

Solution Structure of the C-terminal Domain of A20, the Missing Brick for the Characterization of the Interface between Vaccinia Virus DNA Polymerase and its Processivity Factor

Beate Bersch [1]*, Nicolas Tarbouriech [1], Wim P. Burmeister [1] and Frédéric Iseni [2]

*1 - **Institut de Biologie Structurale,** Univ. Grenoble Alpes, CNRS, CEA, IBS, F-38000 Grenoble, France*
*2 - **Unité de Virologie,** Institut de Recherche Biomédicale des Armées, BP73, F-91223 Brétigny-sur-Orge Cedex, France*

***Correspondence to Beate Bersch:** beate.bersch@ibs.fr (B. Bersch)*
https://doi.org/10.1016/j.jmb.2021.167009
Edited by M.F. Summers

Abstract

Poxviruses are enveloped viruses with a linear, double-stranded DNA genome. Viral DNA synthesis is achieved by a functional DNA polymerase holoenzyme composed of three essential proteins. For vaccinia virus (VACV) these are E9, the catalytic subunit, a family B DNA polymerase, and the heterodimeric processivity factor formed by D4 and A20. The A20 protein links D4 to the catalytic subunit. High-resolution structures have been obtained for the VACV D4 protein in complex with an N-terminal fragment of A20 as well as for E9. In addition, biochemical studies provided evidence that a poxvirus-specific insertion (insert 3) in E9 interacts with the C-terminal residues of A20. Here, we provide solution structures of two different VACV A20 C-terminal constructs containing residues 304–426, fused at their C-terminus to either a BAP (Biotin Acceptor Peptide)-tag or a short peptide containing the helix of E9 insert 3. Together with results from titration studies, these structures shed light on the molecular interface between the catalytic subunit and the processivity factor component A20. The interface comprises hydrophobic residues conserved within the *Chordopoxvirinae* subfamily. Finally, we constructed a HADDOCK model of the VACV $A20_{304-426}$-E9 complex, which is in excellent accordance with previous experimental data.

© **2021 Elsevier Ltd. All rights reserved.**

Introduction

Viruses belonging to the poxviridae family replicate in the cytoplasm of the infected cells. Viral genome synthesis takes place in structures called viral factories and depend almost exclusively on virally encoded-proteins.[1] To date, our current knowledge concerning the molecular mechanisms involved in the replication of the poxvirus genome remains incomplete. However, over the years, genetic and biochemical studies have made it possible to characterize a number of proteins essential for DNA synthesis. For vaccinia virus (VACV), the prototypic member of the Poxviridae, these are: E9, the catalytic subunit of the DNA polymerase; D4, a uracil-DNA glycosylase (UDG)[2] forming with A20 the heterodimeric processivity factor[3]; D5, a DNA-independent nucleoside triphosphatase which contains a putative helicase domain[4] and has primase activity,[5] and finally, I3, a single stranded DNA-binding protein.[6]

E9 is a member of the DNA polymerase family B[7] possessing DNA polymerase and 3′-5′ proofreading exonuclease activities.[8] Interestingly, it was also shown to catalyze annealing of single-stranded DNA,[9] an activity not found in other members of DNA polymerase family B. The end-joining reaction requires the 3′-5′ exonuclease activity of E9 that degrades the extremities of dsDNA to create 5′-ssDNA overhangs.[10] E9 on its own was shown to be distributive under physiological conditions, adding only few nucleotides per binding event.[11] However it becomes highly processive when bound to its cofactor A20-D4 to form the DNA polymerase

holoenzyme E9-A20-D4.[12] A20 links both enzymes, and the DNA binding properties of D4 are believed to increase the association of E9 with the genome template thus rendering the polymerase processive.[12]

A low-resolution structure of the E9-A20-D4 holoenzyme was obtained from small-angle X-ray scattering.[13] High-resolution information on the A20-D4 interface was provided by the crystal structure of the N-terminal fragment of A20 ($A20_{1-50}$) in complex with D4 and by the structure of D4-$A20_{1-50}$ bound to a 10-mer DNA duplex containing an abasic site resulting from the cleavage of an uracil base.[14,15] These data yielded initial information on processivity factor assembly and the coupling of DNA synthesis and base excision repair. The 2.7 Å resolution crystal structure of the full-length VACV DNA polymerase allowed the identification of E9-specific insertions.[13,16] While inserts 1 and 2 (aa 208–233 and 354–434 respectively) are not yet correlated to precise functions, we have shown that insert 3 (aa 567–617) located in the palm domain interacts with a C-terminal domain of A20 ($A20_{304-426}$) and thus serves as the processivity factor binding site.

In this report, we present the solution structure of the A20 C-terminal domain ($A20_{304-426}$). The surface of the domain shows a pocket surrounded by conserved hydrophobic residues. This pocket is shown to be part of the A20-E9 interface using the solution structure of a chimeric protein construct composed of $A20_{304-426}$ fused to the residues of the α-helix of insert 3 ($E9_{576-590}$). Based on this experimentally determined structure we propose a model of the complex between the C-terminal domain of A20 and the full-length E9 DNA polymerase.

Results

Expression of a soluble fragment of A20, $A20_{304-426}$-BAP

The identification of a soluble domain from A20 interacting with the E9 protein has been described previously.[16] The resulting construct, $A20_{304-426}$-BAP, contains the last 123 residues of the A20 protein with a C-terminal biotin acceptor peptide (BAP) tag, connected by a 10-residue linker (Figure S1). $A20_{304-426}$-BAP expresses as a soluble protein in *E. coli* BL21(DE3)-Star cells, yielding up to 30 mg of purified protein per liter of M9 minimal medium.

NMR study of $A20_{304-426}$-BAP

Except for the N-terminal 50 residues in complex with D4, there is only low-resolution structural information available for the VACV A20 protein.[13] 2D-^{1}H,^{15}N correlation spectra of the $A20_{304-426}$-BAP construct show well dispersed correlation peaks, indicating that $A20_{304-426}$-BAP is a folded protein domain (Figure S2). Using a series of 3D NMR experiments, a nearly complete resonance assignment could be obtained. Backbone chemical shifts were analyzed by the TALOS + program[17] yielding 218 backbone dihedral angle constraints for structure calculation. Distance restraints were obtained from four complementary 2D and 3D NOESY spectra, assigned during the ARIA[18] structure calculation process. ARIA derived 2351 unambiguous and 566 ambiguous distance restraints from the available experimental information. The final structural ensemble comprises 20 models obtained after water refinement using ARIA/CNS. During the automatic NOE assignment process, some NOEs were frequently assigned to contacts between the BAP-tag and residues of the A20 protein. Closer inspection of the NMR matrices, however, did not allow unambiguous assignment of the corresponding NOE cross-peaks to residues of the BAP-tag. These peaks were therefore discarded from the structure calculation, leaving the BAP-tag completely unconstrained. The final structural ensemble is shown in Figure 1 and statistics are resumed in Table S1. The well-defined structure is composed of five N-terminal α-helices, a two-stranded, antiparallel β-sheet and a long, C-terminal helix (helix 7) with a kink at the level of residues 405 and 406. The hydrophobic core contains a cluster of nine aromatic amino acids between the β-sheet and helices 4, 5 and 7 (F354, Y363, F368, F377, F387, Y404, F407, F410, F414). The Van-der-Waals surface of the protein shows a pocket surrounded by conserved hydrophobic side chains (residues F354, F377, F407, F410, F414, V372, I379, V384, V411) (Figure 2).

Data of ^{15}N-relaxation experiments were collected on the same protein construct (Figure S3) and consistently show that the N-terminal 10 residues (including the strictly conserved Y307) are flexible. This is probably a consequence of the truncation of the A20 construct. Interestingly, the C-terminal BAP-tag shows a net restriction in its mobility. BAP-tag residues numbered L_{BAP}11-17 are characterized by $\{^{1}H\}^{15}N$-hetNOE values in between 0.4 and 0.55, which is less than the values observed for residues in secondary structure elements (≈ 0.8) but clearly above the $\{^{1}H\}^{15}N$-hetNOEs observed for the residues in the linker (residues L_{BAP}3-7), which are around 0.2. In addition, L_{BAP}:F15 shows an increased transverse relaxation rate that probably results from chemical exchange. Slightly increased transverse relaxation rates (>17 s^{-1}) are also observed for 22 additional residues (Figures S3 and S4), most of which are in proximity of the C-terminus of the A20 domain and the binding pocket. This suggests unspecific and fuzzy interactions of the BAP-tag close to the binding pocket, where the high number of aromatic side chains can induce large chemical shift changes due to local ring current effects. Indeed, removal of the BAP-tag leads to protein

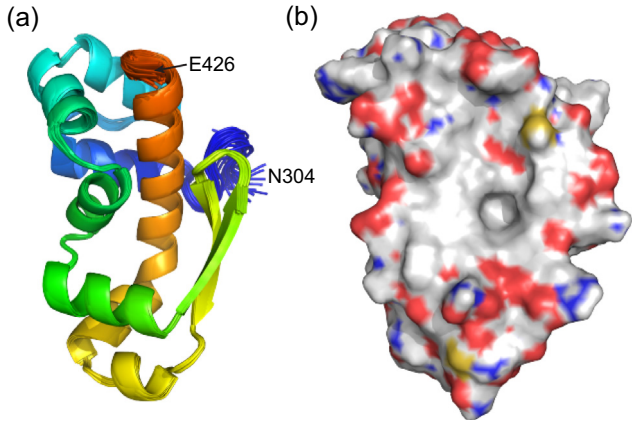


Figure 1. Solution structure of $A20_{304-426}$-BAP. (a) Structural ensemble of 20 structures, colored from blue (N-terminal) to dark orange. (b) Molecular surface of the first conformer colored by atom type (white: H, C; red: O, blue: N; yellow: S). The surface is shown in the same orientation as in (a). Note the well-defined pocket in the center. Flexible residues from the linker and the BAP-tag were omitted from both structure representations.

precipitation (data not shown), further supporting its role in masking the hydrophobic surface of $A20_{304-426}$, thus increasing solubility.

