Smallpox caused by the poxvirus variola virus is a highly lethal disease that marked human history and was eradicated in 1979 thanks to a worldwide mass vaccination campaign. This virus remains a significant threat for public health due to its potential use as a bioterrorism agent and requires further development of antiviral drugs. The viral genome replication machinery appears to be an ideal target, although very little is known about its structure. Vaccinia virus is the prototypic virus of the Orthopoxvirus genus and shares more than 97% amino acid sequence identity with variola virus. Here we studied four essential viral proteins of the replication machinery: the DNA polymerase E9, the processivity factor A20, the uracil-DNA glycosylase D4, and the helicase-primase D5. We present the recombinant expression and biochemical and biophysical characterizations of these proteins and the complexes they form. We show that the A20D4 polymerase cofactor binds to E9 with high affinity, leading to the formation of the A20D4E9 holoenzyme. Small-angle X-ray scattering yielded envelopes for E9, A20D4, and A20D4E9. They showed the elongated shape of the A20D4 cofactor, leading to a 150-Å separation between the polymerase active site of E9 and the DNA-binding site of D4. Electron microscopy showed a 6-fold rotational symmetry of the helicase-primase D5, as observed for other SF3 helicases. These results favor a rolling-circle mechanism of vaccinia virus genome replication similar to the one suggested for tailed bacteriophages.

Poxviruses are a family of large, complex, double-stranded DNA viruses that replicate exclusively in the cytoplasm of the infected cell. Although they infect a wide host range, only 2 viruses from different genera can circulate within the human population, the Molluscipoxvirus molluscum contagiosum virus and the Orthopoxvirus variola virus (VARV). Other poxviruses can be transmitted from animals to humans as zoonoses (1, 2). VARV was the causative agent of smallpox, a highly lethal disease that marked human history. In 1979, the world was declared to be free of smallpox by the World Health Organization, thanks to the mass vaccination campaign initiated nearly 2 centuries previously by prophylactic inoculations with cowpox virus (CPXV) and later with vaccinia virus (VACV) (3). VACV is the prototypic virus of the Orthopoxvirus genus and shows over 97% amino acid sequence identity with VARV (4) and more than 98% identity for the proteins considered in this article. Thus, VACV can be used as a model system for drug development and the study of all aspects of virus replication. Since vaccination was halted in the 1980s, a significant percentage of the world population is now immunologically naïve. Hence, orthopoxviruses remain a potential public health threat due to their potential use as bioterrorism agents or in case of a spread of animal-borne poxviruses, such as monkeypox virus, to humans (5, 6). Vaccination against smallpox may result in a variety of complications, particularly in immunologically depressed patients (7, 8). The development of potent antiviral therapies is required as a first line of defense in the event of human exposure to poxviruses. Cidofovir, an acyclic cytosine phosphonate analogue used for cytomegalovirus infections, targets the VACV DNA polymerase E9 and inhibits orthopoxvirus DNA replication by acting as a chain terminator. Cidofovir is recommended for short-term prophylaxis and emergency treatment (9, 10), despite its low bioavailability, which requires intravenous administration, and its nephrotoxicity. As a consequence, reduced side effects and higher bioavailability are the main goals for further drug developments, which have led, for example, to the cidofovir-derived prodrug CMX001 (11). Drugs targeting other pathways are also under development, such as ST-246, which inhibits virion assembly and dissemination (12).

Poxviruses form large and complex virions that contain the enzymes required for mRNA synthesis together with the genome. The genome is a single circular DNA molecule best described as a linear double-stranded DNA of 130 to 300 kbp terminated at each end with hairpin structures. VACV has a genome of approximately 190 kbp and can express potentially more than 200 proteins. Replication occurs within discrete foci in the cytoplasm, termed viral factories, and can even occur in enucleated cells (13, 14). Self-priming, primer-dependent, and recombination models of VACV genome replication have been proposed and were reviewed by Moss and Silva (15). One model takes into account the hairpins at the ends of the genome (16) and resembles the rolling hairpin strand displacement mechanism proposed for paroviruses (17, 18). The presence of transient concatamer intermediates was demonstrated by the presence of junction fragments and high-molecular-weight DNA (18–20). The evidence of nick formation near the ends of the genome (21, 22) and the resolution of concatemeric forms of DNA by a virus-encoded Holliday junction endonuclease (23, 24) are compatible with this rolling hairpin mechanism. However, there are reports of short nascent VACV DNA segments resembling Okasaki fragments that could be
chased into larger molecules \textit{in vivo} (25). The recent identification of a DNA primase encoded by VACV (26) suggests the use of RNA primers for the initiation of lagging-strand synthesis (27). In addition, both minichromosomes with specific telomere ends (28) and closed circular DNA molecules without poxvirus-specific sequences replicate efficiently in infected cells (29, 30), suggesting an origin-independent mechanism.

