Supporting Information

Coulibaly et al. 10.1073/pnas.0910686106

SI Methods

Production of Polyhedra. We produced AcMNPV polyhedra in Sf21 cells by infection with a baculovirus encoding the G25D mutated polyhedrin protein. This mutation resulted in the formation of fewer and larger crystals in cells, as described previously (1). Polyhedra were produced and purified according to established protocols. G25D AcMNPV polyhedra incorporating selenomethionine were produced for phase determination (2). These crystals diffracted more consistently and to a higher resolution than the corresponding native polyhedra.

Polyhedra were also purified from infected larvae of porina moths (*Wiseana* spp.) collected from pastures in the Otago region (New Zealand). Infected larvae, as determined by light microscope observation of larval fat body, were homogenized in 20 mM phosphate buffer, pH 6.2, using a Polytron MR-PT 1600E homogenizer (Kinetmatica AG). The homogenate was filtered through three layers of muslin, larger debris was removed by centrifugation at 1,000 × *g* for 5 min, and polyhedra were concentrated by centrifugation at 12,000 × *g* for 10 min. Polyhedra were purified on a five-step 45–65% (wt/wt) sucrose gradient at 28,000 × *g* for 1 h, and the polyhedral band was collected. The polyhedra were processed by two successive cycles of wash, followed by centrifugation at 12,000 × *g* for 15 min. The concentration and quality of polyhedra were estimated by optical microscopy on a hemocytometer (3).

The density of G25D AcMNV polyhedra was experimentally determined, and polyhedra were analyzed by scanning electron microscopy, as previously described (4).

Scanning Electron Microscopy. Scanning electron microscopy was performed on Pt-sputter-coated samples (Polaron SC 7640 Sputter Coater, 5 mA, 1.1 kV for 5 min) and examined using an XL30S Field Emission Gun SEM unit (Philips). Electrospray ionization mass spectrometry was carried out using a QStar XL Quadrupole TOF mass spectrometer (Applied Biosystems) with acetic acid/formic acid-dissolved polyhedra. Mass spectrometry analyses identified removal of the initiating methionine in the nucleopolyhedrovirus polyhedrin proteins analyzed but no post-translational modification.

Density Measurement. The density of recombinant AcMNPV G25D polyhedra was investigated following established protocols (4). Briefly, Nycodenz solutions at 50–80% (wt/vol) were prepared in H_2O. The density of each solution was measured in triplicate at 25 °C using a DMA 5000 density meter (Anton Paar GmbH) with an estimated accuracy to within 5×10^{-6} g·cm $^{-3}$. The relative density of Nycodenz at 25 °C against its concentration was used to generate a standard curve. The density of AcMNPV G25D polyhedra was measured using sedimentation centrifugation. Glass Pasteur pipettes were siliconized in 2% (vol/vol) siliconizing solution. The narrow tips (1-mm diameter) were cut to 3 cm in length using a diamond-point pencil, and one end was sealed by gentle heating. Nine microliters of Nycodenz (Axis-Shield PoC AS) solution was pipetted into the sealed Pasteur pipette, which was then placed in a 1.5-mL Eppendorf tube for centrifugation (16,000 × *g*, 30 s at 25 °C) to remove air bubbles. One microliter of polyhedra was carefully laid over the Nycodenz solution using a 0.18-mm diameter extra-long gel-loading tip (CLP Molecular Biology). The filled pipettes were placed in a 1.5-mL Eppendorf tube and subjected to further sedimentation centrifugation (15,000 × *g*, 3 h at 25 °C). The position of centrifuged polyhedra in each Nycodenz solution was determined using a microscope, and the relative density of polyhedra was assessed using the standard curve.

Heavy-Atom Soaks. Heavy-atom solutions were prepared as saturated stock solutions in 50 mM Hepes, pH 7.0, from Hampton Research kits, and a series of systematic dilutions was set up as soaking solutions for each heavy-atom compound. The only exception were soaks with triiodide ions, in which the stock solution was prepared by dissolving 1 g of KI in 4 mL of water and then adding 0.54 g of I_2 (5). For all soaks, 1 μL of diluted AcMNPV-G25D polyhedra (≈100 polyhedra/μL) was transferred to a drop of heavy-atom soaking solution on a single cavity (concave) microscope slide. The cavity was sealed with vacuum grease and a plastic cover slide to prevent evaporation. Heavy-atom soaks were performed at various heavy-atom concentrations for different soaking periods (1 h to overnight) and were kept in the dark before data collection. Soaks used for structure determination were all carried out on selenomethionine crystals because they diffracted more reproducibly than the native ones. The following compounds were used $AgNO_3$ (10 mM, 4 h), K_2PtCl_4 (1/100 of saturation, 8 h), KI/I_2 (1/20 of stock solution, 1 h), and methylmercury phosphate (1/100, 16 h and 1/20 of saturation, 4 h).

