MCP with corresponding residues of the MCP variants (1 being conserved, and 0 being not conserved). **b**, Ribbon diagrams of one MCP molecule showing different structural regions (up) and identity values as indicated in **a** (down).



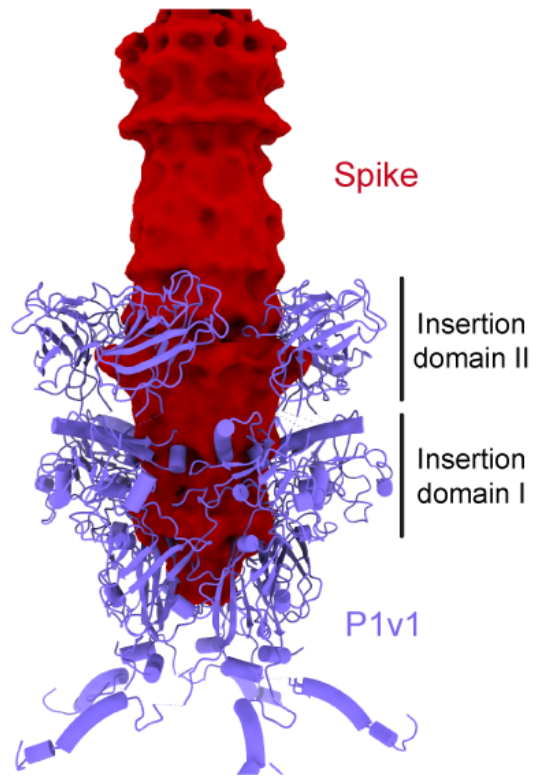
Supplementary Figure 5 | Sequence alignment of P1 and P1v1.

a, Sequence alignment of the penton protein P1 and the penton protein variant P1v1. The identity values were calculated similarly as in Supplementary Fig. 4a. **b**, Ribbon diagrams of one P1 molecule showing the structural regions (up) and identity values as indicated in **a** (down).



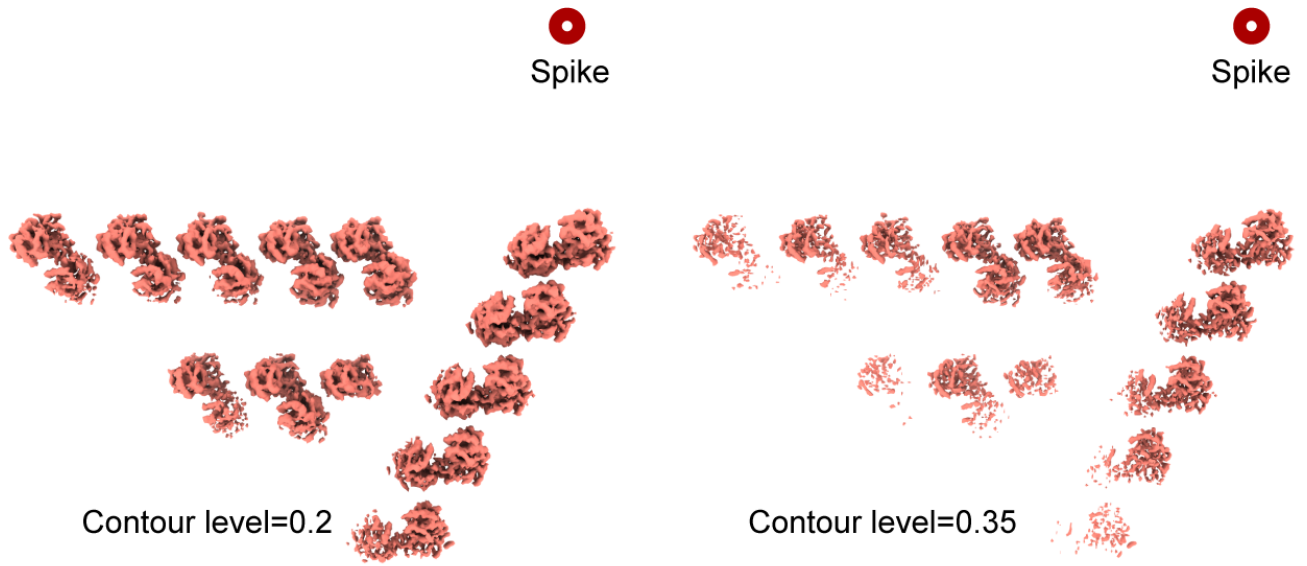
Supplementary Figure 6 | Structural comparisons and sequence identity analyses between pseudo-hexameric and between pentameric capsomers.

