

Combination of deep XLMS with deep learning reveals an ordered rearrangement and assembly of a major protein component of the vaccinia virion

Yeva Mirzakhanyan,¹ Andris Jankevics,² Richard A. Scheltema,² Paul David Gershon¹

AUTHOR AFFILIATIONS See affiliation list on p. 12.

ABSTRACT Vaccinia virus, the prototypical poxvirus and smallpox/monkeypox vaccine, has proven a challenging entity for structural biology, defying many of the approaches leading to molecular and atomic models for other viruses. Via a combination of deep learning and cross-linking mass spectrometry, we have developed an atomic-level model and an integrated processing/assembly pathway for a structural component of the vaccinia virion, protein P4a. Within the pathway, proteolytic separation of the C-terminal P4a-3 segment of P4a triggers a massive conformational rotation within the N-terminal P4a-1 segment that becomes fixed by disulfide-locking while removing a steric block to trimerization of the processing intermediate P4a-1+2. These events trigger the proteolytic separation of P4a-2, allowing the assembly of P4a-1 into a hexagonal lattice that encloses the nascent virion core.

IMPORTANCE An outstanding problem in the understanding of poxvirus biology is the molecular structure of the mature virion. Via deep learning methods combined with chemical cross-linking mass spectrometry, we have addressed the structure and assembly pathway of P4a, a key poxvirus virion core component.

KEYWORDS poxvirus, vaccinia virus, structural biology, mass spectrometry, structural proteomics

Vaccinia, the prototypical member of the poxvirus family, is a large DNA virus with just under 200 genes. It comprised the vaccine used for smallpox eradication and is the current vaccine against monkeypox. On the basis of genomic similarity, all three of these viruses are likely structurally identical. Over the past three decades, substantial insights have been gained into the biology of vaccinia at the functional level: broad roles have been deduced for most of the essential virus genes via phenotypic analysis of natural or engineered conditional mutants (1). The ultrastructure of the vaccinia virion has been apparent in outline since the 1960s, comprising an ellipsoid or “brick-shaped” structure whose core contains the genome with various enzymes and structural proteins, enclosed by at least one lipid envelope with envelope proteins. The core wall is perhaps the virion’s most tangible ultrastructural feature though it likely has substantially less structural rigidity than the capsid of a non-enveloped virus (2–5). Atomic structures for 34 of the 75–80 proteins packaged within the mature virion (MV) can be found, currently, in RCSB Protein Data Bank. A much more limited picture has been achieved of the virion’s internal molecular architecture. Reasons for this include the size and complexity of vaccinia, its polymorphic nature and overall asymmetry, and the presence of at least one phospholipid envelope. As an intact entity, vaccinia has largely defied the classical approaches of X-ray crystallography and cryoelectron microscopy (cryoEM)/tomography.

Editor Anne Moscona, Columbia University Medical College, New York, New York, USA

Address correspondence to Paul David Gershon, pgershon@uci.edu.

The authors declare no conflict of interest.

See the funding table on p. 13.

Received 3 May 2023

Accepted 11 July 2023

Published 30 August 2023

[This article was published on 30 August 2023 with typographical errors in Results, Fig. 1 and its legend, and the supplemental material. The errors were corrected in the current version, posted on 12 September 2023.]

Copyright © 2023 Mirzakhanyan et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Chemical cross-linking with bifunctional reagents (cross-linking mass spectrometry; XLMS) can inform structural biology in multiple ways, for example, by constraining atomic models of individual proteins, guiding the docking of proteins of known 3D structure, and providing protein molecular interaction network models. In the latter regard, our preliminary XLMS study of vaccinia revealed intra-virion protein-protein networks in outline (6). We have fundamentally extended this XLMS data set, resulting in, after stringent thresholding, a saturating data set comprising ~135,000 cross-link-spectral matches (CSMs) representing ~22,000 unique cross-links (XL) on the basis of protein accession, cross-link position and cross-linked peptide Mr. To assemble these into a molecular model, a desirable starting point would be atomic models for at least the major structural protein components of the virion—the very entities for which classical structural biology has been unsuccessful at providing atomic models, perhaps as a result of their higher-order complexity and intrinsic insolubility.

Recent deep learning approaches for the *de novo* prediction of protein structure have been described as having, to a large extent, solved the “protein folding problem.” Initially skeptical, particularly with regard to vaccinia structural proteins—the majority of which are unique to the poxviridae—we generated structural predictions merely as placeholders for XLMS network analysis. However, via exhaustive statistical and other benchmarking studies, we came to regard some structural models as approaching experimental accuracy (manuscript in preparation). In the study described here, a vaccinia major structural protein, P4a, was structurally modeled via deep learning methods. In combination with experimental data (XLMS), the resulting structural models yield insights into processing, conformational change, disulfide rearrangement, and higher-order assembly to the level of the whole-virion.

RESULTS

Vaccinia protein P4a is one of the virion’s most abundant protein components. It is conserved in all known poxviridae with no identifiable sequence or structural homologs outside this family. Conditional mutants under nonpermissive conditions exhibit an interruption of normal virion morphogenesis, reduced virus yield and an accumulation in the cytoplasm of abnormal immature virus (IV)-like particles lacking the normal, dumbbell-shaped core morphology (7–17). During virion morphogenesis, P4a is proteolytically processed at two sites (at residues 614 and 697) marked by the diamino acid AGI (9, 11, 17–23). Processing at these two sites is critical for normal morphogenesis (20, 24–29).

P4a structural models

The three major products of P4a processing are referred to here, in linear order as P4a-1, P4a-2, and P4a-3 (Fig. 1A). P4a-1 and P4a-3 are packaged, while P4a-2 is likely discarded and eliminated at the proteasome (19). Despite an absence of sequence homologs outside of the poxviridae, atomic structures for P4a-precursor, P4a-1 and P4a-3 could be predicted by AlphaFold2 (Fig. 1B; Fig. S1) with high statistical confidence (Fig. S1). An equally confident structure was predicted for the P4a-1 segment of the presumed cleavage intermediate P4a-1+2 (Fig. 1B; Fig. S1). No significant structural homologs were identified to any of the above forms of P4a (data not shown) supporting the uniqueness of this protein to the poxviruses.

Conformational hinging

The AlphaFold2 model for P4a-1 showed two subdomains (denoted here as P4a-1 domains I and II; Fig. 1B) connected by a hinge that could support a substantial (~180°) rotation of domain II with respect to domain I. This rotation was evident by the comparison of models for the P4a-1 segment in precursor versus processed forms (Fig. 1C; Movie S1). Supporting the authentic modeling of both conformations, precursor and product models of the P4a-1 segment showed equally strong average pLDDT scores (Fig. S1).

