

REVIEW

Poxvirus uracil-DNA glycosylase—An unusual member of the family I uracil-DNA glycosylases

Norbert Schormann,¹ Natalia Zhukovskaya,² Gregory Bedwell,³ Manunya Nuth,² Richard Gillilan,⁴ Peter E. Prevelige,³ Robert P. Ricciardi,^{2,5} Surajit Banerjee,⁶ and Debasish Chattopadhyay^{1*}

¹Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama 35294

²Department of Microbiology, School of Dental Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104

³Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294

⁴MacCHESS (Macromolecular Diffraction Facility at CHESS) Cornell University, Ithaca, New York 14853

⁵Abramson Cancer Center, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104

⁶Department of Chemistry and Chemical Biology, Cornell University, and NE-CAT, Argonne, Illinois 60439

Received 12 August 2016; Accepted 27 September 2016

DOI: 10.1002/pro.3058

Published online 29 September 2016 proteinscience.org

Abstract: Uracil-DNA glycosylases are ubiquitous enzymes, which play a key role repairing damages in DNA and in maintaining genomic integrity by catalyzing the first step in the base excision repair pathway. Within the superfamily of uracil-DNA glycosylases family I enzymes or UNGs are specific for recognizing and removing uracil from DNA. These enzymes feature conserved structural folds, active site residues and use common motifs for DNA binding, uracil recognition and catalysis. Within this family the enzymes of poxviruses are unique and most remarkable in terms of amino acid sequences, characteristic motifs and more importantly for their novel non-enzymatic function in DNA replication. UNG of vaccinia virus, also known as D4, is the most extensively characterized UNG of the poxvirus family. D4 forms an unusual heterodimeric processivity factor by attaching to a poxvirus-specific protein A20, which also binds to the DNA polymerase E9 and recruits other proteins necessary for replication. D4 is thus integrated in the DNA polymerase complex, and its DNA-binding and DNA scanning abilities couple DNA processivity and DNA base excision repair at the replication fork. The adaptations necessary for taking on the new function are reflected in the amino acid sequence and the

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Institutes of Health; Grant numbers: 5U01-A1 082211, P41 GM103403; Grant sponsor: U.S. DOE; Grant number: DE-AC02-06CH11357; Grant sponsor: National Science Foundation; Grant number: DMR-0936384; Grant sponsor: National Institute of General Medical Sciences, National Institutes of Health; Grant number: GM-103485.

*Correspondence to: Debasish Chattopadhyay, Department of Medicine, University of Alabama at Birmingham, CBSE-250, 1025 18th Street South, Birmingham, AL 35294. E-mail: dchattop@uabmc.edu

three-dimensional structure of D4. An overview of the current state of the knowledge on the structure-function relationship of D4 is provided here.

Keywords: uracil-DNA glycosylase; poxvirus; processivity factor; DNA polymerase holoenzyme; DNA repair; replication; antiviral agents

Introduction

DNA repair pathways are essentially employed by all organisms to ensure the integrity of their genome. Uracil-DNA glycosylases (UDGs) are important DNA repair enzymes that initiate the first step in the base excision repair pathway for removal of uracil and several other unusual bases from DNA. These mono-functional DNA glycosylases use water as a nucleophile to cleave the *N*-glycosidic bond between the target base and deoxyribose, and as a result release the free base and generate a cytotoxic apurinic/apyrimidinic (AP) site in the DNA. The resulting abasic sites are usually repaired in a following reaction by other enzymes such as AP endonuclease. Currently, the UDG superfamily consists of seven subfamilies.^{1–4} Of these, five families (I–V) share a common fold and two active site motifs (named Motif A and Motif B: catalytic water-activating loop and Leu-intercalation loop, respectively) in the N- and C-terminal domain. These enzymes recognize uracil in DNA and catalyze its excision. While enzymes of families II, III and V can recognize and cleave other bases in addition to uracil, family I enzymes exhibit exquisite specificity for uracil. Family VI enzymes also recognize and remove uracil but use a different structural motif (helix-hairpin-helix). In addition, they are capable of repairing 8-oxoguanine mismatches with adenine and thymine. Members of family VII share the fold and motifs with families I–V enzymes but do not recognize uracil. Instead, they catalyze excision of hypoxanthine from DNA. Families IV, V and VI are found in thermophilic and hyperthermophilic eubacteria and archaea, and they contain a structural iron-sulfur (4Fe-4S) cluster that does not seem to be necessary for activity.⁴

Family I UDGs (also called UNGs) specifically excise uracil from single-stranded (*ss*) and double-stranded (*ds*) DNA with preference for ssU > dsU:G > dsU:A.⁵ Within this family, UNGs of poxviruses are most remarkable for the diversity of their sequences (maximum ~20% sequence identity), and more significantly for their novel function in replication. The overall sequence identity between human UNG and D4 is low (~21% for all D4 residues 1–218; only ~17% for residues 20–204 of the core domain). Vaccinia virus UNG (known as D4, M_r ~25 kDa) is the most extensively studied UNG of the poxvirus family. D4 is integrated into the DNA processivity factor and the DNA polymerase complex, where it plays an essential role in replication but this function is independent of its glycosylase activity.^{6–8}

Thus, although in the DNA polymerase complex the catalytic activity of D4 directly couples DNA replication and base excision repair, the functional significance of D4 relates to its role in the processivity activity.^{8,9}

The poxvirus processivity factor is a heterodimeric 1:1 stoichiometric complex containing D4 and A20.⁷ A20 is a unique poxvirus-specific protein with a molecular weight of 49 kDa. It acts as a linker that bridges D4 with the DNA polymerase E9 and enables interactions of other proteins involved in DNA synthesis such as helicase/primase D5 and late transcription elongation factor H5 at the replication fork.^{7,8,10–12} Recently, a low-resolution structure of the D4:A20:E9 DNA polymerase holoenzyme was determined by using small-angle X-ray scattering (SAXS).¹² This structure provided a framework for understanding the relative orientation of A20, D4 and E9 within the polymerase complex.

