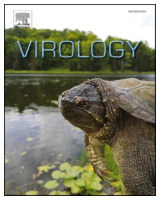
Virology 617 (2026) 110818

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology



# AlphaFold modeling of the white spot syndrome virus polymerase

Tim Skern [a,b,*], Jane Oakey [c]

[a] Max Perutz Labs, Vienna Biocenter Campus (VBC), Dr.-Bohr-Gasse 9, Vienna Biocenter 5, 1030, Vienna, Austria
[b] Medical University of Vienna, Max Perutz Labs, Dr.-Bohr-Gasse 9, Vienna Biocenter 5, 1030, Vienna, Austria
[c] Biosecurity Sciences Laboratory, Queensland Department of Primary Industries, 39 Kessels Road, Coopers Plains, QLD, 4108, Australia

ARTICLE INFO

*Keywords:*
Viral polymerase
Polymerase structure
Viral evolution
Protein modelling
AlphaFold
*Nimaviridae*

ABSTRACT

White spot syndrome virus (WSSV) infects crustaceans, causing severe losses in the global shrimp industry. Several properties, including the DNA genome sequence and the virion morphology, place WSSV as the single member of the *Whispovirus* genus of the *Nimaviridae* family. The DNA polymerase is one of the few gene products of the predicted 184 open reading frames to have been examined. Conserved sequence motifs found in many viral DNA polymerases are found in the WSSV DNA polymerase; nevertheless, the WSSV enzyme remains enigmatic, possessing over 1000 amino acids more than, for example, the DNA polymerase of herpes simplex virus 1 (HSV-1). To examine more closely the WSSV polymerase, we used AlphaFold to generate a structural model and compared it to the DNA polymerases of HSV-1, African swine fever virus and mpox virus. The exonuclease and polymerase domains of the WSSV enzyme were exactly defined based on the equivalence with the other viral enzymes; structurally, the WSSV enzyme appears most closely related to the HSV-1 enzyme. In contrast, the WSSV polymerase N-terminal domain showed an appreciably different architecture. However, the most unusual aspect of the WSSV polymerase is the C-terminal thumb domain which is modelled as two helical domains connected by a flexible acidic loop. This arrangement is quite unrelated to the thumb domains found in the other polymerases and is thus restricted to the WSSV enzyme. Given the uniqueness of such a vital cog of the replication machinery, it will be of interest to examine the structures of further WSSV proteins. (248 words).

## 1. Introduction

White spot syndrome virus (WSSV), the sole member of the *Whispovirus* genus in the *Nimaviridae* family, is an important pathogen of penaeid shrimps. The virus first appeared in shrimps around 1991 in Taiwan (Chen, 1995). Since then, the virus has appeared in many shrimp farms in South-East Asia, the Americas, Central Asia, the Middle East and Australia, causing large scale losses in the shrimp rearing industry (Oakey et al., 2019). WSSV has a double-stranded DNA genome with a size range between 281 and 312 kbp, containing at least 184 major open reading frames. High Throughput Sequencing data shows that the virus has undergone progressive recent evolution including mutations and gene deletion, suggesting that WSSV is adapting to replication in the shrimp (Oakey and Smith, 2018; Zwart et al., 2010).

The relationship of WSSV to other DNA viruses and thus its evolutionary origin remains unclear. Indeed, the *Nimaviridae* family is currently unassigned in higher order viral taxonomy. At present, WSSV is classified taxonomically as lying closest to the *Herpes*- and *Asfavirus* families (Wang et al., 2019), based mostly on the relationship of the polymerase gene products. The polymerase gene of WSSV was identified by (van Hulten et al., 2001; Chen et al., 2002) based on presence of three conserved motifs in the exonuclease domains of viral proteins (Bernad et al., 1989) and seven in viral polymerase domains (Larder et al., 1987). The polymerase gene product of WSSV is unusual in that it encodes around 2000 amino acids, almost twice as many as the herpes simplex virus 1 (HSV-1) and the African swine fever virus (ASFV) polymerases. As a consequence (Chen et al., 2002), recognised that the sequences separating the conserved regions were much longer in the WSSV enzyme than in other viral polymerases from the seven viral families that they examined.

Given that many of the WSSV proteins appear unrelated to known proteins, we have commenced a project to examine whether AlphaFold would be able to predict structures of WSSV proteins and thus help to illuminate their function. To initiate this study, we commenced with the DNA polymerase protein as, on the one hand, the presence of the conserved sequence motifs would serve as an internal control, whilst, on the other hand, the higher number of amino acids present and the lack of identity to other polymerases suggested unusual structural motifs which

* Corresponding author. Max Perutz Labs, Vienna Biocenter Campus (VBC), Dr.-Bohr-Gasse 9, Vienna Biocenter 5, 1030, Vienna, Austria.
*E-mail address:* timothy.skern@muv.ac.at (T. Skern).

https://doi.org/10.1016/j.virol.2026.110818
Received 14 November 2025; Received in revised form 14 January 2026; Accepted 26 January 2026
Available online 28 January 2026
0042-6822/© 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

might be identified by AlphaFold. In addition, a huge body of work on the structure of DNA polymerases is available which can be employed to examine the quality of the models (Steitz, 1999), including the structures of the DNA polymerases of HSV-1, ASFV and mpox virus (MPXV) (Gustavsson et al., 2024; Kuai et al., 2024; Xu et al., 2023).

## 2. Material and methods

### 2.1. Sequence identifiers

The GenBank accession number of the WSSV DNA polymerase amino acid sequence used in the modelling was AAL88907, derived from the nucleotide entry in Genbank with the accession number AF440570 (Leu et al., 2005). We chose this sequence for our studies because it appears to be that of the earliest known isolate of WSSV and because it is also the sequence used by (Chen et al., 2002) in their paper on the WSSV DNA polymerase where it is referred to as the WSSV Taiwan-1 strain. To examine the extra 87 amino acids found in many WSSV genome sequences, we used the sequence with the GenBank accession number AYV99445. This sequence has 100% amino acid identity with the polymerase sequence of the WSSV Taiwan-1 strain, except that it is predicted by the ORF finder used to have 87 additional amino acids at the N-terminus.

### 2.2. Protein structure identifiers

The pdb identifiers of the structures selected to validate the WSSV model are: HSV-1 polymerase, 8ojc and 8oj7; ASFV polymerase, 8ywi; MPXV, 8hoy. All pdb identifiers are publicly available.

### 2.3. Model generation of WSSV polymerase with AlphaFold

Models were generated in May 2025 using the above-mentioned protein sequences on the AlphaFold Server (Abramson et al., 2024). Models were examined for quality using the pLDDT parameters and visualized using PyMOL (The PyMOL Molecular Graphics System, Version 2.4.1 Schrödinger, LLC.; Retrieved from http://www.pymol.org/pymol). Coordinates of the five modelled structures are available via supplementary files 1 to 5.

### 2.4. Comparison and superposition of modelled three-dimensional structures

Comparisons of superpositions of three-dimensional structures were performed using PDBeFold (Krissinel and Henrick, 2004, 2005) and the Dali server (Holm, 2022). We used a local alignment approach to detect related sub-domains of the polymerase proteins, using the rmsd (root mean square deviation) as the metric for the relationship of the $C_\alpha$ atoms.