The mapping of complementation groups of VACV temperature-sensitive mutants, which express early viral proteins and are impaired in DNA synthesis, as well as yeast two-hybrid experiments (31–33) allowed the identification of replication-involved early expressed proteins, including E9 (replicative DNA polymerase), A20 (stochiometric component of the processivity factor), D5 (helicase-primase), B1 (Ser/Thr protein kinase, [34]), and D4 (uracil-DNA glycosylase, or UNG [35]). These proteins are essential for VACV DNA replication, are highly conserved within the \textit{Poxviridae}, and the genes coding for at least three of them (DNA polymerase, helicase-primase, and UNG) are present in all sequenced poxviruses genomes.

Here, we studied the different complexes involving E9, A20, D4, and D5 proteins. E9 (117 kDa) contains both 3′-5′-proof-reading and DNA polymerase activities (36, 37). This protein is also required for the formation of joint molecules and for strand transfer reactions, suggesting a link between replication and recombination (38–40). The polymerase on its own has an intrinsic distributive activity, adding about 10 nucleotides per template-binding event. A processive form exists in infected cells, where E9 is found in complex with A20 and D4 (36).

A20 is a 49-kDa protein that interacts with viral proteins E9, D4, and D5 (41–43). It is assumed that it serves as a bridge between E9 and D4 (41, 43). It has no known catalytic motifs and displays no sequence similarities with previously studied proteins. The N-terminal 25 residues of A20 are necessary and sufficient for interaction with D5. The interaction with D5 involves amino acids 201 to 250 (41). The interaction with E9 has not been mapped.

D4 is a 25-kDa uracil-DNA glycosylase. Its three-dimensional structure was solved by X-ray crystallography (44). While the presence of D4 is required to produce infectious virions (45, 46), its UNG activity is not mandatory (as shown by experiments using catalytically inactive mutants [47]). It has been suggested that the intrinsic, processive DNA-scanning activity of D4 tethers the holoenzyme to the DNA template (48). D5 is a 96-kDa protein expressed early after infection with a nucleoside triphosphatase activity that appears, so far, to be independent of any common nucleic acids oligomers (49). Analysis of the amino acid sequence of D5 revealed an archaeo-eukaryotic primase domain (50) located N-terminally from a superfamily 3 (SF3) helicase domain (51, 52) (Fig. 1). Like many helicases, multimerization appears to be a prerequisite for the ATPase activity of D5 (53). So far, no helicase activity has been observed, although purified D5 is able to synthesize ribonucleotide oligomers on a single-strand DNA template \textit{in vitro}, confirming primase activity (54). Whereas D5 is not essential for the assembly of a processive DNA polymerase complex \textit{per se}, it is required for replication due to the double-stranded nature of the poxviral genome. The protein may be recruited to the replication fork via its interaction with the A20 protein (32).

We produced E9, D4, A20, and D5 in milligram quantities in insect cells. Low-resolution structural information for purified proteins and complexes was obtained by using small-angle X-ray scattering (SAXS) and electron microscopy. This allowed us to propose a first structural model of the replication complex of VACV.

**MATERIALS AND METHODS**

**Cloning and expression.** The sequences of E9L, A20R, D4R, and D5R were from the VACV strain Copenhagen (GenBank accession number M35027.1). E9L and D5R were cloned into the pFastBac HTa vector (Invitrogen). A20R and D4R were both cloned into the pFastBac Dual plasmid (Invitrogen). E9, D5, and D4 were fused to a N-terminal hexahistidine tag that can be cleaved with the TEV (tobacco etch virus) protease. D4 was also fused to a C-terminal (PKPQQFM) tag for \textit{in vitro} fluorescent labeling (55). DH10EMBacAY bacteria transformed with a bacmid carrying a yellow fluorescent protein reporter gene (36) and \textit{Spodoptera frugiperda} (SF21) cells were provided by the EF facility (EMBL Grenoble). SF21 cells were grown in suspension in SF900I-SFM medium ( Gibco) at 27°C following the facility protocols described in reference 57.

\textbf{E9 and A20D4 purification.} Pellets from insect cells were resuspended in 10 volumes of equilibration buffer (20 mM Tris-HCl [pH 7], 300 mM NaCl, 10 mM β-mercaptoethanol, 10 mM imidazole) with Complete protease inhibitor cocktail (Roche). Cells were lysed by sonication, and the lysate was clarified by centrifugation at 48,000 \times g for 20 min at 4°C. The lysate was loaded onto a nickel affinity column (HIS-select; Sigma) previously equilibrated with equilibration buffer. The column was washed with washing buffer (20 mM Tris-HCl [pH 7], 300 mM NaCl, 10 mM β-mercaptoethanol, 25 mM imidazole). Proteins were eluted with elution buffer (20 mM Tris-HCl [pH 7], 300 mM NaCl, 10 mM β-mercaptoethanol, 100 mM imidazole) and concentrated using Amicon centrifugal filter units (Millipore) prior to injection onto a Superdex 200 GL 10/300 column (GE Healthcare) equilibrated with gel filtration buffer (20 mM Tris-HCI [pH 7], 300 mM NaCl, 5 mM diethiothreitol [DTT]). Eluted peaks were analyzed by SDS-PAGE and stained with InstantBlue (Expedeon). The purified proteins were kept at 4°C.