Vaccinia virus D4 has been well characterized both structurally and functionally.^{6–9,12–20} Crystal structures of wild type D4, a D4:uracil complex and five D4 mutants have been reported. A common feature in these structures is the occurrence of a dimeric assembly of D4 subunits. The crystal structures of wild type D4 in complex with a peptide comprising the N-terminal 50 residues of A20 (A20_{1–50}) and the structure of the A20_{1–50}:D4 complex containing a uracil molecule in the binding pocket of D4 have also been determined. Interestingly, parts of the homodimer interface observed in the crystal structures of D4 overlap with the protein-protein interface in the A20_{1–50}:D4 complexes. Since binding of D4 and A20 is essential for processive DNA synthesis, the heterodimer interface has been targeted for the development of protein-protein interaction inhibitors (PPIs) that block viral DNA synthesis and replication.^{21–27} These studies led to the identification of small molecules, which effectively blocked processive DNA synthesis *in vitro* and inhibited viral plaque formation in cell culture.^{24,25,27} Recently, crystal structures of D4 in complex with specific and non-specific DNA have also been reported.^{9,19} These structures offer some insights about DNA-binding and uracil excision by D4. In addition, crystal structures of two D4 mutants with A20_{1–50} and of *wt*D4 in complex with an A20_{1–50} mutant have recently been published.²⁸

Recognition of the unique characteristics of D4 and its novel function in viral replication has attracted considerable interest in the field. In this review, we provide a comprehensive summary of the

Table I. List of All 19 Available D4 Crystal Structures (with PDB Codes)

PDB codes	Construct
Free wild type D4	
(1) 2OWQ	His-tag: MGSSHHHHHSSGLVPRGSH
(2) 5JX8 ^a	His-tag: same as above
(3) 4DOF	His-tag: same as above
(4) 5JX3 (supersedes 2OWR) ^a	cleaved His-tag: GSH remaining
(5) 4DOG	cleaved His-tag: GSH remaining
D4:uracil complex	
(6) 4LZB	His-tag: -MGSSHHHHHSSGLVPRGSH
Free D4 mutants	
(7) 4QC9 ^a (3G171-173)	cleaved His-tag: GSH remaining
(8) 4QCA ^a (Arg167Ala)	cleaved His-tag: GSH remaining
(9) 4IRB ^a (Δ 171-172)	His-tag: MGSSHHHHHSSGLVPRGSH
(10) 5JX0 ^a (Leu110Phe)	His-tag: same as above
(11) 3NT7 ^a (Arg187Val)	His-tag: same as above
D4:DNA complex	
(12) 4QCB	cleaved His-tag: GSH remaining
D4:A20 ₁₋₅₀ complex	
(13) 4OD8	His-tag: MGSSHHHHHSQDP
(14) 4ODA	His-tag: MGSSHHHHHSQDP
A20 ₁₋₅₀ :D4:uracil complex	
(15) 4YGM	His-tag: MGSSHHHHHSQDP
A20 ₁₋₅₀ :D4:DNA complex	
(16) 4YIG	His-tag: MGSSHHHHHSQDP
D4:A20 ₁₋₅₀ (W43A) mutant complex	
(17) 5JKR	His-tag: MGSSHHHHHSQDP
mutant D4(R167A):A20 ₁₋₅₀ complex	
(18) 5JKS	His-tag: MGSSHHHHHSQDP
mutant D4(P173G):A20 ₁₋₅₀ complex	
(19) 5JKT	His-tag: MGSSHHHHHSQDP

^a Previously unpublished.

current state of knowledge on D4. In addition to reviewing published results we have incorporated new experimental data from solution studies and crystal structure analyses of *wt*D4, and included unpublished structural and functional data on several D4 mutants that were generated by altering residues near the homodimer interface. We have also replaced previously deposited crystal structures of *wt*D4 with updated structures refined using either a new data set or using an improved refinement protocol. An up-to-date list of all D4 structures is presented in Table I. Refinement statistics are described in Supporting Information Table I.

Unique Features of D4

So far 11 crystal structures of free D4 (without bound DNA or A20) have been deposited in the Protein Data Bank (see Supporting Information Table II). These structures represent *wt*D4, a *wt*D4:uracil complex and D4 mutants. Crystals used in these studies were grown in different crystallization media and at different temperatures. The D4 crystals belong to orthorhombic, trigonal and hexagonal crystal systems (Supporting Information Table II).

The three-dimensional (3D) structure of D4 consists of the conserved UDG-fold composed of a central four-stranded parallel β -sheet with α -helices on each side.^{15,20} In addition, the D4 structure contains

a two-stranded anti-parallel β -sheet at the N-terminus, and at the C-terminus the protein chain folds back to form a helix-pair and a short anti-parallel β -sheet of two strands. The crystal structure of the D4:uracil complex demonstrated that the overall architecture and the amino acid residues of the uracil-binding pocket are similar in D4 and other UNGs.¹⁷ However, D4 shows variations in the characteristic UDG motifs that are responsible for DNA-binding and glycosylase activity.²⁰ A major distinctive feature of D4 is the substitution of the conserved Leu residue of the Leu-intercalation loop by Arg185. This replacement, which influences the size and structure of the intercalation loop, may offer D4 protection against inhibition by the uracil DNA glycosylase inhibitor (Ugi) protein, which is a natural inactivator of UNGs. It should be noted that unlike other UNGs, the glycosylase activity of D4 is not inhibited by Ugi.^{13,14} Ugi inhibits UNGs from different organisms by capitalizing on the conserved sequences of DNA-binding regions in UNGs. This inhibitory action of Ugi is based on charge and shape complementarity. Ugi inactivates UNG by inserting a β -strand into the conserved DNA binding groove without directly binding to the residues of the uracil binding pocket, thereby functionally mimicking DNA-binding. In the crystal structure of the hUNG:Ugi complex, the protruding Leu residue

Table II. Listing of D4 Interface Residues for Homodimers (D4:D4) and Heterodimers (A20₁₋₅₀:D4)

Homodimer	155-LYCLG K TDFSNIRAKLESPVTTIVGYHAARDRQFEKDRSFEIINVLLELDN-206
Heterodimer	155-LYCLG K TDFSNIRAKLESPVTTIVGYHAARDRQFEKDRSFEIINVLLELDN-206

Shown is the vaccinia virus UNG sequence and highlighted interface residues are based on 4OD8 and 5JX8. D4 interface residues with hydrogen bonds are highlighted in bold (of these residues with side chain hydrogen bonds are underlined), and D4 interface residues with non-bonded contacts are shaded. In addition, there are two salt bridges at the homodimer (D4:D4; 5JX8) interface (Lys160_NZ; Asp205_OD1) and 1 salt bridge at the heterodimer (A20₁₋₅₀:D4; 4OD8) interface (Gln203_OE1; Lys18_NZ).

from the Leu-intercalation loop inserts into the hydrophobic pocket of Ugi [1UGH].²⁹ However, the structural and molecular basis for the lack of inhibition of D4's catalytic activity by Ugi has not been established experimentally.