## 3. Results

### 3.1. An AlphaFold model of the WSSV polymerase

To examine the relationship of the WSSV polymerase to other viral polymerases, we first performed a standard protein blast search of the GenBank database. The search failed to show any identity to other proteins in the database, even when all WSSV sequences in the database were excluded from the search. We therefore entered the protein sequence into the AlphaFold algorithm to investigate the structure of the WSSV polymerase. The algorithm generates as default co-ordinates for

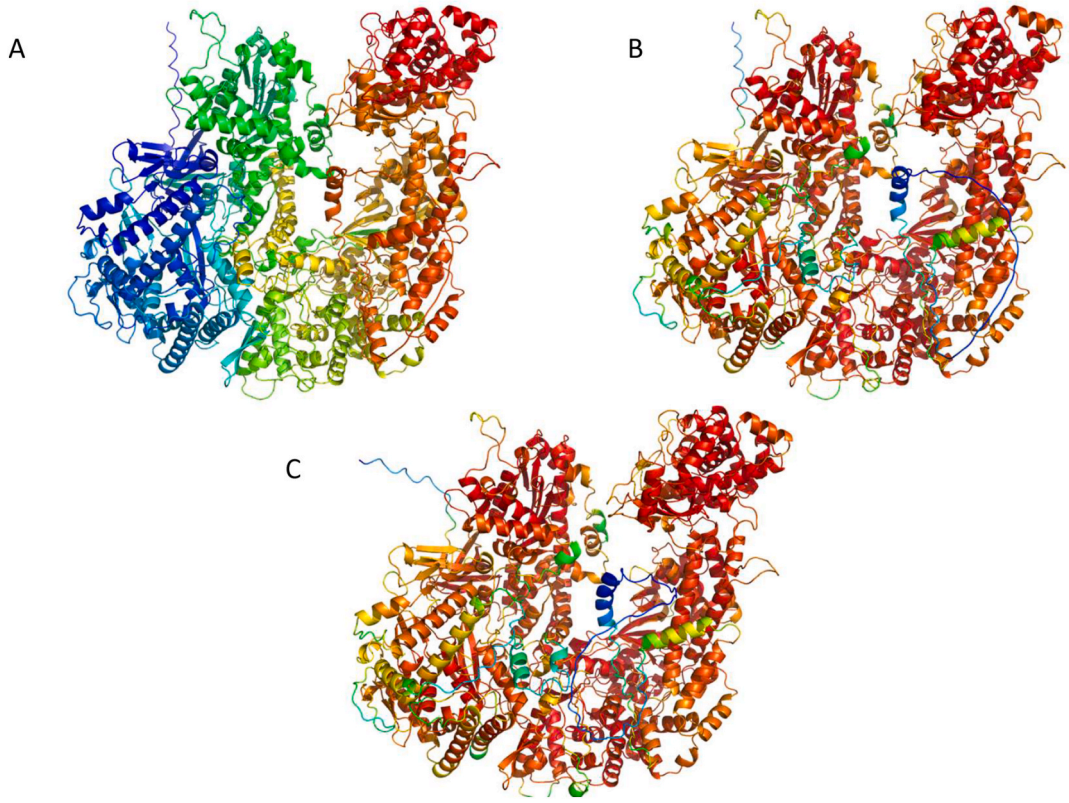



**Fig. 1.** Structural models of the WSSV polymerase
(A) Structural drawing of AlphaFold model 0 of the WSSV polymerase colored as a spectrum from the N-terminal (blue) to the C-terminal (red). (B and C) Models 0 (B) and 4 (C) of the WSSV polymerase colored according to the predicted pLDDT values (highest values red, lowest values blue). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

five models (numbered 0 to 4) for a particular sequence; Fig. 1A shows model 0 generated from the WSSV DNA polymerase sequence, colored as a spectrum from the N-terminal (blue) to the C-terminal (red). To assess the quality of the model, Fig. 1B shows model 0 colored as a spectra using the predicted local Distance Difference Test (pLDDT) values (Jumper et al., 2021). High pLDDT values (colored red) indicate strong confidence in the model whereas low values (blue) indicate low confidence. Overall, the pLDDT values for the model 0 are high, with just a few particular areas with lower values. The lowest values are found in a noticeable loop and helix region towards the C-terminus, most probably reflecting an inability of AlphaFold to model this part of the protein. Indeed, this region is the most varied part of the models generated by AlphaFold, as can be seen in Fig. 1C which shows model 4 similarly colored by the pLDDT values. The loop in blue is in a completely different position. Images of the other three models colored according to pLDDT values are given in the supplementary figures. Furthermore, an image of all five models superposed on each other is shown in the supplementary figures together with a table of the rmsd values of the superpositions and an enlargement of the image of area of the loop and helix region at the C-terminus. All pLDDT values for all five models color-coded according to their domain position are given in a supplementary table.

Given the overall similarity of the models as evidenced by the above comparisons, we arbitrarily chose model 0 for further analysis. The first question that we wished to answer was whether the structure resembled that of a viral polymerase, given the exceptional number (2351) of amino acids present in the WSSV polymerase polypeptide (Chen et al., 2002). The general structure of DNA polymerases has been described by Steitz as that of a right hand (Ollis et al., 1985; Steitz, 1999), recognizable by palm, finger and thumb domains. The images in Fig. 1 indeed fit this description. To investigate the relationship of the model of the WSSV polymerase in more depth, we searched the Dali server (Holm, 2022) to look for structural homologs. Table 1 shows fifteen structures with Z scores higher than 12.0 from the Dali search across the full PDB25 dataset. The PDB25 dataset is a subset of 26414 chains (as of November 2, 2025; http://ekhidna2.biocenter.helsinki.fi/barcosel/tmp//pdb25.list) in the PDB generated by using a 25% sequence identity threshold (Holm, 2022). We employed this dataset to reduce the number of hits to structures of the same molecule. Examination of Table 1 reveals that all of the structures identified are DNA polymerases. Three represent polymerases from DNA viruses, namely ASFV, HSV-1 and MPXV whereas two structures are from the *E. coli* bacteriophage polymerase RB69. The detection of the relationship of the model structure to the polymerases of these three viruses increased our confidence in the model of the WSSV polymerase. Furthermore, the presence of the structure of the RB69 bacteriophage which has been taken as a prototype of type B DNA polymerases (Steitz, 1999; Wang et al., 1997) was a strong indication that the model structure was robust. The DNA polymerases of ASFV, HSV-1 and MPXV all show DNA structures of type B (Gustavsson et al., 2024; Kuai et al., 2024; Xu et al., 2023).

The coordinates determined by the DALI server for the superposition of these three viral structures were downloaded and the superpositions visualized in PyMOL (Fig. 2) The superposition of the HSV-1 polymerase onto the model of the WSSV polymerase (Fig. 2A) was the most straightforward to interpret, as it mostly concerned one domain whereas that of the ASFV enzyme seemed to be spread over two or three domains. The MPXV polymerase superposition was the poorest, in line with its lower Z score. The MPXV polymerase has an extra domain at the bottom of the palm domain compared to the HSV-1 and ASFV enzymes; this domain does not overlap with any domains of the WSSV model.

The part of the HSV-1 polymerase shown in Fig. 2A spans residues 59 to 616, containing the N-terminal and exonuclease domains; the complete domain structure of the HSV-1 protein is shown in Fig. 3A (Gustavsson et al., 2024). The structural alignment from the Dali server revealed the strongest alignments between residues 682 to 1024 of the WSSV polymerase model and residues 362 to 616 of the HSV-1 polymerase, corresponding approximately to the exonuclease domain. This implied that some or all of residues 682 to 1024 of the WSSV polymerase domain represent its exonuclease domain.

