The crystal structure of vaccinia virus protein E2 and perspectives on the prediction of novel viral protein folds

William N.D. Gao[1], Chen Gao[1], Janet E. Deane[2], David C.J. Carpentier[1], Geoffrey L. Smith[1], Stephen C. Graham[1]*

[1]Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK

[2]Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, UK

*Correspondence: Stephen C. Graham, scg34@cam.ac.uk

ORCIDs: JED 0000-0002-4863-0330; SCG 0000-0003-4547-4034

Contributions: Conceptualization: GLS, SCG; Data Curation: SCG; Funding Acquisition: JED, GLS, SCG; Investigation: WNDG, CG, DCJC, SCG; Project Administration: GLS, DCJC, SCG; Resources: JED; Supervision: GLS, DCJC, SCG; Visualization: SCG; Writing – Original Draft Preparation: DCJC, SCG; Writing – Review & Editing: JED, GLS, SCG.

Short title: Structure of vaccinia virus E2

Keywords: poxvirus, alphafold2, structural virology, modelling, deep learning

Abstract

The morphogenesis of vaccinia virus (VACV, family *Poxviridae*), the smallpox vaccine, is a complex process involving multiple distinct cellular membranes and resulting in multiple different forms of infectious virion. Efficient release of enveloped virions, which promote systemic spread of infection within hosts, requires the VACV protein E2 but the molecular basis of E2 function remains unclear and E2 lacks sequence homology to any well-characterised family of proteins. We solved the crystal structure of VACV E2 to 2.3 Å resolution, revealing that it comprises two domains with novel folds: an N-terminal annular (ring) domain and a C-terminal globular (head) domain. The C-terminal head domain displays weak structural homology with cellular (pseudo)kinases but lacks conserved surface residues or kinase features, suggesting that it is not enzymatically active, and possesses a large surface basic patch that might interact with phosphoinositide lipid headgroups. Recent deep learning methods have revolutionised our ability to predict the three-dimensional structures of proteins from primary sequence alone. VACV E2 is an exemplar 'difficult' viral protein target for structure prediction, being comprised of multiple novel domains and lacking sequence homologues outside *Poxviridae*. AlphaFold2 nonetheless succeeds in predicting the structures of the head and ring domains with high and moderate accuracy, respectively, allowing accurate inference of multiple structural properties. The advent of highly accurate virus structure prediction marks a step-change in structural virology and beckons a new era of structurally-informed molecular virology.

Vaccinia virus (VACV) is the prototype member of the *Poxviridae*, a family of DNA viruses producing large and complex enveloped virions [1]. The family includes variola virus, the causative agent of the highly infectious and lethal disease smallpox, and several viruses endemic in a variety of animal species, some linked with increasing incidences of zoonotic spread and disease in humans [2–4]. While a concerted vaccination programme led to the WHO declaring smallpox eradicated in 1980, the potential for re-emergence of poxvirus disease remains and only two drugs, TPOXX and Tembexa, are licenced for the treatment of orthopoxvirus infection.

Orthopoxviruses produce two distinct types of infectious virion, mature virions (MVs, also called intracellular mature virions, IMVs) and enveloped virions (EVs, also known as extracellular enveloped virions, EEVs). MVs are formed in cytoplasmic viral factories, where the genome-containing viral core and lateral bodies are wrapped by a single lipid membrane derived from the endoplasmic reticulum [5]. MVs are highly stable and, when released upon cell lysis, can survive in the environment to mediate horizontal spread to new hosts. However, MVs are susceptible to recognition by host adaptive immune response due to the abundance of conserved viral epitopes on their surface, including components of the virus membrane fusion and entry machinery. Prior to cell lysis a proportion of MVs are trafficked on microtubules to sites enriched in trans-Golgi/early endosome derived membranes, where they are wrapped by two additional envelopes to form intracellular enveloped virus (IEV, also known as wrapped virus, WV). These IEVs recruit the cellular kinesin-1 microtubule-associated motor complex to mediate virion trafficking to the cell periphery [6–9], whereupon the outer IEV envelope fuses with the cell membrane to release EVs onto the cell surface and into the extracellular medium. These EVs play an important role in cell-to-cell and systemic spread of infection within a host [10].

During IEV egress at least three viral proteins are involved in the activation of kinesin-1-dependent transport of IEVs. These include an integral membrane protein A36 and two cytoplasmic proteins, F12 and E2. Kinesin-1 is a tetramer of two heavy chains, comprising the microtubule-binding motor domain and a long coiled-coil dimerization domain, and two light chains, each comprising a tetratricopeptide repeat (TPR) domain and a coiled-coil domain that mediates dimerization plus

heavy-chain association. A36 is associated with the outer IEV envelope [11] and possesses two tryptophan acidic (WE/WD) motifs, conserved in cellular kinesin light chain (KLC) binding proteins [12], that associate with a binding groove in the KLC TPR domain [13]. E2 also associates with KLC, binding to the unstructured C-terminal tail present on a subset of KLC isoforms [14, 15]. E2 and F12 function as a complex and both are essential for IEV egress [16]. The E2:F12 complex may associate with IEVs through an interaction between A36 and F12 [17], though the maintenance of E2:F12-mediated IEV egress in the absence of A36 [18] suggests that E2:F12 may utilise additional/alternative interactions to bind IEVs. The molecular basis by which E2 and F12 regulate the recruitment of kinesin-1 and promote microtubule-based IEV trafficking remains poorly understood.

While conserved across poxviruses, VACV E2 lacks identifiable sequence homology to any other protein family. Viral proteins in general, and poxvirus proteins in particular, can maintain structural homology to proteins of known function in the absence of identifiable sequence similarity [19–21]. We therefore sought to solve the structure of E2 from VACV strain Western Reserve. As extensive attempts to express E2 in bacterial (*Escherichia coli*) and insect cell systems were largely unsuccessful, a mammalian expression system was pursued. Small-scale expression tests using transient transfection of codon optimised E2 into human embryonic kidney (HEK)293T cells confirmed successful purification using cobalt affinity chromatography of VACV E2 tagged at the carboxy terminus with a decahistidine tag. Large-scale expression was thus performed via transient transfection of Freestyle 293F cells cultured in suspension and purification via cobalt affinity chromatography and size exclusion chromatography (SEC) followed by anion exchange chromatography, yielding ~0.75 mg of highly pure E2 per L of cultured cells (Figure 1A). Differential scanning fluorimetry (a.k.a. Thermofluor) confirmed that the protein was folded (Figure 1B), the biphasic melt curve of E2 suggesting the presence of two independently-folded domains. SEC analysis with inline multiangle light scattering (SEC-MALS) confirmed that the protein is predominantly monomeric (Figure 1C), the observed molecular mass (92.3 kDa) being close to expected mass as calculated from the sequence (87.5 kDa).