Data Collection and Processing. Crystal samples for data collection were prepared as described previously (4). Briefly, samples were pipetted onto a MicroMesh Mount (MiTeGen) (6) with a 400-μm diameter mesh and a 25-μm grid size and were allowed to sediment onto the mount surface by gravitational force. Cryoprotectant containing a final concentration of 50% (vol/vol) ethylene glycol was added onto the mount, and excess residual liquid was removed with a paper liquid wick without disturbing the settled crystals. In the case of heavy-atom-soaked polyhedra, back-soaking was carried out by washing the settled polyhedra with cryoprotectant five times to remove any nonspecifically bound heavy atoms to give less background and reduce radiation damage. Mounted polyhedra were flash-frozen in liquid nitrogen before data collection at 100 K. Diffraction experiments were carried out at the Swiss Light Source X06SA microbeamline using the MD2 microfocus diffractometer with a beam of 15 × 5 μm focused on the detector. A typical crystal could be exposed for about 10 s (i.e., 10 images of 1 s with an oscillation of 1°) with unattenuated beam before a drastic decay in the diffracting power. Thus, datasets were assembled from data collected on multiple crystals. In the case of WNPV polyhedra, 19 crystals were included, which resulted in a high redundancy of 15.0. To ensure that radiation damage did not deteriorate the quality of the dataset, criteria for inclusion of more images were that they had scale factor parameters to the reference image (initial image of the best crystal: K = 1 and B = 0) of K >0.5 and B <12 and a R_{sym} <50% as listed by Scalepack (7). The resolution cutoff for the whole dataset was determined as the resolution at which the mean (I/σI) ≈2 and R_{pim} ≈40% (Tables S2 and S3), where R_{pim} is defined as:

$$R_{p.i.m.} = \frac{\sum_{hkl} \sqrt{\frac{1}{n-1}} \sum_{j=1}^{n} |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_{j} I_{hkl,j}}.$$

Structure Determination. Phases were obtained from selenomethionine-substituted crystals and heavy-atom derivatives

(Table S2) at a resolution of 2.8 Å. Sites were identified using SHELXc/d (8) and refined with SHARP (9). The selenomethionine derivatives had the strongest phasing power. Additional sites observed in the KI/I_2 and methyl mercury chloride soaks were treated as selenium sites in SHARP. Despite variable diffracting powers and the absence of heavy-atom sites beyond the four selenomethionines for $SeMet_{Ag}$ and $SeMet_{Pt}$, all datasets mentioned in Table S2 were chosen among data collected from over 200 crystals because they contributed significantly to phasing. This was assessed by the figures of merit, phasing power, and improvement of the quality of the final map when data from these crystals were included. The model was built with the software Coot (10) in the solvent-flattened maps, starting from a scaffold of 159 residues over 245 built by RESOLVE as implemented in PHENIX (11). We initially refined the structure in Refmac5 (12) using data up to a resolution of 2.3 Å, but overfitting was evident from a rapid divergence between R and R_{free}. This may be attributable to the low data-to-parameter ratio resulting from the high protein density of the crystal and a medium resolution. The structure of AcMNPV polyhedra was then refined in PHENIX with one temperature factor per residue and a global group describing the translation, libration, and screw-rotation displacements (TLS) group for the whole protein chain. Finally, we used BUSTER 2.8.0 (13) for the last rounds of refinement. The structure could then be refined with individual B factors for each atom in the structure without signs of overfitting (target of 0.008 Å of rmsd from ideal values). Refinement statistics are summarized in Table S3. Restraints on the disulfide bonds could not be implemented, because the disulfide bonds are formed over a symmetry axis among alternate conformations. However, the geometry of the disulfide bond was close to standard values (distance between the sulfur atoms of 2.15 Å and a torsion angle of 95°).

The structures of WNPV polyhedra were determined by isomorphous molecular replacement and refined with PHENIX at a resolution of 2.18 Å. The same set of "free" reflections was kept throughout refinements of both AcMNPV and WNPV polyhedrins because of the strong isomorphism between the crystals. In the first cycle of refinement, rigid body refinement, simulated annealing (starting temperature of 10,000 K), coordinate refinement, and refinement of one temperature factor per residue were performed in PHENIX. The model was subsequently improved by several rounds of rebuilding and refinement (coordinate refinement and temperature factor refinement as two groups per residue for the main chain and side chain atoms). We used BUSTER 2.8.0 for the three final rounds of refinement, with a target of 0.008 Å of rmsd and individual B factors.

The final model of the WNPV polyhedrin includes residues 10 to 243 with three disordered regions. The first region is a long helical linker (residues 28–47) for which only weak density was visible. The main chain could be traced to the exception of a short break consisting of residues 37–39. The second region is a short loop comprising residues 142–145 for which no electron density is visible. The third region comprises residues 171–202. A short helix (residues 186–194) could be built in clear electron density and resulted in a 1.1% drop in the R_{free}, but connections to the rest of the molecule are absent.

Cysteine 131 (equivalent to C132 in AcMNPV) forms a disulfide bond across a two-fold axis. The side chain of the cysteine is in two alternate conformations, and conformation A in one molecule makes a disulfide bond with the symmetry-related cysteine in conformation B. Occupancies of 0.5 yielded difference maps devoid of large peaks in this region and acceptable side chain temperature factors for Cys-131.

The model for G25D AcMNPV polyhedrin is very similar [rmsd of 0.441 Å over 167 residues (14)] and includes residues 10–245. The only notable difference with WNPV polyhedrin is the absence of well-defined electron density for residues 29–48.

Structure Analysis. Contact areas and numbers of hydrogen bonds were determined using the Protein Interfaces, Surfaces, and Assemblies Service (PISA) at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) for most interfaces (15). The intertrimer interface was determined with the CPP4 program Areaimol (14). The surface complementarity among relevant interfaces was determined by the program Sc with default parameters (16).

We have constructed an alignment of 62 full-length nonredundant sequences of the nucleopolyhedrosis viruses (cutoff of 98% sequence identity for nonredundancy) using ClustalW (17). These proteins share sequence identities with AcMNPV polyhedron, ranging from 34 to 57% for granuloviruses and higher than 75% for α-baculoviruses. Analysis of this multiple-sequence alignment by the Consurf server (18) was mapped onto the structure of the WNPV polyhedrin.

Illustrations were prepared with PyMOL 1.0 (http://www.pymol.org).