The above comparisons confirmed that the WSSV model can be considered to resemble a viral DNA polymerase with an exonuclease domain related to that of HSV-1. During our analysis, we noted that the HSV-1 DNA polymerase structure identified by DALI, 8ojc (Gustavsson et al., 2024), contained several unstructured regions and did not encompass the entire enzyme. As the PDB25 database only generally contains one structure of a particular protein for a particular organism (Holm, 2022), we searched the entire PDB with the entry 8ojc for further, more complete structures of the HSV-1 DNA polymerase and found a set of entries from the same authors as 8ojc (Gustavsson et al., 2024). One of these, with the pdb id 8oj7, spanned 1200 amino acids of the HSV-1 polymerase structure, with only three areas of unmodeled amino acids. We therefore decided to use this structure for further analysis. To confirm and further define the relationship of the exonuclease domains, we used PDBeFold to compare the structures of the exact exonuclease domains of HSV-1 DNA polymerase from the 8oj7 structure (residues 363 to 594) and of residues 682 to 1024 of the WSSV model. The overall rmsd was 2.0 Å over 172 residues (Table 2). The residue corresponding to the last residue of the HSV-1 exonuclease domain in the WSSV enzyme was identified as I1006, indicating that residues 1007 to 1024 are not part of the WSSV exonuclease but part of the small N-terminal domain that follows the exonuclease domain as discussed below. Thus, the exonuclease domain of the WSSV DNA polymerase spans the residues 682 to 1006. Fig. 3B shows the folds of the two exonuclease domains; the aligned secondary structures of the fold of this domain are shown in Fig. 3C. In both cases, the HSV1 exonuclease domain is color-coded brown as in Fig. 3A whereas the WSSV domain is

**Table 1**
Top 15 structures (PDB25 subset) from the Dali server that are related to the AlphaFold model of the WSSV DNA polymerase.

| | Protein | Pdb id | Z | rsmd | lali | nres | % id | Comments |
|---|---|---|---|---|---|---|---|---|
| 1 | African swine fever virus | 8ywi-A | 18.2 | 5.0 | 788 | 980 | 14 | |
| 2 | *E. coli* Phage RB69 | 2p5g-B | 17.7 | 4.5 | 611 | 756 | 16 | |
| 3 | *E. coli* Phage RB69 | 3cfo-A | 17.6 | 4.7 | 652 | 906 | 15 | |
| 4 | Human DNA polymerase alpha catalytic subunit | 8d0b-F | 16.7 | 11.1 | 656 | 915 | 14 | |
| 5 | *S. cerevisiae* DNA polymerase epsilon catalytic subunit A | 6s2e-A | 16.6 | 4.6 | 684 | 1006 | 14 | |
| 6 | Entry declared obsolete | 7m5p-A | 16.6 | 5.3 | 759 | 1097 | 15 | |
| 7 | *S. cerevisiae* DNA polymerase epsilon catalytic subunit A | 6wjv-A | 16.3 | 9.7 | 740 | 1951 | 14 | |
| 8 | HSV-1 DNA polymerase | 8ojc-A | 16.3 | 3.3 | 287 | 395 | 17 | |
| 9 | DNA polymerase zeta catalytic subunit | 8tlq-A | 15.9 | 5.1 | 772 | 1283 | 13 | |
| 10 | *Sulfolobus solfataricus* DNA polymerase I | 1s5j-A | 15.7 | 4.5 | 615 | 727 | 16 | |
| 11 | *E. coli* DNA polymerase II | 3k5o-B | 15.7 | 5.2 | 695 | 744 | 15 | |
| 12 | Human DNA polymerase alpha catalytic subunit | 7opl-A | 15.3 | 7.7 | 638 | 1070 | 13 | |
| 13 | *Thermococcus gorgonarius* DNA polymerase | 7b0f-A | 15.2 | 4.1 | 666 | 721 | 15 | |
| 14 | Mpox virus DNA polymerase | 8hoy-A | 13.8 | 5.0 | 710 | 1006 | 14 | |
| 15 | DNA binding protein TREX1 Homolog Plex9. | 9mrc-A | 12.2 | 2.9 | 158 | 189 | 15 | |

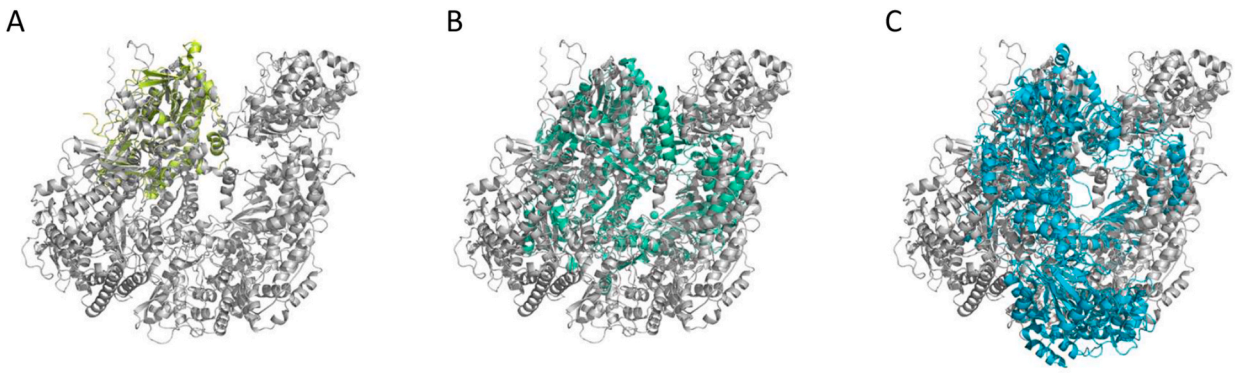



**Fig. 2.** Superposition of the WSSV polymerase model on three viral polymerases as determined by the DALI server
Model 0 of the WSSV polymerase is shown in gray80. The domains of the polymerases determined to overlap by the DALI server are those of HSV-1 (A, colored lemon, pdb id 8ojc), ASFV (B, colored greencyan, pdb id 8ywi) and MPXV (C, colored cyan, pdb id 8hoy). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