\textbf{D5 purification.} D5 was purified under similar conditions as for E9 and A20D4, except that the buffers contained 10% glycerol and 150 mM NaCl. After affinity purification, D5 was not subjected to gel filtration, but the buffer was exchanged for 20 mM Tris-HCl (pH 7), 150 mM NaCl, 10% glycerol, 5 mM DTT, and 1 mM EDTA using Econo 10DG columns (Bio-Rad).

\textbf{SPR.} Surface plasmon resonance (SPR) analysis was conducted with a Biacore X instrument (GE Healthcare) and a CM5 sensor chip. Recombinant E9 was diluted at 0.2 nM in 10 mM sodium acetate (pH 5.5) for coupling to the sensor channel by direct aldehyde coupling following the manufacturer’s protocol (BIApplication handbook). The level of immobilized E9 was around 12,000 resonance units (RUs). All experiments were done at 25°C with a flow rate of 30 μl/min. Purified A20D4 complex was freshly diluted with running buffer (20 mM Tris-HCl [pH 7], 300 mM NaCl) to a final concentration between 50 and 150 nM. The response from channel 1 was baseline corrected by subtracting the response from chan-

![FIG 1 Schematic representation of SF3 helicases. Representation of D5, bacteriophage P4 alpha protein, simian virus 40 large T antigen, and bovine papillomavirus E1 protein, showing their domain organization. Zn-bd, zinc-binding motif; hel-Nt, helicase N-terminal extension; ori binding, origin of replication-binding domain; Zn, zinc finger.](http://jvi.asm.org/content/jvi/81/3/1680/F1)

---

1680 jvi.asm.org Journal of Virology
nel 2, which had been activated and blocked. Experimental data were treated with BIAevalulation 3.0 software. The surface was regenerated by applying pulses of 30 μl regeneration buffer (10 mM glycine [pH 3]), causing the dissociation of the A20D4 complex from E9. Up to 10 regeneration cycles could be done per coupling.

Electron microscopy. Recombinant D5 protein was applied to the clear side of carbon on a carbon-mica interface, washed in a glycerol-free buffer, and stained with 2% (wt/vol) uranyl acetate. Images were recorded under low-dose conditions with a JEOL 1200 EX II microscope at 100 kV and at a nominal ×40,000 magnification. Negatives were digitized on a Zeiss scanner (Photoscan TD) with a pixel size of 3.5 Å at the object scale. A data set of circa 29,000 subframes was semiautomatically selected with EMAN (58), CTF-corrected with CTFFIND3 (59) and Bsoft (60), and subjected to multivariate statistical analysis and classification with IMAGIC-5 (61).

SAXS experiments. (i) Data collection. SAXS experiments were carried out on beamline ID14-3 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, which allows particularly streamlined data collection and data processing (62). Ten successive frames of 10-s exposure times were recorded on a Pilatus 1 M detector (Dectris) in flow mode using 30 μl of sample. Bovine serum albumin (A7906; Sigma) solutions at 5.0 mg/ml and 3.8 mg/ml were measured as references. Up to five different concentrations of each protein sample (8.7, 2.8, 1.2, 0.6, and 0.3 mg/ml for E9; 1.2, 0.6, and 0.3 mg/ml for A20D4; 2, 1.3, 0.7, and 0.3 mg/ml for A20D4E9) were measured. Background scattering from the corresponding buffer was measured before and after each protein measurement and was subtracted using PRIMUS (63). All experiments were performed at 20°C.

(ii) Data processing. The ATSAS suite (64) was used for data processing. For each measurement, scattering curves were based on the average of 10 spectra, excluding outliers. The radius of gyration, \( R_g \), was derived by the Guinier approximation with data up to \( q \times R_g \) of <1.3. \( q \) is the scattering vector, \( q = (4\pi\lambda/\lambda)\sin(\theta/2) \), with \( \lambda \) being the wavelength and \( \theta \) the scattering angle. Radii of gyration calculated for different A20D4 and A20D4E9 concentrations displayed concentration dependence, indicating the presence of attractive interactions between the complexes. High-angle data of the highest concentration and low-angle data of the lowest concentration were combined to generate merged scattering curves with minimized interparticle effects. The high-resolution cutoff was determined from a Porod plot. Pair-distance distribution functions \( P(r) \) were computed with the program GNOM (65). They allowed the determination of the maximum dimension, \( D_{\text{max}} \), of the macromolecule and of an alternative value of \( R_g \) based on the entire scattering curve.

(iii) Ab initio modeling of the overall shapes. Low-resolution models were calculated from the experimental data with DAMMIF (66). Ten low-resolution models obtained from DAMMIF were compared using the program DAMAVER (67) in order to estimate the reproducibility of the \( ab \) initio shape calculation, which is described by the normalized spacial discrepancy index (NSD). An average model was built using DAMFILT (67).