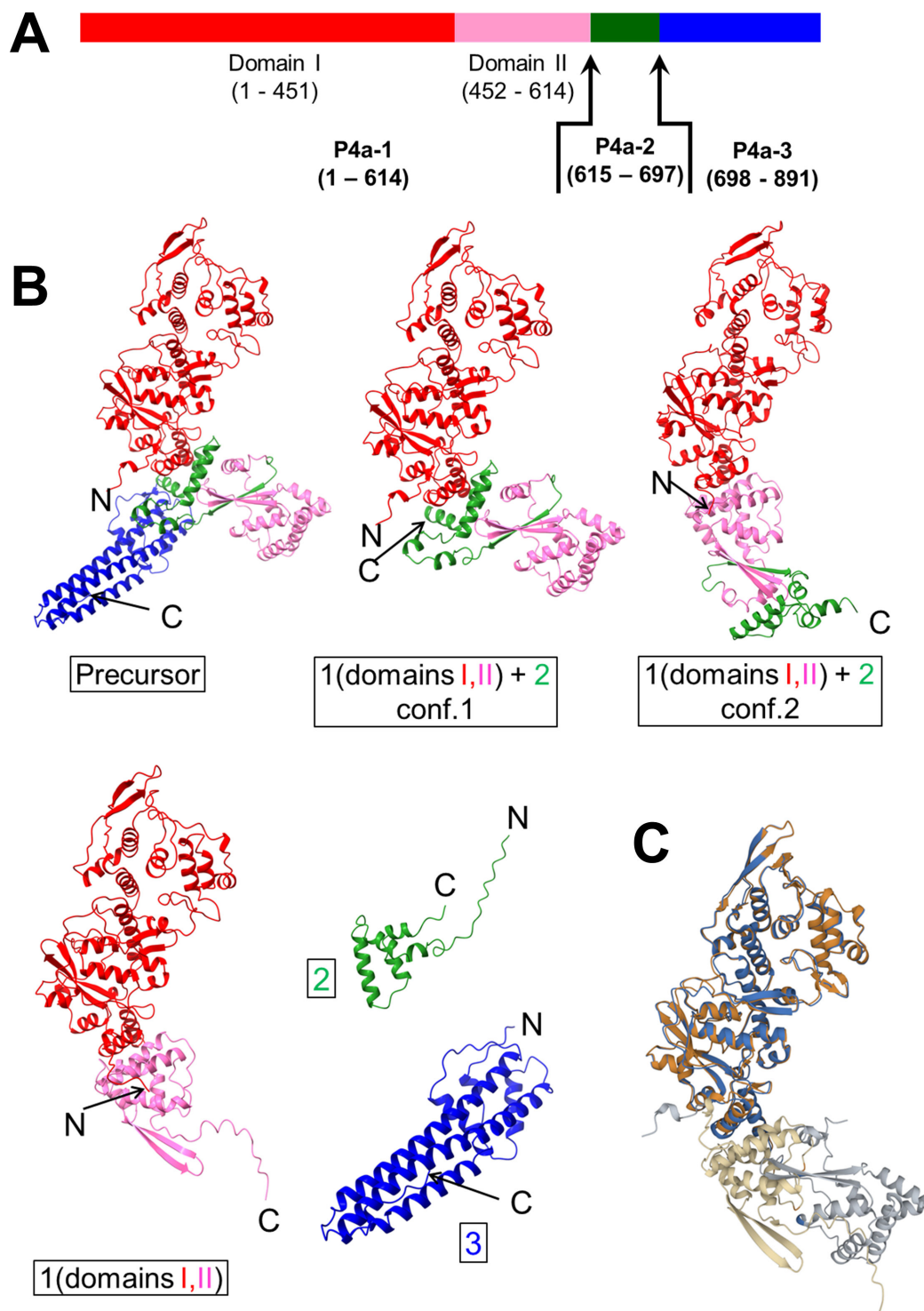


FIG 1 Structural models of P4a. (A) P4a precursor as a linear chain, colored by processed segments (P4a-1, P4a-2, and P4a-3) and domains of P4a-1. Black arrows indicate AGI processing sites. (B) Structural models for P4a precursor and processing products P4a-1, P4a-2, P4a-3, and the two P4a-1+2 intermediate forms. P4a-1 was predicted with two domains denoted here I (red) and II (pink) connected via a hinge. Green and blue: P4a-2 and P4a-3 segments, respectively. Conformational inversion at the hinge corresponded with the removal of P4a-3, the C-terminal processed product of P4a. (C) Aligned AlphaFold2 predictions for the P4a-1 segment from P4a-precursor (residues 1–614; blue: Aligned residues; gray: Residues not in alignment) and P4a-1 processed product (orange: Aligned residues; tan: Residues not in alignment). P4a-1 domain II has rotated ~180°.

Models for mature P4a-1 predicted a single conformation exclusively (Fig. S2), and those for the N-terminal region of P4a precursor spanning segments P4a-1 and P4a-2 showed almost exclusively a single conformation (Fig. S2). Models for the P4a-1+2 cleavage intermediate, however, clustered neatly around a pair of distinct conformations which differed at the hinge (Fig. S2; Fig. 1; Fig. S1), strongly suggesting that this intermediate, specifically, was bistable.

The conformation of mature P4a-1 could be verified from experimental intra-protein XLMS data from intact MV: Of the summed XLMS CSM count of 4,009 available for intra-P4a-1 cross-links, a subset totaling 3,829 (95.5% of the overall total) supported the conformation modeled in mature P4a-1 (Fig. S3), while a subset totaling just 2,265 (only 56.5% of the overall total) supported the conformation of P4a-1 in the P4a precursor model (Fig. S4). Among the distance-violating cross-links in the latter model that were structurally rational in mature P4a-1, the majority lay directly across the P4a-1 hinge, strongly supporting the “post-rotation” conformation of P4a-1 in the virion *in vivo*.

P4a-3 removal as the trigger for conformational hinging

In the precursor conformation, P4a-1 domains I and II are separated from one another by a wedge-like interjection of the P4a-2 segment into the P4a-1 hinge. This interjection comprises the three C-terminal alpha helices of P4a-2 nestled between two helices of P4a-1 domain I and one helix of P4a-1 domain II, along with a three-stranded beta sheet comprising one strand from the N-terminus of P4a-2 and two strands from P4a-1 domain II (Fig. 1B: “Precursor”). This conformation of P4a-2 (and thereby of the overall precursor) seems to be stabilized, externally, by the P4a-3 segment which, while flexible in absolute placement (data not shown), modeled consistently in the vicinity of the hinge in the precursor model (Fig. 1B: “Precursor”) and presents a barrier to P4a-2 movement or rearrangement. Cleavage at P4a’s downstream AG| processing site to release the P4a-3 segment removes the external brace, destabilizing the P4a-2 wedge and allowing hinging to the “post-rotational” P4a-1 product conformation (Fig. 1B; Movie S1).

Disulfide locking

Models were next scanned for cysteine pairs lying within the potential intra-protein disulfide bonding range. One such pair was found, namely Cys31 and Cys569. These two cysteines (which are conserved in all known poxvirus P4a sequences) are located within the P4a-1 segment, one each in domains I and II where they directly span the hinge. While they lay far outside disulfide bonding range in P4a precursor (Fig. S7) and in the P4a-1+2 intermediate conf.1 (prerotation conformation; data not shown), they are within disulfide bonding range in fully processed P4a-1 (Fig. S7) and in the P4a-1+2 intermediate conf.2 (post-rotation conformation). After AG| processing and removal of the P4a-3 product, a potential mechanism for locking of the post-rotational conformation is provided by the vaccinia-encoded and packaged oxidoreductases (30–32), whose activity in disulfide bond stabilization is documented (33).

P4a-1 trimerization

To investigate the higher-order structure of P4a-1 in MV, structural predictions were attempted for P4a-1 dimer, trimer, tetramer, and pentamer using AlphaFold-multimer. Among these, the trimeric model showed a particularly high confidence (Fig. 2A and B) and by far the strongest PAE plot (Fig. 2C; Fig. S8). Moreover, a very high degree of convergence was noted between the 25 trimer models reported by AlphaFold-multimer (Fig. S9 and S2). The trimer model adopted the approximate shape of a cylindrical “candlestick holder” with “opening petals,” with overall dimensions of ~6.8 nm diameter (at the waist) × ~11.5 nm height (Fig. 2A and B). Around the trimer body, the subunits partially enwrap one another at the base (Fig. 2D). Views of the trimer from above and below each suggest a hexagonal outline (Fig. 2E and F). The upper and lower hexagons superimpose on one another “in phase” as a hexagonal prism shape (Fig. 2G), with a ~60° twist within each individual subunit.