a, Backbone superposition of all the six types of pseudo-hexameric capsomers. **b**, Surface identity between the MCP and all the MCP variants. The identity values were calculated in the same way as in Supplementary Fig. 4a. **c**, Backbone superposition of the type I and type II pentameric capsomers. **d**, Surface identity between the type I and type II pentameric capsomers. The identity values were calculated using the same way as in Supplementary Fig. 5a. **e**, Backbone superposition of the three types of fiber-attached capsomers. **f**, Surface identities between any two types of fiber-attached capsomers. The identity values were calculated similarly as in Supplementary Fig. 4a.






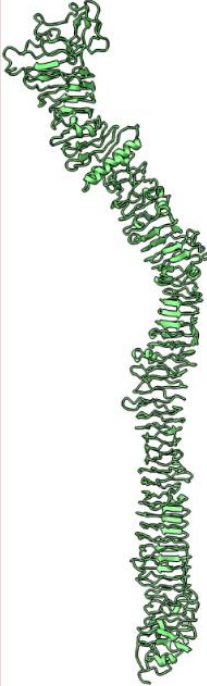
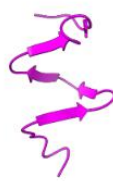
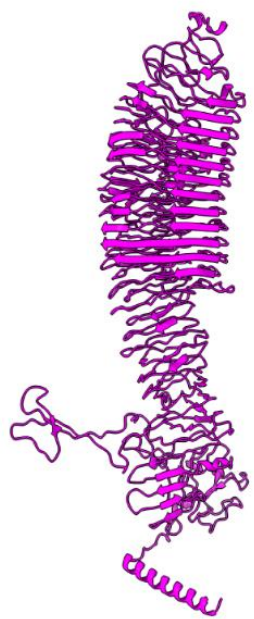
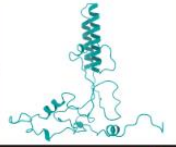
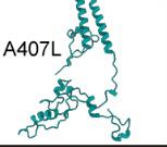


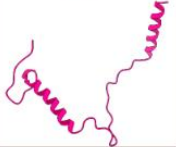
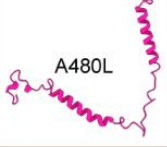


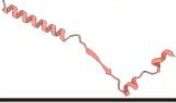

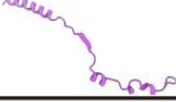

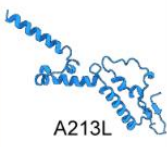

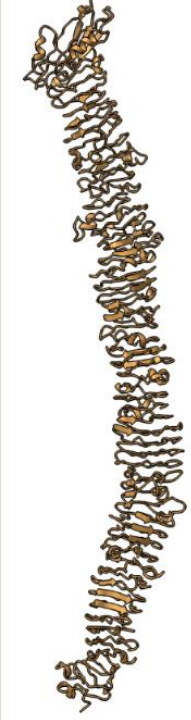

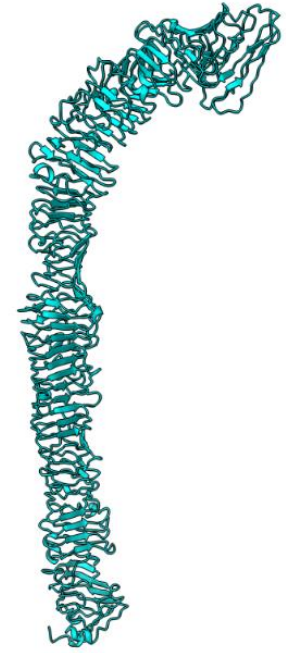
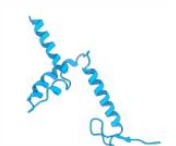