Construction of the structural models. PyMOL (DeLano Scientific LLC) was used to extract the A20D4 low-resolution envelope from the (A20D4)2 model, to place molecular models in low-resolution envelopes and to manipulate structures. Envelopes were combined using COLLAGE (from the SITUS suite) (68). By using a BLAST search against the PDB (69), the best homologues of E9 with known crystal structures were identified. A sequence alignment of the four most representative hits was generated using Kalign (70) and manually modified using data from a structural model, to place molecular models in low-resolution envelopes (A20D4E9). A model of E9 based on the herpes simplex polymerase was built where stretches of residues with no equivalent in E9 were deleted. The relative positions of domains were adjusted to fit experimental SAXS data by using CRYOSOL (73). The model was positioned into the low-resolution envelopes of E9 and A20D4E9 by using the SUPCOMB (74) and Situs (68) programs.

DLS and calculations of the hydrodynamic radius. Dynamic light scattering (DLS) was performed at 20°C on a Proteinsolutions instrument (Wyatt Technology) on 20-μl samples of the protein solutions used for SAXS data collection. Data were analyzed with the Dynamics software (v5.26.41). Hydrodynamic radii were calculated from bead models obtained from \( ab \) initio modeling by using HYDROPOR (75).

RESULTS

Expression of A20D4, E9, and D5 using the baculovirus-insect cell system. The A20 protein on its own has never been successfully expressed and probably needs the presence of the D4 cofactor to be properly folded (42). We coexpressed A20 and D4 in the baculovirus-insect cell system using a pFastBacDual vector. E9 and D5 were expressed individually using the pFastBac H TA vector. After optimization of buffer conditions and purification protocol, we could produce these proteins in the milligram range, starting from 1 to 2 liters of insect cell culture (Fig. 2A). We checked that the recombinant D4, E9, and D5 proteins retained their respective uracil-DNA glycosylase, replication/recombination, and ATPase activities (data not shown).

With SDS-PAGE, purified A20D4 showed the expected 50- and 28-kDa bands for A20 and D4 (Fig. 2A, lanes 2 and 3) with a constant intensity ratio suggesting a 1:1 stoichiometry. The pres-
ence of D4 was confirmed by anti-His Western blotting (data not shown). The A20D4 complex appeared stable at high ionic strength, suggesting a hydrophobic nature for the interaction, as described previously (43, 48). A20D4 was only soluble in a narrow pH range (pH 7 to 8) and at high salt concentration (above 300 mM). The tendency to aggregate may be due to the absence of its partner protein E9. Whereas dilute preparations eluted as a 75-kDa protein from a gel filtration column (Fig. 2A, lane 3, and B), more-concentrated preparations showed a predominant peak at 150 kDa (Fig. 2A, lane 2, and B). This high-molecular-mass species seems to be composed of the same ratio of A20 and D4 as the small species (Fig. 2A, lanes 2 and 3), suggesting a dimer of heterodimers (A20D4), (theoretical mass of 156 kDa).

In the gel filtration step, E9 eluted as expected, at approximately 120 kDa (calculated mass of 117 kDa) (Fig. 2A, lane 5). Its identity was confirmed by an anti-His Western blotting and by mass spectrometry (data not shown). Dynamic light scattering experiments showed the monodisperse nature of recombinant E9. The hydrodynamic radius was 4.7 nm (Table 1), corresponding to an estimated molecular mass of 120 kDa, confirming that E9 is a monomeric protein.

Up to 2 mg of pure D5 migrating with a molecular mass of 95 kDa on SDS-PAGE (calculated mass of 96 kDa [Fig. 2A, lane 7]) was obtained after running a single nickel affinity column. The presence of glycerol in the buffer appeared to be required for solubility. As the purity of the sample after the first affinity column was satisfactory and most of the protein was lost in the gel filtration, this step was omitted for D5.

**D5 is a hexamer in solution.** Negative stain electron microscopy of purified D5 showed a population of compact particles with a diameter of about 13 nm and an inner hole with a diameter of 3 nm (Fig. 3). The eigenvector analysis of the images of particles after centering clearly indicated the predominance of a 6-fold rotational symmetry. A class average representing a top view of a ring-shaped object with a prominent handedness and 6-fold symmetry is shown as an inset in Fig. 3.