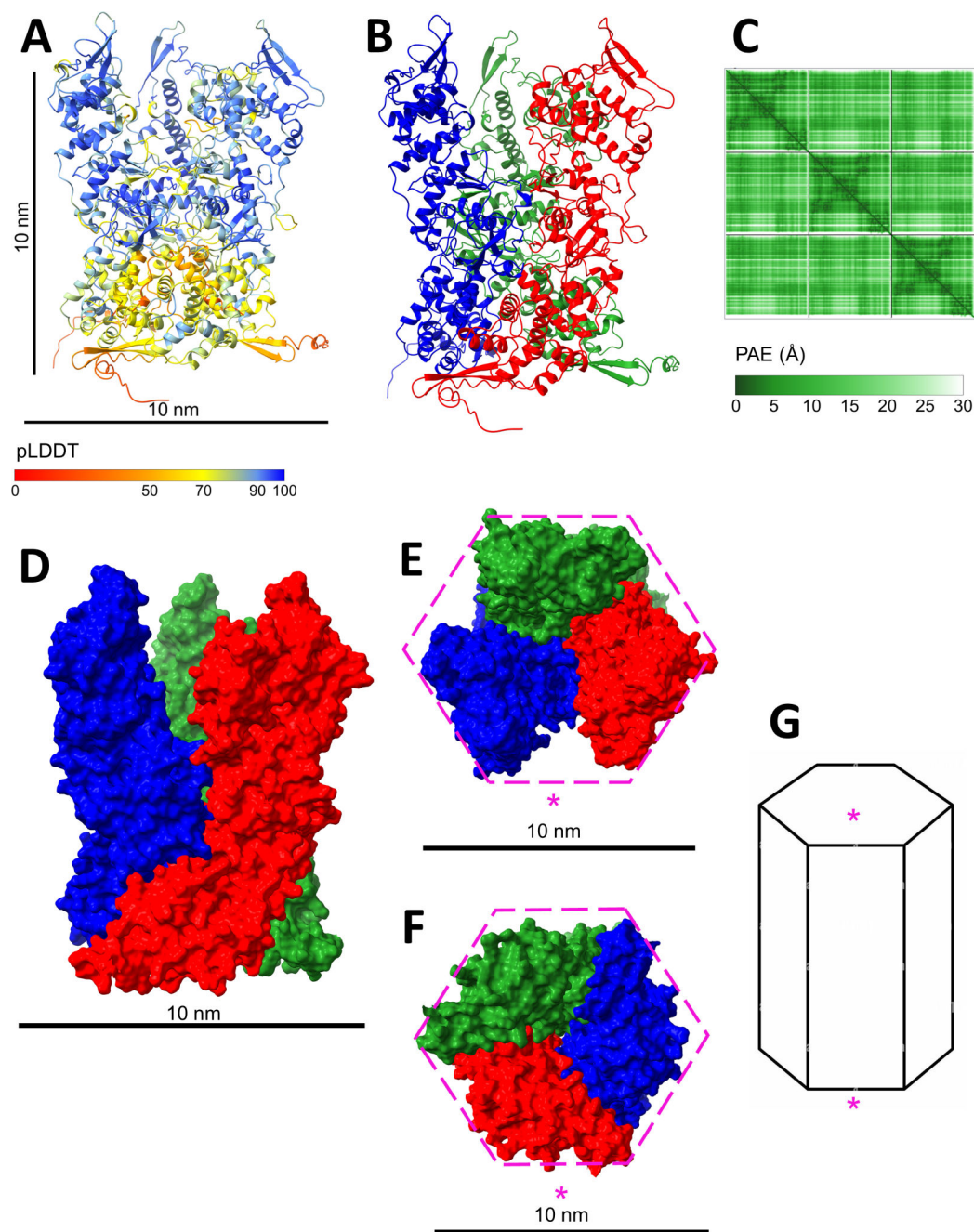


FIG 2 P4a-1 trimer. (A) Colored pLDDT (scale bars: 10 nm), (B) Colored by P4a-1 chain within the trimer. (C) PAE plot for trimer. Color scale is in Angstrom units. (D–F) Trimer in spacefill from side, above, and below, respectively (colored as in Fig 2B) showing the mutual enwrapping of subunits at the base, and an outline approximating to hexagonal geometry (residues 600–614 excluded from view). (G) The trimer model can be fit within a hexagonal prism geometry (upper and lower hexagons are in phase), with a $\sim 60^\circ$ twist along the body of each subunit such that the starred side, viewed from above (E) corresponds to the starred side viewed from below (F).

P4a-3 removal is both necessary and sufficient for P4a-1 trimer formation

All attempts to model a trimer for the full-length P4a precursor failed (data not shown), consistent with which the P4a-3 segment presents a fundamental steric block to subunit mating (Fig. 1B, data not shown). However, trimer models could be readily obtained for both P4a-1 and the P4a-1+2 intermediate (in both pre- and post-rotation conformations

about the P4a-1 hinge; Fig. S10). We conclude that the removal of the P4a-3 segment is both necessary and sufficient for P4a-1 trimer formation.

Trimer molecular model is consistent with cryoEM imaging

Since P4a-1 is considered to lie on the outer surface of the virion core wall (34), we scrutinized published reports for the dimensions of core wall ultrastructural features that might correlate with P4a-1 trimers. Early EM and cryoEM studies of virion cores reported a “palisade” layer of spikes or “pegs” on the outer core wall surface (35–38). Measurements of the spike dimensions varied in these early reports, perhaps due to limitations in instrument capabilities, with spike lengths ranging from 100 to 200 Å (36) and diameters of 50–100 Å (39). A very recent cryoEM study (released during the final preparation of this manuscript) of Vaccinia IVs and MVs by Hernandez-Gonzalez et al. (40) measured the palisade layer thickness at 12.5 nm. These measurements agreed closely with the dimensions of our P4a-1 trimer model (Fig. 2). Furthermore, the P4a-1 trimer model’s size and shape were comparable with the numerous scattered features visible in Fig. 8 of reference 38 that could now be identified as probable dissociated P4a-1 trimers (Fig. S11). The above data were consistent with P4a-1 trimer being the primary distinguishable component of the “palisade” spikes.

Higher-order P4a-1 trimer structure and assembly

We next investigated the higher-order organization of P4a-1 homotrimers, using XLMS data for guidance: Identical peptides cross-linked to one another (homomultimer XL) can have only arisen if the cross-link bridges homomultimer subunits. Our homomultimer XL data set for P4a-1 from intact MV, showed a summed CSM count of 779. While the discrete homotrimer model (above, Fig. 2) satisfied 95.5% of the CSM count for all intra-P4a-1 cross-links (Fig. S12), it could satisfy only 44.7% of the 779 homomultimer CSMs. The remaining 55.3% represented, apparently, undefined higher-order assemblies of P4a-1. Most striking among these was cross-link 366-366, which accounted for nearly half of all P4a-1 homomultimer CSMs, and was detected strongly for all cross-linker types in our data set including those with the shortest cross-linking distances (e.g., PhoX; just 23 Å). Other higher-order homomultimer XL detected in intact MV included 508-508 and 557-557 (CSM count = 66 and 6, respectively).

The trimer’s hexagonal outline (above; Fig. 2E through G) suggested a higher-order organization comprising some form of hexagonal array. Arranging individual trimers (Fig. 3A) within the simplest conceivable hexagonal array yielded three potential rotationally symmetrical trimer-hexamer models (Fig. 3B). Of these, Fig. 3B(iii) could not satisfy the 366–366 homomultimer cross-link at all. Figure 3B(ii) could satisfy neither 366–366 with cross-linkers PhoX or DSG, nor 508–508 with cross-linkers PhoX, DSA, DSS, or BSPEG5, presenting a fatal limitation for this model. The model of Fig. 3B(i), however, could satisfy 366–366 for all cross-linker types (with a cross-linking distance of just 21.6 Å, Fig. S13) as well as 508–508 and all other higher-order homomultimer XL. This model (Fig. 3B(i)) accounted for 97.2% of all homomultimer XL observed for P4a-1 from MV (either intact or de-enveloped) and was selected as a basis for higher-order modeling (Fig. 3C). Higher-order models comprised a trimer-hexamer whose central hole was filled by a 7th trimer (trimer-heptamer; Fig. 3C(i)) or a tessellation of either abutting or fused trimer-hexamers (Fig. 3C(ii), (iii), respectively). Albeit the trimer-heptamer model could not be fully saturated with 366–366 homomultimer XL (Fig. 3C(i)), this by itself did not render the model invalid. The trimer-heptamer model did, however, lead to an entirely “closed” core wall rather than the more open lattice with trimer-hexamers (Fig. 3C(ii), (iii)), with the former reminiscent of a rigid capsid rather than the flexible, collapsible sac that is more characteristic of the vaccinia core (2, 3).