Crystal structures of D4 display a dimeric assembly of subunits related by crystallographic symmetry or non-crystallographic symmetry (NCS) and an apparent extension of the central β-sheet. The resulting dimers were previously designated as type II dimers.¹⁵ The dimer interface between the crystallographic symmetry-related molecules is significantly larger than in the NCS-related dimers, wherein a shift of ~21° in the relative orientation of the subunits results in fewer interactions between them.¹⁹

The *wt*D4 structure in space group P2₁2₁2₁ [5JX3, supersedes 2OWR] contains eight subunits in the asymmetric unit that are arranged as four dimers AB, CD, EF and GH. Each pair is related by NCS. On the other hand, in space group P3₂21 [5JX8] the crystallographic two-fold symmetry generates the homodimers AA' and BB' with a more extended interface. This homodimer interface is predominantly apolar accounting for ~72% of the total buried surface area (residue ranges of 155–180 and 187–206). The D4 interface residues are listed in Table II. Analysis of the homodimer interfaces for all D4 structures was performed using PDBePISA (<https://www.ebi.ac.uk/pdbe/pisa/>) and the InterPro-Surf protein-protein interaction server (<http://curie.utmb.edu/usercomplex.html>). Results of the analyses, which include calculation of buried surface area with polar and apolar contribution, and the number of interface residues, hydrogen bonds and salt bridges, are listed in Supporting Information Table III.

Although the homodimeric form of D4 is unlikely to be physiologically relevant the presence of a dimeric packing in D4 crystals of different unit cell composition and symmetry is interesting, especially since UNGs are known to be monomeric proteins. At lower protein concentrations D4 also remains as a monomer in solution. However, at higher protein concentration it tends to form dimers.¹⁹ Results of the coupled size exclusion chromatography-multi-angle light scattering experiments indicated the presence of a fast concentration-dependent monomer-dimer equilibrium in solution of D4.¹⁹ We used

analytical ultracentrifugation (AUC) to determine the dissociation constant (K_D) of the monomer-dimer equilibrium for *wt*D4 (Supporting Information Table IV and Figs. S1 and S2; method described in Supporting Information). Sedimentation velocity and sedimentation equilibrium experiments were employed, and fitting of the sedimentation boundaries provided the K_D (15 μM), the stoichiometry ($n = 2$) and the experimental molecular weights for a monomer ($M_r = 25.2$ kDa) and a dimer ($M_r = 49.7$ kDa). In addition, results of the solution studies using dynamic light scattering (DLS) and SAXS also suggest concentration-dependent dimeric and oligomeric associations in preparations of *wt*D4. At protein concentrations greater than 80 μM the average molecular weight of D4 indicated the presence of mixtures of different oligomers (Supporting Information Tables V and VI; methods described in Supporting Information).

Recent studies revealed that 20 out of 32 residues that form the homodimer interface are also involved in D4's interactions with A20 (see Table II).¹⁹ The A20-binding interface consists of D4 residues 160–180 and 189–206 (labeled Region 1 and Region 2, respectively, in Table III). Amino acid sequences for these segments are nearly identical in UNGs of poxviruses but differ from their counterparts in other organisms. In these two regions only 7 of 39 residues (18%) are identical between human UNG and D4. Since A20 proteins are found only in poxviruses, the characteristic sequences in D4 and other poxvirus UNGs indicate an evolutionary adaptation required for the formation of the heterodimeric processivity factor in these viruses.

Processivity Factor: D4 and A20

Most replicative DNA polymerases employ additional proteins to ensure that the template remains attached for continuous DNA synthesis.³⁰ These proteins are known as processivity factors. In the absence of processivity proteins, DNA polymerases fail to synthesize long DNA strands continuously. For example, vaccinia virus DNA polymerase E9 alone can synthesize only short stretches of ~10 nucleotides per template binding event.^{8,31} Poxviruses utilize a novel heterodimeric processivity factor composed of two proteins, A20 and D4. Both A20 and D4 are essential, and together with E9 they are

Table III. Sequence Alignment of Two Specific Regions in UNG Enzymes

	Region 1	Region 2
Human	247-SYAOKKGS <u>AID</u> DRKRHHVLOTA-267	278-FFGCRHESKTNELLQKSG-295
GADMO	247-SYAHKK <u>GATI</u> DRKRHHV <u>LQAV</u> -267	278-FLGCKHFSKANGLLKLSG-295
ECOLI	166-SHA <u>QKKGAII</u> DKQRHHV <u>LKAP</u> -184	197-FFGCNH <u>FVLANQWLE</u> QRG-214
MYCTU	170-RDADT <u>LKPM</u> LAAGNCVAIE <u>SP</u> -190	201-FFGSRPFSRANE <u>LLV</u> GMG-218
DEIRA	185-SYARKKK <u>LIT</u> GKNHV <u>VIESG</u> -205	215-FFGTRPFSKTNEA <u>LEKAG</u> -232
HSV-1	190-THAQ <u>N-AIRPDPRV</u> HCV <u>LKFS</u> -209	219-FGTCQH <u>FLVANRYL</u> ETRS-236
EBV	168-AKAGDK <u>ASLINSKKH</u> LVL <u>TSQ</u> -188	206-FLGN <u>NHFVLANNF</u> REKG-223
VACV	160-KT <u>DFSNIRAKL</u> E <u>SPVTTIVGY</u> -180	189-FLKD <u>RFSFEIINV</u> L <u>ELDN</u> -206
VARV	160-KT <u>DFSNIRAKL</u> E <u>SPVTTIVGY</u> -180	189-FEKD <u>RSFEIINV</u> L <u>ELDN</u> -206
MPXV	160-KT <u>DFSNIRAKL</u> E <u>SPVTTIVGY</u> -180	189-FEKD <u>RSFEIINV</u> L <u>ELDN</u> -206
CPXV	160-KT <u>DFSNIRAKL</u> E <u>SPVTTIVGY</u> -180	189-FEKD <u>RSFEIINV</u> L <u>ELDN</u> -206
CMPV	160-KT <u>DFSNIRAKL</u> E <u>SPVTTIVGY</u> -180	189-FSKD <u>SFEAINIL</u> KING-206
SPPV	160-KT <u>DFSNIKSII</u> E <u>VPVTTIIGY</u> -180	189-FDKD <u>RAFEVIN</u> V <u>ELDN</u> -206
SWPV	160-KT <u>DFSNIKSILDTPITTIVGY</u> -180	

GADMO: *Gadus morhua* (Atlantic cod); ECOLI: *Escherichia coli*; MYCTU: *Mycobacterium tuberculosis*; DEIRA: *Deinococcus radiodurans*; HSV-1: Herpes Simplex virus 1; EBV: Epstein-Barr virus; VACV: Vaccinia virus; VARV: Variola major virus; MPXV: Monkeypox virus; CPXV: Cowpox virus; CMPV: Camelpox virus; SPPV: Sheeppox virus; SWPV: Swinepox virus. VACV, VARV, MPXV, CPXV and CMPV belong to the Genus Orthopoxvirus. SPPV belongs to the Genus Capripoxvirus. SWPV belongs to the Genus Suipoxvirus.