**A20D4 and E9 interact with high affinity to form a stable A20D4E9 complex in vitro.** We obtained the heterotrimeric complex by mixing purified A20D4 complex and E9 protein at room temperature. The resulting complex could be purified by size exclusion chromatography. A species of approximately 200 kDa composed of A20, D4, and E9 was observed (Fig. 4A and B). SDS-PAGE image analysis using ImageJ software (76) suggested a 1:1:1 stoichiometry (measured stoichiometric ratio between 1 and 1.3 for each band). In order to determine the affinity of the A20D4 complex for E9, we used SPR. E9 was covalently bound to the sensor chip, and different concentrations of A20D4 were injected. The association curves, shown in Fig. 4C, could be fitted with a single exponential, whereas dissociation curves were fitted with a double exponential. Residuals (Fig. 4D) show the adequacy of the models. Although the on-rates were strongly concentration dependent (see Table S1 in the supplemental material). It is likely that the concentration-dependent dimerization of A20D4 described above reduces the available concentration of E9-binding sites at higher concentrations, if we assume that the (A20D4)2 dimerization site overlaps with the E9-binding site on A20D4 (discussed below). The faster component of the biphasic dissociation with a constant proportion of about 30% (see Table S1) may indicate the presence of a less stable A20D4-E9 complex intermediate. By using the association rate determined at the lowest concentration of A20D4 (50 nM) and the slow dissociation rate of 3 nM was calculated, based on our assumption that all the A20D4 is heterodimeric at this concentration. This value corresponds to a very tight interaction between A20D4 and E9. However, in SPR experiments with immobilized A20D4 on Ni-nitrilotriacetic acid chips, we could not observe any interaction between

### Table 1 Parameters derived from SAXS and ab initio models

<table>
<thead>
<tr>
<th>Protein</th>
<th>Massa (kDa)</th>
<th>No. of residues</th>
<th>Rg (nm)</th>
<th>Dmax (nm)</th>
<th>χ2/d</th>
<th>NSDc</th>
<th>Rh (nm)</th>
<th>Modelf</th>
<th>DLSg</th>
</tr>
</thead>
<tbody>
<tr>
<td>E9</td>
<td>117</td>
<td>1,006</td>
<td>3.51</td>
<td>3.61</td>
<td>11</td>
<td>3.35</td>
<td>0.72</td>
<td>4.4</td>
<td>4.7</td>
</tr>
<tr>
<td>(A20D4)2</td>
<td>156</td>
<td>1,288</td>
<td>6.53</td>
<td>6.60</td>
<td>24</td>
<td>5.53</td>
<td>0.73</td>
<td>6.2</td>
<td>5.8</td>
</tr>
<tr>
<td>A20D4E9</td>
<td>191</td>
<td>1,650</td>
<td>7.03</td>
<td>7.12</td>
<td>23</td>
<td>4.52</td>
<td>0.85</td>
<td>6.6</td>
<td>6.5</td>
</tr>
</tbody>
</table>

*Massa, calculated from the amino acid sequence.
Rg, calculated from the averaged bead model.
Dmax, calculated from the pair distribution function.
χ2/d, Average discrepancy between the experimental scattering curve and the values calculated from the models obtained from DAMMIF.
NSDc, Average of the normalized spatial discrepancies (NSD), calculated with DAMAVER for the different models generated by DAMMIF. An NSD of <1 indicates similarity between individual models.
Rh, Hydrodynamic radius, calculated from the averaged bead model.
DLSg, Hydrodynamic radius, determined by DLS measurements.
D5 and A20D4 or A20D4E9 under the experimental conditions, which clearly showed the A20D4-E9 interaction (data not shown).

**Low-resolution structures of A20D4, E9, and the A20D4E9 complex by SAXS.** Overall dimensions of the different components of the replication complex were extracted (Table 1), and low-resolution models of the components were built (Fig. 5A, B, and C) using the ATSAS program suite (64). The model for E9 is the most reliable, based on the agreement of the individual envelopes with the scattering curve ($\chi^2$ in Table 1) and the similarity of individual reconstructions (normalized spatial discrepancies in Table 1).

The polymerase E9 displays a “half-avocado” shape, with a prominent central depression (Fig. 5A; see also Fig. S1A in the supplemental material). The reconstruction of A20D4 leads to a very elongated model. The position of the peak in the gel filtration chromatography step used for SAXS sample preparation and the concentrations used for SAXS speak in favor of the presence of a (A20D4), heterotetramer, so we imposed 2-fold symmetry during the final reconstruction (Fig. 5B; see also Fig. S1B). The heterotrimeric A20D4E9 complex has an elongated shape with a box-shaped extremity (Fig. 5C; see also Fig. S1C). The hydrodynamic radii calculated from the low-resolution models agree well with experimental radii determined from the DLS data (Table 1).

**DISCUSSION**

Using the baculovirus-insect cell expression system, we produced and purified the VACV DNA polymerase E9, the helicase-primase D5, and the complex formed by A20 and the uracil-DNA glycosylase D4. The work presented here allows us to propose a low-resolution structural model of the VACV DNA replication machinery.

DNA helicases are grouped in 6 superfamilies. Based on the sequence of RNA and DNA virus helicases, the VACV D5 was assigned to superfamily 3 (SF3) (52, 77). SF3 helicases are hexameric or double hexameric rings moving on single-stranded substrates in the 3'–to–5’ direction (78). Hexameric helicase rings have been observed for adeno-associated virus type 2 Rep40 protein (79), and simian virus 40 (SV40) large T antigen and papillomavirus E1 protein (80). All these members of the SF3 family carry additional origin of replication binding domains instead of primase domains, as found with D5 (Fig. 1), bacteriophage P4 α protein (81), and related proteins in iridoviruses and mimiviruses (50). It had been demonstrated that D5 needs multimerization for its activity (53), but the exact state of multimerization was still unproven. Based on electron microscopy image analysis, we have shown for the first time a clear hexameric organization for the helicase-primase D5, in agreement with its classification in the SF3 family. The diameter of the hexamer (13 nm) appears slightly larger than the one derived from the crystal structures of the helicase domains of SV40 T antigen (12.5 nm) or papillomavirus E1 (10.5 nm).