1. Lin GY, Zhong J, Wang XZ (2000) Abnormal formation of polyhedra resulting from a single mutation in the polyhedrin gene of Autographa californica multicapsid nucleopolyhedrovirus. *J Invertebr Pathol* 76(1):13–19.
2. Bellizzi JJ, Widom J, Kemp CW, Clardy J (1999) Producing selenomethionine-labeled proteins with a baculovirus expression vector system. *Structure* 7(11):R263–R267.
3. Sadler T, Ward V, Glare T, Kalmakoff J (1998) Examination of New Zealand's endemic Wiseana nucleopolyhedrovirus by analysis of the viral polyhedrin gene. *Arch Virol* 143(12):2273–2288.
4. Chiu E (2008) Biophysical and structural studies of insect virus polyhedra. PhD thesis (University of Auckland, Auckland).
5. Evans G, Bricogne G (2002) Triiodide derivatization and combinatorial counter-ion replacement: Two methods for enhancing phasing signal using laboratory Cu Kalpha X-ray equipment. *Acta Crystallogr D* 58(Pt 6 Pt 2):976–991.
6. Thorne RE, Stum Z, Kmetko J, O'Neill K, Gillilan R (2003) Microfabricated mounts for high-throughput macromolecular cryocrystallography. *J Appl Crystallogr* 36(6):1455–1460.
7. Otwinowski Z, Minor W (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, ed Elsevier (Academic, San Diego), Vol 276, pp 307–325.
8. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64(Pt 1):112–122.
9. Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* 364(2):215–230.
10. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60(Pt 12 Pt 1):2126–2132.
11. Zwart PH, et al. (2008) Automated structure solution with the PHENIX suite. *Methods Mol Biol* 426:419–435.
12. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* 53(Pt 3):240–255.
13. Blanc E, et al. (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr D* 60(Pt 12 Pt 1):2210–2221.
14. Collaborative Computational Project N (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D* 50(Pt 5):760–763.
15. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372(3):774–797.
16. Lawrence MC, Colman PM (1993) Shape complementarity at protein/protein interfaces. *J Mol Biol* 234(4):946–950.
17. Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics* Chapter 2:Unit 2 3.
18. Landau M, et al. (2005) ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* 33(Web Server issue):W299–W302.

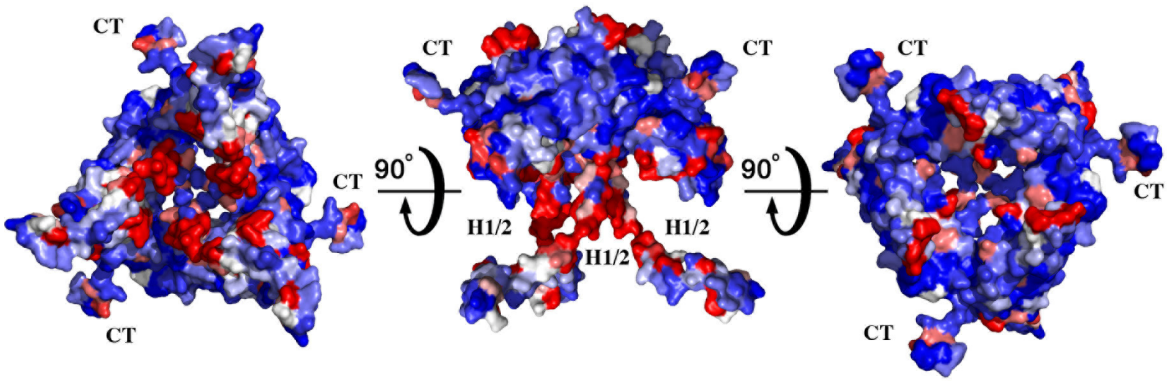

Fig. S1. Conserved surfaces of the baculovirus polyhedrin. Surface representation of the polyhedrin trimer with residues colored according to their conservation in a blue-white-red gradient from the most conserved to the most variable as determined by Consurf from a nonredundant alignment of 62 sequences of polyhedrin and granulin proteins.