Residues in these two regions are part of the protein-protein interface in VACV D4 homodimers, which overlaps with the protein-protein interface in VACV A20:D4 heterodimers. No analogous interface exists in any of the non-poxvirus UNGs (only 7 of 39 residues total or 18% are identical when human and VACV UNGs are compared).

Highlighted by shading are the hydrophobic residues at the dimer interface of VACV D4 that are well conserved among poxviruses but not in UNG enzymes of other species. Underlined are hydrophobic residues of UNGs from other species.

sufficient for processive DNA synthesis and replication.³² A20 is a 49 kDa protein (426 amino acids), which may be unstable or unstructured except in the presence of D4.^{8,12} Until recently, knowledge of the A20:D4 interactions was entirely based on the results of mutagenesis studies, which revealed that the N-terminal 25 residues of A20 were necessary and sufficient for binding to D4, although binding was enhanced with A20 residues 1–50.¹¹ In addition, a central region of A20 (residues 185–191) was also identified to be important for interaction with D4.^{7,8} On the other hand, little was known about the D4 residues responsible for binding to A20. A single Gly179Arg mutation in D4 impaired its interaction with A20.^{7,8} Ishii & Moss showed that deletion of 25 residues from its N- or C-terminus abolished D4's ability to interact with A20.¹¹ However, it is not clear if deletion of terminal residues affected protein folding. As discussed in the previous section the N- and C-terminal extensions are characteristic for D4, and may be important for stability of the protein.

It is clear that in the presence of A20, D4 forms a heterodimer, but the binding affinity for the A20:D4 interaction has not been measured because so far A20 could only be produced when co-expressed with D4.^{12,19} The first structural data showing interaction between these two proteins emerged from SAXS analysis.¹² For this study full length A20 was co-expressed with D4 in the baculovirus-insect cell system and the 1:1 stoichiometric complex was purified. The A20:D4 complex was soluble only in a narrow pH range of 7–8. A high salt concentration of above 300 mM was

required for the stability of the complex; this observation reaffirmed earlier suggestions that the interactions between A20 and D4 were predominantly hydrophobic. At a low protein concentration, the complex remained in a heterodimeric state but at higher concentrations a hetero-tetramer (a dimer of heterodimers) of ~150 kDa molecular weight was formed. The SAXS low-resolution envelope for the processivity factor showed an elongated shape for the (A20:D4)₂ hetero-tetramer in which A20 (of each dimer) forms the interface in the tetramer (D4:A20:A20:D4; see Fig. 1).¹² Upon addition of E9, the heterotrimeric D4:A20:E9 complex of 1:1:1 stoichiometry was formed with high affinity ($K_D < 3 \text{ nM}$). As shown in Figure 1, in this complex, E9 binds to A20 on the opposite face of the D4-binding site.¹² This model is supported by the shape complementarity in fitting the low-resolution envelopes for A20:D4 and D4:A20:E9, and is in agreement with previous reports of the absence of a direct interaction between D4 and E9.

A detailed picture of the binding of D4 and A20 was captured in the high resolution crystal structures of the A20_{1–50}:D4 complex [4OD8, 4ODA].¹⁹ A monodisperse 1:1 heterodimeric complex of D4 and A20_{1–50} was obtained by co-expressing both proteins in *E. coli*. The structure of the A20_{1–50}:D4 complex validated the role of the N-terminal domain of A20 in binding to D4 and demonstrated that a portion of the homodimer interface observed in crystal structures of D4 was engaged in interactions with A20.¹⁹ An extensive flat contact surface (1890 Å² buried surface) is formed at the heterodimer interface. D4 residues 167–180 and 191–206,

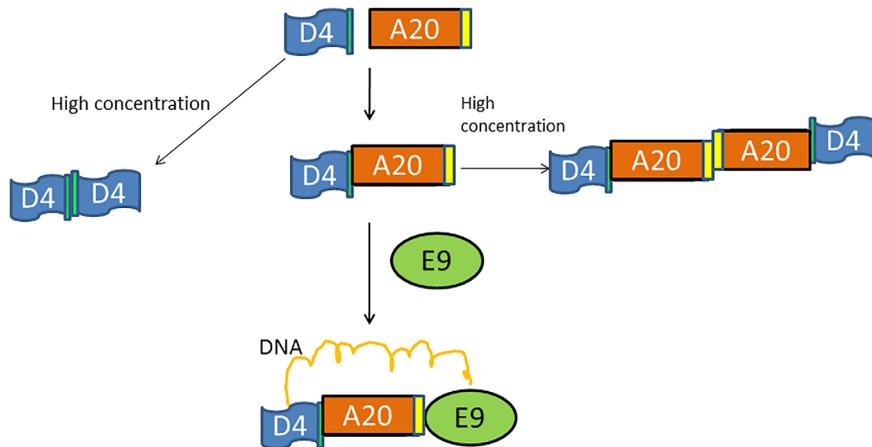


Figure 1. Schematic cartoon of processivity factor and DNA polymerase holoenzyme. A schematic drawing shows the association of D4, A20 and E9 in the poxvirus DNA polymerase holoenzyme. D4 exhibits a concentration dependent monomer-dimer equilibrium. In the presence of A20, a heterodimer is formed. The D4 homodimer interface (shown in green) partially overlaps with the heterodimer interface in A20:D4. This heterodimer also self-associates at higher concentration forming a dimer of heterodimers. The interface between the dimers in the tetramer (shown in yellow) is formed by A20 subunits. In the heterotrimeric polymerase holoenzyme E9 binds to A20 but not to D4. In the polymerase complex DNA contacts both D4 and E9.

and A20 residues 1–14 and 40–47 form the interface.¹⁹ The protein-protein interface is essentially hydrophobic (accounting for ~74% of the total buried surface area). Mainly, six residues from D4 (Arg167, Pro173, Ile197, Val200, Leu201 and Leu204) and twelve amino acids of A20 (Met1, Thr2, Leu7, Leu10, Leu14, Tyr17, Tyr42, Trp43, Lys44, Ile45, Gly46 and Val47) contribute to this hydrophobic surface. Six hydrogen bonds are formed between four D4 residues (Thr175, Thr176, Arg193 and Ser194) and two A20 residues (Ser4 and Ser40). Binding of A20 has minimum impact on the overall structure of D4 and the major conformational changes in D4 were localized in the protein interface area. The sequences of interface residues in the D4 homodimer and A20_{1–50}:D4 heterodimer are shown in Table II, while the structures of the binding surfaces are compared in Figures 2 and 3.