Based on the structure of the papillomavirus SF3 helicase–single-stranded DNA (ssDNA) complex (PDB 2GXA [80]), we suggest that the D5 ring-like helicase structure encircles the leading strand, excluding the lagging strand. The primase domains located at the N terminus (Fig. 1) would be positioned on the side of the ring where they could act on the lagging strand.

The other principal component of the replication machinery is the heterotrimeric complex formed by A20D4 and E9. We confirmed its 1:1:1 stoichiometry in vitro and showed strong binding ($K_d < 3$ nM) of E9 to the A20D4 subcomplex. The A20D4 interaction cannot be studied, as A20 alone cannot be produced and appears to be stable only in complex with D4. On the other hand, we were unable to identify an interaction between A20D4E9 (or A20D4) and D5. The interactions described in the literature (32, 41) may be transient or of low affinity.

In the absence of E9, the purified A20D4 recombinant complex showed a self-interaction leading to the predominant (A20D4)$_2$ form. At low concentrations, D4 on its own is a monomer in solution (48), whereas at high concentrations dimerization has been observed (44), which utilizes a hydrophobic patch on its surface. As the A20D4E9 holoenzyme does not show self-association even at high concentrations, this suggests that the hydrophobic patch of D4 implicated in concentration-dependent dimeriza-
tion is not exposed in the ternary complex. This implies that this surface on D4 is involved in A20 binding, as D4 does not bind to E9. It is very likely that the surface on A20 which binds to E9 in the ternary complex probably through hydrophobic interactions, as indicated by its stability at high salt concentrations (43), has a tendency to interact with itself leading to the formation of (A20D4)₂. As a consequence, A20 should be located in the center of the complex around the 2-fold axis.

Using SAXS, we built low-resolution envelopes of the (A20D4)₂ complex, the E9 polymerase, and the A20D4E9 complex. The envelope of the ternary complex can be interpreted using the high-resolution X-ray structure of D4 and the envelopes of (A20D4)₂ and E9. E9 is located in the bulky head of the A20D4E9 envelope, whereas the elongated handle must correspond to the A20D4 complex whose envelope was extracted manually from that of the (A20D4)₂ complex. It is possible to fit the E9 envelope and the A20D4 envelope into the A20D4E9 low-resolution envelope with a good overall complementarity of the shapes (Fig. 5D).

A20 was placed between E9 and D4, with no direct contact between E9 and D4. This agrees with the hypothesis that the dimer interface of the (A20D4)₂ complex corresponds to the E9-binding site, as well as with the literature reporting no E9-D4 interactions. A20 may act as a spacer separating the polymerase and proofreading activities of E9 from the DNA-binding and UNG activities of D4. The distance between these two active sites is approximately 150 Å, a distance large enough to accommodate 50 to 60 bp of double-helical DNA. It is possible that the DNA-binding activity of D4 tethers E9 more tightly to the DNA substrate while still allowing rapid migration along the DNA, as observed for UNGs.