Interaction of $A20_{304-426}$-BAP with a synthetic peptide derived from E9

Previous work identified residues from both E9 and A20, which are essential for the formation of the A20-E9 complex: residues E9:L578 and E9:I582 and A20:F414.[16] E9:L578 and E9:I582 are part of a helix, situated in the poxvirus-specific insert 3 identified from sequence and structure analyses.[16] A synthetic 16-residue peptide (E9-i3) was made that covered the range of this helix, i.e. residues E9:N576 to Y591 (Figure S1). E9:L587 and E9:L588 were replaced by alanine in order to increase the solubility and the helical propensity of the synthetic peptide.[20] The E9 structure shows that these leucine residues are not exposed on the same side of the helix than the critical E9:L578 and E9:I582 residues.[16] ^{15}N-labeled $A20_{304-426}$-BAP was titrated with increasing concentrations of the E9-i3 peptide. The spectral changes observed during the titration experiment are shown in Figure S5. Some correlation peaks change their position gradually with increasing peptide concentration, while others (A20:F354, A20:D355, L_{BAP}:F15) are broadened beyond detection at a peptide to protein ratio of 1:1. Such a behavior indicates that the peptide binds to $A20_{304-426}$-BAP with a relatively low affinity. No saturation is observed up to a molar ratio of 8:1 at a protein concentration of 0.15 mM. The fit of the weighted chemical shift difference with respect to the peptide to protein ratio, assuming a single binding site, yields a dissociation constant of 55 μM. Figure 3(a) shows the chemical shift difference measured at a peptide to protein ratio of 4:1 as a function of the protein sequence, and as a projection on the protein surface (Figure 3(b)). Residues A20:353–355, 367, 373, 374, 376, and 379 are the most affected by the presence of the peptide. Significant effects are also observed for residues of the C-terminal part of helix 7 (residues A20:415–426), close to the C-terminus where the BAP-tag is attached. Remarkably, residues from the BAP-tag (residues L_{BAP}:F15, E16, A17, and K19) also show significant chemical shift differences upon the addition of the E9-i3 peptide. This confirms that the BAP-tag interacts with the A20 part of the construct, as already suggested from the analysis of the ^{15}N relaxation data (see above). Addition of the peptide therefore leads to the displacement of the BAP-tag from the hypothetical A20-E9 interaction site. Thus, the calculated dissociation constant is an apparent one, rather determined by the relative affinities of the BAP-tag and the E9-i3 peptide. The projection of the chemical shift differences on the surface of the $A20_{304-426}$ domain shows that addition of the peptide mainly affects residues in proximity of the hydrophobic binding pocket (Figure 3(b)). Residues that show the highest chemical shift variations are situated in helices 4 and 5 as well as within the β-sheet.

F414A-$A20_{304-426}$-BAP does not interact with the E9-derived peptide

We previously showed that the A20:F414A mutant displayed much weaker binding to E9. An analogous mutation was introduced in the $A20_{304-426}$-BAP construct, yielding the F414A-$A20_{304-426}$-BAP protein.[16] This protein could be produced and purified under the same conditions as $A20_{304-426}$-BAP. Figure S6 shows the superposition of two ^{1}H,^{15}N correlation spectra of $A20_{304-426}$-BAP and F414A-$A20_{304-426}$-BAP. Even if many peak positions

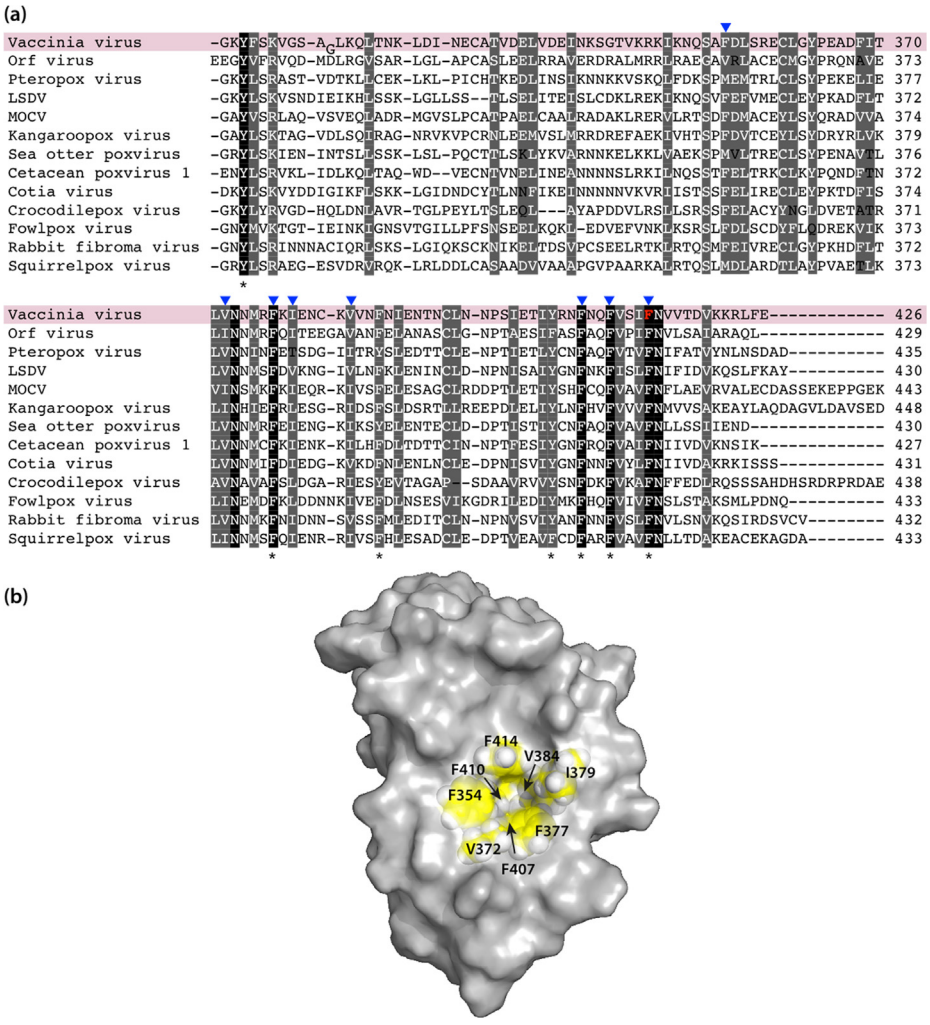


Figure 2. Sequence conservation and surface cavity of $A20_{304-426}$-BAP surrounded by hydrophobic amino acids. (a) The sequence of VACV $A20_{304-426}$ (shown on a colored background) was aligned against the UniRef50 database using PSI-BLAST.[19] The name of the representative species of each UniRef50 cluster is indicated. LSDV: Lumpy skin disease virus; MOCV: *Molluscum contagiosum* virus. Strictly conserved residues are in white on a black background, partially conserved residues are on a grey background. Asterisks indicate conserved aromatic amino acids. The conserved F414, which has been mutated to alanine is shown in red. Triangles indicate residues that form the hydrophobic pocket. (b) The cavity is lined by two layers of hydrophobic side chains. F354, V372, F377, I379 and F414 form the hydrophobic surface surrounding the pocket, whereas F410, F407 and V384 are located below. V411, at the bottom of the pocket, is not indicated in the figure.

differ, the overall distribution suggests that the two proteins have a similar structure. A20:F414 is located within the above-mentioned cluster of aromatic residues. Its substitution by an alanine may lead to a structural rearrangement within this cluster that induces large chemical shift differences due to ring current effects. Interestingly, the addition of an 11-fold excess of the E9-i3 peptide does not induce any significant chemical shift variation, as shown in Figure S7. It can therefore be concluded that the A20:F414A mutant maintains its structure but does not interact with the E9-i3 peptide under our experimental conditions. This mutation probably leads to a significant modification of the geometry of the binding pocket, thus hindering the interaction with the E9-i3 helix.