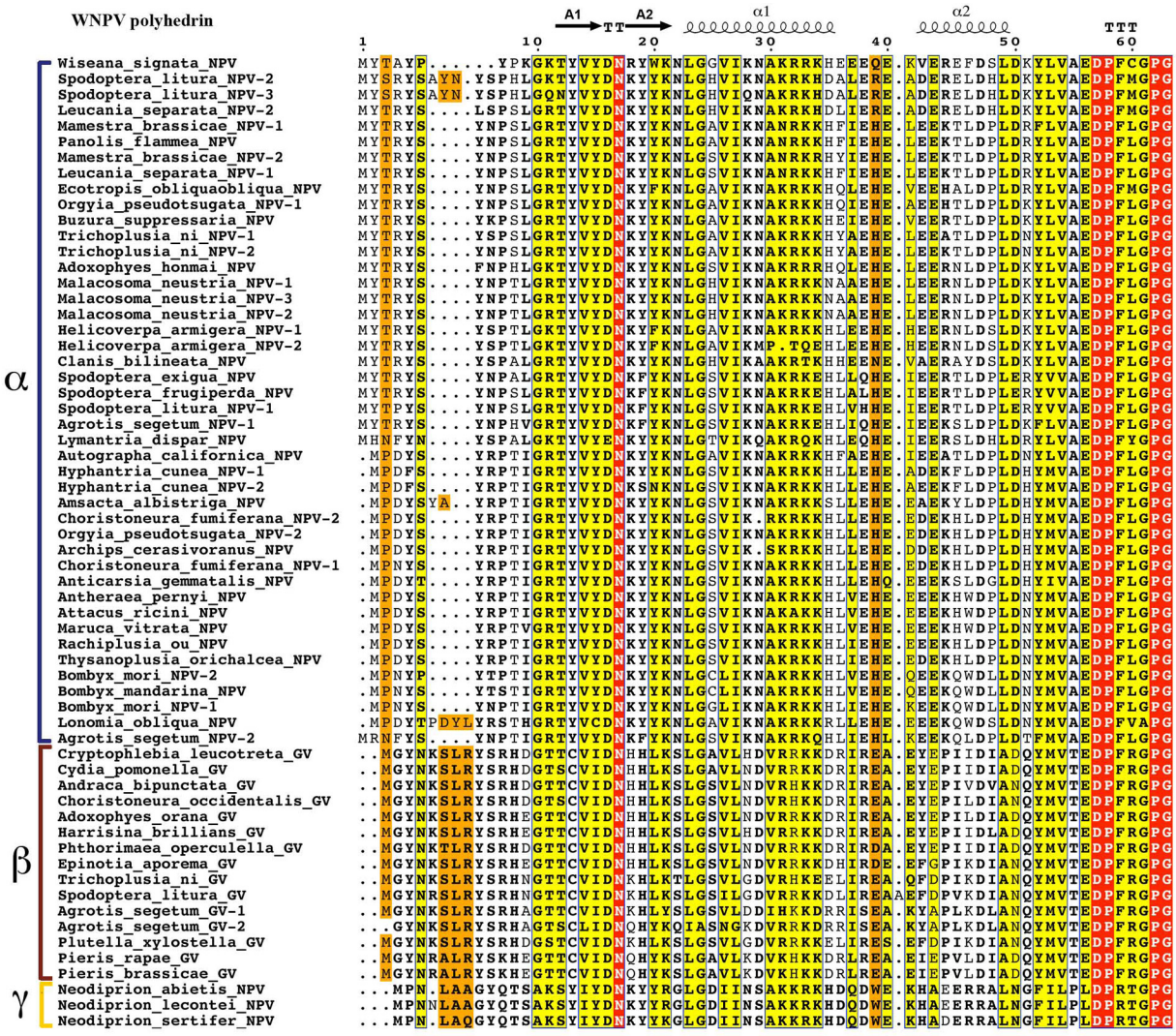

Fig. S2. Sequence alignment of 62 nonredundant polyhedrin and granulin sequences. Sequences were aligned with ClustalW and processed with ESPript [Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 31(13):3320–3323] as three groups representing α-, β-, and γ-baculoviruses. Numbering and secondary structures correspond to the WNPV polyhedrin. Residues in white characters with red box are strictly conserved, residues in black bold characters are similar within a group, residues in blue frames filled in with yellow are similar across groups, and residues in orange box are dissimilar among conserved groups. Secondary structure elements calculated by the DSSP program are represented above the alignments.