Supplementary Figure 7 | Association between the spike and type II pentameric capsomer.

The structures of the spike (dark red) and the type II pentameric capsomer (cornflower blue) are shown as cryo-EM density map and ribbon models, respectively.



Supplementary Figure 8 | Cryo-EM densities of the 25 copies of the finger proteins displayed with two contour levels.

A dark red circle is used to indicate the position of the viral spike in the unique vertex.

	<i>Ab initio</i> models	<i>In silico</i> models		<i>Ab initio</i> models	<i>In silico</i> models		<i>Ab initio</i> models	<i>In silico</i> models
P1v1		 A533R	P20		 A018L	P22		 A122/123R
P9 Unique vertex		 A407L						
P15		 A484L						
P16		 A480L						
P17 copy 1		 A314R						
P17 copy 2								
P17 copy 3								
P17 copy 4								
P18 copy 1		 A213L	P21		 A014R	P23		 A025/027/029L
P18 copy 2								
P19		 A286R						

Supplementary Figure 9 | Structure comparisons between the *ab initio* models and their *in silico* model candidates.

The *ab initio* models and the *in silico* models were manually built using the cryo-EM density maps and computed using AlphaFold² and/or RoseTTAFold³, respectively.

Supplementary Tables

Supplementary Table 1 | Information on identified proteins.

Protein name	Gene name	Icosahedral asymmetric units where this protein is present	Copy number within one virion
MCP	<i>a430l</i>	All	4,905
MCPv1 ^a	<i>a622l</i>	1	40
MCPv2 ^a	<i>a383r</i>	1	5
MCPv3 ^a	<i>a011l</i>	1, 5, 7, 10, 11	55
MCPv4 ^a	<i>a010r</i>	1, 5, 7, 11	20
MCPv5 ^a	<i>a558l</i>	4	15
P1	<i>a310l</i>	2–12	55
P1v1 ^a	<i>a533r</i>	1	5
P2	<i>a342l</i>	All	60
P3	<i>a523r</i>	All	180
P4	<i>a572r</i>	All	240
P5	<i>a526r</i>	All	60
P6	<i>a203r</i>	All	60
P7	<i>a262/263l</i>	2–12	55
P8	<i>a644r</i>	All	60
P9	<i>a407l</i>	All	60
P10	<i>a454l</i>	2–12	110
P11	<i>a352l</i>	All	720
P12	<i>a139l</i>	All	60
P13	<i>a421r</i>	All	60
P14	<i>a500l</i>	2–12	55
P15 ^a	<i>a484l</i>	1	5
P16 ^a	<i>a480l</i>	1	5
P17 ^a	<i>a314r</i>	1	20
P18 ^a	<i>a213l</i>	1	10
P19 ^a	<i>a286r</i>	1–3	125
P20 ^a	<i>a018l</i>	1, 5, 7, 11	40
P21 ^a	<i>a014r</i>	1, 5, 7, 11	20
P22 ^a	<i>a122/123r</i>	4	15
P23 ^a	<i>a025/027/029l</i>	10	15

^a Previously unresolved proteins.

Supplementary Table 2 | Model validation statistics on previously unresolved proteins.