VACV E2 at a concentration of 10.85 mg/mL was subjected to nanolitre crystallisation trials, crystals being obtained when 200 nL of protein was mixed with an equal volume of reservoir solution and equilibrated against an 80 µL reservoir of 0.1 M sodium formate pH 7.0, 12% (w/v) polyethylene glycol 3350 at 20°C. Crystals were cryoprotected by brief immersion in reservoir solution supplemented with 25% (v/v) glycerol before plunge cryocooling and diffraction data were recorded at Diamond Light Source beamline I04-1 to 2.7 Å resolution. These data were provisionally assigned to space group $P\,2_12_12_1$ with unit cell dimensions α = 78.4 Å, β = 92.3 Å and γ = 146.2 Å. Unfortunately, all crystals obtained in the first crystallisation experiment were consumed in the process of obtaining this diffraction dataset and extensive attempts to reproduce crystallisation under these conditions were unsuccessful, precluding solution of the crystallographic phase problem by experimental methods. Attempts to solve the structure of E2 via molecular replacement using *ab initio* models generated by the I-TASSER modelling server [22], which was state-of-the-art at the time (in 2016), were unsuccessful. Exhaustive molecular replacement searches using 68,087 folded domains drawn from across the protein data bank (PDB) [23] were also unsuccessful, suggesting that E2 either possessed a novel fold or that the structural homology to the (multiple) domains was insufficient to facilitate structure solution.

Extensive sparse matrix screening eventually identified new conditions for the crystallisation of VACV E2 (Figure 1D), these new crystals sharing the same space group and unit cell dimensions as the crystal that was collected previously. The structure of VACV E2 was solved by single isomorphous

replacement with anomalous scattering using an ethylmercurithiosalicylate (EMTS, a.k.a. Thimerosal) derivative and the structure was refined to 2.3 Å resolution (Table 1). The structure of VACV E2 comprises two folded domains: an N-terminal annular (ring) domain spanning residues 1–454 and C-terminal compact globular (head) domain spanning residues 455–732 (Figure 2A). The final 5 amino acids of E2 (FKSSK) and all residues of the cloning tag were not visible in electron density and are presumed disordered. E2 is primarily α-helical, with only three short β-sheets being evident at the apex of the head domain (Figure 2B). Structural homology searches against the PDB performed using DALI [24] and PDBePISA [25] did not identify any significant structural homologues of the ring domain. The head domain does share weak homology to protein and glycan kinase domains, and to the bacterial SidJ pseudokinase domain that possesses glutamylase activity [26]. However, the overall structural correspondence is low, with less than half of the domain structurally aligned, and key kinase catalytic motifs such as the glycine-rich loop and catalytic lysine and aspartic acid residues are not conserved. Mapping the conservation of E2 sequence across poxviruses onto the structure [27] does not reveal any surface patches of high conservation as would be expected at an enzyme active site. This suggests that the E2 head domain lacks catalytic activity and that any similarity between this head domain and (pseudo)kinases is either spurious and/or vestigial, potentially representing a cellular (pseudo)kinase domain acquired by an ancestral poxvirus that has subsequently evolved toward a novel function [20, 21]. Searches performed against a database of 23,391 predicted structures for the human proteome [28] using DALI failed to identify significant structural homologues for the head and ring domains, frustrating attempts to infer E2 function by analogy to human proteins.

The N-terminal ring domain of VACV E2 is particularly striking, forming a central aperture that is 2.3 nm (23 Å) wide at its narrowest point (Figure 2A). While this is compatible with the diameter of B-DNA (2 nm), the radius is smaller than observed for DNA-binding proteins like PCNA and the inner surface of the ring domain lacks the positive electrostatic potential (Figure 2C) that would be expected for a DNA-binding protein [29]. The central aperture is too narrow to accommodate actin filaments (~6 nm diameter) or microtubules (~25 nm diameter), suggesting that the ring domain does not encircle cytoskeletal elements to promote VACV EV transport during infection. While VACV E2 is an acidic protein (theoretical isoelectric point 5.43 [30]) the head domain possesses a large basic patch on its surface (Figure 2C). Given the functional role of E2 in associating with intraluminal vesicles (ILVs) comprised of virions surrounded by trans-Golgi/early endosomal derived membranes, which are defined in part by their specific complement of phosphoinositides, it is tempting to speculate that the basic patch on the surface of the E2 head domain acts as a membrane recognition motif. This speculation is supported by the similarity of this basic patch to the phosphoinositide-binding surfaces of cellular domains that are known to promote membrane binding via recognition of specific phosphatidylinositol phosphates (Figure S1). Other noteworthy features of E2 include N-terminal acetylation of the initiator methionine (Figure 2D), which was verified by mass spectrometry, and the presence of a disulphide bond between Cys residues 496 and 535 in the head domain (Figure 2E) despite having included reducing agent (1 mM DTT) in the SEC purification buffer. Intramolecular disulphide bonding within cytosolic VACV proteins has been observed before [21] and its functional relevance remains unknown, although we note that a recent preprint has implicated redox proteins present in VACV lateral bodies in counteracting cellular oxidative stress generated during infection [31]. In summary, the structure of VACV E2 has provided some potential functional insights but, given the lack of structural homology to other well-characterised proteins, definitive mechanistic information has remained elusive.

Recently the application of deep learning technologies to the prediction of protein structures has made the resultant models significantly more accurate with regards to the backbone conformation

(overall protein fold), although the side chain conformation prediction remains more challenging [32]. The lack of sequence identity between VACV E2 homologues and any other protein family, and the novelty of both the ring and head domains, makes E2 a particularly difficult target for structural prediction. Models of VACV E2 were thus generated using two leading structure prediction packages, AlphaFold2 (AF2) [33] and RoseTTAFold (RTF) [34]. Predictions were performed using the sequence of VACV E2 as an input and all default parameters via a locally installed version of AF2 (version 2.0.1) or the Robetta web server (https://robetta.bakerlab.org/), and the top five models obtained using each program have been deposited in the University of Cambridge Data Repository (https://doi.org/10.17863/CAM.77496). Both AF2 and RTF accurately predicted the presence of two domains in the E2 structure, successfully identifying that the head of E2 would have a compact globular fold while the ring would have an extended helical conformation.