NMR study of a chimeric $A20_{304-426}$-i3 protein

The previous experiments did not allow determining the relative orientation of the two binding partners. The detailed structural characterization of the A20-E9 interface required a stable complex between $A20_{304-426}$ and E9-i3. Based on the NMR studies of $A20_{304-426}$-BAP, we sought to replace the BAP-tag by the sequence of the helix of E9-i3. Two constructs were designed in which the E9-i3 helix was connected to

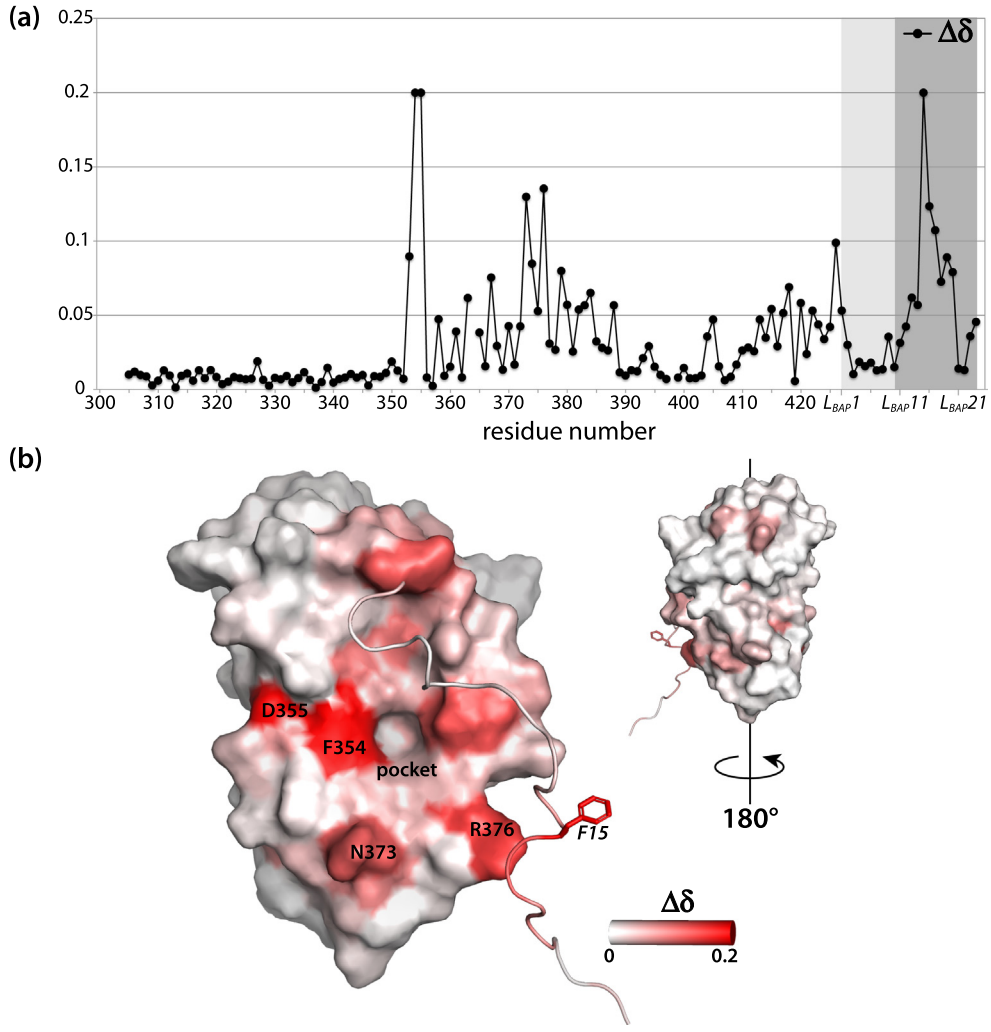


Figure 3. Chemical shift differences observed during titration of $A20_{304\text{-}426}$-BAP with the E9-i3 peptide. (a) Weighted chemical shift difference ($\Delta\delta = [(\Delta H)^2+(0.15 \times \Delta N)^2]^{1/2}$) calculated from the spectra obtained at peptide to protein ratios of 0 and 4. Peaks that broadened beyond detection were given a value of 0.2. (b) Weighted chemical shift differences plotted on the molecular surface of the $A20_{304\text{-}426}$-BAP protein (1st conformer). The backbone of the flexible BAP-tag (residues $L_{BAP}1$-21) is shown in cartoon format in an arbitrary conformation together with the side chain of residue L_{BAP}-F15. Colors range from white to red for $\Delta\delta$ values between 0 and 0.2. The protein is represented in the same orientation as in Figure 1. The insert shows the molecular surface after a 180° rotation.

$A20_{304\text{-}426}$ by the following linker sequences: (1) NASGNGSGGGS and (2) NASGS. E9:L587 of the original E9-i3 sequence was replaced by an alanine for increased solubility and E9:K590 is followed by a serine at the C-terminus. Whereas the protein with the short linker resulted in low yield and mainly insoluble protein, the first construct, named $A20_{304\text{-}426}$-i3, could be expressed and purified with a final yield comparable to $A20_{304\text{-}426}$-BAP. The sequence of this construct is shown in Figure S1 and the corresponding $^{1}H,^{15}N$ correlation spectrum is shown in Figure S8. The chemical shift differences with respect to $A20_{304\text{-}426}$-BAP were plotted as a function of the protein sequence (Figure S9). Remarkably, comparison of the BT-TROSY spectra of $A20_{304\text{-}426}$-BAP in presence of the E9-i3 peptide and of $A20_{304\text{-}426}$-i3 reveals that the same peaks are affected in a similar way by the presence of the E9-i3 helix. However, the effect is much larger for $A20_{304\text{-}426}$-i3 compared to $A20_{304\text{-}426}$-BAP in presence of 4 molar equivalents of E9-i3 peptide (Figure S10). This strongly suggests that the E9-i3 helix does interact with the $A20_{304\text{-}426}$ similarly to the isolated E9-i3 peptide. Further evidence for a stable interaction is provided by the analysis of the ^{15}N-relaxation rates measured for $A20_{304\text{-}426}$-i3 (Figure S11). $\{^{1}H\}^{15}N$-hetNOE values measured for residues between positions E9:579 and 585 are in between 0.71 and 0.81 and are in the same order of magnitude as the well-ordered A20 part of the

chimeric $A20_{304-426}$-i3 construct (unfortunately, peaks corresponding to residues E9:576–578, 580, and 581 were not resolved and could not be included in the analysis). This differs from what was seen for $A20_{304-426}$-BAP (Figure S3). Decreasing $\{^1H\}^{15}N$-hetNOE values beyond residue E9:585 indicate that the C-terminal part of the E9-i3 helical sequence is more flexible and may only partially adopt a helical structure. Increased R_2 rate constants were observed for residues A20:N373, R376, E9:E579, and N584, probably due to chemical exchange. Residues E9:579 and 584 are close to the leucine and isoleucine residues that have been shown to be involved in the A20-E9 interaction (E9:L578 and I582).[16] Residues A20:N373 and R376 showed large chemical shift differences when the $A20_{304-426}$-BAP protein was titrated with the E9-i3 peptide (Figure 3) and are probably also involved in the interaction with E9.

231 dihedral angle constraints were derived from backbone chemical shifts and used as input into an ab initio 3D structure calculation. Semi-automatic analysis of four 2D and 3D NOESY experiments allowed extraction of 1965 unambiguous and 776 ambiguous distance constraints. The final structural ensemble composed of 20 structures, obtained after refinement in explicit water, is shown in Figure 4(a) and structural statistics are given in Table S2. The core of the protein closely resembles the $A20_{304-426}$-BAP structure (Figure 1(a)). The two representative conformers of each ensemble can be superimposed with an RMSD of 0.83 Å (measured for 88 backbone atoms within the secondary structure elements), as shown in Figure S12. 16 NOEs were identified between the A20 and the E9-i3 parts of the chimeric construct that allowed constraining the relative position of the E9-i3 helix with respect to A20. Figure 4(b) shows details of the interface between the A20 part and the E9-i3 component. It can be seen that the side chain of E9:L578 inserts into the hydrophobic pocket where it makes contacts with A20:F354, V372, F377, I379, F407, F410, and F414. On the same hydrophobic side of the helix, E9:I582 makes contacts to A20:F414, I379 and C382, while E9:L586 is close to A20:C382. On the opposite side, the helix is stabilized by salt bridges between E9:R577 and A20:E390, and between E9:E581 and A20:R376. Note that residues A20:F414 as well as E9:L578 and I582 had previously been shown to be required for a stable intermolecular interaction. Likewise, the double-mutant E9:E580R-E581R showed a 10-fold lower affinity with respect to wt-E9. Thus, the presented $A20_{304-426}$-i3 structure obtained ab initio and without any assumptions on the possible molecular interface agrees nicely with previously reported biochemical findings.[16]

HADDOCK-generated model of the complex between VACV E9 and $A20_{304-426}$

The size of the VACV E9 protein (117 kDa) precludes a direct analysis of the E9:$A20_{304-426}$ complex by NMR. However, as the interface in the chimeric $A20_{304-426}$-i3 construct was in good agreement with the previous experimental data, we decided to model the complex by high ambiguity driven protein–protein docking, using the HADDOCK software.[21] The docking was guided by 16 unambiguous distance restraints, derived from NOEs observed between the A20 and the E9 components within $A20_{304-426}$-i3. One hundred generated complexes were further refined in explicit water and clustered according to the positional