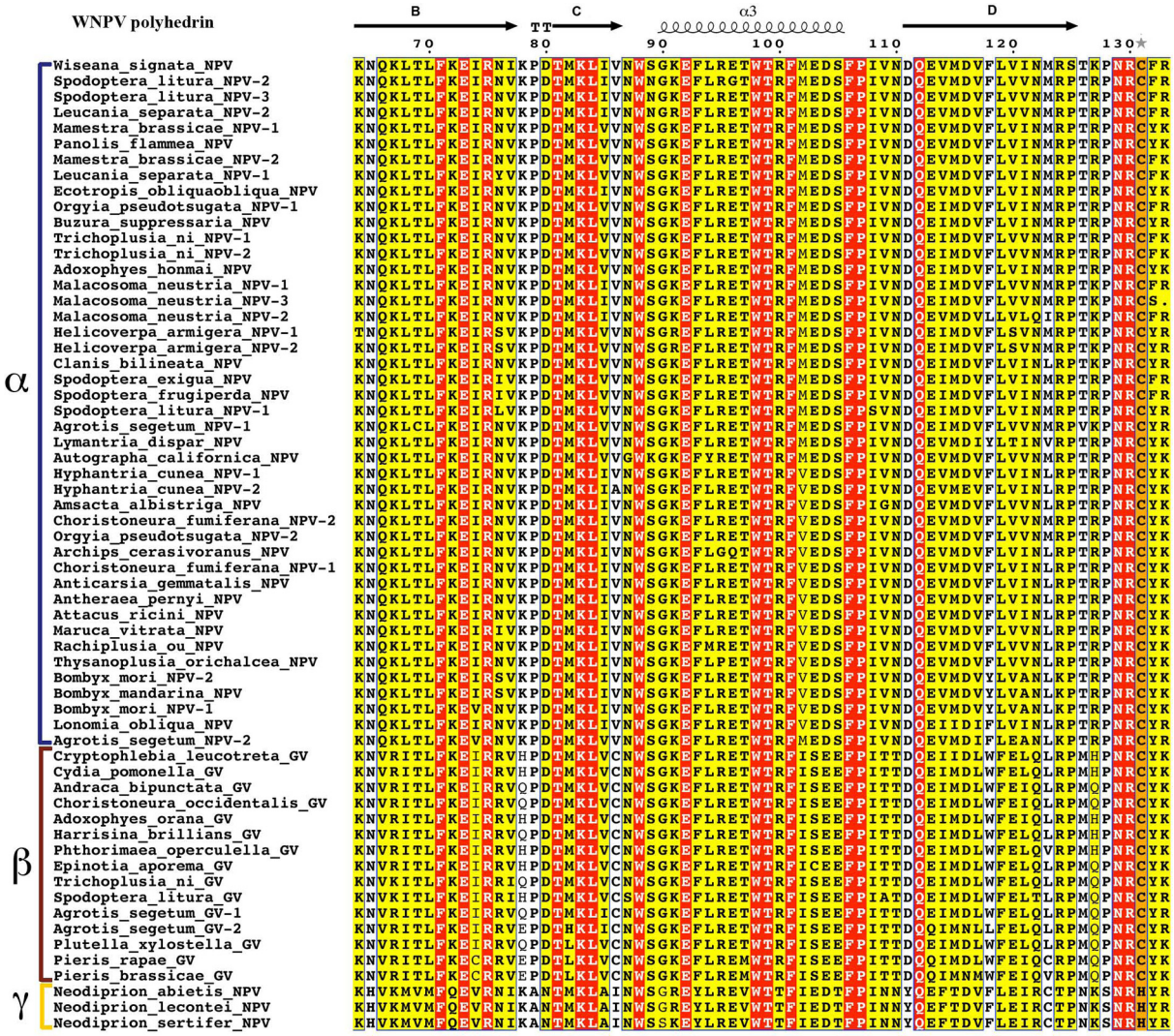

Fig. S2. Continued.

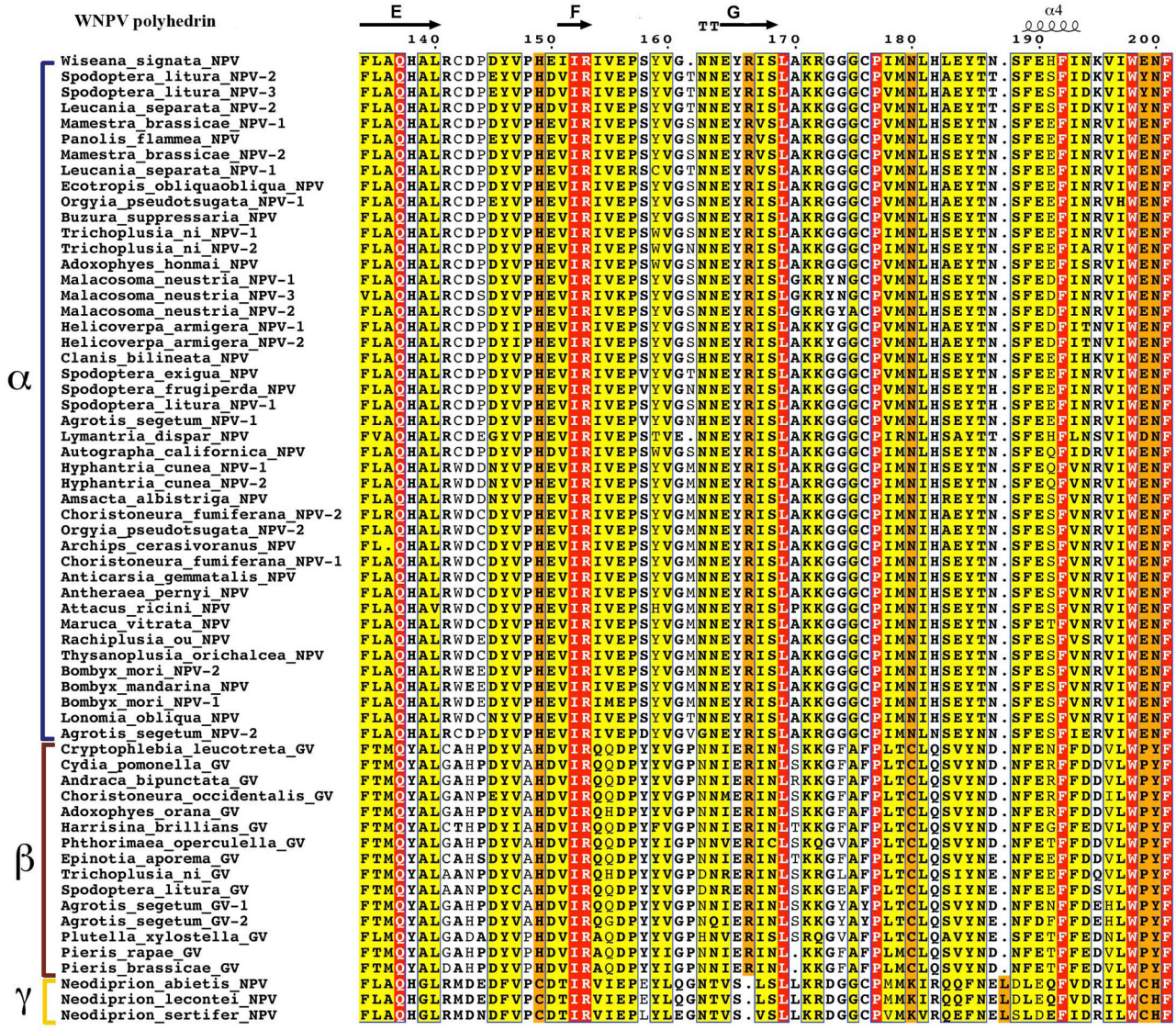

Fig. S2. Continued.