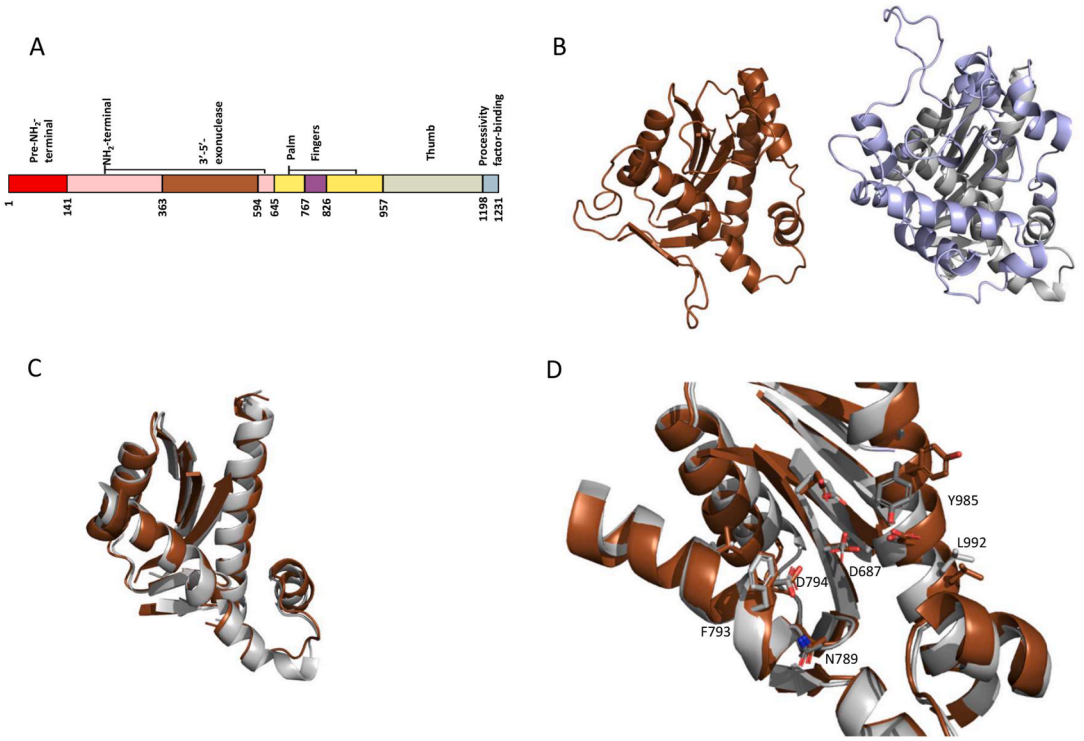



**Fig. 3.** Relationship of the WSSV polymerase model exonuclease domain to that of the HSV-1 polymerase (A) Schematic domain structure of the HSV-1 DNA polymerase (redrawn after (Gustavsson et al., 2024)). (B) Comparison of the exonuclease domains of HSV-1 (colored brown, pdb id 8oj7) and WSSV (aligned regions to the HSV-1 colored gray80, non-aligned regions in light blue) oriented as in Fig. 1. (C) Superposition of the aligned exonuclease folds colored as (B). (D) As (C), but rotated −50°on the y axis and zoomed to show the alignment of the conserved amino acids. The side-chains of the conserved amino acids between the HSV-1 and WSSV polymerases (see Table 4) are shown as sticks with oxygen atoms red, nitrogen blue and carbon brown or gray, respectively. Selected residues of the WSSV exonuclease domain are labelled. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

in gray; non-aligned residues in the WSSV structure are colored in light blue in Fig. 3B. The large non-aligned extensions in the WSSV polymerase give rise to the presence of 94 more residues in this domain than the HSV-1 enzyme (Table 3).

The exonuclease domains of viral polymerases were characterized by Salas et al. (Bernad et al., 1989) as having three blocks of conserved sequences at the active site. Table 4 shows the conserved amino acids identified by the group of Salas, along with the corresponding amino acids in the WSSV exonuclease domain that we were able to identify. All of these residues were also identified by Chen et al. (2002) as being conserved in viral DNA polymerases. The image in Fig. 3D shows that the $C_\alpha$ atoms of these residues superpose, although in some cases the side-chains in the model of WSSV are displayed in a different rotamer to those seen in the HSV-1 structure. All residues identified in the active site of the exonuclease domain have high pLDDT values, giving further confidence that the model represents a viral polymerase.

To further investigate the relationship of the two polymerases, we were able, using the superposition of the exonuclease domains, to superpose the full structure of the HSV-1 polymerase (Fig. 4A, colored according to the domains shown in Fig. 3A) onto the model of the WSSV polymerase (Fig. 4B). This allowed us to search for further domains in the WSSV polymerase that were similar in structure to the HSV-1 molecule. Close examination of Fig. 4B showed overlap in the N-terminal domain (pink) of some β-strands as well as one α-helix and β-strands

**Table 2**
Parameters for the superposition of the WSSV model on the indicated domains and DNA polymerases.

| Domain | HSV-1 | | ASFV | | *MPXV* | |
|---|---|---|---|---|---|---|
| | Rmsd (Å) | No aligned residues | Rmsd (Å) | No aligned residues | Rmsd (Å) | No aligned residues |
| N-terminal part 1 | 4.1 | 130 | 3.5 | 140 | 3.7 | 108 |
| Exonuclease | 2.0 | 172 | 2.0 | 182 | 2.3 | 175 |
| N-terminal part 2 | 2.0 | 39 | 1.78 | 39 | 1.7 | 16 |
| Palm and Fingers | 1.8 | 194 | 2.3 | 203 | 2.4 | 213 |

**Table 3**
Summary of the residue counts of the polymerase domains of the four viral DNA polymerases examined. Values for HSV-1, ASFV and MXPV are taken from (Gustavsson et al., 2024; Kuai et al., 2024; Xu et al., 2023), respectively.

| Domain | WSSV | | HSV-1 | ASFV | MXPV |
|---|---|---|---|---|---|
| | Residues | Residues (n) | Residues (n) | | |
| Pre-terminal | 1–17 | 17 | 141 | | |
| N-terminal part 1 | 18–681 | 664 | 222 | 205 | 157 |
| Exonuclease | 682–1006 | 325 | 231 | 230 | 339 |
| N-terminal part 2 | 1007–1049 | 43 | 51 | 49 | 27 |
| Palm and Fingers | 1050–1589 | 540 | 302 | 306 | 307 |
| Thumb | 1590–2351 | 762 | 241 | 205 | 177 |
| Processivity binding factor | | | 37 | | |

**Table 4**
Conserved residues in the active sites of the exonuclease and polymerase domains of the HSV-1 and WSSV polymerases.

| | Exonuclease domain | Polymerase domain | |
|---|---|---|---|
| HSV-1 | WSSV | HSV-1 | WSSV |
| D368 | D687 | D717 | D1079 |
| E370 | E688 | L721 | L1083 |
| N466 | N789 | Y722 | Y1084 |
| F470 | F793 | P723 | P1085 |
| D471 | D794 | K811 | K1417 |
| L475 | V798 | N815 | N1421 |
| Y577 | Y985 | S816 | S1422 |
| D581 | D989 | Y818 | Y1424 |
| L584 | L992 | F820 | H1426 |
| | | D886 | D1499 |
| | | T887 | T1500 |
| | | D888 | D1501 |
| | | S889 | S1502 |

in the palm and fingers domain (yellow and purple). In contrast, there appeared to be little identity either with the pre-N-terminal region (red) or with the thumb region (green).

We first focused our attention on the palm and fingers domain, that is the domain that binds and incorporates the next nucleotide into the growing chain, as we predicted that these domains would be conserved in a similar way to the exonuclease domains. To this end, we made a pdb file of the palm and fingers domain of the HSV-1 polymerase (residues 691 to 957, residues 644 to 690 are disordered in the 8oj7 structure) and examined the WSSV residues that could be superposed to it. In this way, we could identify residues 1050 to 1589 as the palm and fingers domain of WSSV. The superposition is shown in Fig. 4C. Satisfyingly, the fingers domains of the two structures were well aligned, especially N- and C-termini at the start and end of the helices, allowing us to define the finger domain of WSSV as residues 1369 to 1430.

There are large extensions in the WSSV enzyme between residues 1147 and 1361 as well as 1092 and 1140. A comparison of the two structures with PDBeFold showed an rmsd of 1.8 Å over 194 residues (Table 2) with 10 aligned secondary structure elements.

Fig. 4D shows the overall domains of both palm and fingers domains without the extensions noted above. To confirm the above assignment of the WSSV palm and fingers domain, we again searched for conserved active site amino acids identified by (Bernad et al., 1987). Table 4 shows these amino acids along with the corresponding amino acids in the WSSV palm and fingers domain that we were able to identify; again, all these residues were noted as being conserved by (Chen et al., 2002). The image in Fig. 4E shows that the $C_\alpha$ atoms of these residues superpose, again with some differences in the rotamers. The presence of the dATP moiety and the magnesium atom from the 8oj7 structure of HSV-1 in the superposition in Fig. 4D underscore the presence of the catalytic residues. The conserved residues at the C-terminus of the second of the characteristic helices of the fingers domain (K1417, N1421, S1422 and Y1424) are of particular interest as these four amino acids represent the only sequence identity at the primary sequence level in the fingers region of the two polymerases. Nevertheless, AlphaFold predicts two finger-like helices, even if their orientation is somewhat different to those of the HSV-1 enzyme.