Between the two trimer-hexamer interface models (abutted or fused, Fig. 3C(ii), (iii), respectively), only the fully fused model (Fig. 3C(iii)), when extended to a larger lattice (Fig. 3D), could fully saturate the lattice with “holes” (Fig. 3D) to combine uniform

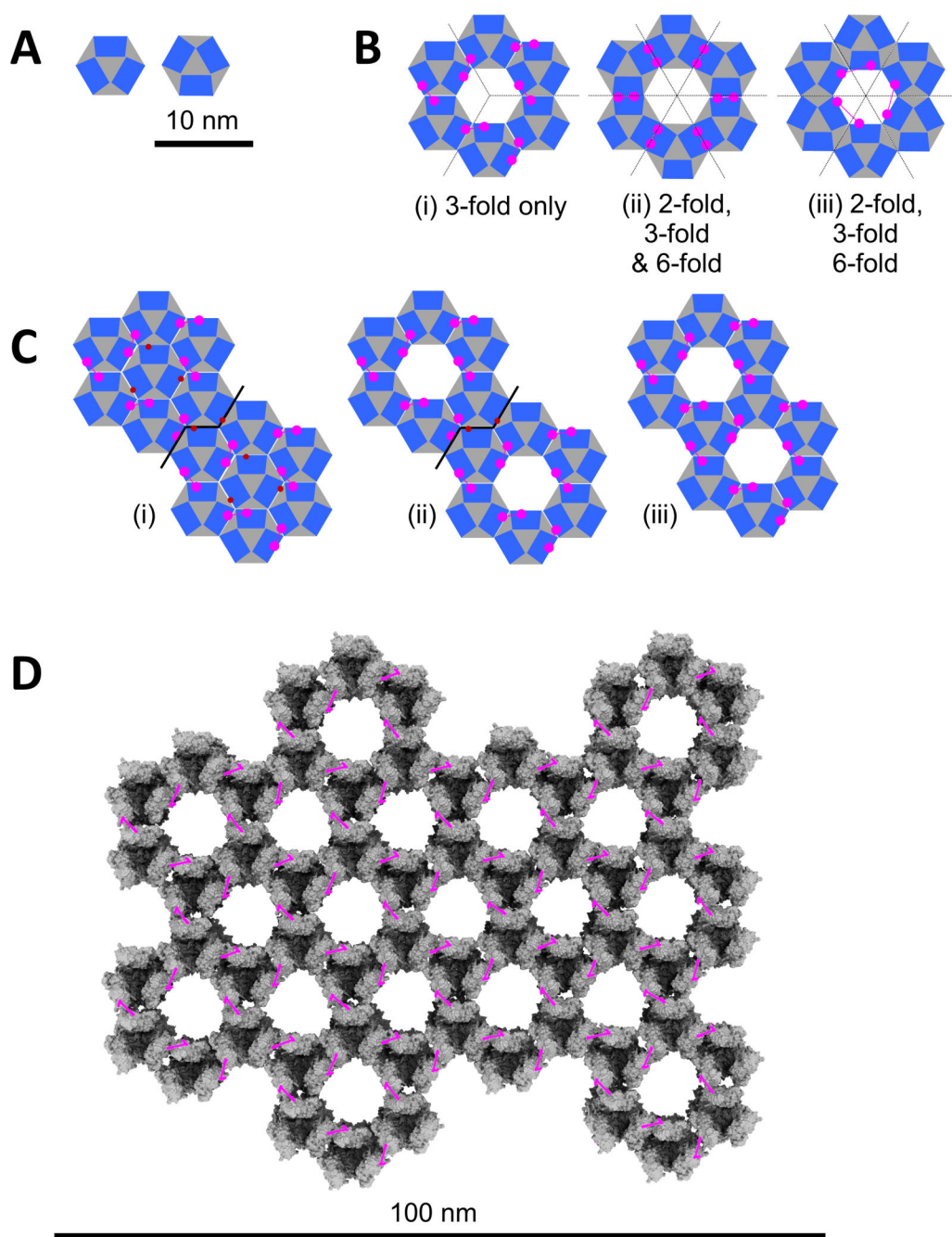


FIG 3 Hexamer-of-trimers models: Genesis, in schematic form. (A) The two rotamers of an individual trimer (60° difference). (B) Rotationally symmetrical hexamer-of-trimers models based on the trimer's hexagonal outline (symmetries are indicated). In (i), all trimers are in the same rotational orientation, while in (ii) and (iii), the two forms in panel (A) alternate. Internal trimer-trimer 366–366 cross-links are indicated in Magenta (the outward-facing six subunits are unsaturated providing potential hooks to neighboring trimer-hexamers). Cross-link distances favor model (i) over models (ii) and (iii). (C) Trimer-heptamer (i) and trimer-hexamer interface schemes (ii, iii) based on the model in panel B(i). Schemes (i) and (ii) do not saturate all interfacial 366–366 XL (unsatisfied cross-linkable sites shown brown), while (iii) does. (D) Lattice based on the model in panel C(iii). 366–366 cross-links, which reside below the upper surface of the protein when viewed from above (Fig. S14), have been projected over the surface to emphasize their locations. The Euclidian plane tessellation of panel (D) has regular $p6m$ symmetry (one-uniform) with periodic (H and K) = (3,0). The low confidence C-terminal tails (residues 600–614) were removed for clarity.

strength in all directions with maximum flexibility. Also unique to this model: Every trimer fell at the vertex of three fused trimer-hexamers, every trimer-trimer interface separated two holes in the lattice, and three sides of every trimer faced a hole. Moreover, the highest order trimer cluster was simply a trimer-dimer: In contrast to model C(i) and (ii) or mixed forms (not shown), there were no higher-order clusters. Finally, model C(iii) was unique in providing a 366–366 cross-link for every P4a-1 subunit, uniformly saturating the lattice with this XL (Fig. 3D). Model C(iii) was also fully saturated with 508–508 and all other higher-order homomultimer XL. For all the above reasons, model C(iii) was the favored building block for a full core wall lattice (Fig. 3D).

The higher-order lattice model is consistent with cryoEM imaging

We next attempted to reconcile the trimer-hexamer molecular model with cryoEM images of the intact virion core wall. Dubochet (38), Cyrklaff (41), and Hernandez-Gonzalez (40), reported patches of hexagonal features on the surface of the core wall, identified as the “palisade” coated with pegs (41), all of which fit and scaled to our P4a-1 lattice (Fig. S14). Pegs visible in relief around the edges of the virion core in cryoEM images seemed to have a periodicity of ~8.5 nm, comparable to that of both the hexagonally arranged features in the same cryoEM images (38, 41) and individual P4a-1 trimers in our lattice molecular model (Fig. S14). At a specific imaging orientation (Fig. 3D), trimer electron density may be expected to sum through sequential rows of trimers to yield the “palisade” effect.

Published cryoEM images show hexagonal features only discontinuously across the core (38, 41). This could arise from overlaying features (such as bent or distorted pegs), the technicalities of cryosectioning, or because the P4a-1 lattice is discontinuous. To establish whether the lattice can be reasonably considered to cover the entire surface of the core, we derived the anticipated experimental mass of P4a-1 in the virion based on known virion mass, the proportion of total mass comprising protein, and proportion of total virion protein comprising P4a-1 (Supp. Doc. 1). This calculation indicated that 12.4% of the overall virion mass, or 404 MDa, comprised P4a-1. In an independent calculation for the theoretical mass of P4a-1 in the virion if it were enclosed entirely by our lattice, we calculated the surface area of a virion core (taking the mean core size from a number of published images) and the total number and then mass of P4a-1 molecules that could coat it fully according to the lattice of Fig. 3D. This totaled 6,064 molecules or 430 MDa of P4a-1 (Supp. Doc1). From the similarity of the two values (404 and 430 mDa), we conclude that the modeled P4a-1 lattice likely encloses the entire surface of the virion core, albeit with, perhaps, localized regions of disruption. In this regard, for example, “pore-like” features on the surface of the core wall (34, 40, 41) may contribute to the disruption of the P4a-1 lattice.