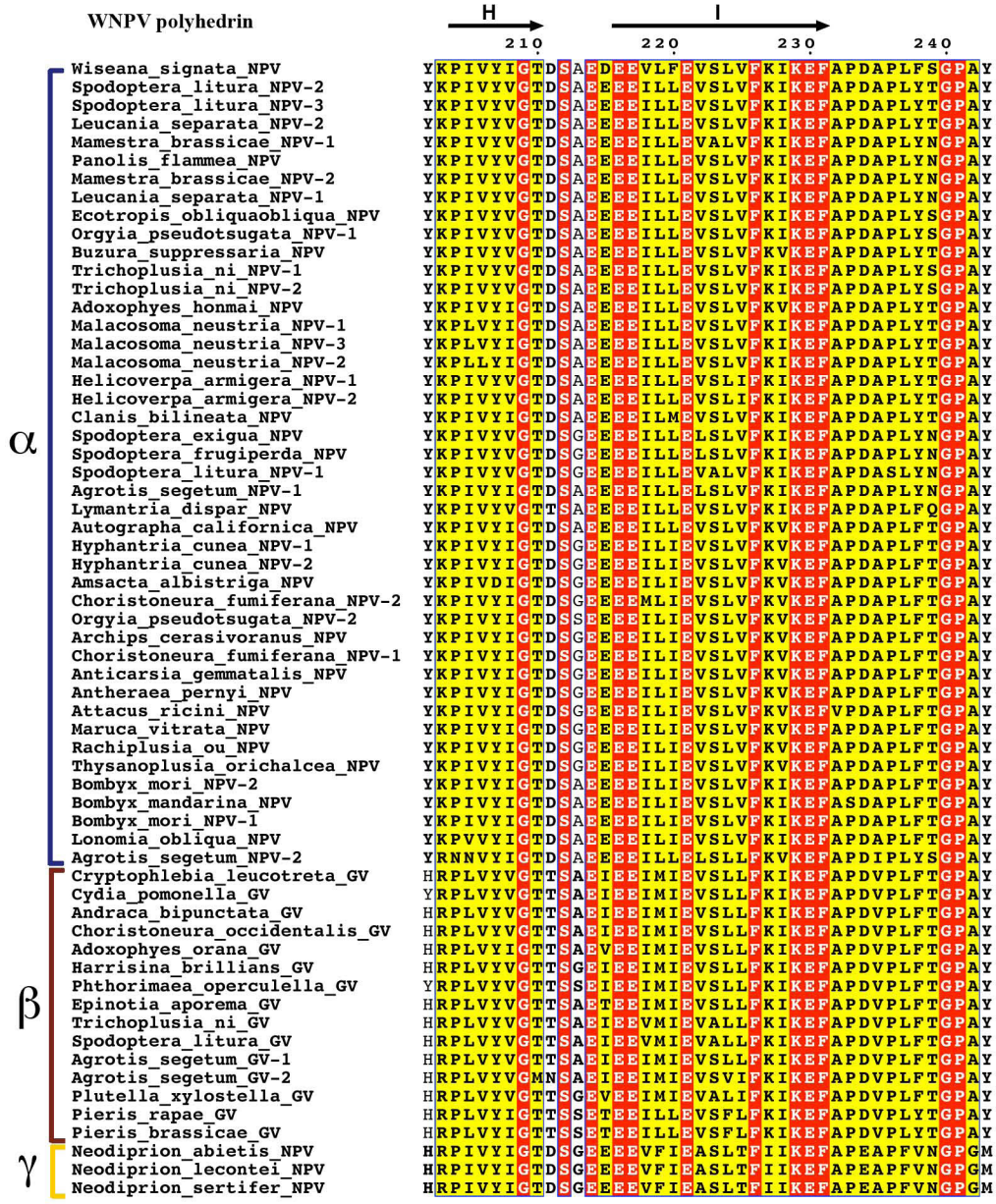

Fig. S2. Continued.