Returning to the overall fold comparison in Fig. 4B, we now turned our attention to the N-terminal domain, which in the DNA polymerases of HSV-1, is divided in two parts, interrupted by the exonuclease domain (Fig. 3A). The folds for the HSV-1 and WSSV N-terminal domains are shown in Fig. 5A. The data for the superposition for part 1 (N-terminal to the exonuclease domain) are much poorer (4.1 Å over 130 residues, Table 2) than for the exonuclease and palm and fingers domain. This is reflected in the limited number of superposed secondary structure elements seen in the superposition in Fig. 5A, which demonstrates the structure of the larger part of the HSV-1 N-terminal domain and the structurally equivalent parts of the WSSV polymerase. Nevertheless, there is clearly some equivalence in this region, shown with the superposition with Fig. 5B. In the absence of further equivalence, we determined the N-terminal end of the WSSV N-terminal domain to be residue 18 and the C-terminal end to be residue 681. This domain in WSSV contains 664 residues, 442 more than the HSV-1 enzyme.

Examination of the second, smaller part of the N-terminal domain which bridges the exonuclease domain with the polymerase domain (Fig. 5C) shows that three of the four predicted helices of the WSSV enzyme overlap with those of the HSV-1 enzyme, suggesting that the structure of this region must be conserved to position the exonuclease domain relative to that of the polymerase domain, as can be seen in Fig. 4B. Comparison with PDBeFold shows an rmsd of 2.0 Å over 39 residues (Table 2). This domain of the HSV-1 enzyme comprises 51 residues whilst the WSSV enzyme has 43 (residues 1007–1049; Table 3).

### 3.2. An N-terminal extension in the polymerase protein of certain WSSV strains

At present, the exact N-terminus of the WSSV DNA polymerase has not been exactly defined at the protein level. In some sequences, algorithms that search for open reading frames have identified an extra 87 amino acids at the N-terminus of the WSSV DNA polymerase (for example, the GenBank accession: AYV99445.1). We were interested to see whether this extra sequence could show a defined structure. We therefore also analyzed the protein product of AYV99445 with AlphaFold. However, these amino acids were modelled as a loop with very low pLDDT. Consequently, we did not incorporate them into any of our models.

### 3.3. Regions lacking equivalence

The above results show that we were able to find structural equivalence with the N-terminal, exonuclease and palm and finger domains of

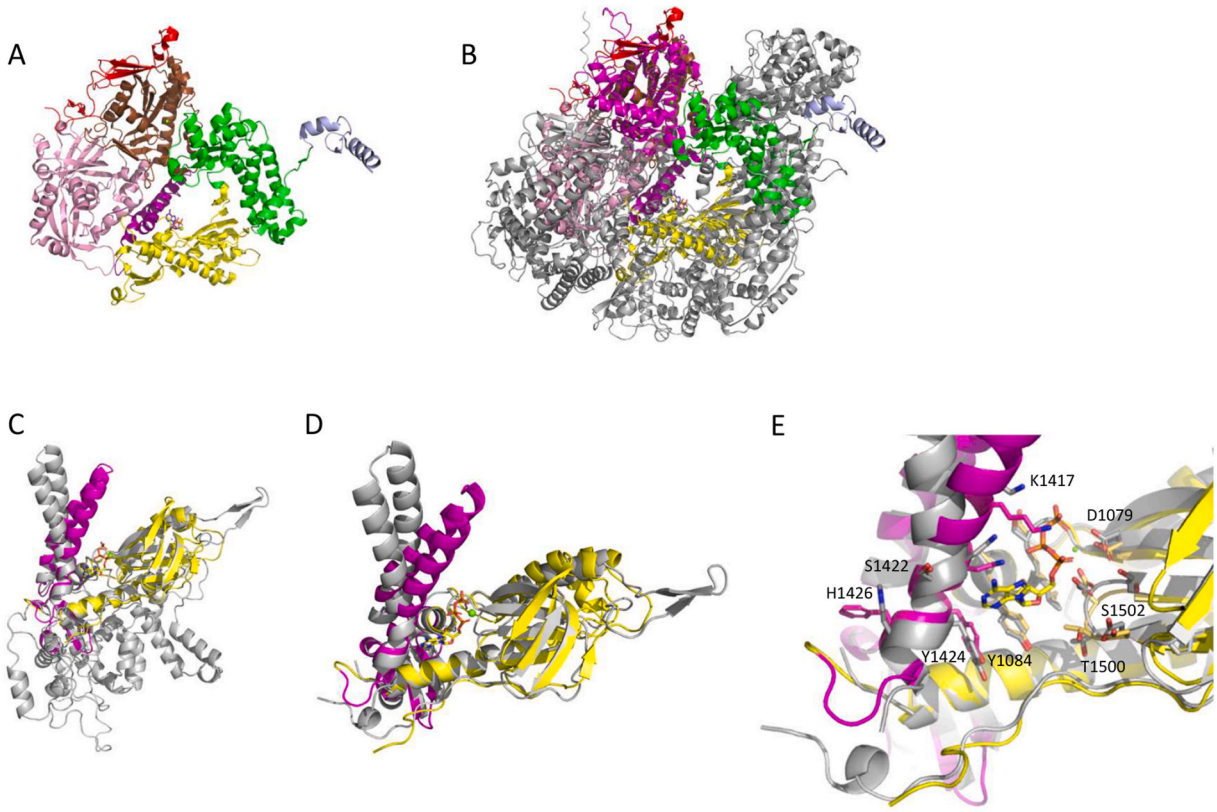



**Fig. 4.** Superposition of the WSSV DNA polymerase model onto the structure of the HSV-1 DNA polymerase and relationship of the palm and fingers domains (A) The structure of HSV-1 DNA polymerase (8ojc) colored according to the scheme in Fig. 3A. (B) Superposition of the WSSV DNA polymerase model (colored as gray80) superposed onto the structure of the HSV-1 DNA polymerase (colored as in (A)) using the exonuclease domains for the superposition. (C) Superposition of the palm and fingers domains of the two structures. The HSV-1 domain is colored yellow (palm) and purple (fingers), according to Fig. 3A; the WSSV domain is colored gray80. The dATP and $Mg^{2+}$ ion from the HSV-1 are shown as sticks (carbon, yellow, oxygen, red, nitrogen, blue, phosphorus, orange) and a green sphere, respectively. (D) The common fold of the two palm and fingers domains, colored as (C). (E) Zoom of (D) with the side-chains of the conserved amino acids (see Table 4) shown as sticks with oxygen atoms red, nitrogen blue and carbon yellow or purple for HSV-1 and gray for WSSV. Selected residues of the WSSV polymerase are labelled. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