Correspondence of P4a trimer lattice to the external scaffold

Vaccinia morphogenesis initiates with the insertion of virion envelope proteins A17 and A14 into the ER membrane. This membrane is subsequently fractured, allowing A17 to associate with trimers of Vaccinia external scaffold protein D13, resulting in spherical IVs coated with the D13 external scaffold (1, 43). Deep-etch EM has shown that the external scaffold forms a hexagon-like honeycomb lattice across the entire surface of IV, with pentameric and heptameric defects (43). The known experimental structure of the external scaffold also shows D13 trimers forming a hexagonal trimer-hexamer lattice (42). During maturation from IV to MV, the A17 N-terminus is cleaved by Vaccinia protease I7, leading to D13 scaffold release. Either simultaneously or soon thereafter, the A17 C-terminus and core structural proteins (including P4a) are processed, resulting in condensation of the core wall and palisade layer (20, 24–29). Hexagonal features of the palisade layer, visualized by cryoEM (40, 41) are reminiscent of the honeycomb D13 lattice. Our P4a-1 trimer-hexamer lattice model was, dimensionally, entirely coincident with the atomic structure of the D13 lattice (Fig. 4). In some manner, the external

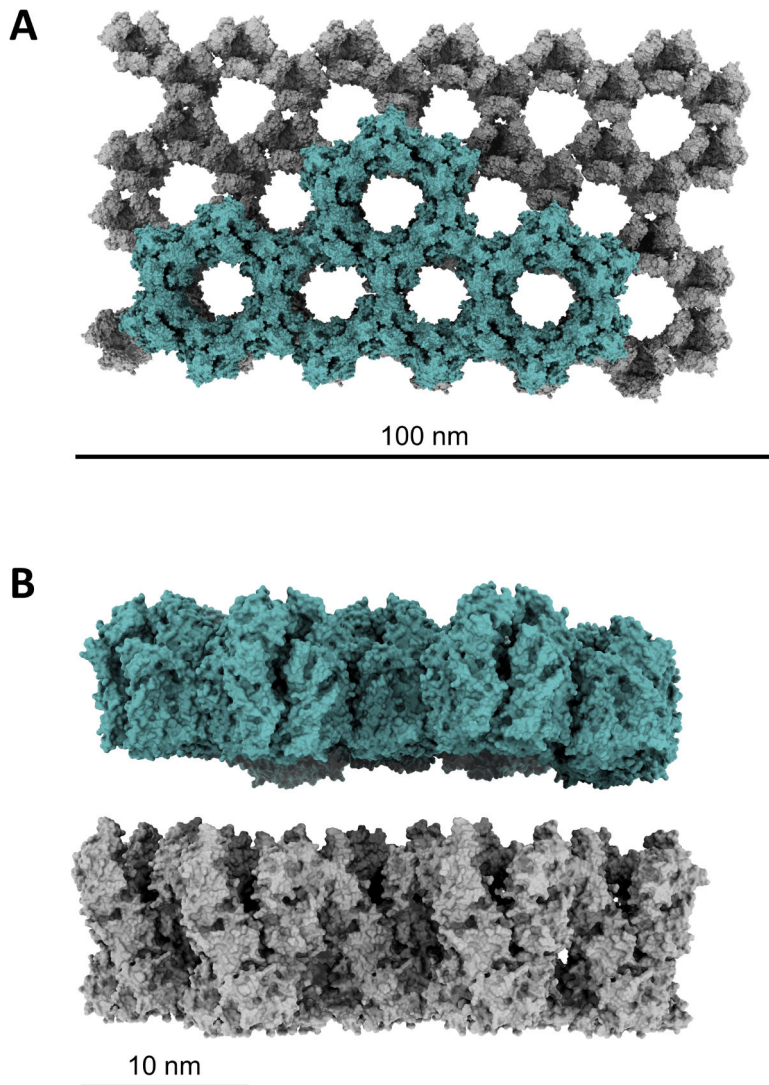


FIG 4 Modeled lattice of P4a-1 is co-dimensional with the D13 external scaffold (42). Gray: P4a-1. Turquoise: D13. View from above (A) and side in higher zoom (B). In panel (B), P4a-1 and D13 lattices are separated by an arbitrary distance of 5 nm.

scaffold, through its interaction with A17, may template condensation of the P4a-1 lattice during virion morphogenesis, for subsequent core wall assembly.

DISCUSSION

There are currently atomic-level structures for all or parts of 34 of the ~75 packaged Vaccinia gene products. However, with perhaps one exception, namely the packaged multisubunit Vaccinia RNA polymerase (44), how these proteins condense into higher-order assemblies has eluded analysis. Here, we have investigated Vaccinia major structural protein P4a, which accounts for 12.4% of the total virion mass. P4a precursor is transported to IVs during assembly where it is proteolytically processed at two AGI sites (19). P4a-1 and P4a-3 are packaged in MVs, while P4a-2 is discarded (19). The biological and functional significance of the two processing events has not been previously elucidated. Moreover, how P4a-1 forms the core wall and/or palisade layer associated with the core wall has remained unresolved. Here, we have combined deep-learning protein structure prediction with a deep XLMS data set comprising ~135,000 cross-link spectral matches via 10 distinct chemical cross-linkers to yield ~22,000 unique-mass

cross-linked peptide pairs. The predicted structure for P4a-1 monomer was consistent with all intra-protein cross-links (Fig. 1; Fig. S3). Homomultimer models were optimal at the trimer level (Fig. 2; Fig. S8 and S13) and the resulting trimer model was consistent with published low-resolution cryoEM images (Fig. S11) but did not accommodate all cross-links of the inter-subunit type. Building a hexamer of trimers of a specific topology, however, allowed essentially all inter-subunit cross-links to be satisfied and they remained satisfied upon building the trimer-hexamer to a full core wall lattice (Fig. 3; Fig. S13 and S14).

We suggest a five-step ordered/concerted assembly pathway for P4a-1 (Fig. 5; Movie S2). In the initial P4a precursor-monomer conformation (prior to step 1), the P4a-3 segment presents a steric block to trimerization and the P4a-1 trimerization interface is conformationally split. P4a-3 detachment by AGI processing initiates the pathway (step 1), relieving the block to trimerization, while also destabilizing the precursor conformation of the P4a-1 segment leading to the massive conformational rotation of P4a-1 domain II (Fig. 5, steps 2 and 3). We cannot distinguish the temporal order of trimerization and conformational inversion without evidence that either is dependent on the other: Models for both P4a-1+2 monomer and trimer showed both pre- and post-rotation conformations (Fig. 5, step2-lower or step3-upper, Fig. S10 and S2), and both conformations were able to trimerize (Fig. 5, step2-upper or step3-lower, Fig. S10). Perhaps the two events occur simultaneously. After conformational inversion, P4a-1 domains I and II pack together (Fig. 1B) and become conformationally locked by disulfide bond formation between Cys31 and Cys569 (Fig. S7). In addition, P4a-1 domain I and II helices now pack together allowing new interactions between them (Fig. 1B). The P4a-2 segment is now isolated at the end of the molecule (Fig. 5, steps 2/3). Since a trimer could be modeled for not only fully processed P4a-1 but also the P4a-1+2 intermediate (Fig. 5, before/after step 4) albeit with a lower confidence PAE, excision of the P4a-2 segment is not an absolute prerequisite for P4a-1 trimerization. P4a-2 excision is, however, a prerequisite for higher-order trimer assembly: after conformational inversion in the P4a1+2 trimer (Fig. 5, step3-upper) removes one steric block to higher-order assembly (Fig. 5, step 3), a steric block remains after step three due to presence of the P4a-2 segment on the outside surfaces of P4a-1+2 trimer subunits coincident with sites of higher-order assembly. In step 4 (Fig. 5), release of the P4a-2 segment via processing at