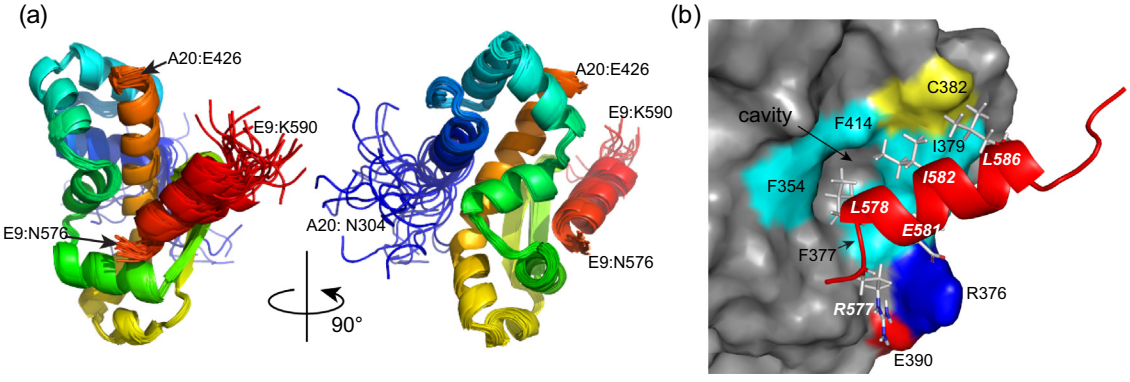


Figure 4. Solution structure of $A20_{304-426}$-i3. (a) Structural ensemble of 20 structures, colored from blue (N-terminal) to dark orange. Flexible residues from the linker were omitted for clarity. The C-terminally fused E9-i3 helix (N576-K590) is shown in red. (b) Interface between $A20_{304-426}$ and the E9-i3 helix. The solvent-accessible surface of $A20_{304-426}$ from the $A20_{304-426}$-i3 structure is colored in grey, the E9-i3 helix of $A20_{304-426}$-i3 is shown as cartoon in red. The linker was omitted from the representation. Side chains of the interacting residues from the E9-i3 helix are shown as sticks and are numbered in white italic. Residues of $A20_{304-426}$ are labeled in black and the surface is colored as follows: positively and negatively charged residues are shown in blue and red, respectively, F354, F377, I379 and F414 are shown in cyan and C382 in yellow.

interface ligand RMSD (iL-RMSD).[22] Ninety-nine of the resulting structures belonged to a single cluster and were characterized by HADDOCK scores between −128 and −58 The ten best-scoring complexes were submitted to the PDB-Dev (accession code: **PDBDEV_00000075**). Table S3 lists HADDOCK-derived properties for the ten selected complexes. The best-scoring model of the complex is shown in Figure 5(a) while Figure 5(b) presents the interaction interface in more detail. Three different stretches of E9 are involved in contacts with the $A20_{304-426}$ domain: E9:Y74-D77, T575-K590 and N607-I609. Interestingly, they belong to two poxvirus-specific inserts: insert 0 (residues E9:67–82) and insert 3 (residues E9:567–617). Note that only the insert 3 helix (residues E9:576 to 590) was present in the $A20_{304-426}$-i3 construct and gave rise to unambiguous distance constraints. The helix runs diagonally across the interface and is not aligned with any of the helices of $A20_{304-426}$ (Figure 4). The contacts involving E9 residues beyond this helix have been established by molecular modeling with HADDOCK. As previously observed for the chimeric $A20_{304-426}$-i3 protein, E9:L578 is sitting in the hydrophobic pocket and is in close contact (<5 Å) with A20:F354, V372, F377, V384, F407, F410, and F414. These residues surround the pocket observed in the $A20_{304-426}$-BAP protein (Figure 2(b)). In close vicinity, E9:I582 is in contact with A20:I379 and A20:F414. In addition, several interactions between polar residues contribute to the stabilization of the complex. Salt-bridges are observed between residues E9:R577 and A20:E390, E9:E581 and A20:R376 and between E9:D77 and A20:K349.

The ten best-scoring complexes were further analyzed using the PRODIGY (PROtein binDIng enerGY prediction) webserver,[23,24] that yields an estimation of the binding energy of protein – protein complexes from their molecular structure. Table S4 shows that the predicted binding energy of the best-scoring complex is 10 kcal mol^{-1}. This corresponds to a dissociation constant of 48 nM at 25 °C. This value is in the range of the experimentally determined K_d of 23 nM.[16]

Discussion

The interface between E9 and $A20_{304-426}$ is formed both by polar and non-polar contacts, involving the poxvirus-specific inserts 0 and 3 of E9. Analysis of the interface together with sequence alignments of $A20_{304-426}$ (Figure 2(a)) and E9 (Figure 6(a)) reveal that the main contacts involve conserved residues. The central residue is E9:L578 at the N-terminal end of the insert 3 helix, that fits into the hydrophobic pocket of $A20_{304-426}$. This pocket is surrounded by a number of conserved hydrophobic residues (Figure 2(b)), many of which also establish contacts with E9:

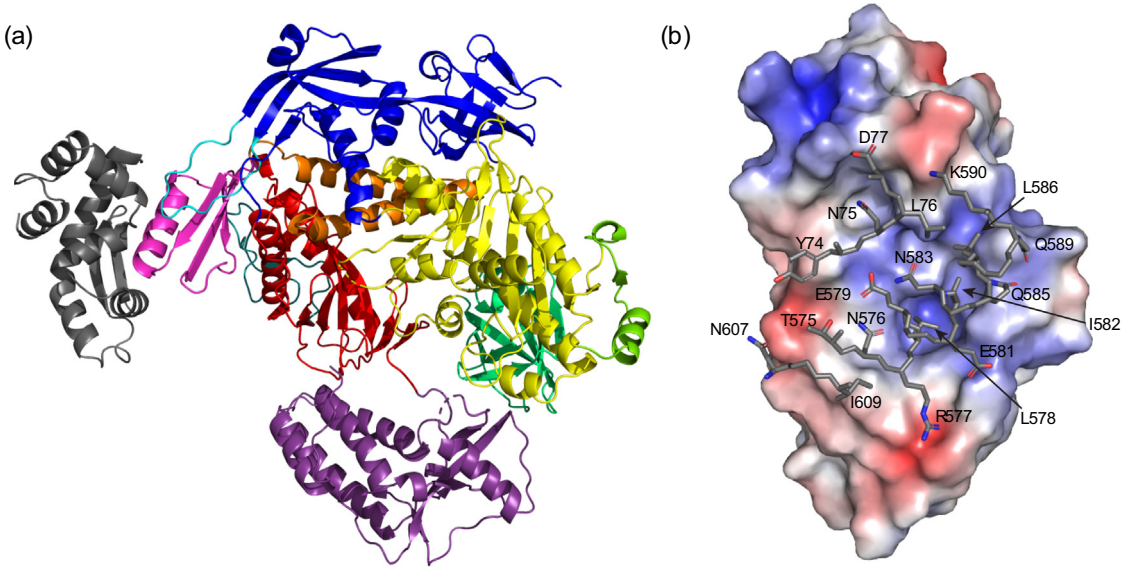


Figure 5. HADDOCK model of the E9-$A20_{304-426}$ complex. (a) The structure of the best-scoring complex from HADDOCK is shown. The VACV DNA polymerase E9 is colored according to its structural domains: N-terminal domain (blue), thumb (violet), palm (red), finger (orange), exonuclease domain (yellow). VACV-specific inserts are indicated as follows: insert 0 (cyan), 1 (chartreuse green), 2 (mint green), 3 (magenta), and 4 (teal). $A20_{304-426}$ is shown in grey. (b) Details of the A20-E9 interface of the best-scoring HADDOCK generated complex. $A20_{304-426}$ is shown as the molecular surface, colored according to the electrostatic potential ranging from −5 kT (red) to + 5 kT (blue). Three regions of E9 are shown in stick representation: Y74-D77, T575-K590, and N607-I609. Backbone atoms are in grey and side chains are coloured by element type. Only side chains of residues that establish contacts with $A20_{304-426}$ are labelled and shown.