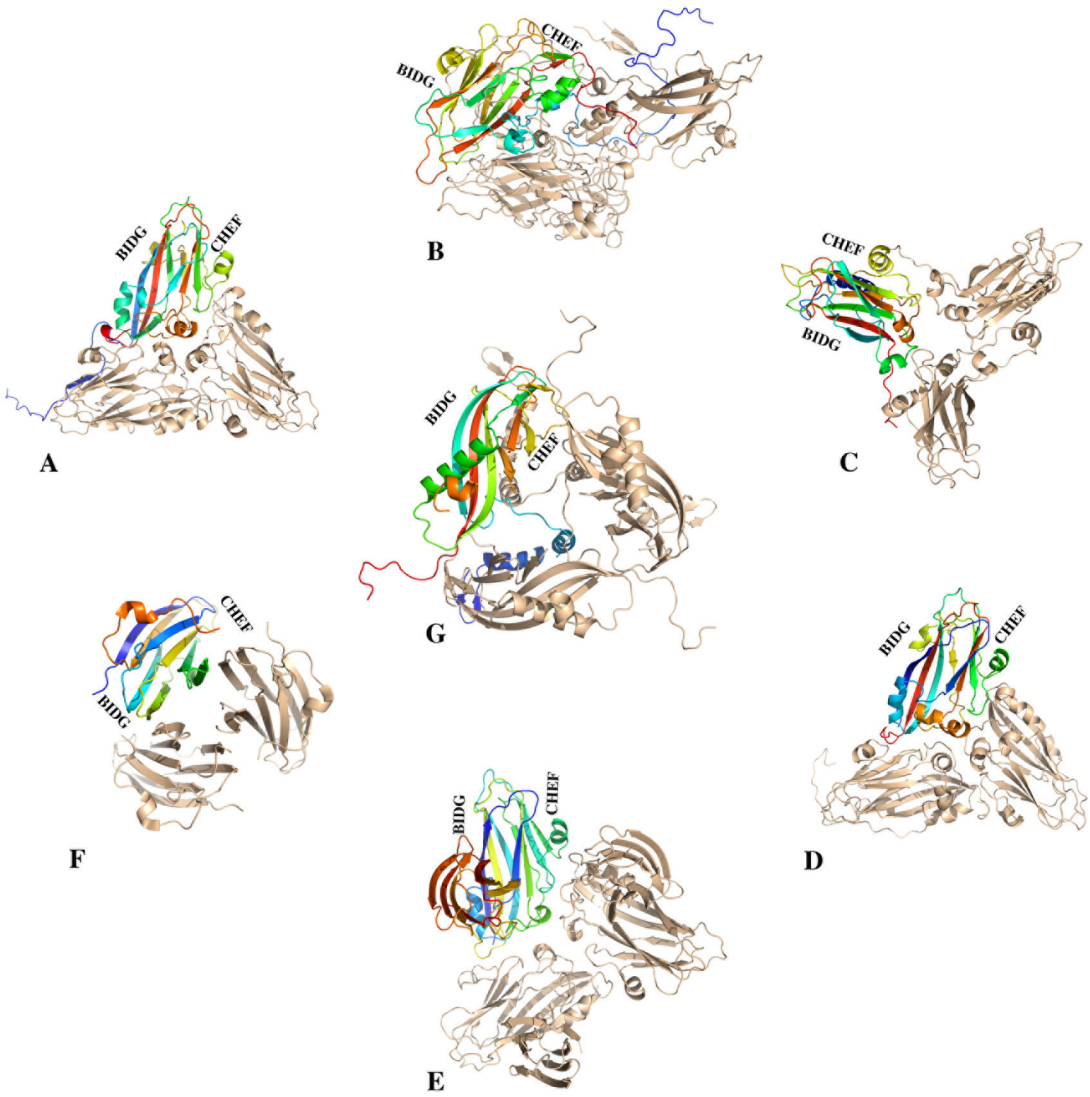

Fig. S3. Limited structural similarity of the baculovirus polyhedrin with jelly-roll containing proteins. Cartoon representation of the top six Dali (1) hits (cf. Table S1 for details). By decreasing the *z* score, the hits are as follows: the capsid of the rice yellow mottle virus (*A*), the coat protein of coxsackievirus b3 (*B*), the coat protein of cucumber mosaic virus (*C*), the coat protein of the carnation mottle virus (*D*), the capsid protein of the tobacco necrosis virus (*E*), and the *Yersinia* pseudotuberculosis-derived mitogen (YPM) (*F*). (*G*) Baculovirus polyhedrin trimer is represented in the center of the figure for comparison. The asymmetrical unit is represented for the viral proteins [obtained from Viperdb2 (2)], and the biological unit is represented for YPM. A protein subunit is highlighted with a blue-green-yellow-red gradient from the N terminus to C terminus, and the rest of the molecules are shown in pale brown. The BIDG and CHEF β-sheets are also indicated.

1. Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v. 3. *Bioinformatics* 24(23):2780–2781.
2. Carrillo-Tripp M, et al. (2009) VIPERdb2: An enhanced and web API enabled relational database for structural virology. *Nucleic Acids Res* 37(Database issue):D436–D442.

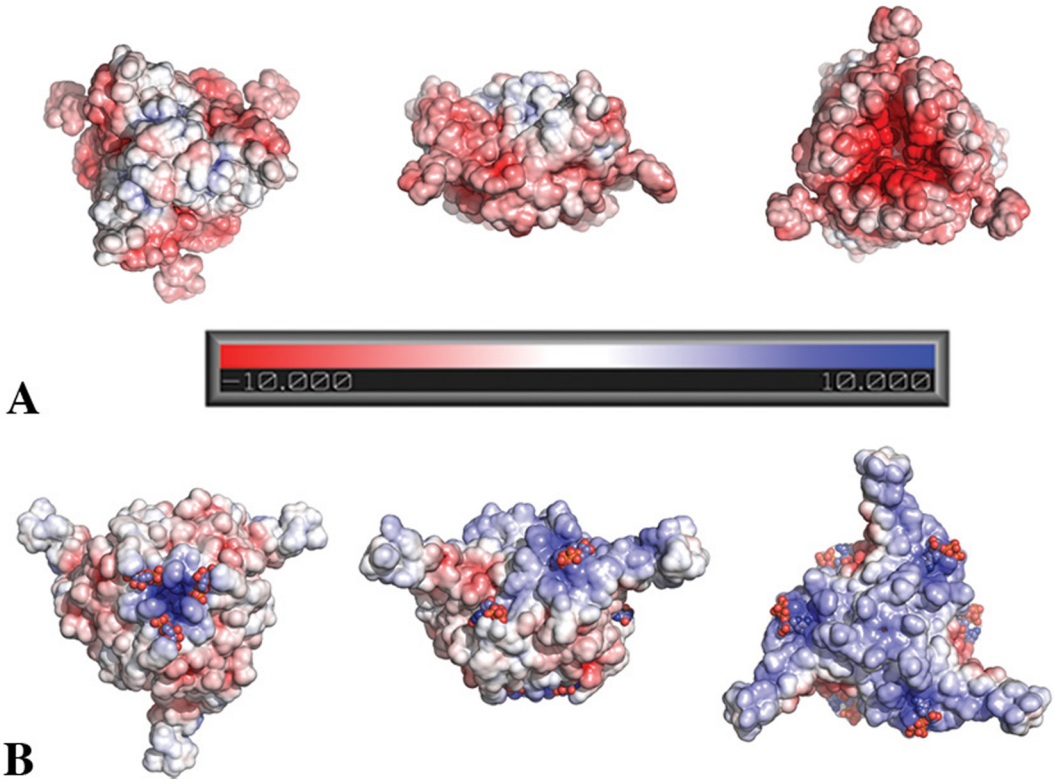

Fig. S4. Comparison of the electrostatic potential at the surface of baculovirus and cypovirus polyhedra. The electrostatic potential was calculated using APBS software [Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98:10037–10041] with standard settings (310 K, no ions, nonlinear Poisson–Boltzmann equation, 1.4-Å solvent radius, protein/solvent dielectric constants of 2.0 and 80.0, respectively). The structure of G25D AcMNPV polyhedrin was used here, and residues for which side chains could not be modeled were omitted from the calculations. The electrostatic potential is represented on the solvent-accessible surface by a red-white-blue gradient for the −10 to +10 *kT* range.