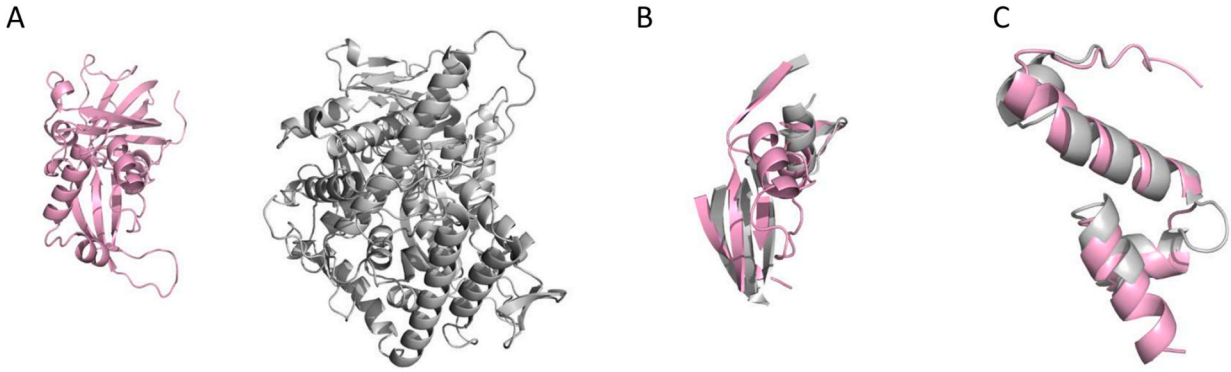



**Fig. 5.** Comparison of the N-terminal domains of HSV-1 and WSSV. (A) Structural models of the larger of the N-terminal domains of the HSV-1 (pink) and WSSV (gray80) polymerases. (B) Superposition of the common fold of the two larger N-terminal domains. (C) Superposition of the common fold of the two smaller N-terminal domains (zoomed for clarity). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the model of the WSSV polymerase to the HSV-1 polymerase structure with the pdb id 8oj7. We were however unable to find any equivalence to the domains we considered as the pre-N-terminal domain (residues 1–17) and the thumb domain (residues 1590 to 2351). Further, there was no sign of a processivity factor binding domain found at the C-terminus of the HSV-1 polymerase (Fig. 3A).

The thumb domain showed not only the greatest structural divergence but also the greatest difference in the number of residues, namely 762 compared to 241 in HSV-1 (Table 3). The thumb domain (Fig. 6A) of the WSSV enzyme is modelled by AlphaFold as two, largely helical domains joined by 110 residues (1907–2106) featuring a long flexible loop and two helices. The two helices show high pLDDT values whereas the 110 residues joining them show low pLDDT values. This arrangement of the thumb domain is quite different to that of the HSV-1 polymerase (Fig. 6A). Of the 110 residues, 28 are acidic, that is either aspartate or glutamate, and 13 are glycine (Fig. 6B). We also used the DALI server to

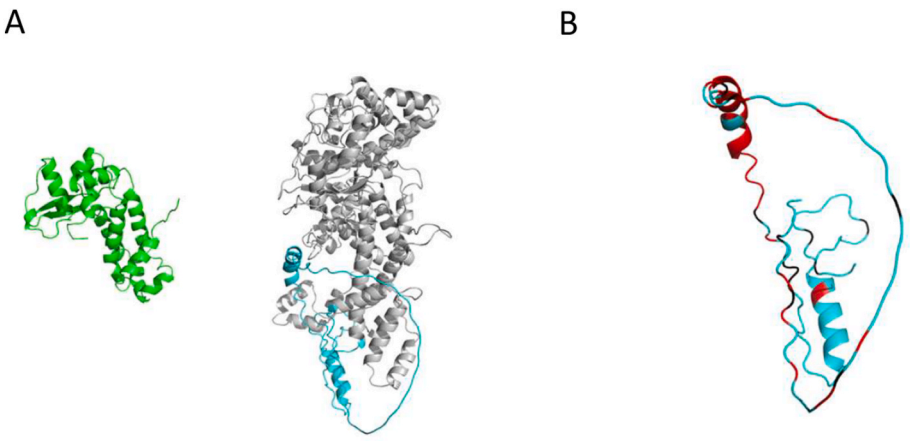



**Fig. 6.** The unusual thumb domain of WSSV DNA polymerase. (A) Structural models of the thumb domains of HSV-1 (green) and WSSV (gray80) DNA polymerases. The flexible loop und helix region of the WSSV DNA polymerase is in cyan. (B) The acidic nature of the flexible linker and helix region of the WSSV thumb domain. Acidic residues are marked in red, glycine residues in black, the model is rotated slightly relative to A so show the coils near the lower helix. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

search for structural identities of the pre-N-terminal and thumb domains without success. Searches using the thumb domain in its entirety or as separate domains were negative.

### *3.4. Relationship with AFSV and MPXV polymerases*

Having thoroughly examined the relationship of the WSSV polymerase to the HSV-1 polymerase, we returned to the relationship with the ASFV and the MPXV polymerases as revealed in the original DALI search. Fig. 7 compares the structures of the exonuclease (Fig. 7A), palm and fingers (Fig. 7B) and N-terminal part 2 (Fig. 7C) domains of the HSV-1, ASFV and MPXV polymerases with those of the WSSV polymerase model. Tables 2 and 3 show respectively the details of the superpositions of the domains and the variation in the number of residues in each domain of the four viral polymerases.

## 4. Discussion

The use of the AlphaFold algorithm to model the WSSV DNA polymerase allowed us to expand on the work of (Chen et al., 2002; van Hulten et al., 2001) to identify the individual domains of the enzyme and visualize their structures. Fig. 8 shows these domains of the WSSV polymerase color-coded according the scheme used in Fig. 3A for the HSV-1 enzyme. Chen et al., (2002) showed that the conserved polymerase regions in the WSSV enzyme were spaced differently to all the other DNA polymerases that they examined. The work here defines the limits of each of the domains and illustrates their putative structures. The highest levels of equivalence are found in the exonuclease and polymerase domains whereas that of the N-terminal domains is the lowest. There is no equivalence in the thumb domain and the pre-N-terminal domain is apparently absent. Furthermore, this domain is also not