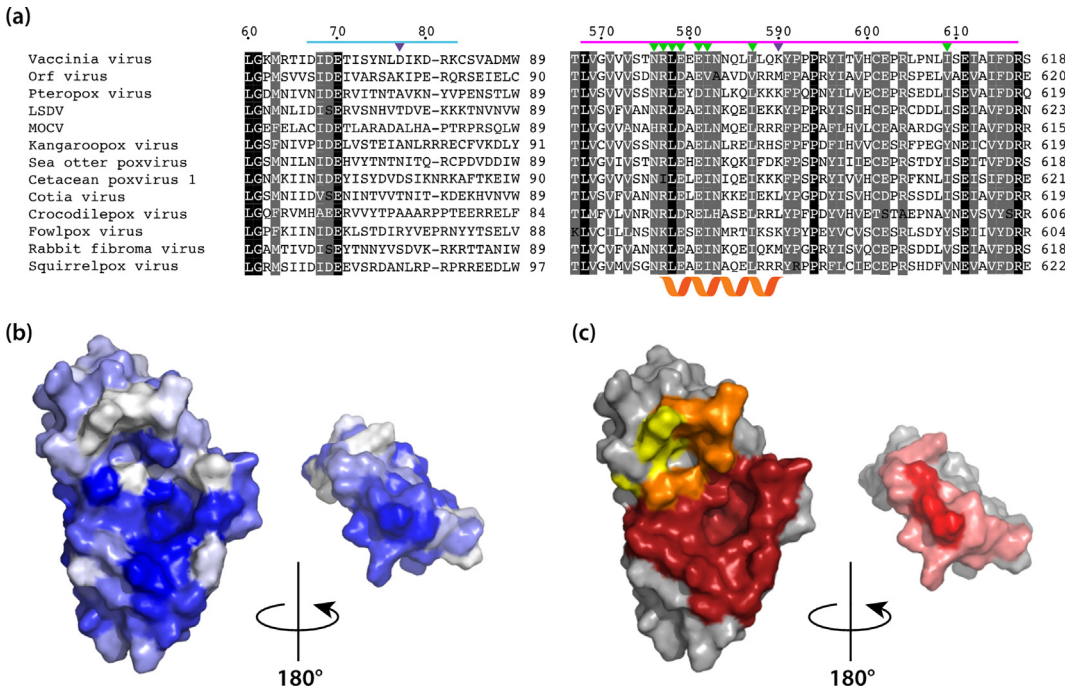


Figure 6. E9 sequence conservation and E9-A20 interface in the HADDOCK-docked model. (a) Sequence of vaccinia virus E9 inserts 0 (cyan) and 3 (magenta) are shown aligned to sequences from the same viral species as in Figure 2. LSDV: Lumpy skin disease virus; MOCV: *Molluscum contagiosum* virus. Strictly conserved residues are in white on a black background, partially conserved residues are on a grey background. Conserved residues involved in important contacts with $A20_{304\text{-}426}$ are labelled with a green triangle. Violet triangles indicate non-conserved amino acids involved in polar interactions. (b) Amino acid conservation mapped onto the contact surfaces of $A20_{304\text{-}426}$ (left) and E9-i3 (right). Colours range from white, no conservation, to blue, strictly conserved. (c) Intermolecular contacts mapped onto the accessible surfaces of $A20_{304\text{-}426}$ (left) and E9-i3 (right). $A20_{304\text{-}426}$ residues involved in contacts with E9 are shown in dark red (contacts with insert 3), yellow (contacts with insert 0) or orange (contacts with both insert 0 and insert 3). E9-i3 residues involved in contacts with $A20_{304\text{-}426}$ are indicated in pink or in red for L578, I582 and L586. The surface of the best-scoring HADDOCK model is shown.

L578 (A20:F354, V372, F377, I379, F410, F414). Additional hydrophobic contacts involve E9:I582 and, to a lesser extent, E9:L586, located on the same side of the insert 3 helix. E9:I582 is in contact with the conserved A20:I379 and F414. On E9, the strictly conserved L578 is surrounded by conserved hydrophilic residues (N576, R577, E579 and E581). At the N-terminal end of the insert 3 helix, the strictly conserved E9:N576 stacks onto the aromatic ring of A20:F354. Four salt bridges further stabilize the E9-$A20_{304\text{-}426}$ interface: E9:D77-A20:K349 (involving insert 0), E9:R577–A20:E390, E9:E581–A20:R376, and E9:K590–A20:E426. E9:K590 is located at the end of the insert 3 helix and anchors its C-terminal end to the C-terminal carboxyl group of A20. The salt-bridges involving E9:R577 and E9:E581 have also been observed within the $A20_{304\text{-}426}$-i3 solution structure described herein. As shown in Figure 6 (a), E9:E581 is conserved throughout the chordopoxvirus. A second conserved glutamate, E9:579, contacts the backbone of A20:F354 and the neighbouring A20:S352 and A20:A353. The strictly conserved A20:N373 makes a hydrogen bond with the backbone of E9:R577, but also forms contacts with E9:I609, which is located at the beginning of the central β-strand of the sheet being part of insert 3 (Figure 5). In addition to the multiple contacts involving insert 3, residues E9:74–77 of insert 0 are likely to contribute to the interface, in particular E9:Y74 and L76, but are not conserved at the sequence level (Figure 6(a)).

It is interesting to note that the HADDOCK generated model complex shows many more intermolecular contacts than what was observed in the experimental $A20_{304\text{-}426}$-i3 structure. Indeed, comparison of the molecular surfaces of A20 either with the BAP-tag, with the i3 extension or within the HADDOCK model reveals that the hydrophobic surface pocket increases with increasing number of contacts (Figure S13). For $A20_{304\text{-}426}$-BAP, no explicit experimental information on the occupancy of the E9 binding site could be obtained. However, ^{15}N relaxation and peptide titration data strongly suggest that the BAP-tag interacts in a rather unspecific way with the hydrophobic pocket of $A20_{304\text{-}426}$-BAP. The solution structure of the $A20_{304\text{-}426}$-i3 construct

clearly shows insertion of E9:L578 into the hydrophobic pocket and also reveals contacts between E9:I582 and E9:L586 with A20:I379 and C382. As a consequence, a slight reorientation of the F354 and F377 side chains that restrict the hydrophobic pocket in the $A20_{304-426}$-BAP structure is observed in $A20_{304-426}$-i3 (Figure S13). Comparing the experimental $A20_{304-426}$-i3 structure and the HADDOCK model complex, we observed that the angle formed between the insert 3 helix and the long helix H7 in the $A20_{304-426}$ protein is not the same. In addition, the insert 3 helix within $A20_{304-426}$-i3 is shorter and ^{15}N relaxation data indicate increased flexibility towards its C-terminal end, suggesting less favored interaction for residues beyond E9:A587 that have $\{^1H\}^{15}N$-hetNOE values below 0.5 (Figures S11 and S13). As a consequence, in the HADDOCK-generated model, the binding pocket for E9:I582 is more pronounced. This results from a different orientation of the A20:C382 side chain. PRODIGY analysis of the interface between $A20_{304-426}$ and the E9-i3 helix in the $A20_{304-426}$-i3 solution structure gave an estimated K_d of 1.34 ± 0.5 μM which is compatible with the apparent K_d of 55 μM obtained by the NMR titration of $A20_{304-426}$-BAP with the E9-i3 peptide. As the BAP-tag is required for protein solubility, we were not able to experimentally determine the dissociation constant for untagged $A20_{304-426}$ in complex with the E9-i3 peptide. In any case, these values are much higher than the K_d in the tens of nanomolar range measured for the E9-$A20_{304-426}$-BAP interaction[16] or calculated for the best-scoring HADDOCK-generated complex. Therefore, the additional intermolecular contacts identified in the HADDOCK-generated complexes that involve residues outside of the insert 3 helix (residues 74 to 77 of insert 0, N607 and I609 of insert 3) seem to be essential for a stable E9-$A20_{304-426}$ complex.

It is possible to model the E9-$A20_{304-426}$ complex interacting with a template and primer strand of DNA using the model of E9 in complex with dsDNA presented previously.[16] Conserved hydrophobic residues on the face of $A20_{304-426}$ opposite of the E9 binding pocket correspond most likely to residues of the hydrophobic core, which got exposed by the truncation of A20. They may indicate the direction in which the structure of intact A20 would extend. The surface they form is located below the first ordered residue S313 (Figure S14). Consequently, one may suggest that the extended A20 protein positions the DNA glycosylase D4 on the DNA thus tethering E9 to its substrate (Figure 7(a)). This model can be compared with other DNA holoenzyme structures. Indeed, the VACV DNA polymerase fulfils the two functions of leading strand and lagging strand synthesis, activities also shared by pol ε and pol δ in eukaryotes. Recently, sparked by the development of single-particle cryo-electron microscopy, a number of family B DNA polymerase holoenzyme structures composed of the catalytic subunit, accessory proteins and most often DNA became available. These are the human[25] and yeast[28] polymerase δ bound to PCNA (Proliferating Cell Nuclear Antigen) co-factor Figure 7(b), the yeast polymerase ζ,[29] closely related to pol δ, and the yeast polymerase ε[27] (Figure 7(c)). Looking at processivity factor binding modes, these DNA polymerases display important differences with respect to the VACV DNA polymerase holoenzyme. The human polymerase δ structure shows that PCNA interacts with a PIP (Processivity factor Interacting Peptide) motif at the C-terminal extension of the catalytic subunit (Figure 7(b)) in contrast, the polymerase ε, which probably results from a duplication of the gene coding the polymerase catalytic subunit (Figure 7(c)), has it's PIP motif located in a disordered linker between the catalytic N-terminal domain and the inactive C-terminal domain. At this position, PCNA would bind similarly to what is observed in human polymerase δ.[27] Remarkably, these C-terminal extensions are missing in VACV E9.