As no crystal structure for E9 is available so far, we compared the VACV polymerase to other class II DNA polymerases in order to obtain a model for E9 which could be positioned inside the A20D4E9 envelope. All DNA polymerases retrieved from a BLAST search against the PDB had a similar overall structure (see Fig. S2 in the supplemental material); however, they appear to adopt either an open conformation (herpes simplex virus [HSV], PDB 2GV9, subunit A) or a closed conformation (Saccharomyces cerevisiae, PDB 3K59). By careful study of the sequence and the structure alignments, we found 4 insertions in E9 specific for poxviruses, compared to the other DNA polymerases (Fig. 6). The sequence of HSV polymerase is longer than most other polymerases of the same family (1,235 residues, compared to 773 to 1,097 residues). It has several extra domains: an N- and a C-terminal extension as well as 6 inserts within the sequence, some of which are not visible in the crystal structure. We based our model for E9 on the HSV polymerase because of its similarity to VACV E9, based on amino acid sequence and size (1,006 residues for E9). Some insertions of the HSV polymerase are also shared with E9.
FIG 6 Sequence and structure comparisons of E9 homologues. Alignment of *T. gorgonarius* (THEG, PDB 1TGO), *Sulfolobus solfataricus* (SULS, PDB 1S5J), *S. cerevisiae* (SACC, PDB 3IAY), HSV1 (PDB 2GV9), and VACV DNA polymerases sequences. Secondary structures based on the crystal structure or on the consensus of secondary structure predictions (for VACV) are shown in red ([H]9252-sheet) and blue ([H]9251-helix), and disordered residues (not visible in the crystallographic structures) are shown in lowercase letters. Specific exonuclease or polymerase motifs are underlined, and within the motifs, residues that are identical to the VACV E9 sequence are indicated by a gray background. Outside the motifs, strict conservation is indicated by a gray background. Yeast-, HSV-, and archaeon-specific inserts are highlighted with a yellow background, and VACV-specific inserts are highlighted with a green background. HSV and VACV inserts are numbered.
order to obtain a better fit with the SAXS data, we generated a “closed” version of the HSV polymerase by moving the palm and thumb domains toward the exonuclease domain and the N-terminal domain, as observed in the yeast and archaeal polymerase structures. We removed residues up to residue 167 of the HSV N-terminal extra domain, residues 321 to 332 (HSV insert 1), and 982 to 1006 (HSV insert 5) from the HSV structure. Finally, we modeled residues 639 to 703, which are missing in the HSV structure, based on the yeast and archaeal polymerases, where this part of the sequence is visible in the structure. The HSV inserts 3 (residues 640 to 699) and 6 (residues 1079 to 1135) as well as the C-terminal extension are disordered and, by consequence, absent from the structure. The modeling of the poxvirus-specific inserts was not possible, as there is no sequence similarity with known polymerases. Concerning the position of D4, the situation is less obvious, as there is no sequence similarity with known polymerases. It should be noted that the poxvirus inserts, comprising a total of 110 residues, all point toward the back or side of the polymerase, away from the DNA-binding pocket (Fig. 7A). This suggests a rather structural role for these insertions. The comparison of theoretical scattering curves for polymerases from either HSV, *Thermococcus gorgonarius*, or the model of VACV polymerase described above with the experimental curve showed that the VACV model fit best (Fig. 7B). When we fitted the VACV polymerase model into the E9 low-resolution envelope, we obtained a preferential orientation (Fig. 7A). There remains clearly enough unoccupied volume inside the envelope obtained from SAXS to accommodate the extra residues from the poxvirus-specific inserts (Fig. 7A). If we conserve this orientation of the polymerase in the model of the A20D4E9 complex (Fig. 5D), this positions the poxvirus-specific inserts 1 and 2 next to A20. Such an orientation would be different from the association of the HSV polymerase with its processivity factor, UL42, via its C-terminal region (82, 83).

The structure of D4 can be positioned at the opposite end of the A20D4E9 envelope, based on the absence of a direct D4–E9 contact and the role of A20 as a bridge, as suggested in recent experiments (43, 48). Since it has been shown that the A20–D4 interaction implicates the predominantly hydrophobic residues in the 25 N-terminal residues of A20 (48), we assume that similarly hydrophobic residues on D4 are required for the contact. Such a surface is only present at the level of the dimeric interface of D4, as observed in the crystal structures (44) (see Fig. S3 in the supplemental material). It involves hydrophobic residues (Ile166, Ala168, Leu170, Val174, IleValGly177–179, IleIle197–198, ValLeuLeu200–202, and Leu204), which are not shared with other UNGs. In addition, the only known D4 mutant deficient in its interaction with A20 carries a Gly179Arg mutation (43), which introduces a charged and bulky arginine residue into this region (see Fig. S4 in the supplemental material). These different arguments all support a model where the dimer interface observed in the D4 crystal is the most likely contact surface for A20. When we positioned D4 at the tip of the protruding arm of the A20D4E9 SAXS envelope with this interface oriented toward E9, we created a gap which had to be filled by A20. Furthermore, since the UNG activity of D4 is still functional within the replication complex (48) (even if this is not mandatory for processive replication), we suppose that the DNA-binding site of D4 remains exposed in the holoenzyme to bind and repair DNA. Based on these hypotheses, we propose an approximate orientation for the D4 protein within the A20D4E9 complex (Fig. 8A). The superposition of the D4 structure with the structure of human UNG in complex with DNA (84) allows us to position the DNA relative to D4 and, furthermore, to obtain a plausible position of D4 in the holoenzyme envelope (Fig. 8A; see also Fig. S5A in the supplemental material). In the same way as numerous type B polymerase structures are available in complex with DNA (RB69 bacteriophage, PDB 1IG9 [85]; *T. gorgonarius*, PDB 2VWJ [86]; *S. cerevisiae*, PDB 3IAY [87]) the position of the DNA next to the polymerase active site can be modeled. As the *S. cerevisiae* DNA polymerase (87) is part of our alignment, we used its structure in complex with DNA to position the DNA relative to the envelope of E9 and, indirectly, relative to the holoenzyme (Fig. 8A; see also Fig. S5B).

Starting from this model of the interaction of the holoenzyme with DNA, we analyzed four different alternatives concerning leading-strand replication (see Fig. S5 in the supplemental material) which differed in the position of D4 relative to the polymerase E9 and in the position of D5 on either the leading or the lagging strand. As D5 moves in the 3′→5′ direction, which has been inferred from the classification of D5 within the SF3 helicases, models where D5 is placed on the lagging strand (see Fig. S5A and D in the supplemental material) are excluded. As a consequence, the mechanism differs from the one proposed for tailed bacteriophages, which use SF4 helicases, moving in the opposite direction compared to SF3 helicases (88), and which thus act on the lagging strand.