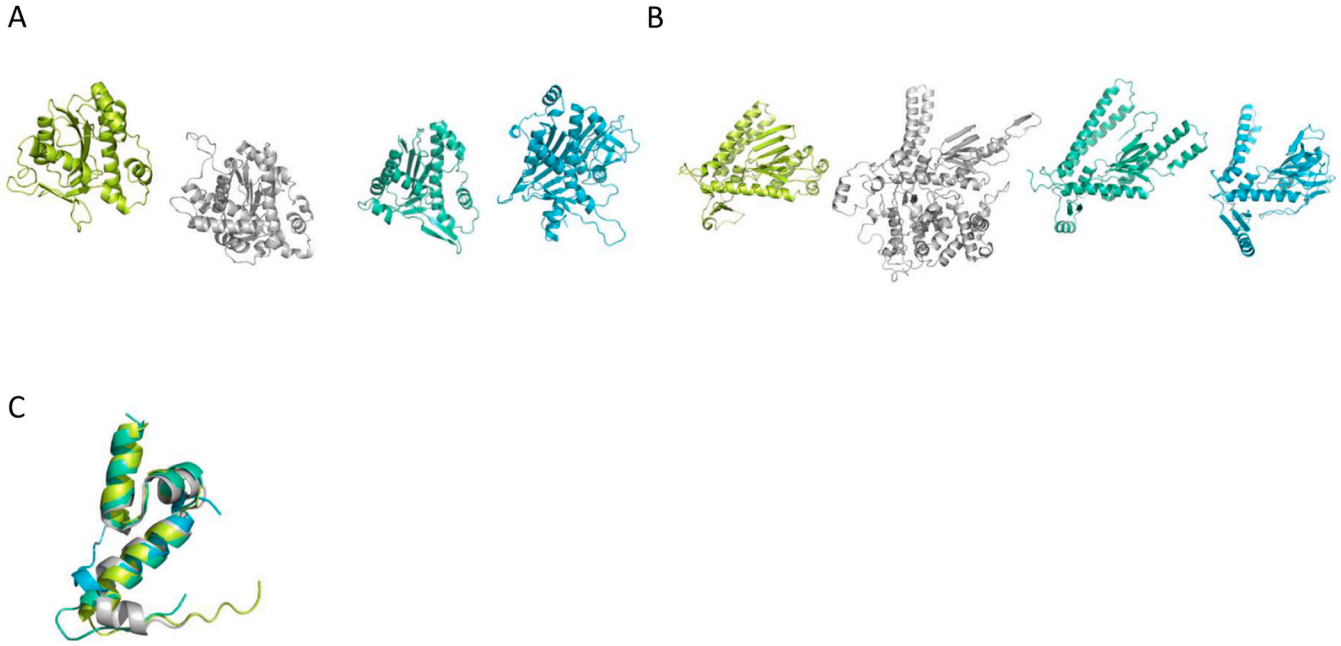



**Fig. 7.** Comparison of the domain structures of the DNA polymerases of HSV-1, ASFV and MPXV with the structural model of the WSSV DNA polymerase
Structural models of the DNA polymerases of HSV-1 (lemon), WSSV (gray80), ASFV (greencyan) and MPXV (cyan). (A) Exonuclease domains. (B) Palm and fingers domain. (C) Smaller parts of the N-terminal domain. The domains are oriented as in Fig. 1. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

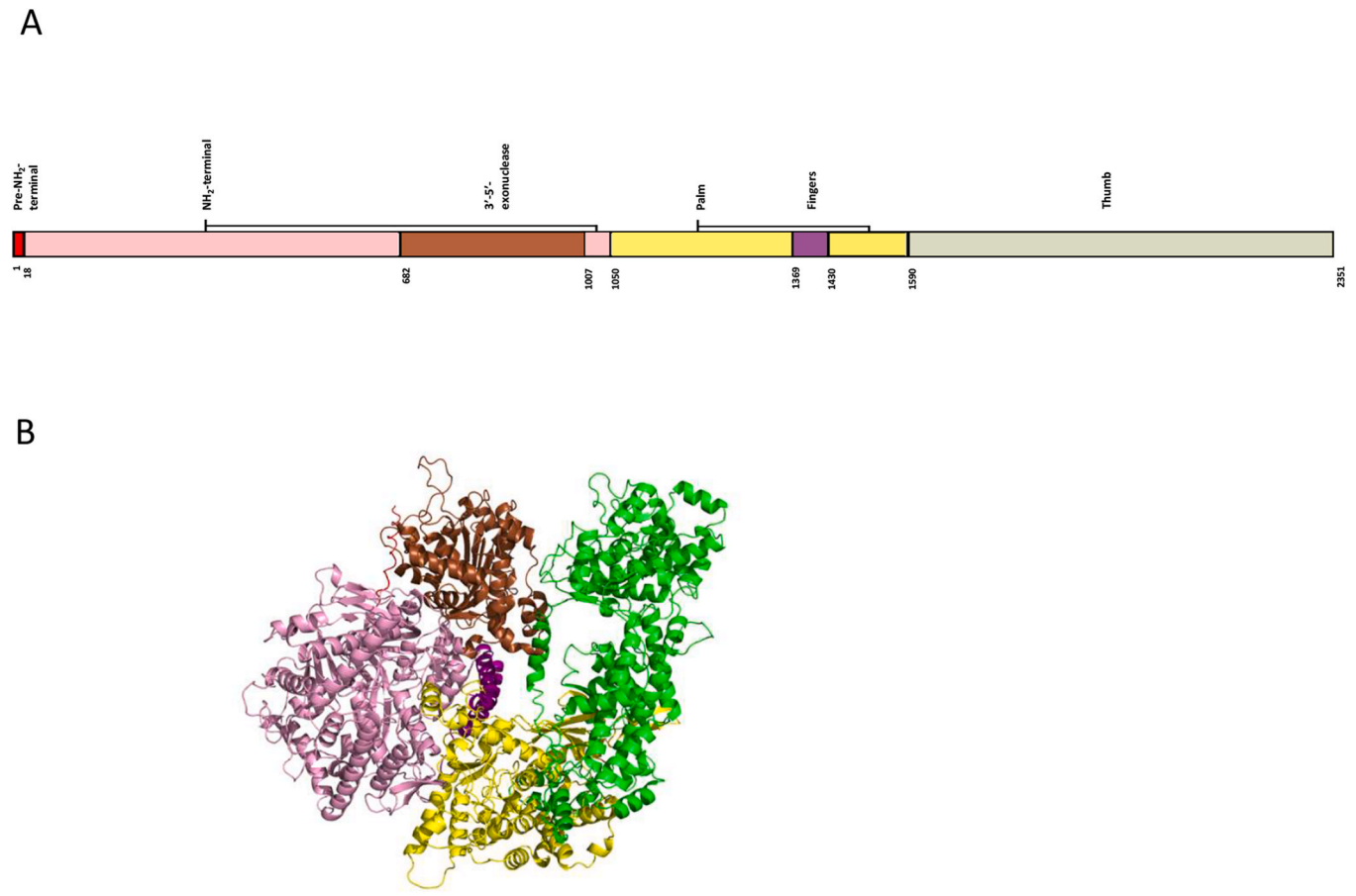



**Fig. 8.** Domain structure of the WSSV DNA polymerase (A) Schematic domain structure of the WSSV DNA polymerase. (B) PyMOL drawing of the model of WSSV DNA polymerase colored as in (B).

detected in WSSV protein sequences that show an 87 amino acid extension on the N-terminus of the DNA polymerase. For our analysis, we used the sequence of one of the first strains to have been isolated. As WSSV has been reported to have undergone a quite rapid genome reduction coupled with other mutational events since its entry into shrimps, we compared 55 sequences of the WSSV DNA polymerase generated over the last 30 years and noted only seven sites with a polymorphism and one site with a conserved mutation that does affect the amino acid sequence. We are confident therefore that our model is not affected by any temporal mutations that might have occurred during the adaptation to shrimps.

The annotation of the domains shows that AlphaFold modelling can be used to propose the structures of viral proteins, even when the molecule comprises more than twice as many amino acids as the equivalent proteins of other viruses and contains domains that appear unrelated to other viral proteins. Further, the modelling indicates in which regions the extra amino acids are located and illustrates which domains are not equivalent to the known domains. This is especially the case with the thumb domains of the WSSV enzyme being quite unrelated to those of the three viral polymerases that we examined as well as the RB69 bacteriophage enzyme which has been taken as a prototype of type B DNA polymerases (data not shown).

What could be the function of the unusually large and unusually modelled C-terminal thumb domain and the extended N-terminal domain? We searched the DALI server separately with each of the three parts of the thumb domain but failed to find any significant structural identity with other protein structures. The two structured regions of the thumb domain could be involved in protein-protein interactions during DNA replication. For the loop domain with its two helices, the lack of secondary structure elements indicates that it is a non-structured region of the protein which may potentially become structured upon binding to an interaction partner. Compared to the other viruses examined here, the WSSV infects the penaeid shrimp, an organism that has larval stages. It therefore seems possible that the WSSV polymerase must be capable of interacting with different proteins, depending on the life-cycle stage of the shrimp, suggesting that the extensions on the N-terminal domain and the presence of the large thumb domain may provide platforms for binding proteins characteristic for each life cycle stage. The role of the N-terminal domain in the HSV-1 enzyme appears unclear but there are indications that it may be involved in DNA binding (Gustavsson et al., 2024). The thumb domain of DNA polymerases has been observed to play roles in binding the DNA, processivity of synthesis and translocation following synthesis (Steitz, 1999). It seems therefore plausible that the unusual thumb domain of may be involved in some or all of these activities. Seitz also noted that the thumb domains of polymerase I, polymerase α and reverse transcriptase enzymes are different, suggesting that the thumb domain of the WSSV enzyme will have important roles in the above activities. It will thus be of great interest to try to solve the structure of the WSSV polymerase as a whole or of just the C-terminal thumb domain and to search for interaction partners of the polymerase in cells infected with the virus. Together, the unusual structures of the extended N-terminal domain and the thumb domain suggest that the control of DNA replication in WSSV may differ from that in mammalian viruses.