Detailed analysis shows that predictions of the head domain by both AF2 and RTF were more accurate than for the ring domain (Figure 3A–D), despite both programs assigning lower confidence to the predictions of this domain (Figure S2). Superpositions using SSM [25] and LGA (cutoff = 4 Å) [35] demonstrate that the head domain from the top ranked AF2 model can be superposed on the equivalent region of the experimental structure (residues 455–732, 278 residues in total) with a root-mean-squared deviation (rmsd) of 1.12 Å across 261 C^{α} atoms and a Global Distance Test Total Score (GDT_TS) of 87.2, which is remarkably accurate (Figure 3A). The largest discrepancy is at residues 477–492, which form part of the extended surface loop between the first helix and sheet of the head domain (Figure 3A). The prediction by RTF is less accurate (3.08 Å rmsd across 277 C^{α} atoms, GDT_TS = 49.8), consistent with accurate prediction of the gross topology but significant differences in the relative orientation of secondary structural elements (Figure 3B). With regards to the ring domain (residues 1–454), the top two AF2 models accurately predicted the 'closed' conformation of the E2 ring domain (Figure 3C) whereas the top RTF model predicted an 'open' conformation of the ring domain, the correct 'closed' conformation being observed in the second-ranked model (Figure 3D). The AF2 prediction was again more accurate (2.39 Å rmsd across 433 C^{α} atoms, GDT_TS = 68.3) than that of RTF (4.11 Å across 203 C^{α} atoms, GDT_TS = 36.7, for the top ranked model and 3.44 across 402 C^{α} atoms, GDT_TS = 45.3, for the second ranked model with a 'closed' ring). Furthermore, RTF demonstrated more heterogeneity in the conformation of the head domain relative to the ring. Neither AF2 nor RTF predicted the same relative domain orientation as seen in the experimental structure (Figure 3E), although it is possible that the two domains of E2 move relative to each other in solution and thus sample additional conformations not observed in the crystal structure.

While the above analysis confirms that the AF2 model is closer to the crystal structure of E2 than the RTF model, the obvious use-case of *ab initio* modelling for molecular virologists is in situations where a reference crystal structure is not known. Such structural models can generate functional hypotheses by identifying structural homology to proteins of known function. They can also inform site-directed mutagenesis experiments by identifying surface-exposed residues and prominent surface features such as charged, hydrophobic or conserved patches. The question thus arises: *How useful is the AF2 model as a basis for generating hypotheses and designing mutations to test E2 function?* As mentioned above, the structures of the E2 ring and head domains do not share significant structural homology with other domains, but queries of the PDB with the two domains of the AF2 model using DALI [24] (after release of the PDB-deposited E2 structural coordinates) succeeded in identifying structural homology for each domain to the E2 crystal structure (Z = 26.5 and 31.1 for ring and head domains, respectively), suggesting that these domain models would have identified significant structural homologues should they have existed. The surface charge of the AF2 model is very similar to that of the crystal structure (compare Figure 3G with Figure 2C), although we

note that the large basic patch on the head domain is less prominent due to the different conformation of residues 477–492 in the AF2 model. Furthermore, the percentage solvent-accessibility of each residue as calculated using the CCP4 program AREAIMOL [36, 37] was similar between the crystal structure and AF2 model (Figure 3E; Spearman nonparametric correlation coefficient $\rho = 0.9253$, $P < 0.0001$, as calculated using Prism7 [GraphPad]). Surprisingly, correlation was higher for the ring domain ($\rho = 0.9475$) than for the globular head domain ($\rho = 0.8945$), perhaps owing to its higher surface-area to volume ratio, but overall the AF2 model is clearly capable of predicting those residues that are buried in the core of the protein and should be avoided for mutagenic studies.

In addition to accelerating the generation and testing of functional hypotheses, high-quality structural models can simplify the process of solving macromolecular structures. Molecular replacement is a technique used for solution of the crystallographic phase problem whereby initial phases for a diffraction dataset are obtained from the atomic coordinates of a structure with a highly similar fold [38], rather than using heavy-atoms and/or anomalous scatterers to phase the structure as was done for E2. Molecular replacement represents a stringent test for the 'value added' by structural models [39], which has increased dramatically with the advent of deep learning techniques for protein structure prediction [32, 40, 41], and so the best-ranked models of E2 obtained from AF2 and RTF were tested for their ability to solve the structure of E2. The molecular replacement phasing experiment was performed using phenix.phaser [42] with a single search model (100% sequence identity to target structure) with per-residue confidence scores being converted to estimated B factors by phenix.voyager, success being indicated by a translation function Z (TFZ) score > 8 [43]. The AF2 and RTF full-length models could not be used to solve the structure, nor could the ring domains of these models. The use of structural ensembles, where multiple models are superposed, can improve the signal in molecular replacement experiments by upweighting the phase contribution of structural regions confidently predicted to have the same conformation across multiple models (and are thus more likely to be correct) and downweighting the contribution from variable regions [42]. However, ensembles of the ring domains from the top 5 AF2 and RTF models generated by phenix.ensembler were also incapable of solving the structure. Similarly, the head domain of the top RTF model, or an ensemble of the head domains from the top 5 models, could not solve the structure. However, the head domain of the AF2 model could be successfully positioned in the crystallographic asymmetric unit (TFZ = 19.5) and using this model as a fixed component allowed positioning of the AF2 model ring domain (TFZ = 14.1). Subsequent automated structure completion using phenix.autobuild [44] confirmed that these molecular replacement solutions provided sufficient phase information for successful structure determination, the autobuild model comprising 737 residues and 242 ordered solvent molecules with a R_{free} of 0.336 and rmsd of 0.81 Å across 697 Cα atoms when compared to the refined and deposited structure.

In conclusion, we have solved the crystal structure of VACV E2 to 2.3 Å resolution. E2 comprises two novel domains, an N-terminal annular (ring) domain and C-terminal head domain. While the fold of the head domain shares weak homology with cellular (pseudo)kinases and contains a large basic surface patch that may bind phosphoinositide headgroups, the lack of strong structural homology hampers attempts to infer E2 function by analogy to other proteins. Being a multi-domain protein with novel domain folds and limited availability of homologous protein sequences, VACV E2 is an excellent test for the ability of modern deep-learning algorithms to predict 'difficult' viral protein structures. The results are impressive, with both AF2 and RTF correctly predicting the overall fold of both domains and AF2 predicting the head domain with very high precision. AF2 models will prove a significant resource for the molecular virology community, allowing the identification of structural

homologies in the absence of identifiable sequence homology, the exploration of protein surface features, and accelerating the experimental determination of novel viral protein structures.