DNA-Binding and DNA Repair

Overview

The mechanism of the glycosylase activity of UNGs and the proposed “pinch, push, plug and pull” scheme for base-flipping have been discussed previously.³³ Briefly, UNGs bind to DNA non-specifically and depend on the specificity of the uracil-binding pocket for recognition of uracil. If uracil is encountered, dU is flipped out of the base stack. Conserved serine and proline residues in three loops (“Pro-rich loop,” “Gly-Ser loop,” “Leu-intercalation loop”) bend the DNA by compressing the DNA phosphate backbone (“pinch”). Next, the Leu residue of the “Leu-intercalation loop” penetrates the minor groove (“push”), which effectively increases the lifetime of the flipped-out nucleotide in the active site (“plug”). After N-glycosidic bond cleavage, the Leu residue is retracted (“pull”) and the uracil base moves deeper into the active site while the

abasic nucleotide relaxes to a less-strained conformational state.

The high efficiency with which UNG enzymes detect rare uracil bases in DNA is intriguing. Studies indicate that UNGs do not actively participate in base pair opening in search of uracil. Instead, they passively recognize extra-helical nucleotides resulting from the thermally induced opening of base pairs.³⁴ Only specific bases that can fit into the binding pocket in the active site are cleaved by the glycosylase activity of the enzymes. A model favoring a combination of hopping and sliding on DNA has been proposed for UNG activity.^{35–37} In this model, UNG in its “closed” conformation binds non-specifically to the DNA phosphate backbone and undergoes one-dimensional sliding along the DNA for a short distance of ~10 bp.³⁵ To efficiently scan long genomic DNAs for damage UNGs use the hopping mode in which the enzyme dissociates from the DNA intermittently, and 3D diffusion through bulk solvent.^{35–37} Partial eversion of the sampled base may occur but only dU allows precise conformational adjustments of the enzyme-substrate complex that lead to the correct conformation of uracil in the enzyme’s active site after cleavage of the N-glycosidic bond. The flipping of dU can be actively orchestrated by the enzyme or can occur passively as a result of spontaneous flip.³⁷

UNGs undergo an “open-to-close” conformational change during productive cleavage of uracil from DNA substrates as was described for hUNG.³⁸ A recent study shows that under physiological conditions electrostatic interactions are major contributors to non-specific DNA-binding by UNGs, and specific DNA-binding is driven mainly by non-electrostatic interactions.³⁶ In addition, ionic interactions found in non-specific complexes are also preserved in the

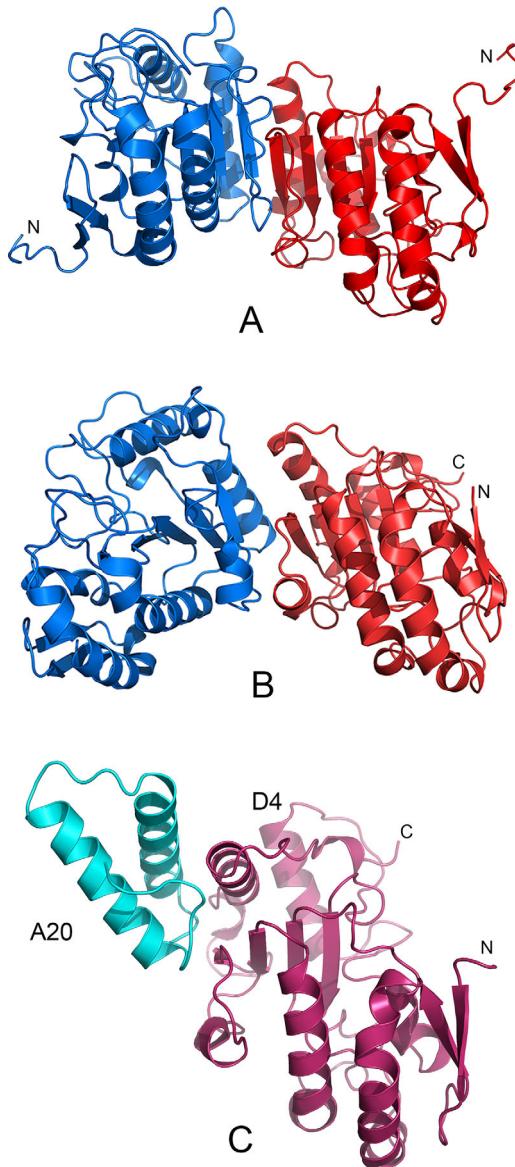


Figure 2. Homodimeric and heterodimeric packing. A. Homodimer [5JX8]. B. Homodimer [5JX3]. C. Heterodimer [4OD8]. The figure represents an overview of the homodimeric D4:D4 (5JX8: subunit A red, subunit A' marine blue; 5JX3: subunit A firebrick red, subunit B marine blue) and the heterodimeric A20₁₋₅₀:D4 (as in 4OD8: D4 pink, A20₁₋₅₀ cyan) packing. The viewpoint is the same. Notice the shift at the homodimer interface of the NCS related subunits in 5JX3 compared to symmetry-related subunits in 5JX8 resulting in a smaller interface and fewer protein-protein interactions. When D4 binds to A20 a significant contribution to heterodimeric binding is contributed by the stacking interaction of D4 residues Arg167 and Pro173 with A20 residue Trp43.

specific DNA complexes.^{34,36,37} Due to the predominance of ionic interactions non-specific DNA binding by proteins is more sensitive to high-salt concentration as compared to specific DNA binding.³⁴

The DNA glycosylase activity of D4 exhibits a stringent tolerance for ionic strength. Typically, enzymatic analyses of D4 have been performed at 60 mM NaCl.⁸ Studies by Scaramozzino *et al.* and

Duraffour *et al.* indicated that the glycosylase activity of D4 was essentially abolished at > 60 mM NaCl and 7.5 mM MgCl₂.^{14,39} The binding affinity of D4 for dsDNA (U-G substrate) and ssDNA (PS-U substrate) in the absence of salt was in the nanomolar range (K_M of 30 and 8.7 nM, respectively) but at 60 mM NaCl the affinities were reduced 100-fold