The comparisons of the model of the WSSV polymerase with the known structures of the polymerases of HSV-1, ASFV and MPXV showed appreciable structural equivalences. Examination of Table 2 showed that the lowest rmsd between the exonuclease and palm and fingers domains of the model were with the HSV-1 DNA polymerase. Furthermore, we chose the HSV-1 enzyme as the basis for our comparisons

because it had the lowest rmsd score based on the initial DALI search. The drawings of the three domains of the four polymerases in Fig. 7 also strengthens the notion that the HSV-1 enzyme is structurally closest to that of WSSV. This structural convergence of the WSSV enzyme to the HSV-1 enzyme should not however be taken to imply any taxonomic relatedness. Indeed, the phylogenetic tree of Chen et al. (2002) which, based on genome sequences, places the WSSV polymerase closer to the enzymes of the *Poxviridae* and *Asfarviridae* families.

We have been careful not to over-interpret our analysis and comparisons with the model structure of the WSSV enzyme as we are aware of certain weaknesses of the AlphaFold algorithm. It is possible that AlphaFold was trained for instance more on the HSV-1 DNA polymerase structure which would bias the modelling towards that enzyme. Further, the AlphaFold is known to have a weakness in loop prediction and parts of polypeptides that are unstructured, as exemplified by the different modelling of the long flexible loop in the thumb domain of the WSSV polymerase. Consequently, we did not base any inferences on this region or on any regions with long loops.

## 5. Conclusion

Together, our work demonstrates that the modelling of the DNA polymerase protein of WSSV with AlphaFold can illuminate its relationship to other viral polymerases, even though the amino acid identity is extremely low. The similarity to the polymerases of three DNA viruses supports the designation of this gene product as the DNA polymerase. The differences in the N-terminal and thumb however indicate a different evolutionary trajectory for the WSSV enzyme. It will thus be of interest to confirm the model of the WSSV enzyme by structural studies and then search for viral polymerases that show such non-canonical thumb domains and to elucidate the role of the flexible acidic loop that connects the two domains. Such studies may also aid in defining the origin of the WSSV and the organism from which it jumped into shrimps around 1991.

## CRediT authorship contribution statement

**Tim Skern:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis, Conceptualization. **Jane Oakey:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization.

## Funding statement

TS was supported by core funding from both the Medical University of Vienna and the University of Vienna.

## 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 data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2026.110818.

## Data availability

The data is available in the supplementary information.

## References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Zemgulyte, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A.I., Cowie, A., Figurnov, M., Fuchs, F.B., Gladman, H., Jain, R., Khan, Y.A., Low, C.M.R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E.D., Zielinski, M., Zidek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., Jumper, J.M., 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630, 493–500.

Bernad, A., Blanco, L., Lazaro, J.M., Martin, G., Salas, M., 1989. A conserved 3'——5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. Cell 59, 219–228.

Bernad, A., Zaballos, A., Salas, M., Blanco, L., 1987. Structural and functional relationships between prokaryotic and eukaryotic DNA polymerases. EMBO J. 6, 4219–4225.

Chen, H., 1995. Current status of shrimp aquaculture in Taiwan. In: Browdy, C.L., Hopkins, J.S. (Eds.), Swimming Through Troubled Water. Proceedings of the Special Session on Shrimp Farming. Aquaculture '95. World Aquaculture Societ, Baton Rouge, Lousiana, pp. 29–34.

Chen, L.L., Wang, H.C., Huang, C.J., Peng, S.E., Chen, Y.G., Lin, S.J., Chen, W.Y., Dai, C. F., Yu, H.T., Wang, C.H., Lo, C.F., Kou, G.H., 2002. Transcriptional analysis of the DNA polymerase gene of shrimp white spot syndrome virus. Virology 301, 136–147.

Gustavsson, E., Grunewald, K., Elias, P., Hallberg, B.M., 2024. Dynamics of the Herpes simplex virus DNA polymerase holoenzyme during DNA synthesis and proof-reading revealed by Cryo-EM. Nucleic Acids Res. 52, 7292–7304.

Holm, L., 2022. Dali server: structural unification of protein families. Nucleic Acids Res. 50, W210–W215.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589.

Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D Biol. Crystallogr. 60, 2256–2268.

Krissinel, E., Henrick, K., 2005. Multiple Alignment of Protein Structures in Three Dimensions. Springer, Berlin.

Kuai, L., Sun, J., Peng, Q., Zhao, X., Yuan, B., Liu, S., Bi, Y., Shi, Y., 2024. Cryo-EM structure of DNA polymerase of African swine fever virus. Nucleic Acids Res. 52, 10717–10729.

Larder, B.A., Kemp, S.D., Darby, G., 1987. Related functional domains in virus DNA polymerases. EMBO J. 6, 169–175.

Leu, J.H., Tsai, J.M., Wang, H.C., Wang, A.H., Wang, C.H., Kou, G.H., Lo, C.F., 2005. The unique stacked rings in the nucleocapsid of the white spot syndrome virus virion are formed by the major structural protein VP664, the largest viral structural protein ever found. J. Virol. 79, 140–149.

Oakey, J., Smith, C., 2018. Complete genome sequence of a white spot syndrome virus associated with a disease incursion in Australia. Aquaculture 484, 152–159.

Oakey, J., Smith, C., Underwood, D., Afsharnasab, M., Alday-Sanz, V., Dhar, A., Sivakumar, S., Sahul Hameed, A.S., Beattie, K., Crook, A., 2019. Global distribution of white spot syndrome virus genotypes determined using a novel genotyping assay. Arch. Virol. 164, 2061–2082.

Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G., Steitz, T.A., 1985. Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP. Nature 313, 762–766.

Steitz, T.A., 1999. DNA polymerases: structural diversity and common mechanisms. J. Biol. Chem. 274, 17395–17398.

van Hulten, M.C., Witteveldt, J., Peters, S., Kloosterboer, N., Tarchini, R., Fiers, M., Sandbrink, H., Lankhorst, R.K., Vlak, J.M., 2001. The white spot syndrome virus DNA genome sequence. Virology 286, 7–22.

Wang, H.C., Hirono, I., Maningas, M.B.B., Somboonwiwat, K., Stentiford, G., 2019. ICTV virus taxonomy profile: Nimaviridae. Ictv Report, C. J. Gen. Virol. 100, 1053–1054.

Wang, J., Sattar, A.K., Wang, C.C., Karam, J.D., Konigsberg, W.H., Steitz, T.A., 1997. Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. Cell 89, 1087–1099.

Xu, Y., Wu, Y., Zhang, Y., Fan, R., Yang, Y., Li, D., Zhu, S., Yang, B., Zhang, Z., Dong, C., 2023. Cryo-EM structures of human monkeypox viral replication complexes with and without DNA duplex. Cell Res. 33, 479–482.

Zwart, M.P., Dieu, B.T., Hemerik, L., Vlak, J.M., 2010. Evolutionary trajectory of white spot syndrome virus (WSSV) genome shrinkage during spread in Asia. PLoS One 5, e13400.