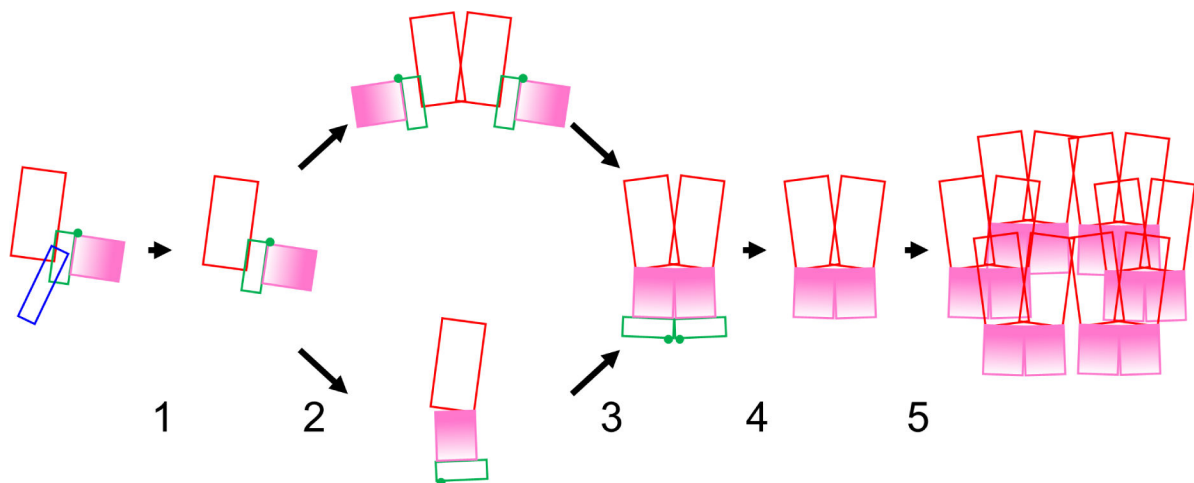


FIG 5 Suggested five-step pathway of P4a processing, conformational change and assembly. Coloration (as in Fig. 1A): P4a-1 domain I red, P4a-1 domain II pink, P4a-2 green, and P4a-3 blue. P4a-1 domain II is shown with gradient fill to emphasize its conformational inversion at steps 2/3. Green spot: P4a-1|two processing site. In step 1, P4a-3 removal permits both conformational inversion and trimerization of the P4a-1+2 intermediate (steps 2 and 3). The order of steps 2 and 3 is undefined hence the upper and lower loops of the pathway. In step 4, P4a-1|two processing permits higher-order P4a trimer assembly (step 5). For simplicity, only two of the three trimer subunits are shown in multimer schematics. For details see text.

the 1|2 AG| site allows higher-order assembly into a trimer-hexamer lattice (Fig. 5, step 5). The overall pathway (Fig. 5) is animated in Movie S2.

In studies with wild-type MV in the infected cell (19), anti-P4a-2 antibody could detect the P4a-1+2 intermediate but not P4a-2+3, suggesting that processing to remove P4a-3 precedes processing to split P4a-2 from P4a-1. Our pathway reflects this and now can be rationalized as the deferral of higher-order assembly until after trimerization has occurred. A model in which this order of processing events is regulated sterically (steric occlusion of the 1|two processing site until after 2|three processing and the resulting conformational inversion) is not viable since both P4a processing sites appear to be solvent exposed in all conformers and intermediates (Fig. 1). We therefore suggest an alternative model in which the three 1|two processing sites of a trimer must be brought into spatial proximity (at a point in space beneath the trimer's center; Fig. 5, step 3) for 1|two processing to occur. This would be the case if, for example, the processing protease is spatially restricted instead of freely diffusible, and would then defer P4a-2 removal (the last block to higher-order trimer assembly) to the final portion of the pathway.

MATERIALS AND METHODS

Protein structure prediction, *de novo*

Monomer and multimer structure predictions for vaccinia protein P4a (precursor and processing products) were run on local installations of AlphaFold2 and AlphaFold-multimer, using a non-docker setup (https://github.com/kalininalab/alphafold_non_docker). Models shown here represent the top-ranked prediction for each protein or protein complex. Side chains were present in models but not rendered except where such rendering was chosen. For generating 20 models from AF2 (e.g., Fig. S3) where it would otherwise produce only produce five models AlphaFold2 was run four times and the resulting models were pooled, or AlphaFold-multimer was run for the target protein. For each proteoform (P4a precursor, P4a-1+2 intermediate, mature P4a-1, and P4a-1 trimer), the top-ranked structure was compared individually against the other 19 models using ChimeraX MatchMaker. The RMSD (Å) between all-carbon atom pairs, per comparison, was reported as a data point, with 19 total data points for each model set.

Structure predictions through RoseTTAFold were run on the Robetta server (<https://robeta.bakerlab.org/>) using standard parameters.

Scoring of structural models

Average pLDDT scores for structures predicted with AlphaFold2 were obtained by averaging pLDDT values for all-carbon atoms (representing all atoms) from the top predicted model for each structure (in many cases, some top models were either identically or almost identically scored).

Visualization

Reported structures were visualized and measured, and images were generated using UCSF ChimeraX (45, 46). PAE plots of AlphaFold-multimer structures were visualized by the "AlphaFold Error Plot" tool provided by ChimeraX.

Cross-linking mass spectrometry

Virion proteins were cross-linked, prepared for XL-MS, and analyzed as described in greater detail in reference (6). Briefly, bis(succinimidyl) penta(ethylene glycol) (BSPEG5), bis(succinimidyl) nona(ethylene glycol) (BSPEG9), adipic acid dihydrazide (ADH), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC), and n-hydroxysuccinimide (NHS) were purchased from ThermoFisher, Inc. Isotopically coded disuccinimidyl suberate (DSS), bis(sulfosuccinimidyl)suberate (BS3), disuccinimidyl glutarate (DSG), and disuccinimidyl adipate (DSA) were purchased from Creative Molecules. 3,5-bis(((2,5-dioxopyrrolidin-1-yl)oxy) carbonyl)phenyl)phosphonic acid (PhoX) was provided by the

Scheltema lab and later purchased from Bruker and prepared as described in reference (45). Vaccinia virus was grown in HeLa S3 cells (ATCC CCL-2.2) and purified over a 36% sucrose cushion, followed by two 24–40% sucrose gradients. The harvested virus was washed with 100 mM triethylammonium bicarbonate buffer (TEAB) pH 8.5 and cross-linked as intact virions or after brief uncoating treatment (0.05% NP40, 40 mM TCEP, 100 mM TEAB, pH 8.5). Appropriate XL concentrations were determined by SDS-PAGE. XL reactions were quenched with ammonium bicarbonate or removed by spin desalting into ammonium bicarbonate.

Cross-linked virus samples were disaggregated by various methods, including an in-lab modified FASP protocol and urea denaturation. Cleavage of solubilized proteins into peptides employed various reagents/reagent combinations: Trypsin, ArgC, GluC, AspN, LysC, Trypsin+LysC, Trypsin+GluC, Trypsin+AspN, ArgC+AspN, ArgC+GluC, AspN+GluC, or CNBr+Trypsin. All cleaved samples were acidified with formic acid (FA) to 2% FA final concentration and desalted by C18/SCX as described (47). Peptides were eluted with 5% NH₄OH, 80% CH₂CN, 0.1% FA (Buffer X) or with a six-step ammonium acetate gradient in 20% CH₃CH, 0.5% FA, followed by a final elution with Buffer X. Samples were dried under vacuum and reconstituted in 0.1% FA in water for MS. The nanoLC-MS/MS method used is described in reference (6).

Instrument raw files were converted to mgf or mzML using MSConvert by ProteoWizard. Cross-links were identified by the programs described in reference (6) with the addition of pLink2, Kojak2, MetaMorpheus, and XlinkX. The in-house code described in reference (6), with substantial upgrades including for large data sets, was used to consolidate and assess XLMS data and calculate CSM counts. The latter represents total CSMs for a particular cross-link or set of cross-links.

Cross-link networks were rendered using CrosslinkViewer (48). Solvent-accessible surface distances between cross-linked residues were calculated using Topolink (49). Models of cross-linked pairs with through-space (Euclidean) distances indicated were rendered using ChimeraX.