Acknowledgements

The authors thank Dr Len Packman (University of Cambridge) for performing mass spectrometry experiments, Mr Ben Butt (University of Cambridge) for assistance with installing AlphaFold2, SBGrid for performing the Wide-Search Molecular Replacement experiment, and Tristan Croll (University of Cambridge) for reading the draft manuscript. We thank Diamond Light Source for access to beamlines I04-1 (mx11235) and I04 (mx15916).

Funding information

Remote access to Diamond Light Source was supported in part by the EU FP7 infrastructure grant BIOSTRUCT-X (Contract No. 283570). A Titan V graphics card used for this research was donated by the NVIDIA Corporation. This work was supported by a Wellcome Trust Principal Research Fellowship (090315) to GLS, an MRC research grant (MR/R010536/1) to GLS and DCJC, a Royal Society University Research Fellowship (UF100371) to JED and a Sir Henry Dale Fellowship (098406/Z/12/B), jointly funded by the Wellcome Trust and the Royal Society, to SCG.

Conflict of interest

The authors declare that there are no conflicts of interest.

Table 1. X-ray diffraction data collection and structure refinement. Crystals of VACV E2 were grown by sitting drop vapour diffusion against 80 µL reservoirs, crystallisation drops containing 200 nL 9.7 mg/mL E2 plus 120 nL 50 mM ADA pH 6.0, 50 mM ADA pH 6.5, 8% v/v MPD (*Low-resolution native*), 120 nL 50 mM ADA pH 6.0, 50 mM ADA pH 6.5, 8% v/v MPD (*EMTS soak*), or 60 nL 50 mM ADA pH 6.5, 50 mM ADA pH 7.0, 10% v/v MPD (*High-resolution native*). Heavy atom derivitisation was achieved by soaking crystals for 90 min in reservoir solution supplemented with 1 mM ethylmercurithiosalicylate (EMTS) and 25% v/v glycerol. All crystals were cryoprotected by rapid transfer to reservoir solution supplemented with 25% v/v glycerol before plunge cryocooling in liquid nitrogen. Diffraction data were recorded at Diamond Light Source beamline I04 and processed using DIALS [45] as implemented in the xia2 [46] autoprocessing pipeline. The structure of VACV E2 was solved via single isomorphous replacement with anomalous scattering (SIRAS) by CRANK2 [47] using the *low-resolution native* and *EMTS soak* datasets. The substructure comprised six mercury atoms with occupancies ranging between 0.86 and 0.27 and the overall figure of merit was 0.198/0.394 (overall/lowest resolution shell) following initial phasing, rising to 0.345/0.470 after density modification and to 0.574/0.779 after iterative automated model building. The initial model comprised 729 residues in 10 fragments with R = 0.361, R_{free} = 0.412. This model was used to phase the *high-resolution native* data and the model was completed and refined using COOT [48], ISOLDE [49], autoBUSTER [50] and phenix.refine [51] in consultation with MolProbity [52] and the validation tools present in COOT [48]. Values in parentheses refer to the high-resolution shell. The atomic coordinates and structure factors have been deposited in the Protein Data Bank [53] with accession code 7PHY and the original diffraction data are available from the University of Cambridge Data Repository (https://doi.org/10.17863/CAM.74391).

Data collection	*Low-resolution native*	*EMTS soak*	*High-resolution native*
Wavelength (Å)	0.9795	0.9795	0.9795
Space group	$P\,2_12_12_1$	$P\,2_12_12_1$	$P\,2_12_12_1$
Cell dimensions (*a*, *b*, *c*) (Å)	77.57, 90.73, 146.32	78.08, 90.52, 144.23	77.17, 90.93, 147.20
Resolution range (Å)	146.3–2.5 (2.51–2.47)	59.1–3.1 (3.11–3.06)	39.2–2.3 (2.34–2.30)
Completeness (%)	100.0 (99.8)	100.0 (98.6)	100.0 (100.0)
Multiplicity	13.0 (13.4)	12.8 (12.9)	13.2 (12.4)
$CC_{1/2}$	0.997 (0.618)	0.999 (0.509)	1.00 (0.404)
Mean I/σ(I)	15.98 (2.48)	8.6 (2.3)	19.0 (0.6)
R_{merge}	0.098 (0.893)	0.214 (0.937)	0.074 (2.422)
R_{meas}	0.102 (0.928)	0.223 (0.976)	0.077 (2.526)
R_{pim}	0.028 (0.251)	0.062 (0.269)	0.021 (0.708)
Anomalous completeness (%)	100.0 (99.8)	99.9 (98.4)	100.0 (99.7)
Anomalous multiplicity	6.9 (6.9)	6.8 (6.8)	6.9 (6.5)
Anomalous $CC_{1/2}$	-0.019 (-0.015)	0.280 (-0.010)	-0.136 (0.021)
Wilson B factor (Å^2)	52.0	68.2	58.9
Refinement			
Resolution range (Å)			35.5–2.3 (2.35–2.30)
Reflections			
Working set			44,190 (2543)
Test set			2396 (153)
R			0.1943 (0.4000)
R_{free}			0.2370 (0.4195)
No. of atoms			

Protein			6005
Solvent			249
Other[a]			41
Root mean square deviation			
Bond length (Å)			0.008
Bond angle (°)			0.866
Ramachandran favoured (%)			97.81
Ramachandran outliers (%)			0.14
Clash score			2.46
Poor rotamers (%)			0.58
Mean B value (Å^2)			75.71

[a]The N terminus of E2 was modelled as an N-acetylated methionine and five ordered glycerol molecules were observed in the structure.