Protein name	Protein sequence length	Chain ID^a	Ordered part	CC^b	Resolution (FSC=0.5)^c	Resolution (FSC=0.25)^c
MCPv1	520	bu	2–518	0.83	4.1	3.8
		bv	2–518	0.80	4.1	3.8
		bw	2–518	0.81	4.1	3.7
		bx	5–518	0.84	4.1	3.7
		by	2–518	0.83	4.1	3.7
		bz	11–518	0.80	4.2	3.8
		bA	2–517	0.82	4.2	3.7
		bB	2–518	0.81	4.2	3.9
MCPv2	486	bC	2–486	0.73	4.2	3.8
MCPv3	403	dd	3–403	0.82	4.1	3.8
		dp	3–403	0.82	4.1	3.8
		dr	3–403	0.81	4.1	3.8
		df	3–403	0.82	4.2	3.8
		dj	3–403	0.80	4.2	3.8
		bD	3–403	0.82	4.2	3.8
		dl	3–403	0.77	4.2	3.9
		cZ	3–403	0.80	4.2	3.8
		dk	3–403	0.80	4.2	3.9
		bF	3–403	0.81	4.3	3.9
		cX	3–403	0.79	4.3	3.9
MCPv4	401	dq	3–399	0.83	4.2	3.8
		de	3–399	0.83	4.2	3.9
		bE	3–399	0.82	4.2	3.9
		cY	3–399	0.79	4.3	3.9
MCPv5	400	cR	2–400	0.81	4.2	3.8
		cT	2–400	0.82	4.2	3.9
		cS	2–400	0.81	4.3	3.9
P1v1	530	bG	2–203, 210–367, 370–479, 491–524	0.67	4.3	3.8
P15	155	ch	2–153	0.74	4.1	3.8
P16	93	ci	8–87	0.72	4.5	3.5
P17	80	ck	2–58	0.75	4.3	3.6
		cm	5–67	0.74	4.4	3.7
		cj	5–63	0.74	4.4	3.6
		cl	4–58	0.74	4.5	3.8
P18	148	cn	49–146	0.76	4.1	3.6
		co	50–144	0.75	4.2	3.6
P19	378	cv	48–339	0.74	4.4	4.0
		cx	48–339	0.74	4.4	4.1
		cL	48–339	0.74	4.5	4.1
		ct	48–339	0.76	4.6	4.2
		cJ	48–339	0.76	4.6	4.2
		cw	60–339	0.73	4.6	4.2
		cH	48–339	0.75	4.6	4.2
		cF	48–339	0.75	4.7	4.2

		cr	48–339	0.75	4.7	4.3
		cp	48–339	0.73	4.7	4.2
		cB	48–339	0.74	4.7	4.2
		cy	60–339	0.76	6.8	4.3
		cD	48–339	0.73	7.0	4.3
		cE	60–339	0.75	7.1	4.4
		cN	48–339	0.7	7.5	4.3
		cI	61–339	0.75	7.7	4.4
		cu	60–339	0.73	7.9	4.5
		cC	61–339	0.72	7.9	4.3
		cG	58–339	0.74	7.9	4.5
		cK	61–339	0.72	7.9	4.6
		cq	58–339	0.69	8.1	4.6
		cz	48–339	0.69	8.1	4.4
		cs	65–339	0.75	8.4	4.5
		cA	61–339	0.62	8.6	6.9
		cM	63–339	0.66	8.6	5.7
P20	1335	da	6–37	0.81	4.1	3.4
		dc	7–34	0.83	4.2	3.4
		cO	7–36	0.81	4.3	3.5
		ds	6–37	0.80	4.3	3.8
		du	7–36	0.75	4.3	3.5
		cQ	7–37	0.81	4.3	3.4
		dg	6–37	0.79	4.4	3.6
		di	7–35	0.76	4.4	3.6
P21	1369	cP	6–40	0.80	4.2	3.5
		dh	6–40	0.79	4.3	3.7
		dt	6–40	0.79	4.4	3.6
		db	7–36	0.79	4.5	3.6
P22	1343	cU	6–48	0.79	4.3	3.9
		cV	6–48	0.78	4.3	3.9
		cW	6–48	0.78	4.4	3.6
P23	1359	dm	10–34	0.65	4.5	3.7
		dn	10–34	0.68	4.5	3.3
		do	10–34	0.59	4.9	3.6

^a Chain ID in the coordinate file (PDB ID: 8H2I).