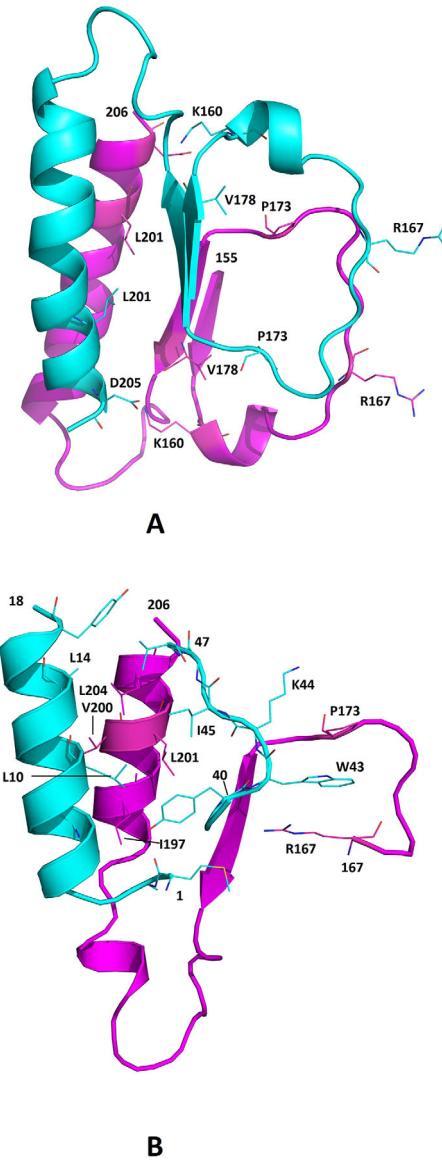


Figure 3. Homodimeric and heterodimeric interfaces. A. A close-up view of the homodimeric D4:D4 interface [5JX8] generated by crystallographic symmetry is shown as a cartoon representation (D4 subunit A: magenta; subunit A': cyan). Some important interface residues in D4 are labeled and shown as line models (subunit A: C magenta, O red, N blue; subunit A': C cyan, O red, N blue). The termini of the interacting secondary structure elements on D4 are also labeled. B. A close-up view of the heterodimeric A20₁₋₅₀:D4 interface [4OD8] is shown as a cartoon representation (D4 in magenta; A20₁₋₅₀ in cyan). Important interface residues in D4 and A20 are shown as line models (D4: C magenta, O red, N blue; C cyan, O red, N blue). The termini of the interacting secondary structure elements on D4 and A20₁₋₅₀ are also labeled.

(K_M of 2.6 and 0.5 μM , respectively). In both cases D4's catalytic activity was a fraction (2–3 orders of magnitude slower) of the activity of hUNG. In addition, at 150 mM NaCl the dissociation constant (K_D) for specific and non-specific complexes dropped to $> 50 \mu M$ for both dsDNA and ssDNA.⁹ At this salt concentration, only DNA substrates containing abasic sites bind with low micromolar or sub-micromolar affinity (K_D of 0.39–8.0 μM). Combined, these data emphasize the functional significance of non-specific DNA binding by D4.

D4-DNA complexes

Two crystal structures of vaccinia virus D4 in complex with dsDNA have been described. The crystal structure of D4 bound to a non-specific dsDNA represented the first UNG structure with truly undamaged DNA [4QCB].²⁰ The only other UNG complex with any non-specific DNA was reported for hUNG bound to a damaged or destabilized non-specific DNA [2OXM].³⁴ The second D4:DNA complex captured an uracil-containing specific dsDNA bound to the A20₁₋₅₀:D4 complex [4YIG].⁹ The DNA construct in this D4 specific DNA complex is similar to that used for the hUNG:DNA complex described previously [1SSP],⁴⁰ and thus allows us to compare DNA binding by human and poxvirus UNGs. Moreover, availability of D4 structures in complex with both specific and non-specific DNA make it possible to observe the associated structural transitions in D4. It should be noted that crystals of both DNA complexes of D4 were obtained at low salt concentrations (non-specific complex: 60 mM KCl; specific complex: < 100 mM NaCl).

D4 in complex with undamaged dsDNA [4QCB]

The non-specific DNA complex was prepared using a self-complementary blunt-end 12 nucleotide construct with the sequence 5'-GCA AAC GTT TGC-3' (forward strand)/3'-CGT TTG CAA ACG-5' (reverse strand). In the crystal structure the DNA duplex is located between two D4 subunits. The two D4 subunits interact with a different 4-nucleotide segment in each strand of the dsDNA but there is little interaction between the two protein subunits. D4 subunit B binds to nucleotides 5–8 [—ACGT—] of DNA subunit C and D4 subunit A interacts with nucleotides 23–26 [—AAC—] of complementary DNA subunit D. An overview of the D4 complex with the undamaged non-specific dsDNA (subunits ACD) is shown in Figure 4. Figure 4(A) provides a schematic diagram of the DNA double helix and shows the interacting protein residues with hydrogen bonds (< 3.9 Å). Figure 4(B) displays a cartoon representation of the DNA-binding loops and residues in the D4 complex with non-specific DNA.

Although the protein subunits bind to different regions (and sequence) of DNA, the same thirteen

residues of each D4 subunit form the interface with four nucleotides from each DNA strand and bury a total of 580–590 Å² of the solvent-accessible surface area (Supporting Information Table VII). One of the thirteen interface residues (Ala183) and two additional residues of each D4 subunit also bind to the opposite DNA strand but the buried surface for this interface is only ~230 Å². The fifteen D4 residues that participate in protein-DNA interactions are distributed in the “water-activating loop” (Ile67, Pro71), the “Pro-rich loop” (Lys87), the “extended DNA-binding loop” (Gly128, Glu129, Thr130, and Lys131), the “Gly-Ser loop” (Gly159, Lys160, Thr161, Asp162, and Asn165) and the “Leu-intercalation loop” (Tyr180, His181 and Ala183). The main chain N atoms of D4 interface residues Thr130, Lys160, Thr161 and His181, side chain NZ atom of Lys131 and the side chain hydroxyl OH atom of Tyr180 are involved in hydrogen bonding with the phosphate backbone. No interactions with the base or the ribose sugars are observed. The electrostatic potential of the contact surface between D4 and the DNA illustrates the charge complementarity of the positively charged DNA binding surface of D4 and the negatively charged DNA phosphate backbone.^{9,20} Presently, we cannot compare binding of undamaged DNA to D4 and hUNG since a structure of the latter with undamaged non-specific DNA is not available. Comparison of sequence motifs used for DNA-binding and catalysis show that among the D4 residues involved in hydrogen bonding to DNA, only the active site residue His181 is conserved in hUNG (His268). Of the fifteen residues of D4 that form the interface with DNA only two, Gly159 and His181, are conserved in hUNG (Gly246 and His268). Two conserved characteristic active site motifs, defined as Motif A and Motif B, have previously been identified in family I UNGs.¹ These two motifs represent the “water-activating loop” and “Leu-intercalation loop,” respectively. As shown in Table IV, Motifs A and B are conserved in human, *E. coli* and Herpes Simplex Virus UNGs but differ in D4. Although the active site residues that form the uracil-binding pocket (Asp68, Tyr70, Phe79, Asn120, His181: in D4) are identical in all UNGs across the species, catalytic residue Arg185 (in D4) substitutes for the conserved leucine residue (Table IV).