The archaeal polymerases of the family B DNA polymerase also interact with their ring-shaped PCNA processivity factor through a PIP-motif at the extreme C-terminus.[30] The same organisation is found in the DNA polymerases of the herpesvirus family, where the processivity factors are incomplete rings derived from PCNA circles. The polymerase of human cytomegalovirus belongs to this family and is also the only polymerase where an interaction with a Uracil N-glycosylase, UL114, has been described, but which has not been characterized or localized.[31]

Therefore, VACV DNA polymerase appears to be unique having the processivity factor binding site located between the palm and finger domains in the poxvirus-specific insert 3 and by using D4, a protein involved in DNA repair, as an integral part of the processivity factor.

Figure 7 further indicates that yeast ε and human δ polymerases bind accessory proteins on the opposite site of the VACV processivity factor-interacting insert 3. Interestingly, at the same location, VACV E9 has additional poxvirus-specific inserts (inserts 1 and 2). It is tempting to speculate that these inserts have evolved towards a specific function that may be fulfilled by the interaction with additional, still unknown accessory proteins. The understanding of the global structure of the poxvirus DNA polymerase holoenzyme has to await results from additional structural studies.

There have been attempts to develop small molecules as anti-poxvirus drugs that interfere with the assembly of the holoenzyme and thus with DNA replication, an essential step in viral infection. They focused essentially on the A20-D4 interface.[32,33] The relatively small size of the E9-A20 interface centered on the binding pocket of A20, makes the E9-A20 interaction an additional

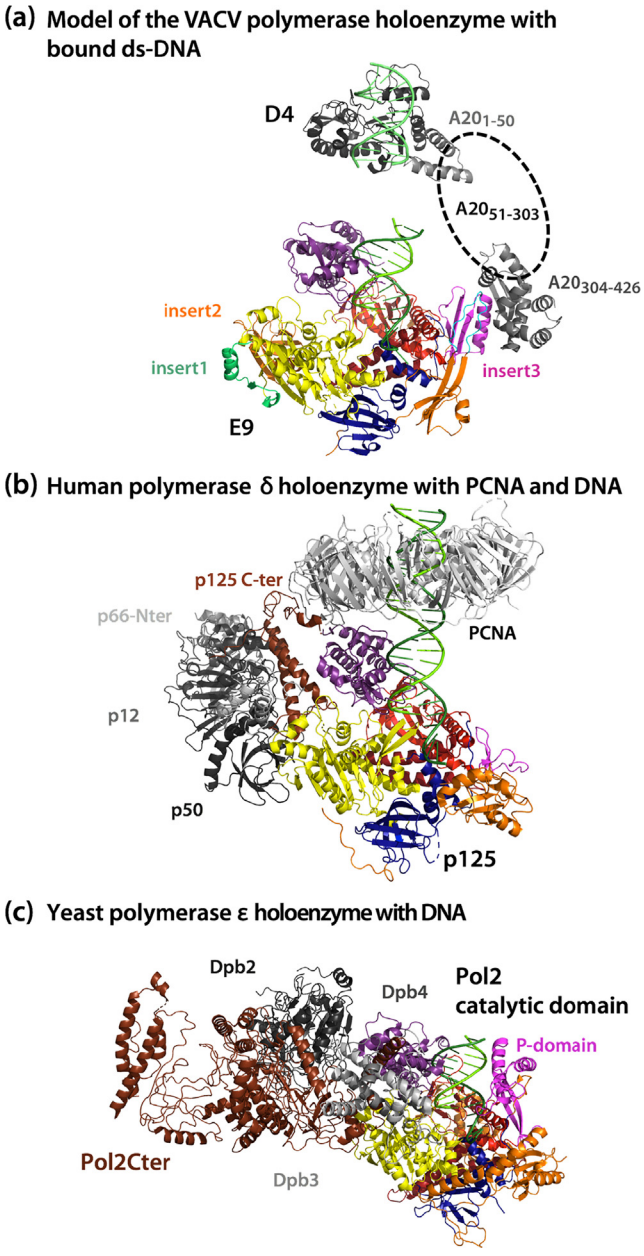


Figure 7. Structural comparison of family B polymerases with accessory subunits and processivity factors. Structures have been superposed using the catalytically active polymerase subunit. These subunits are coloured according to their domains (N-terminal: blue; exonuclease: yellow; palm: red; fingers: brick; thumb: violet; C-terminal domains: chocolate). Insertions into the canonical domains of family B polymerases are shown in orange, whereas domains at the location of VACV insert 3 are shown in magenta. The subunits of the polymerases are shown in black and grey. The template DNA strand is shown in dark green, the primer strand in yellow green. (a) VACV E9 polymerase. $A20_{304-426}$ from the HADDOCK model has been added to the modelled complex with template and primer strand[16] using a superposition based on insert 3. Inserts 1 and 2 are coloured in mint green and orange, respectively. The uracil DNA glycosylase D4 bound to the A20 N-terminal domain ($A20_{1-50}$) and DNA is shown in grey. (b) Model of the human polymerase pol δ holoenzyme (pdb entry **6TNY**,[25]) in complex with PCNA (light grey). Note that PCNA only binds through the PIP motif located on the C-terminal domain (brown) of the catalytic subunit p125. (c) The holoenzyme of yeast polymerase ε in complex with dsDNA combining models from pdb entry **6WJV**,[26] and pdb entry **4M8O**.[27] The inactive domain at the C-terminus of the catalytic subunit of Pol2 is shown in brown.

target. Sequence conservation and the hydrophobic property of the cavity seem to predestinate this site for the development of small-molecule inhibitors opening up new possibilities for the design of specific molecules binding tightly to this pocket, thus interfering with the holoenzyme assembly.

Material and methods

Plasmids and peptides

The DNA sequence coding for the VACV A20 C-terminal domain was derived from the full-length VACV A20 DNA using the ESPRIT technology[34] as described elsewhere.[16] The resulting plasmid, pESPRIT002-A20, carries the gene sequence coding for residues 304 to 426 with an N-terminal His_6-tag followed by a TEV cleavage site and a C-terminal BAP-tag, used in the ESPRIT screen.[34] Insertion of the F414A mutation was achieved by mutagenesis as described before.[16] The $A20_{304-426}$-i3 construct was obtained by PCR amplification of the pESPRIT002-A20 plasmid using the following primers: AACCAGCTGGCTCTGCAGAA AAGTtaatagtccggataatagg and GTTGATTTCTTC TTCCAGACGGTTAGAcccaccgccactcccatttcc (capital letters corresponding to E9-i3 helix coding sequence, replacing the BAP tag sequence, see Figure S1). The resulting linear PCR product was circularized to obtain the final mutated plasmid.

The synthetic E9-i3 peptide was provided by GeneCust (Boynes, France).

Protein expression and purification

E. coli BL21(DE3)-Star cells were transformed with the corresponding plasmid. Isotopically labeled proteins were produced in M9 minimal mineral medium (pH 7.8), supplemented with 0.1 mM $MnCl_2$, 0.05 mM $FeCl_3$, 0.05 mM $ZnSO_4$, a vitamin solution, 30 mg/L kanamycin, and $^{15}NH_4Cl$ (1 g/L) and [$^{13}C_6$]-glucose (2 g/L) as the sole nitrogen and carbon sources, respectively. For ^{15}N-labelled protein, glucose concentration was 4 g/L. When the OD_{600} reached a value between 0.6 and 0.8, cell cultures were placed in a cold room for 30 min before protein production was induced with 0.2 mM IPTG (isopropyl β-D-1-thiogalactopyranoside). Cells were then grown at 20 °C for 16 h, harvested by centrifugation (5000*g*, 15 min), suspended in Buffer A (50 mM Tris-HCl, 300 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol, pH 7.4) and broken in a Microfluidizer (Microfluidics), thoroughly avoiding heating of the sample as this was found to lead to protein aggregation. Cell debris and insoluble proteins were discarded by centrifugation at 45,000*g* for 40 min. The supernatant was collected and loaded on a 3 mL Ni-NTA affinity column pre-equilibrated with Buffer A. The column was washed successively with 5 column volumes (CV) of Buffer A supplemented with 5 mM imidazole, and 10 CV of Buffer A with 20 mM imidazole, before the desired protein was eluted in 5 CV each of Buffer A with 100 mM and 300 mM imidazole. The resulting solution was concentrated to 10 mL on an Amicon Ultrafiltration device (15 mL, MWCO 10 kDa) and dialyzed twice against 250 mL of Buffer A. The His-tag was cleaved overnight at 4 °C in presence of 0.5 mM DTT (dithiothreitol) by adding 1 mg of TEV protease for 50 mg of purified $A20_{304-426}$ protein. Cleaved protein was then recovered in the flow-through and wash solution of a second, 3 mL Ni-NTA column. The protein was concentrated to 5 mg/mL and stored at −80 °C. For the NMR sample preparation, 5 mg of protein were concentrated to 0.3–0.5 mL on a 4 mL Amicon Ultrafiltration device (MWCO 10 kDa). The buffer was exchanged to Buffer D (50 mM potassium phosphate, 300 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP), pH 6.0) using either a NAP-5 column or by dialysis against 2x100 mL of Buffer D. For the preparation of NMR samples in D_2O, the protein in Buffer D was lyophilized in a SpeedVac vacuum concentrator and resuspended in D_2O. For most NMR experiments, the protein concentration was adjusted to 0.8–1 mM. When possible, NMR tubes were sealed in order to avoid oxidation of the protein.