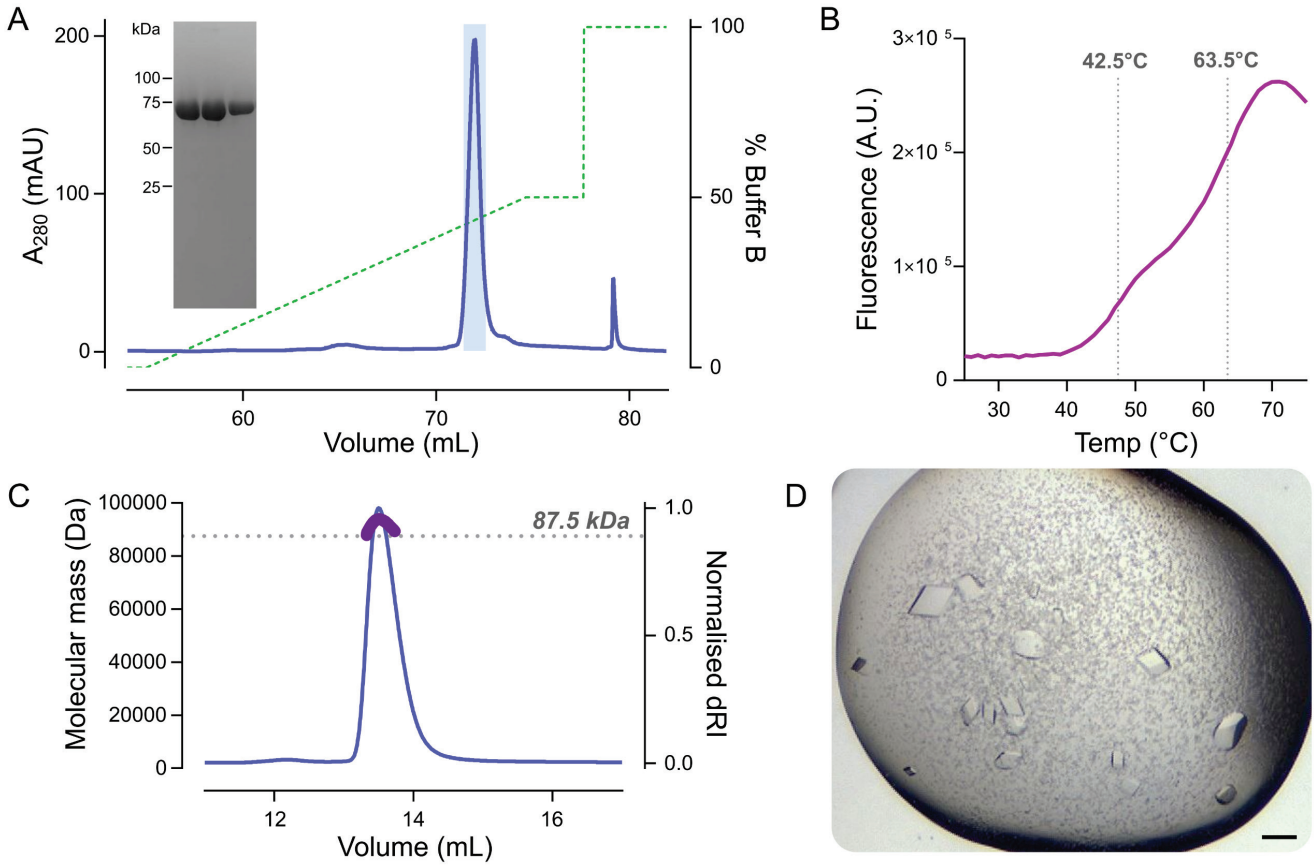

Figure 1. Purification, characterisation and crystallisation of VACV E2. (A) Preparative anion exchange chromatography. VACV E2 was expressed in Freestyle 293F cells and grown in Freestyle 293 medium (ThermoFisher) as per the manufacturer's instructions, by transfection of pcDNA3 encoding VACV E2 with a C-terminal A_3H_{10} tag mixed in a 1:2 ratio with 25 kDa branched polyethylenimine (PEI), adding 1 µg DNA and 2 µg PEI per mL of cultured cells. Cells were cultured for 40 h in a humidified 8% CO_2 atmosphere at 37°C before being harvested by centrifugation, washed thrice with ice-cold PBS, resuspended in lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl supplemented with protease inhibitors [Roche]) and lysed by five passages through a 23G needle. Lysates were clarified by centrifugation (40,000×g, 40 min, 4°C) before being applied to a 5 mL HiTrap TALON Crude Co^{2+} affinity column (Cytiva) and purified with elution in 200 mM imidazole as per the manufacturer's instructions. Pooled eluate was further purified by size-exclusion chromatography (SEC) using a Superdex 200 10/300 GL column (Cytiva) equilibrated in SEC buffer (20 mM Tris pH 8.0, 200 mM NaCl, 1 mM DTT). As eluted protein retained contaminants, E2 was further purified by anion exchange chromatography with a MonoQ 5/50 GL column (Cytiva) using a linear gradient of 0–500 mM NaCl (green dashed line) in 20 mM Tris pH 8.0, protein elution being monitored using UV absorbance (blue line). Peak fractions containing VACV E2 that were pooled and used for subsequent analysis are highlighted (light blue) and SDS-PAGE of these fractions shows VACV E2 to be highly pure. (B) Differential scanning fluorimetry of VACV E2. Purified E2 (4 µg) was mixed with 1× Protein Thermal Shift dye (Applied Biosystems) in a final volume of 20 µL and heated from 25 to 95°C at 1 degree per 30 s, with fluorescence (purple curve) being monitored at each increment. Two inflection points are visible (grey dotted lines), consistent with biphasic melting. (C) SEC with inline multi-angle light scattering (SEC-MALS) shows VACV E2 to be predominantly monomeric. Purified E2 (100 µg) was injected onto a Superdex 200 10/300 GL column (Cytiva) equilibrated in SEC buffer at 0.4 mL/min at room temperature with inline measurement of static light scattering (DAWN 8+, Wyatt Technology), differential refractive index (dRI; Optilab T-rEX, Wyatt Technology), and 280 nm absorbance (Agilent 1260 UV, Agilent Technologies). The normalised

dRI is shown (thin blue line), as is the molecular mass of the peak (thick purple line) as calculated using ASTRA6 (Wyatt Technology) assuming a protein dn/dc of 0.186. The calculated mass (92.3 kDa) is in good agreement with theoretical mass of VACV E2 with a C-terminal A_3H_{10} tag (87.5 kDa; dotted grey line), confirming that the protein is predominantly monomeric. (D) Crystals of VACV E2, grown by sitting drop vapour diffusion. 200 nL of 9.7 mg/mL E2 was mixed with 60 nL of reservoir solution (50 mM ADA (N-(2-acetamido)iminodiacetic acid) pH 6.5, 50 mM ADA pH 7.0, 10% v/v 2-methyl-2,4-pentanediol [MPD]) and equilibrated against 80 µL reservoirs at 20°C, crystals growing within 21 days. Scale bar = 100 µm.