D4 in complex with uracil-containing specific dsDNA [4YIG]

This crystal structure provides a snapshot of the post-uracil-excision state of D4. The structure reveals how D4 exploits the altered motifs to perform the glycosylase activity, an evolutionarily conserved mechanism for DNA repair, yet maintains distinct conformational features that may be important for its essential role in DNA synthesis. The complex was crystallized by mixing A20₁₋₅₀:D4 and a 10-mer DNA duplex (forward

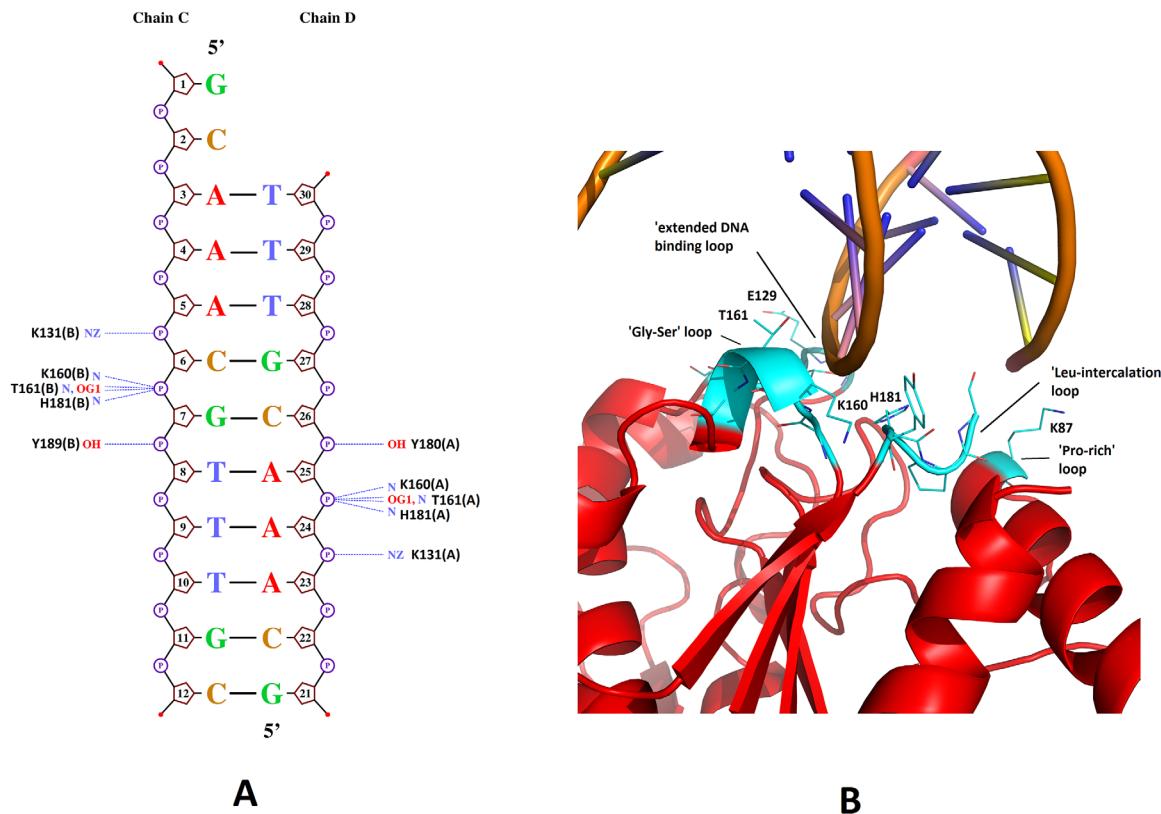


Figure 4. DNA binding: non-specific DNA complex of D4 [4QCB]. **A.** Schematic diagram of D4:DNA interactions. The forward strand of the dsDNA runs from top (5') to bottom (3') on the left side while the reverse strand on the right side runs in the opposite direction. DNA bases are represented in one-letter code (A red; T blue; C orange; G green) and the base-pairs are connected by a solid black line. The DNA backbone (sugars are depicted as brown pentagons, phosphates as purple circles) is drawn adjacent to the DNA bases. Nucleotide numbering is shown inside the sugars. Blue (dashed) lines represent hydrogen-bonding contacts ($<3.9 \text{ \AA}$) as defined in PDBePISA (<https://www.ebi.ac.uk/pdbe/pisa/>). Protein-DNA interactions are plotted on either side of the DNA stands, and interacting protein residues are represented by their residue name (one-letter code) and residue number (with chain ID in parenthesis). For hydrogen bonding interactions the atom name (O red, N blue) is also shown. On the left side are D4 residues of subunit B that show hydrogen bonds ($< 3.9 \text{ \AA}$) with the phosphate backbone of nucleotides 5-8 (-ACGT-) in DNA strand C. On the right side are the corresponding D4 residues of subunit A with hydrogen bonds to the phosphate backbone of nucleotides 23-26 (-AACAC-) in reverse DNA strand D. Notice that there are no non-bonded contacts shown [3.5 \AA cut-off as in Figs. 6(A) and 7(A)]. **B.** Overview of non-specific D4:DNA Complex [4QCB]. The figure shows D4 in complex with non-specific dsDNA (subunits A, C, D; 4QCB) in cartoon representation (D4: red; DNA: orange). Interface residues of D4 from four DNA binding loops ("Pro-rich loop": 87; "extended DNA-binding loop": 128-132; "Gly-Ser loop": 158-162; "Leu-intercalation loop": 180-183) are highlighted as line models (color code: C cyan, O red, N blue). Several residues in each motif and the motifs are labeled.

strand: 5'-CTG TUA TCT T-3'; reverse strand: 3'-GAC AAT AGA A-5') containing a uracil base (underlined) in one strand. In the crystal structure the uracil base was trapped in the uracil binding pocket of D4 suggesting that base excision occurred during incubation and/or crystallization of the complex. Analysis of this structure, the A20₁₋₅₀:D4:uracil complex [4YGM] and our D4:uracil complex [4LZB] emphasizes similar

binding to uracil irrespective of the origin of uracil (through catalysis of uracil-containing DNA or co-crystallization with uracil). The dsDNA containing the abasic site (resulting from the catalytic activity) was bound to D4 at the conserved DNA binding position and nearly perpendicular to A20₁₋₅₀ (see Fig. 5). The "Leu-intercalation loop" of D4 is inserted into the DNA minor groove.