NMR spectroscopy

NMR data were acquired on 600, 700, 850 or 950 MHz Bruker Avance III spectrometers equipped with cryogenic probes. Sample temperature was set to 25 °C. Data were processed with Topspin 3.6.2 and analyzed with CCPNMR 2.4.[35]

Sequence-specific resonance assignment

All NMR experiments were set up using NMRlib.[36] The following experiments were performed for the backbone assignment of $A20_{304-426}$-BAP and $A20_{304-426}$-i3 resonances: 2D $^1H,^{15}N$-BEST-TROSY and $^1H,^{13}C$-CT-HSQC experiments, 3D BEST-TROSY HNCO, 3D BEST-TROSY HNCACB, 3D BEST-TROSY HNcoCACB, 3D BEST-TROSY HNCA, 3D BEST-TROSY HNcoCA.[37] Side chain resonances were assigned using a 3D hCCH-TOCSY experiment in combination with information from 3D-^{15}N-edited NOESY-HSQC experiments. Aromatic ^1H resonances were assigned using 2D-NOESY and 2D-TOCSY experiments acquired on a sample prepared in D_2O.

Constraints for structure calculation

Backbone dihedral angles were derived from chemical shifts using the TALOS + program.[17] Distance constraints were obtained from the analysis of the following NOESY experiments, acquired on

$A20_{304-426}$-BAP and $A20_{304-426}$-i3: 3D ^{15}N-edited NOESY-HSQC, 3D ^{13}C-edited HSQC-NOESY, 3D Ch_HC methyl-selective NOESY. In addition, 2D-NOESY experiments were recorded for D_2O samples. Peak picking of the standard NOE experiments was performed by the ATNOS algorithm[38] within the UNIO10 program.[39] The 3D Ch_HC, methyl selective NOESY experiment was analyzed manually.

^{15}N relaxation experiments

^{15}N relaxation experiments (T_1, T_2, {^{1}H}^{15}N-hetNOE) were performed using standard pulse schemes[40] implemented in NMRlib. For T_1 and T_2 measurements, 11 time points were recorded with relaxation delays varying from 0 to 1.8 s for T_1 and from 9 to 238 ms for T_2 series. For the {^{1}H}^{15}N-hetNOE, the recycle delay was set to 5 s and the saturation time to 3 s. Reference and saturation spectra were recorded in an interleaved manner.

Structure calculation

The UNIO10 program suite[39] was used for initial NOE assignment[41] and structure calculation to check for the consistency of the data. Peak lists generated by UNIO were then used as input data for structure calculation using ARIA 2.3[18] interfaced to CCPNMR 2.4[35] and CNS 1.21.[42] In addition, manually assigned NOEs from the Ch_HC methyl-selective NOESY experiment were transformed into distance constraints that were recalibrated during the ARIA iterations. One thousand structures were calculated in the eighth ARIA iteration, from which 20 structures with the lowest total energy were selected and refined in explicit water to give the final structural ensemble.

Peptide titration

Small amounts of the synthetic peptide were suspended in Buffer D to a final concentration of 2 mM. The pH was adjusted by addition of a few microliters of 1 M K_2HPO_4. Samples were prepared by mixing $A20_{304-426}$-BAP or F414A-$A20_{304-426}$-BAP solution with different amounts of peptide and Buffer D so that the final protein concentration was 0.15 mM and the peptide to protein ratio varied from 0 to 8 for $A20_{304-426}$-BAP and from 0 to 11 for F414A-$A20_{304-426}$-BAP. For each sample / titration point, a ^{1}H,^{15}N-BEST-TROSY experiment was acquired at 600 MHz with an experimental time of 40 min.

High ambiguity driven protein–protein docking of the $A20_{304-426}$/E9 complex

The high ambiguity driven protein–protein docking software HADDOCK 2.4[21] was used for the modeling of the $A20_{304-426}$/E9 complex. The crystal structure of E9 (PDB ID **5N2E**, alternate conformation A) as well as the NMR ensemble of 20 $A20_{304-426}$-i3 conformers were used as input coordinates. For the latter, residues belonging to the linker and the E9-i3 helix were deleted. The following residues were defined as active residues: A20:353, 354, 369, 372, 373, 376, 377, 379, 382, 384, and 414; E9:576, 578, 582. Passive residues were automatically defined by the HADDOCK program. In addition, 15 unambiguous intermolecular distances were introduced that were derived from NOEs observed between the A20 and the E9-i3 part of the $A20_{304-426}$-i3 construct. A20:303-312 were set as flexible whereas semi-flexible residues were determined by the HADDOCK software from the analysis of the interface of the docked complexes. 4000 complexes were determined by rigid-body docking and minimization. 400 of those were then submitted to semi-flexible simulated annealing refinement, during which the interface packing is optimized by introducing flexibility. Finally, the hundred best-scoring complexes were further refined in explicit water. The final models were clustered based on the positional interface ligand RMSD[22] using a cut-off of 2 Å.

Structural data deposition

Experimental NMR data and molecular coordinates have been deposited at the wwPDB data bank and at the BMRB. $A20_{304-426}$-BAP accession code(s): PDB ID **6ZYC**, BMRB ID **34545**. $A20_{304-426}$-i3 accession codes: PDB ID **6ZXP**, BMRB ID **34544**. The HADDOCK-generated model of the E9-$A20_{304-426}$ complex is available at the PDB-DEV database. Accession code: **PDBDEV_00000075**.

CRediT authorship contribution statement

Beate Bersch: Conceptualization, Methodology, Investigation, Visualization, Writing - original draft, Project administration. **Nicolas Tarbouriech:** Conceptualization, Methodology, Writing - review & editing. **Wim P. Burmeister:** Conceptualization, Writing - review & editing, Funding acquisition. **Frédéric Iseni:** Writing - review & editing, Conceptualization, Project administration, Funding acquisition.

Acknowledgements

We thank Anaïs Branchard for help with protein expression. Alexandre Bonvin is gratefully acknowledged for his advice and help with the HADDOCK software. FI is supported by research grants from the Service de Santé des Armées and the Direction Générale pour l'Armement. This work is supported by the French National

Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02) and by the Fond de la Recherche Médicale (grant ''Equipe 2017'' DEQ20170336754). Financial support from the TGIR-RMN-THC Fr3050 CNRS for conducting the research is gratefully acknowledged. This work used the platforms of the Grenoble Instruct-ERIC center (ISBG ; UMS 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-0005-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2021.167009.

Received 11 March 2021;
Accepted 17 April 2021;
Available online 24 April 2021

Keywords:
NMR;
protein-protein interaction;
poxvirus;
DNA replication;
holoenzyme