Feature size measurement (EM)

Spike lengths and diameters were measured using the “tape” feature of ChimeraX: images from publications by these groups were saved as TIFF files, opened in ChimeraX, and spike lengths and diameters were measured with the ChimeraX Tape feature. The provided scale bars were used to convert the ChimeraX measurement values to Å.

ACKNOWLEDGMENTS

The authors thank Michael Hoopmann (ISB, Seattle) for advice on use of the “Kojak2” search engine.

This study was supported by the National Institutes of Health grant 1R01GM134144-01 (P.D.G.) and National Institutes of Health instrumentation grant S10 OD016328 (P.D.G.).

Conceptualization: Y.M., P.D.G.; Methodology: Y.M., A.J., R.S., P.D.G.; Investigation: Y.M., A.J., R.S., P.D.G.; Formal Analysis: Y.M., P.D.G.; Visualization: Y.M., P.D.G.; Funding acquisition: P.D.G.; Project administration: P.D.G.; Supervision: R.S., P.D.G.; Writing—original draft: Y.M., P.D.G.; Writing—review and editing: Y.M., P.D.G.

The authors declare that they have no competing interests.

AUTHOR AFFILIATIONS

¹Department of Molecular Biology and Biochemistry, University of California Irvine, Irvine, California, USA

²Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, the Netherlands

AUTHOR ORCID*s*

Yeva Mirzakhanyan  <http://orcid.org/0000-0002-4276-8937>

Paul David Gershon  <http://orcid.org/0000-0002-4904-9193>

FUNDING

Funder	Grant(s)	Author(s)
HHS National Institutes of Health (NIH)	1R01GM134144-01	Paul David Gershon

AUTHOR CONTRIBUTIONS

Yeva Mirzakhanyan, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing | Andris Jankevics, Investigation, Methodology | Richard A. Scheltama, Investigation, Methodology, Supervision | Paul David Gershon, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

All structural model coordinates and FASTA files are available on [DRYAD](#). All additional materials are available from the authors by reasonable request.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Document S1 (mbio01135-23 rev Supplementary Document 1.docx). Calculation of P4a-1 coverage of virion core wall surface.

Supplemental figures (mbio01135-23 rev Supplementary File 2.docx). Figures S1 through S14.

Movie S1 (mBio01135-23-s0003.mp4). Proposed conformation change in P4a during AG| processing to the three known products.

Movie S2 (mBio01135-23-s0004.mp4). Proposed overall pathway of P4a processing, conformation change, and assembly.

REFERENCES

- Moss B. 2013. Poxviridae, p 2129–2159. In *Fields virology*, 6th ed. Wolters Kluwer | Lippincott, Williams & Wilkins, Philadelphia.
- Malkin AJ, McPherson A, Gershon PD. 2003. Structure of intracellular mature vaccinia virus visualized by *in situ* atomic force microscopy. *J Virol* 77:6332–6340. <https://doi.org/10.1128/jvi.77.11.6332-6340.2003>
- Kuznetsov Y, Gershon PD, McPherson A. 2008. Atomic force microscopy investigation of vaccinia virus structure. *J Virol* 82:7551–7566. <https://doi.org/10.1128/JVI.00016-08>
- Schmidt FI, Bleck CKE, Reh L, Novy K, Wollscheid B, Helenius A, Stahlberg H, Mercer J. 2013. Vaccinia virus entry is followed by core activation and proteasome-mediated release of the immunomodulatory effector VH1 from lateral bodies. *Cell Rep* 4:464–476. <https://doi.org/10.1016/j.celrep.2013.06.028>
- Pedersen K, Snijder EJ, Schleich S, Roos N, Griffiths G, Locker JK. 2000. Characterization of vaccinia virus intracellular cores: implications for viral uncoating and core structure. *J Virol* 74:3525–3536. <https://doi.org/10.1128/jvi.74.8.3525-3536.2000>
- Mirzakhanyan Y, Gershon P. 2019. The vaccinia virion: filling the gap between atomic and ultrastructure. *PLoS Pathog* 15:e1007508. <https://doi.org/10.1371/journal.ppat.1007508>
- Jesus DM, Moussatche N, McFadden BBD, Nielsen CP, D'Costa SM, Condit RC. 2015. Vaccinia virus protein A3 is required for the production of normal immature virions and for the encapsidation of the nucleocapsid protein L4. *Virology* 481:1–12. <https://doi.org/10.1016/j.virol.2015.02.020>
- Risco C, Rodríguez JR, Demkowicz W, Heljasvaara R, Carrascosa JL, Esteban M, Rodríguez D. 1999. The vaccinia virus 39-kDa protein forms a stable complex with the p4a/4a major core protein early in morphogenesis. *Virology* 265:375–386. <https://doi.org/10.1006/viro.1999.0046>
- VanSlyke JK, Franke CA, Hruby DE. 1991. Proteolytic maturation of vaccinia virus core proteins: identification of a conserved motif at the N termini of the 4b and 25K virion proteins. *J Gen Virol* 72 (Pt 2):411–416. <https://doi.org/10.1099/0022-1317-72-2-411>
- Heljasvaara R, Rodríguez D, Risco C, Carrascosa JL, Esteban M, Rodríguez JR. 2001. The major core protein p4a (A10L gene) of vaccinia virus is essential for correct assembly of viral DNA into the nucleoprotein complex to form immature viral particle. *J Virol* 75:5778–5795. <https://doi.org/10.1128/JVI.75.13.5778-5795.2001>
- Lee P, Hruby DE. 1995. Analysis of the role of the amino-terminal peptide of vaccinia virus structural protein precursors during proteolytic processing. *Virology* 207:229–233. <https://doi.org/10.1006/viro.1995.1069>