Table S1. Structural comparison

	Origin	PDB ID	*z* score	rmsd	Identity, %	Number of aligned residues, N_{algn}
Capsid of the rice yellow mottle virus	*Sobemovirus*	1f2n	6.3	6.1	8	115
Coxsackievirus coat protein	*Picornaviridae*	1cov	6.1	6.2	5	125
Cucumber mosaic virus coat protein	*Bromoviridae*	1f15	5.9	6.1	11	132
Carnation mottle virus coat protein	*Tombusviridae*	1opo	5.8	6.1	7	118
Tobacco necrosis virus capsid protein	*Tobamovirus*	1c8n	5.7	6.0	11	112
Yersinia pseudo-tuberculosis-derived mitogen	*Yersinia* (*Enterobacteriaceae*)	1pm4	5.6	2.9	12	85

This table lists the results of a Dali search for structural similarities between the WNPV polyhedrin and proteins from the PDB sorted according to *z* scores (*z* scores >5.5). Hits corresponding to orthologues in the same virus family are omitted for clarity (e.g., capsids of the numerous members of the *Picornaviridae* family).

Table S2. Data collection and phasing statistics

	G25D AcMNPV	SeMet	$SeMet2_{Ag}$	$SeMet3_{Pt}$	KI/I_2	CH_3Hg phosphate
Data collection						
Space group	I23	I23	I23	I23	I23	I23
Cell dimensions						
a = b = c, Å	103.158	103.207	103.136	103.20	103.242	103.515
Resolution, Å	2.8 (2.95–2.8)	2.92 (3.07–2.92)	2.90 (3.06–2.90)	2.80 (2.95–2.80)	2.40 (2.53–2.40)	2.90 (3.06–2.90)
R_{merge}	10.6 (20.2)	9.3 (34.1)	21.5 (53.8)	9.3 (21.1)	9.0 (28.3)	13.6 (68.1)
Mean, $I/\sigma I$	13.1 (4.5)	10.7 (3.6)	9.7 (4.0)	8.3 (3.6)	7.6 (3.2)	4.4 (1.9)
Completeness, %	98.1 (93.9)	97.2 (93.5)	95.4 (94.4)	84.2 (76.5)	89.1 (93.4)	35.3 (24.8)
Redundancy	5.7 (2.2)	4.6 (4.2)	10.6 (7.4)	1.9 (1.5)	2.0 (1.9)	1.3 (1.1)
Phasing						
Heavy-atom sites/occupancy B	HA1	HA1	HA1	HA1	HA1	
	0.74/47.4	0.70/36.5	0.96/54.2	0.72/45.8	0.68/20.1	
		HA2	HA2	HA2	HA2	HA2
		0.71/37.5	0.66/25.9	0.71/22.8	0.61/25.6	0.76/42.9
		HA3	HA3	HA3	HA3	HA3
		0.85/67.4	0.79/41.4	0.72/44.3	0.75/62.4	0.67/14.1
		HA4	HA4	HA4	HA4	HA6
		0.54/68.3	0.38/53.1	1.29/155.1	0.56/70.1	0.70/49.7
					HA5	HA7
					0.28/23.1	0.59/29.9
						HA8
						2.21/174.3
FOM 18.2–2.8 Å (3.4–2.8 Å)						
Acentric	0.44 (0.29)					
Centric	0.50 (0.33)					
Phasing power acentric/centric		0.97/0.76	1.00/0.86	1.14/1.01	1.54/1.39	1.03/0.75

Highest resolution shell is shown in parenthesis.

HA1 (0.593, 0.911, 0.175), HA2 (0.617, 0.874, 0.091), HA3 (0.335, 0.128, 0.258), HA4 (0.406, 0.416, 0.108), HA5 (0.577, 0.230, 0.411), HA6 (0.942, 0.773, 0.307), HA7 (0.468, 0.870, 0.452), HA8 (0.310, 0.732, 0.275).

Table S3. Data collection and refinement statistics

	G25D AcMNPV	WNPV
Data collection		
No. crystals	5	19
Space group	I23	I23
Cell dimensions		
a = b = c, Å	103.183	102.437
Resolution, Å	20–2.30 (2.42–2.30)	30–2.17 (2.29–2.17)
R_{merge}, %	14.8 (54.6)	34.9 (*)
R_{pim}, %	6.1 (23.0)	9.2 (36.7)
Mean, I/σ I	10.2 (3.8)	8.0 (2.3)
Completeness, %	99.7 (99.1)	99.9 (100.0)
Redundancy	6.6 (6.3)	15.0 (14.5)
Refinement		
Resolution, Å	18.84–2.30 (2.57–2.30)	20.91–2.17 (2.43–2.17)
No. reflections	8,240 (811)	9,620 (947)
R_{work}/R_{free}	16.05/21.40 (16.38/23.46)	16.12/18.99 (19.05/22.44)
No. atoms	1,655	1,710
Protein	1,577	1,619
Ions	1× ethylene glycol	1× SO_4^{2-}
Water	74	86
B-factors		
Protein	24.08	20.83
Water	31.70	31.55
rmsd		
Bond lengths, Å	0.008	0.008
Bond angles, °	1.04	1.04

*R_{merge} values higher than 100% are reported as 0.000 by Scalepack. Given the large number of crystals and high redundancy, R_{pim} is a more appropriate measure of data quality than R_{merge} (see *Methods* for details on the processing strategy).