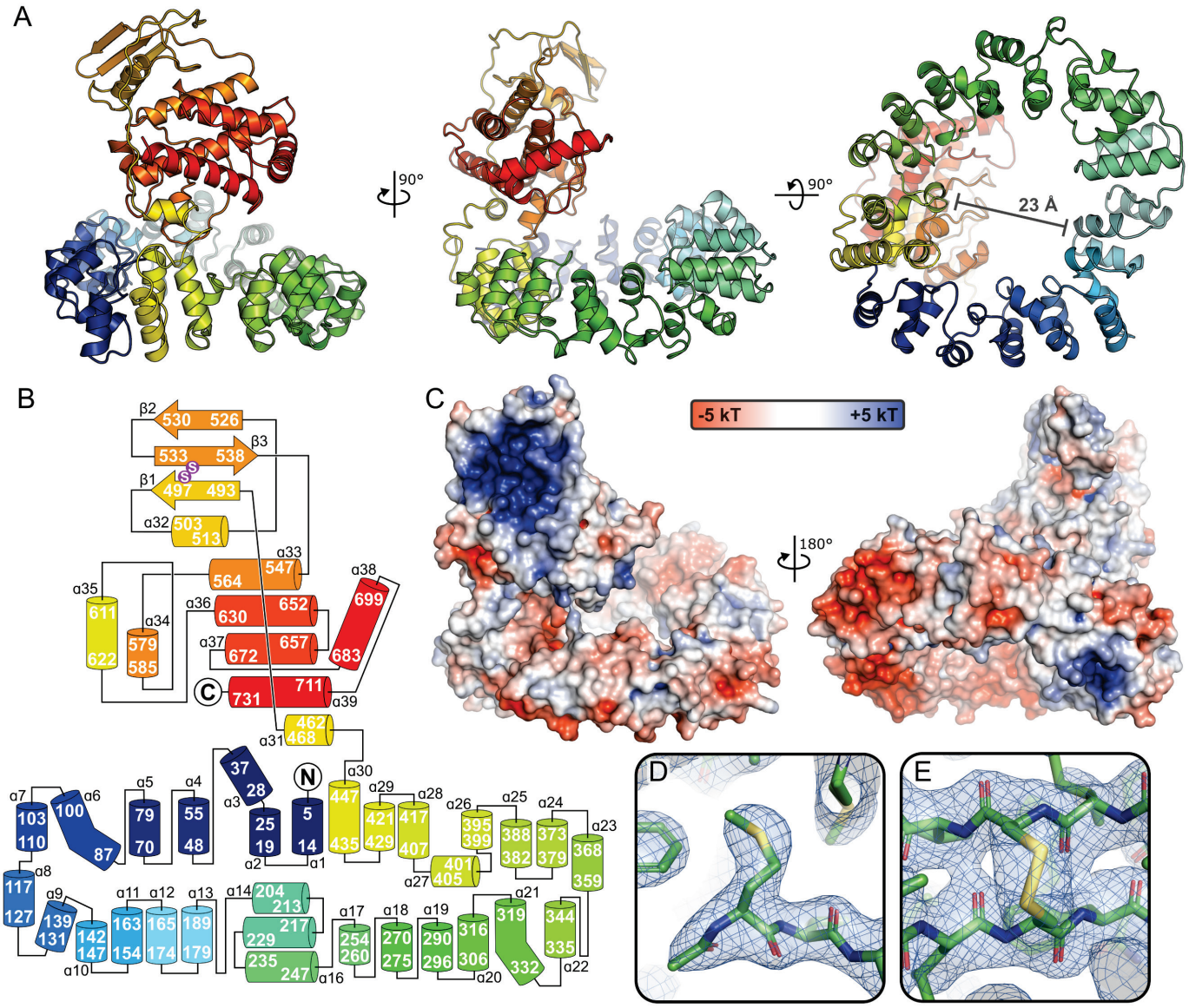

Figure 2. VACV E2 comprises novel N-terminal annular (ring) and C-terminal globular (head) domains. (A) The structure of VACV E2 is shown in three orthogonal views in ribbon representation, rainbow coloured from blue (N terminus) to red (C terminus). Molecular images were generated using PyMOL (Schrödinger LLC). The aperture of the ring domain is 23 Å wide at its narrowest point. (B) Schematic representation of VACV E2, with secondary structural elements coloured as in (A). Helices and sheets are shown as cylinders and arrows, respectively, with start and end residues shown. Sulphur residues that participate in a disulphide bond are shown in purple. (C) Molecular surface of E2 coloured by electrostatic potential from red (-5 kT) to blue (+5 kT), as calculated by APBS [54]. E2 is shown in two views, the left being rotated around the vertical and horizontal axes by approximately 15° from the middle panel of (A) to better illustrate the strong basic patch on the head domain and the lack of strong charge lining the centre of the ring domain. (D) The N-acetylated initiator methionine of E2 is shown in stick representation, with the final refined $2F_O$-F_C electron density map (1.2 σ) being shown as a blue semi-transparent mesh surface. (E) The disulphide bond between Cys residues 496 and 535 is shown in $2F_O$-F_C electron density (1.2 σ).