12. Rodriguez D, Bárcena M, Möbius W, Schleich S, Esteban M, Geerts WJC, Koster AJ, Griffiths G, Locker JK. 2006. A vaccinia virus lacking A10L: viral core proteins accumulate on structures derived from the endoplasmic reticulum. *Cell Microbiol* 8:427–437. <https://doi.org/10.1111/j.1462-5822.2005.00632.x>
13. Williams O, Wolffe EJ, Weisberg AS, Merchinsky M. 1999. Vaccinia virus WR gene A5L is required for morphogenesis of mature virions. *J Virol* 73:4590–4599. <https://doi.org/10.1128/JVI.73.6.4590-4599.1999>
14. Wilcock D, Smith GL. 1996. Vaccinia virions lacking core protein VP8 are deficient in early transcription. *J Virol* 70:934–943. <https://doi.org/10.1128/JVI.70.2.934-943.1996>
15. Wilcock D, Smith GL. 1994. Vaccinia virus core protein VP8 is required for virus infectivity, but not for core protein processing or for INV and EEV formation. *Virology* 202:294–304. <https://doi.org/10.1006/viro.1994.1346>
16. Jesus DM, Moussatche N, Condit RC. 2014. Vaccinia virus mutations in the L4R gene encoding a virion structural protein produce abnormal mature particles lacking a nucleocapsid. *J Virol* 88:14017–14029. <https://doi.org/10.1128/JVI.02126-14>
17. Yang SJ, Hruby DE. 2007. Vaccinia virus A12L protein and its AG/A proteolysis play an important role in viral morphogenic transition. *Virol J* 4:73. <https://doi.org/10.1186/1743-422X-4-73>
18. Vanslyke JK, Whitehead SS, Wilson EM, Hruby DE. 1991. The multi-step proteolytic maturation pathway utilized by vaccinia virus P4a protein: a degenerate conserved cleavage motif within core proteins. *Virology* 183:467–478. [https://doi.org/10.1016/0042-6822\(91\)90976-i](https://doi.org/10.1016/0042-6822(91)90976-i)
19. Whitehead SS, Bersani NA, Hruby DE. 1995. Physical and molecular genetic analysis of the multistep proteolytic maturation pathway utilized by vaccinia virus P4a protein. *J Gen Virol* 76 (Pt 3):717–721. <https://doi.org/10.1099/0022-1317-76-3-717>
20. Ansaiah-Sobrinho C, Moss B. 2004. Role of the I7 protein in proteolytic processing of vaccinia virus membrane and core components. *J Virol* 78:6335–6343. <https://doi.org/10.1128/JVI.78.12.6335-6343.2004>
21. Lee P, Hruby DE. 1994. Proteolytic cleavage of vaccinia virus virion proteins. Mutational analysis of the specificity determinants. *J Biol Chem* 269:8616–8622. [https://doi.org/10.1016/S0021-9258\(17\)37239-3](https://doi.org/10.1016/S0021-9258(17)37239-3)
22. Yang SJ. 2007. Characterization of vaccinia virus A12L protein proteolysis and its participation in virus assembly. *Virol J* 4:78. <https://doi.org/10.1186/1743-422X-4-78>
23. Sarov I, Joklik WK. 1972. Characterization of intermediates in the uncoating of vaccinia virus DNA. *Virology* 50:593–602. [https://doi.org/10.1016/0042-6822\(72\)90410-2](https://doi.org/10.1016/0042-6822(72)90410-2)
24. Kane EM, Shuman S. 1993. Vaccinia virus morphogenesis is blocked by a temperature-sensitive mutation in the I7 gene that encodes a virion component. *J Virol* 67:2689–2698. <https://doi.org/10.1128/JVI.67.5.2689-2698.1993>
25. Katz E, Moss B. 1970. Formation of a vaccinia virus structural polypeptide from a higher molecular weight precursor: inhibition by rifampicin. *Proc Natl Acad Sci U S A* 66:677–684. <https://doi.org/10.1073/pnas.66.3.677>
26. Moss B, Rosenblum EN. 1973. Letter: protein cleavage and poxvirus morphogenesis: tryptic peptide analysis of core precursors accumulated by blocking assembly with rifampicin. *J Mol Biol* 81:267–269. [https://doi.org/10.1016/0022-2836\(73\)90195-2](https://doi.org/10.1016/0022-2836(73)90195-2)
27. Miner JN, Hruby DE. 1989. Rifampicin prevents virosome localization of L65, an essential vaccinia virus polypeptide. *Virology* 170:227–237. [https://doi.org/10.1016/0042-6822\(89\)90370-x](https://doi.org/10.1016/0042-6822(89)90370-x)
28. Villarreal EC, Roseman NA, Hruby DE. 1984. Isolation of vaccinia virus mutants capable of replicating independently of the host cell nucleus. *J Virol* 51:359–366. <https://doi.org/10.1128/JVI.51.2.359-366.1984>
29. Child SJ, Franke CA, Hruby DE. 1988. Inhibition of vaccinia virus replication by nicotinamide: evidence for ADP-ribosylation of viral proteins. *Virus Res* 9:119–132. [https://doi.org/10.1016/0168-1702\(88\)90027-5](https://doi.org/10.1016/0168-1702(88)90027-5)
30. Senkevich TG, White CL, Weisberg A, Granek JA, Wolffe EJ, Koonin EV, Moss B. 2002. Expression of the vaccinia virus A2.5L redox protein is required for virion morphogenesis. *Virology* 300:296–303. <https://doi.org/10.1006/viro.2002.1608>
31. White CL, Senkevich TG, Moss B. 2002. Vaccinia virus G4L glutaredoxin is an essential intermediate of a cytoplasmic disulfide bond pathway required for virion assembly. *J Virol* 76:467–472. <https://doi.org/10.1128/jvi.76.2.467-472.2002>
32. White CL, Weisberg AS, Moss B. 2000. A glutaredoxin, encoded by the G4L gene of vaccinia virus, is essential for virion morphogenesis. *J Virol* 74:9175–9183. <https://doi.org/10.1128/jvi.74.19.9175-9183.2000>
33. Senkevich TG, White CL, Koonin EV, Moss B. 2002. Complete pathway for protein disulfide bond formation encoded by poxviruses. *Proc Natl Acad Sci U S A* 99:6667–6672. <https://doi.org/10.1073/pnas.062163799>
34. Moussatche N, Condit RC. 2015. Fine structure of the vaccinia virion determined by controlled degradation and immunolocalization. *Virology* 475:204–218. <https://doi.org/10.1016/j.virol.2014.11.020>
35. Westwood JC, Harris WJ, Zwartouw HT, Titmuss DH, Appleyard G. 1964. Studies on the structure of vaccinia virus. *J Gen Microbiol* 34:67–78. <https://doi.org/10.1099/00221287-34-1-67>
36. Easterbrook KB. 1966. Controlled degradation of vaccinia virions *in vitro*: an electron microscopic study. *J Ultrastruct Res* 14:484–496. [https://doi.org/10.1016/s0022-5320\(66\)80077-1](https://doi.org/10.1016/s0022-5320(66)80077-1)
37. Stern W, Dales S. 1976. Biogenesis of vaccinia: isolation and characterization of a surface component that elicits antibody suppressing infectivity and cell-cell fusion. *Virology* 75:232–241. [https://doi.org/10.1016/0042-6822\(76\)90022-2](https://doi.org/10.1016/0042-6822(76)90022-2)
38. Dubochet J, Adrian M, Richter K, Garces J, Wittek R. 1994. Structure of intracellular mature vaccinia virus observed by cryoelectron microscopy. *J Virol* 68:1935–1941. <https://doi.org/10.1128/JVI.68.3.1935-1941.1994>
39. Roos N, Cyrklaff M, Cudmore S, Blasco R, Krijnse-Locker J, Griffiths G. 1996. A novel Immunogold Cryoelectron microscopic approach to investigate the structure of the intracellular and extracellular forms of vaccinia virus. *EMBO J* 15:2343–2355. <https://doi.org/10.1002/j.1460-2075.1996.tb00590.x>
40. Hernandez-Gonzalez M, Calcraft T, Nans A, Rosenthal PB, Way M. 2023. A succession of two viral lattices drives vaccinia virus assembly. *PLoS Biol* 21:e3002005. <https://doi.org/10.1371/journal.pbio.3002005>
41. Cyrklaff M, Risco C, Fernández JJ, Jiménez MV, Estéban M, Baumeister W, Carrascosa JL. 2005. Cryo-electron tomography of vaccinia virus. *Proc Natl Acad Sci U S A* 102:2772–2777. <https://doi.org/10.1073/pnas.0409825102>
42. Hyun J, Matsunami H, Kim TG, Wolf M. 2022. Assembly mechanism of the pleomorphic immature poxvirus scaffold. *Nat Commun* 13:1704. <https://doi.org/10.1038/s41467-022-29305-5>
43. Heuser J. 2005. “Deep-etch EM reveals that the early poxvirus envelope is a single membrane bilayer stabilized by a geodetic “honeycomb” surface coat”. *J Cell Biol* 169:269–283. <https://doi.org/10.1083/jcb.200412169>
44. Grimm C, Hillen HS, Bedenk K, Bartuli J, Neyer S, Zhang Q, Hüttenhofer A, Erlacher M, Dienemann C, Schlosser A, Urlaub H, Böttcher B, Szalay AA, Cramer P, Fischer U. 2019. Structural basis of poxvirus transcription: vaccinia RNA polymerase complexes. *Cell* 179:1537–1550. <https://doi.org/10.1016/j.cell.2019.11.024>
45. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE. 2018. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci* 27:14–25. <https://doi.org/10.1002/pro.3235>
46. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2021. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci* 30:70–82. <https://doi.org/10.1002/pro.3943>
47. Rappsilber J, Mann M, Ishihama Y. 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2:1896–1906. <https://doi.org/10.1038/nprot.2007.261>
48. Combe CW, Fischer L, Rappsilber J. 2015. xiNET: cross-link network maps with residue resolution. *Mol Cell Proteomics* 14:1137–1147. <https://doi.org/10.1074/mcp.O114.042259>
49. Ferrari AJR, Clasen MA, Kurt L, Carvalho PC, Gozzo FC, Martínez L. 2019. Topolink: evaluation of structural models using chemical crosslinking distance constraints. *Bioinformatics* 35:3169–3170. <https://doi.org/10.1093/bioinformatics/btz014>