Concerning the position of D4, the situation is less obvious, as D4 either can interact with the DNA ahead of the site of polymerase.
limitations to that study: the use of 70-base minicircles appears through abasic sites. However, there are some methodological prior to the encounter with E9, which is not able to progress template strand. This suggested that they may have been excised experiments that were rather in favor of a position of D4 ahead of the product in the presence of uracil observed by Boyle et al. (48) of uracil excision (89). The different migration of the reaction DNA, as shown in Fig. S5C and D of the supplemental material. However, the higher affinity for ssDNA before strand separation, as presented in Fig. S5A and B of the parental (black) and newly synthesized (pink) DNA is modeled. The orientation cannot yet be inferred with certainty. The components of the VACV replication machinery have been modeled using crystallographic structures whenever possible (except for A20), which have been fitted into the SAXS envelope. The association of D4 with DNA is based on the human UNG-DNA complex (84); the S. cerevisiae DNA polymerase δ (87) in complex with DNA has been used to model E9, and dsDNA has been built to link the two active sites. D5 was modeled using the simian virus 40 SF3 helicase domain (large T antigen, PDB 1SVM [90]) combined with 6 copies of the Sulfolobus islandicus primase (PDB IRNI [91]) linked to the N terminals of the simian virus 40 SF3 helicase. This model provides a first glimpse of the relative position and size of the partners involved in the poxviral replication mechanism (Fig. 8A).

The orientation of SF3 helicases on ssDNA places the primase domain ahead of the helicase domain, as shown in Fig. 8A. This implies that the primase domain acts in trans on the lagging strand, whereas the helicase encircles the leading strand. This mode of action has never been described for helicase-primases so far. It would also be expected for bacteriophage P4 α protein, which shows a similar domain organization (Fig. 1) and has proven helicase, primase, and Ori-binding activities (81).

In summary, from our study we favor a model for poxvirus replication in which D5 opens the replication fork, given that SF3 helicases often have origin-binding and replication-initiating activities. A first priming event will allow the polymerase complex to start the replication of the leading strand, whereas the primase activity of D5 will deposit primers on the lagging strand, which will be used for its replication. In this way, in a rolling-circle type of replication, concatemeric viral genomes will be obtained, which have to be processed further in order to excise RNA primers, to close the gaps between Okazaki fragments, and to process the concatemers into circular viral genomes. The required enzymatic activities are not clearly attributed yet among the various viral proteins with putative roles in replication. Despite the different positioning of the helicase, this model is similar to the one proposed for the replication of tailed bacteriophages (88). Two experiments would be major steps toward the confirmation of this model: the demonstration of the helicase action of D5, probably hampered so far by the lack of a loader protein, which has not yet been identified, and the clear demonstration of lagging strand synthesis.

The principal structural results obtained in this study are the following. The hexameric organization of the D5 helicase-primase and the elongated shape of the A20D4E9 holoenzyme with a considerable separation between polymerase and UNG active sites will allow the design of further experiments in order to verify the model presented above. This is facilitated by the availability of purified proteins of the replication machinery in milligram quan-

FIG 8 Model of the VACV replication fork. (A) Structural model showing sizes and positions of the different partners. The low-resolution envelope of the A20-D4-E9 complex (light pink) and the envelope corresponding to A20 (orange) are shown. Whenever possible, individual proteins have been modeled using crystal structures of the E9 homologue S. cerevisiae DNA polymerase (cyan), of D4 (pink), and a hexameric model of the SV40 helicase (light green), where 6 copies of S. islandicus primase (dark green) have been linked to the N termini of the helicase. Parental (black) and newly synthesized (pink) DNA is modeled. (B) Schematic showing the D5 (green), E9 (blue), A20 (orange), and D4 (purple) proteins. DNA is represented using the same colors as for panel A.
ties. Analysis of the D5-A20 interaction and the precise mapping of A20-binding sites on E9 and D4 will be of significant interest and may help to obtain high-resolution structures of the different components, which in turn can initiate the design of new antiviral drugs against variola virus.

ACKNOWLEDGMENTS

We thank Frédéric Garzoni from the Berger group at EMBL Grenoble for assistance and helpful discussions. We thank Petra Pernet, ESFR, for assistance in SAXS data collection and the ESFR for providing beamtime. We thank the EM platform of the Partnership for Structural Biology (Grenoble) for access to EM equipment.

C.S. was a Direction Générale de l’Armement and Centre National de la Recherche Scientifique PhD Fellow. The work was supported by the Fond d’intervention of Université Joseph Fourier, the Cluster Infectiologie of the Region Rhône-Alpes, the Service de Santé des Armées, the Délegation Générale pour l’Armement, and the Gen-Homme network.

REFERENCES

glycosylase interacts with the A20 protein to form a heterodimeric processivity factor for the viral DNA polymerase. J. Biol. Chem. 281:3439–3451.