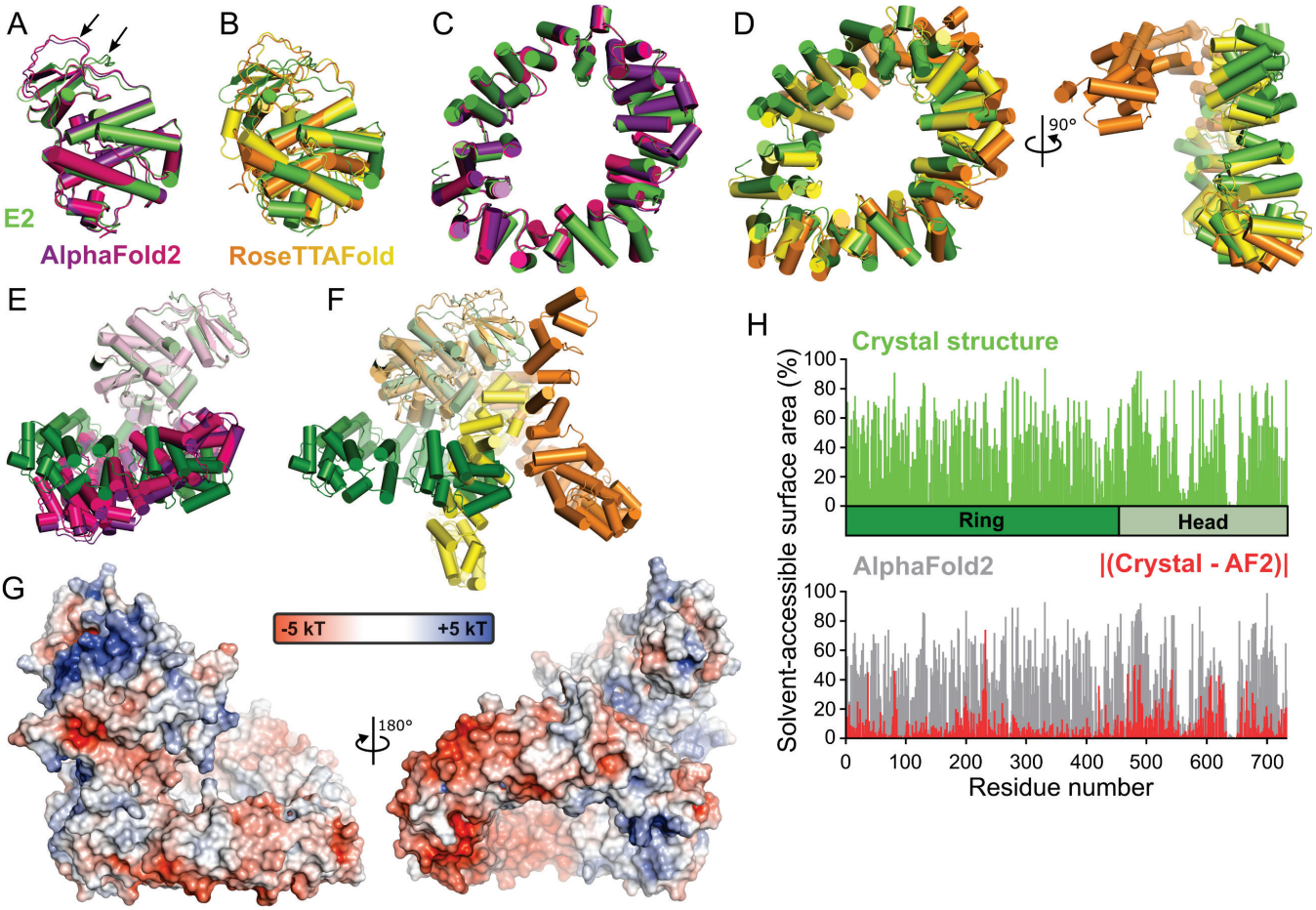

Figure 3. Assessment of prediction of the VACV E2 structure by AlphaFold2 (AF2) and RoseTTAFold (RTF). All superpositions were performed using SSM [25]. (A) Superposition of the head domain from the E2 crystal structure (green) with the top two models from AF2 (purple and pink, respectively). The loop between residues 477–492, where the backbone conformation of the models differs significantly from the crystal structure, is denoted with an arrow. (B) Superposition of the head domain from the E2 crystal structure (green) with the top two models from RTF (orange and yellow, respectively). (C) Superposition of the ring domain from the E2 crystal structure with the top two models from AF2, coloured as in (A). (D) Superposition of the ring domain from the E2 crystal structure with the top two models from RTF, shown in two orthogonal views and coloured as in (B). (E) and (F) Orientation of the E2 ring domain relative to the head domain in the top two AF2 (E; purple and pink) or RTF (F; orange and yellow) models compared with the E2 crystal structure (green). (G) Molecular surface of the top AF2 model of E2 coloured by electrostatic potential from red (-5 kT) to blue (+5 kT), as calculated by APBS [54]. E2 AF2 model is oriented as in Figure 2C. (H) Percent solvent accessibility of residues in the E2 crystal structure (green, top) or AF2 model (grey, bottom) as calculated using AREAIMOL [36, 37]. The absolute difference between calculated accessibility for the crystal structure and AF2 model is shown in red.

References

1. **Moss B, Smith GL**. Poxviridae. In: *Fields Virology*. Wolters Kluwer Heath; 2021. pp. 573–613.

2. **Shchelkunov SN**. An increasing danger of zoonotic orthopoxvirus infections. *PLoS Pathog* 2013;9:e1003756.

3. **Hobson G, Adamson J, Adler H, Firth R, Gould S, *et al.*** Family cluster of three cases of monkeypox imported from Nigeria to the United Kingdom, May 2021. *Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull*;26. Epub ahead of print August 2021. DOI: 10.2807/1560-7917.ES.2021.26.32.2100745.

4. **Mohammadpour R, Champour M, Tuteja F, Mostafavi E**. Zoonotic implications of camel diseases in Iran. *Vet Med Sci* 2020;6:359–381.

5. **Moss B**. Origin of the poxviral membrane: A 50-year-old riddle. *PLoS Pathog* 2018;14:e1007002.

6. **Hollinshead M, Rodger G, Van Eijl H, Law M, Hollinshead R, *et al.*** Vaccinia virus utilizes microtubules for movement to the cell surface. *J Cell Biol* 2001;154:389–402.

7. **Ward BM, Moss B**. Vaccinia virus intracellular movement is associated with microtubules and independent of actin tails. *J Virol* 2001;75:11651–11663.

8. **Rietdorf J, Ploubidou A, Reckmann I, Holmström A, Frischknecht F, *et al.*** Kinesin-dependent movement on microtubules precedes actin-based motility of vaccinia virus. *Nat Cell Biol* 2001;3:992–1000.

9. **Geada MM, Galindo I, Lorenzo MM, Perdiguero B, Blasco R**. Movements of vaccinia virus intracellular enveloped virions with GFP tagged to the F13L envelope protein. *J Gen Virol* 2001;82:2747–2760.

10. **Payne LG**. Significance of extracellular enveloped virus in the in vitro and in vivo dissemination of vaccinia. *J Gen Virol* 1980;50:89–100.

11. **van Eijl H, Hollinshead M, Smith GL**. The vaccinia virus A36R protein is a type Ib membrane protein present on intracellular but not extracellular enveloped virus particles. *Virology* 2000;271:26–36.

12. **Morgan GW, Hollinshead M, Ferguson BJ, Murphy BJ, Carpentier DCJ, *et al.*** Vaccinia protein F12 has structural similarity to kinesin light chain and contains a motor binding motif required for virion export. *PLoS Pathog* 2010;6:e1000785.