^b Real-space correlation coefficient between the atomic models and their density maps.

^c Model-to-map FSCs between the atomic models and their density maps.

Supplementary Table 3 | Protein sequence identities between MCP, P1 and their variants.

Variants	Protein sequence identity between this variant with MCP
MCPv1	42.8%
MCPv2	28.1%
MCPv3	38.0%
MCPv4	35.5%
MCPv5	36.8%

Protein sequence identity between this variant with P1	
P1v1	10.6%

The protein sequence identity is calculated according to the ratio of the number of matched residues to the sequence length of MCP or P1.

Supplementary Table 4 | Information on identified glycosylation sites.

Protein	Glycosylation sites
MCP	AN ₂₈₀ IP GN ₃₀₂ TG GN ₃₉₉ TE AN ₄₀₆ TA
MCPv1	SN ₉₅ IA SN ₁₂₄ LI SN ₁₃₁ VD AN ₁₃₉ AV SN ₁₅₂ VV GN ₁₆₄ TG AN ₁₇₇ LR
MCPv4	AN ₃₅₇ VY
P1	AN ₂₁ TT GN ₉₃ VF GN ₁₂₉ VL GN ₁₃₇ EH
P1v1	GN ₉₁ IR
P20	GN ₂₂ LS GN ₃₁ GA
P21	GN ₂₁ LV GN ₃₂ GG
P22	GN ₂₃ IA AN ₂₆ VI DN ₃₁ GN GN ₃₃ VI GN ₄₆ GA
P23	GN ₂₄ VF GN ₃₃ AS

Supplementary Table 5 | Primers used in this work.

PBCV-1 Gene Primer	Sequence
PBCV-1 <i>a533r</i> F Primer	5'-GCGCAAAC TTTATTCAAGGAC-3'
PBCV-1 <i>a533r</i> R Primer	5'-TAATCCCCAGAGTATCCAGGG-3'
PBCV-1 <i>a383r</i> F Primer	5'-CTCGGAGAAATCTTCTCGC-3'
PBCV-1 <i>a383r</i> F Primer (2)	5'-CGGATGTTTACGTTGACATG-3'
PBCV-1 <i>a383r</i> R Primer	5'-CTAAACCCGCACCTAAACC-3'

Supplementary Table 6 | Cryo-EM data collection, refinement and validation statistics.

Map	PBCV-1 C5 (EMD-34438, PDB: 8H2I)
<hr/>	
Data collection and processing	
Microscope	FEI Titan Krios
Magnification	18,000
Voltage (kV)	300
Detector	Gatan K2 Summit detector
Recording mode	Super-resolution or counting mode
Dose rate (e ⁻ /(pixel·s))	~8
Frame exposure time (ms)	200
Movie micrograph exposure time (s)	~8
Total dose (e ⁻ /Å ²)	~24
Defocus range (μm)	1.0–4.0
Pixel size (Å)	1.62 (physical pixel size)
Symmetry imposed	C5
Map resolution (Å)	3.8–3.9
FSC threshold	0.143
Map sharpening <i>B</i> -factor (Å ²)	-120
Refinement	
Initial model used (PDB code)	6NCL
R.m.s. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.43
Validation	
MolProbity score	1.37
Clashscore	2.49
Poor rotamers (%)	0.14
Ramachandran plot	
Favored (%)	95.22
Allowed (%)	4.40
Disallowed (%)	0.38

Supplementary References

- 1 Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
- 2 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- 3 Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871-876 (2021).