Abbreviations used:
VACV, vaccinia virus; BAP, Biotin Acceptor Peptide

References

1. Moss, B. (2013). Poxviridae. In Fields Virol. 2 (Fields, B.M., Knipe, D.M. & Howley, P.M. eds), Wolters Kluwer Health/ Lippincott Williams & Wilkins, Philadelphia, pp. 2129–2159.
2. Upton, C., Stuart, D.T., Mcfadden, G., (1993). Identification of a poxvirus gene encoding a uracil DNA glycosylase. *Proc. Natl. Acad. Sci. USA*, **90**, 4518–4522. https://doi.org/10.1073/pnas.90.10.4518.
3. Boyle, K.A., Stanitsa, E.S., Greseth, M.D., Lindgren, J.K., Traktman, P., (2011). Evaluation of the role of the vaccinia virus uracil DNA glycosylase and A20 proteins as intrinsic components of the DNA polymerase holoenzyme. *J. Biol. Chem.*, **286**, 24702–24713. https://doi.org/10.1074/jbc.M111.222216.
4. Boyle, K.A., Arps, L., Traktman, P., (2007). Biochemical and genetic analysis of the vaccinia virus D5 protein: multimerization-dependent ATPase activity is required to support viral DNA replication. *J. Virol.*, **81**, 844–859. https://doi.org/10.1128/jvi.02217-06.
5. De Silva, F.S., Lewis, W., Berglund, P., Koonin, E.V., Moss, B., (2007). Poxvirus DNA primase. *Proc. Natl. Acad. Sci. USA*, **104**, 18724–18729. https://doi.org/10.1073/pnas.0709276104.
6. Rochester, S.C., Traktman, P., (1998). Characterization of the single-stranded DNA binding protein encoded by the vaccinia virus I3 gene. *J. Virol.*, **72**, 2917–2926. https://doi.org/10.1128/jvi.72.4.2917-2926.1998.
7. Filée, J., Forterre, P., Sen-Lin, T., Laurent, J., (2002). Evolution of DNA polymerase families: Evidences for multiple gene exchange between cellular and viral proteins. *J. Mol. Evol.*, **54**, 763–773. https://doi.org/10.1007/s00239-001-0078-x.
8. Challberg, M.D., Englund, P.T., (1979). Purification and properties of the deoxyribonucleic acid polymerase induced by vaccinia virus. *J. Biol. Chem.*, **254**, 7812–7819.
9. Willer, D.O., Yao, X.D., Mann, M.J., Evans, D.H., (2000). In vitro concatemer formation catalyzed by vaccinia virus DNA polymerase. *Virology.*, **278**, 562–569. https://doi.org/10.1006/viro.2000.0686.
10. Gammon, D.B., Evans, D.H., (2009). The 3′-to-5′ exonuclease activity of vaccinia virus DNA polymerase is essential and plays a role in promoting virus genetic recombination. *J. Virol.*, **83**, 4236–4250. https://doi.org/10.1128/jvi.02255-08.
11. McDonald, W.F., Traktman, P., (1994). Vaccinia virus DNA polymerase. In vitro analysis of parameters affecting processivity. *J. Biol. Chem.*, **269**, 31190–31197.
12. Stanitsa, E.S., Arps, L., Traktman, P., (2006). Vaccinia virus uracil DNA glycosylase interacts with the A20 protein to form a heterodimeric processivity factor for the viral DNA polymerase. *J. Biol. Chem.*, **281**, 3439–3451. https://doi.org/10.1074/jbc.M511239200.
13. Sèle, C., Gabel, F., Gutsche, I., Ivanov, I., Burmeister, W. P., Iseni, F., Tarbouriech, N., (2013). Low-resolution structure of vaccinia virus DNA replication machinery. *J. Virol.*, **87**, 1679–1689. https://doi.org/10.1128/JVI.01533-12.
14. Contesto-Richefeu, C., Tarbouriech, N., Brazzolotto, X., Betzi, S., Morelli, X., Burmeister, W.P., Iseni, F., (2014). Crystal Structure of the vaccinia virus DNA polymerase holoenzyme subunit D4 in complex with the A20 N-terminal domain. *PLoS Pathog.*, **10**, https://doi.org/10.1371/journal.ppat.1003978 e1003978.
15. Burmeister, W.P., Tarbouriech, N., Fender, P., Contesto-Richefeu, C., Peyrefitte, C.N., Iseni, F., (2015). Crystal structure of the vaccinia virus uracil-DNA glycosylase in complex with DNA. *J. Biol. Chem.*, **290**, 17923–17934. https://doi.org/10.1074/jbc.M115.648352.
16. Tarbouriech, N., Ducournau, C., Hutin, S., Mas, P.J., Man, P., Forest, E., Hart, D.J., Peyrefitte, C.N., Burmeister, W. P., Iseni, F., (2017). The vaccinia virus DNA polymerase structure provides insights into the mode of processivity factor binding. *Nature Commun.*, **8**, 1455. https://doi.org/10.1038/s41467-017-01542-z.
17. Shen, Y., Delaglio, F., Cornilescu, G., Bax, A., (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, **44**, 213–223. https://doi.org/10.1007/s10858-009-9333-z.
18. Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T.E., Nilges, M., (2007). ARIA2: automated

NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, **23**, 381–382. https://doi.org/10.1093/bioinformatics/btl589.
19. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402. https://doi.org/10.1093/nar/25.17.3389.
20. Muñoz, V., Serrano, L., (1997). Development of the multiple sequence approximation within the AGADIR Model of α-helix formation: Comparison with zimm-bragg and lifson-roig formalisms. *Biopolymers*, **41**, 495–509. https://doi.org/10.1002/(sici)1097-0282(19970415)41:5<495::aid-bip2>3.0.co;2-h.
21. Van Zundert, G.C.P., Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastritis, P.L., Karaca, E., Melquiond, A.S.J., Van Dijk, M., De Vries, S.J., Bonvin, A.M.J.J., (2016). The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.*, **428**, 720–725. https://doi.org/10.1016/j.jmb.2015.09.014.
22. Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastritis, P., Karaca, E., Melquiond, A.S.J., Bonvin, A.M.J.J., (2012). Clustering biomolecular complexes by residue contacts similarity. *Proteins Struct. Funct. Bioinforma.*, **80**, 1810–1817. https://doi.org/10.1002/prot.24078.
23. Vangone, A., Bonvin, A.M.J.J., (2015). Contacts-based prediction of binding affinity in protein–protein complexes. *Elife*, **4** https://doi.org/10.7554/eLife.07454.
24. Xue, L.C., Rodrigues, J.P., Kastritis, P.L., Bonvin, A.M., Vangone, A., (2016). PRODIGY: A web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics*, **32**, 3676–3678. https://doi.org/10.1093/bioinformatics/btw514.
25. Lancey, C., Tehseen, M., Raducanu, V.S., Rashid, F., Merino, N., Ragan, T.J., Savva, C.G., Zaher, M.S., Shirbini, A., Blanco, F.J., Hamdan, S.M., De Biasio, A., (2020). Structure of the processive human Pol δ holoenzyme. *Nature Commun.*, **11**, 1109. https://doi.org/10.1038/s41467-020-14898-6.
26. Hogg, M., Osterman, P., Bylund, G.O., Ganai, R.A., Lundström, E.B., Sauer-Eriksson, A.E., Johansson, E., (2014). Structural basis for processive DNA synthesis by yeast DNA polymerase É. *Nature Struct. Mol. Biol.*, **21**, 49–55. https://doi.org/10.1038/nsmb.2712.
27. Yuan, Z., Georgescu, R., Schauer, G.D., O'Donnell, M.E., Li, H., (2020). Structure of the polymerase ε holoenzyme and atomic model of the leading strand replisome. *Nature Commun.*, **11**, 3156. https://doi.org/10.1038/s41467-020-16910-5.
28. Zheng, F., Georgescu, R.E., Li, H., O'Donnell, M.E., (2020). Structure of eukaryotic DNA polymerase δ bound to the PCNA clamp while encircling DNA. *Proc. Natl. Acad. Sci. USA*, **117**, 30344–30353. https://doi.org/10.1073/pnas.2017637117.
29. Malik, R., Kopylov, M., Gomez-Llorente, Y., Jain, R., Johnson, R.E., Prakash, L., Prakash, S., Ubarretxena-Belandia, I., Aggarwal, A.K., (2020). Structure and mechanism of B-family DNA polymerase ζ specialized for translesion DNA synthesis. *Nature Struct. Mol. Biol.*, **27**, 913–924. https://doi.org/10.1038/s41594-020-0476-7.
30. Tori, K., Kimizu, M., Ishino, S., Ishino, Y., (2007). DNA polymerases BI and D from the hyperthermophilic archaeon Pyrococcus furiosus both bind to proliferating cell nuclear antigen with their C-terminal PIP-box motifs. *J. Bacteriol.*, **189**, 5652–5657. https://doi.org/10.1128/JB.00073-07.
31. Strang, B.L., Coen, D.M., (2010). Interaction of the human cytomegalovirus uracil DNA glycosylase UL114 with the viral DNA polymerase catalytic subunit UL54. *J. Gen. Virol.*, **91**, 2029–2033. https://doi.org/10.1099/vir.0.022160-0.
32. Nuth, M., Guan, H., Zhukovskaya, N., Saw, Y.L., Ricciardi, R.P., (2013). Design of potent poxvirus inhibitors of the heterodimeric processivity factor required for viral replication. *J. Med. Chem.*, **56**, 3235–3246. https://doi.org/10.1021/jm301735k.
33. Flusin, O., Saccucci, L., Contesto-Richefeu, C., Hamdi, A., Bardou, C., Poyot, T., Peinnequin, A., Crance, J.M., Colas, P., Iseni, F., (2012). A small molecule screen in yeast identifies inhibitors targeting protein-protein interactions within the vaccinia virus replication complex. *Antiviral Res.*, **96**, 187–195. https://doi.org/10.1016/j.antiviral.2012.07.010.
34. Mas, P.J. & Hart, D.J. (2017). ESPRIT: A method for defining soluble expression constructs in poorly understood gene sequences. In Methods Mol. Biol. Humana Press Inc., pp. 45–63. https://doi.org/10.1007/978-1-4939-6887-9_4.
35. Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., Laue, E.D., (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Genet.*, **59**, 687–696. https://doi.org/10.1002/prot.20449.
36. Favier, A., Brutscher, B., (2019). NMRlib: user-friendly pulse sequence tools for Bruker NMR spectrometers. *J. Biomol. NMR*, **73**, 199–211. https://doi.org/10.1007/s10858-019-00249-1.
37. Favier, A., Brutscher, B., (2011). Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J. Biomol. NMR*, **49**, 9–15. https://doi.org/10.1007/s10858-010-9461-5.
38. Herrmann, T., Güntert, P., Wüthrich, K., (2002). Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR*, **24**, 171–189 http://www.ncbi.nlm.nih.gov/pubmed/12522306.
39. Guerry, P., Herrmann, T., (2012). Comprehensive automation for NMR structure determination of proteins. *Methods Mol. Biol.*, **831**, 429–451. https://doi.org/10.1007/978-1-61779-480-3_22.
40. Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., Kay, L.E., (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*, **33**, 5984–6003 http://www.ncbi.nlm.nih.gov/pubmed/7514039.
41. Herrmann, T., Güntert, P., Wüthrich, K., (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.*, **319**, 209–227 http://www.ncbi.nlm.nih.gov/pubmed/12051947.
42. Brunger, A.T., (2007). Version 1.2 of the crystallography and nmr system. *Nature Protoc.*, **2**, 2728–2733. https://doi.org/10.1038/nprot.2007.406.