13. **Dodding MP, Mitter R, Humphries AC, Way M**. A kinesin-1 binding motif in vaccinia virus that is widespread throughout the human genome. *EMBO J* 2011;30:4523–4538.

14. **Carpentier DCJ, Gao WND, Ewles H, Morgan GW, Smith GL**. Vaccinia virus protein complex F12/E2 interacts with kinesin light chain isoform 2 to engage the kinesin-1 motor complex. *PLoS Pathog* 2015;11:e1004723.

15. **Gao WND, Carpentier DCJ, Ewles HA, Lee S-A, Smith GL**. Vaccinia virus proteins A36 and F12/E2 show strong preferences for different kinesin light chain isoforms. *Traffic* 2017;18:505–518.

16. **Dodding MP, Newsome TP, Collinson LM, Edwards C, Way M**. An E2-F12 complex is required for intracellular enveloped virus morphogenesis during vaccinia infection. *Cell Microbiol* 2009;11:808–824.

17. **Johnston SC, Ward BM**. Vaccinia virus protein F12 associates with intracellular enveloped virions through an interaction with A36. *J Virol* 2009;83:1708–1717.

18. **Carpentier DCJ, Van Loggerenberg A, Dieckmann NMG, Smith GL**. Vaccinia virus egress mediated by virus protein A36 is reliant on the F12 protein. *J Gen Virol* 2017;98:1500–1514.

19. **Butt BG, Owen DJ, Jeffries CM, Ivanova L, Hill CH, *et al.*** Insights into herpesvirus assembly from the structure of the pUL7:pUL51 complex. *eLife* 2020;9:53789.

20. **Neidel S, Maluquer de Motes C, Mansur DS, Strnadova P, Smith GL, *et al.*** Vaccinia virus protein A49 is an unexpected member of the B-cell Lymphoma (Bcl)-2 protein family. *J Biol Chem* 2015;290:5991–6002.

21. **Graham SC, Bahar MW, Cooray S, Chen RA-J, Whalen DM, *et al.*** Vaccinia virus proteins A52 and B14 Share a Bcl-2-like fold but have evolved to inhibit NF-kappaB rather than apoptosis. *PLoS Pathog* 2008;4:e1000128.

22. **Yang J, Zhang Y**. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* 2015;43:W174-181.

23. **Stokes-Rees I, Sliz P**. Protein structure determination by exhaustive search of Protein Data Bank derived databases. *Proc Natl Acad Sci U S A* 2010;107:21476–21481.

24. **Holm L**. Using Dali for Protein Structure Comparison. *Methods Mol Biol Clifton NJ* 2020;2112:29–42.

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

26. **Sulpizio A, Minelli ME, Wan M, Burrowes PD, Wu X, *et al.*** Protein polyglutamylation catalyzed by the bacterial calmodulin-dependent pseudokinase SidJ. *eLife* 2019;8:e51162.

27. **Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, *et al.*** ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* 2016;44:W344-350.

28. **Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, *et al.*** Highly accurate protein structure prediction for the human proteome. *Nature* 2021;596:590–596.

29. **Kelman Z, O'Donnell M**. Structural and functional similarities of prokaryotic and eukaryotic DNA polymerase sliding clamps. *Nucleic Acids Res* 1995;23:3613–3620.

30. **Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, *et al.*** Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol Clifton NJ* 1999;112:531–552.

31. **Bidgood SR, Novy K, Collopy A, Albrecht D, Krause M, *et al.*** *Poxviruses package viral redox proteins in lateral bodies and modulate the host oxidative response*. Preprint; Microbiology. Epub ahead of print 9 December 2020. DOI: 10.1101/2020.12.09.418319.

32. **Pereira J, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM, *et al.*** High-accuracy protein structure prediction in CASP14. *Proteins*. Epub ahead of print 3 July 2021. DOI: 10.1002/prot.26171.

33. **Jumper J, Evans R, Pritzel A, Green T, Figurnov M, *et al.*** Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583–589.

34. **Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, *et al.*** Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 2021;373:871–876.

35. **Zemla A**. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res* 2003;31:3370–3374.

36. **Lee B, Richards FM**. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 1971;55:379–400.

37. **Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, *et al.*** Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 2011;67:235–242.

38. **Rossmann MG**. The molecular replacement method. *Acta Crystallogr A* 1990;46:73–82.

39. **Read RJ, Chavali G**. Assessment of CASP7 predictions in the high accuracy template-based modeling category. *Proteins* 2007;69 Suppl 8:27–37.

40. **Croll TI, Sammito MD, Kryshtafovych A, Read RJ**. Evaluation of template-based modeling in CASP13. *Proteins* 2019;87:1113–1127.

41. **McCoy AJ, Sammito MD, Read RJ**. *Possible Implications of AlphaFold2 for Crystallographic Phasing by Molecular Replacement*. Preprint; Biophysics. Epub ahead of print 18 May 2021. DOI: 10.1101/2021.05.18.444614.

42. **McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, *et al.*** Phaser crystallographic software. *J Appl Crystallogr* 2007;40:658–674.

43. **Oeffner RD, Bunkóczi G, McCoy AJ, Read RJ**. Improved estimates of coordinate error for molecular replacement. *Acta Crystallogr D Biol Crystallogr* 2013;69:2209–2215.

44. **Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, *et al.*** Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* 2008;64:61–69.

45. **Winter G, Waterman DG, Parkhurst JM, Brewster AS, Gildea RJ, *et al.*** DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr Sect Struct Biol* 2018;74:85–97.

46. **Winter G**. *xia2* : an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* 2010;43:186–190.

47. **Skubák P, Pannu NS**. Automatic protein structure solution from weak X-ray data. *Nat Commun* 2013;4:2777.

48. **Emsley P, Lohkamp B, Scott WG, Cowtan K**. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010;66:486–501.

49. **Croll TI**. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr Sect Struct Biol* 2018;74:519–530.

50. **Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., Paciorek W.,Roversi P, Sharff A., Smart O.S., Vonrhein C., Womack T.O.** *BUSTER*. Cambridge, United Kingdom: Global Phasing Ltd.

51. **Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, *et al.*** Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr Sect Struct Biol* 2019;75:861–877.

52. **Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, *et al.*** MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 2010;66:12–21.

53. **Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, *et al.*** The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–242.

54. **Jurrus E, Engel D, Star K, Monson K, Brandi J, *et al.*** Improvements to the APBS biomolecular solvation software suite. *Protein Sci* 2018;27:112–128.