Table IV. Comparison of the Two Active Site Motifs and Active Site Residues in UNG Enzymes

	Motif A	Motif B	Active site residues
1AKZ (Human)	143 -GQDPYH-148	268 -HPSPLSVYR-276	D145,Y147,F158,N204,H268,L272
1UDG (HSV)	86 -GQDPYH -91	210 -HPSPLSKV -217	D88,Y90,F101,N147,H210,L214
1EUG (<i>E. coli</i>)	62 -GQDPYH -67	187-HPSPLSAHR-195	D64,Y66,F77,N123,H187,L191
2OWQ (Vaccinia)	66 -GIDPYP -71	181 -HPAARDR -187	D68,Y70,F79,N120,H181,R185

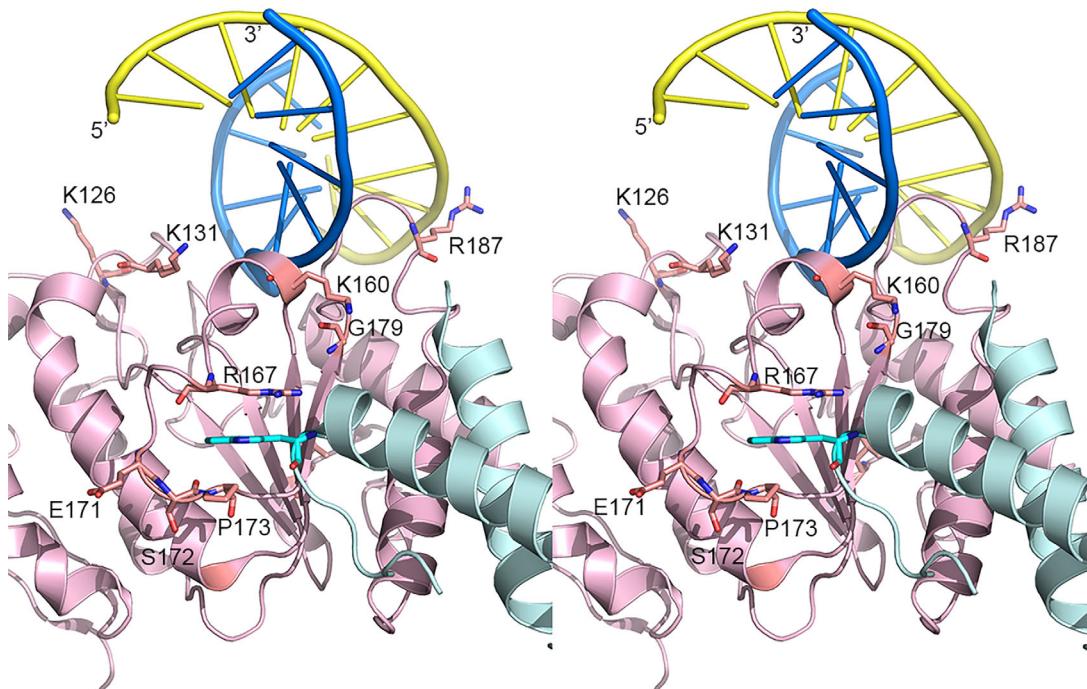


Figure 5. Binding of A20 and DNA to D4. The stereo figure shows specific uracil-containing dsDNA bound to the complex of D4 with A20₁₋₅₀ [4YIG]. The cartoon representation (A20₁₋₅₀: cyan; D4: salmon; DNA: blue and yellow) clearly demonstrates binding of DNA to D4 in the A20₁₋₅₀:D4:DNA complex. The locations of the mutations that lead to reduction of activity in supporting DNA synthesis (R167A, Δ171-172, 3G171-173, K126V, K131V, K160V, P173G, G179R, R187V) are shown as stick models (C salmon, O red, N blue) and are labeled. Residue Trp43 of A20₁₋₅₀ that provides stacking interactions with D4 residues Arg167 and Pro173 is also shown as stick model (C cyan, O red, N blue).

A total of 29 residues belonging to six regions in D4 (“catalytic water-activating loop,” “Pro-rich loop,” “extended DNA-binding loop,” “Gly-Ser loop,” “Leu-intercalation loop” and “uracil-specificity β-strand”) are involved in interactions with DNA and the cleaved uracil (Table V; Supporting Information Table VII). The first five of these areas are the same as those involved in interactions with non-specific DNA discussed in the previous section. Sequence comparison of the hUNG and D4 residues in the DNA binding motifs and the catalytic domain shows 25% sequence identity (12 out of 48 residues in regions 1-6 are identical; see Table V).

In general UNG’s interactions with specific DNA can be separated as: (1) interactions with the uracil-containing strand, (2) interactions with the abasic site, (3) interactions with the complementary opposing strand, and (4) interactions with uracil in the binding pocket.

Nineteen D4 residues form the interface with five nucleotides (3-4 [—GT—], and 6-8 [—ATC—]) adjacent to the abasic nucleotide 2-deoxy-5-phosphono-ribose (ORP) in the middle of the forward strand (subunit C). This D4:DNA interface buries a total of 1140 Å² of the solvent-accessible surface area. Five D4 residues (Lys160 N, Thr161 N, Tyr180 OH, His181 N, and Arg185 NH2) form hydrogen bonds with the phosphate backbone of nucleotides 6-8 ([—ATC—]). Among these only His181 is conserved in hUNG. The side

chain of Thr161 interacts with the deoxyribose sugar of nucleotide 7, and the side chain NZ atom of Lys87 forms a hydrogen bond with the thymine base of nucleotide 4.

A total of 12 D4 residues belong to the D4-DNA interface with the baseless nucleotide ORP. The side chain hydroxyl group of Thr130 makes hydrogen bonds with the deoxyribose sugar, while main-chain and side-chain of Ser88 are hydrogen bonded to the phosphate backbone of ORP.

The cleaved uracil located in the uracil binding pocket is in contact with eleven D4 residues as observed for the complexes of A20₁₋₅₀:D4:uracil [4YGM] and D4:uracil [4LZB]. Six of these residues belong to the “water-activating loop.” In addition, Pro78, Phe79, Ser88, Asn120 and His181 are also involved in interactions with uracil. Hydrogen bonds of uracil atoms O2, O4 and N3 are formed with residues Ile67, Asp68, Phe79 and Asn120. Uracil binding in the pocket through catalysis of uracil-containing DNA or introduction by soaking or co-crystallization only differs by the movement of catalytic residue His181 that allows a hydrogen bond with the O2 atom of uracil. This is indicative of productive DNA binding. In the human enzyme [1SSP] similar interactions are observed, which illustrates the conserved nature of the uracil binding pocket among UNG enzymes.

In addition, six D4 interface residues show contacts with six nucleotides of the opposing strand

Table V. Listing of DNA Interface Residues for Human and Vaccinia UNG in Six Specific Regions that Include the Motifs for DNA Binding and Catalysis

	Region 1	Region 2	Region 3
Human [ISSP]	143-G <u>QDPYH</u> GPNQ-152	165-PPPP <u>SLE</u> -171	201-LLL-N-204
VACV [4YIG]	66 -G <u>IDPYPKDGT</u> -75	84 -FTK <u>KSIK</u> -90	117-I <u>PWN</u> -120
VACV [4QCB]	66 -G <u>IDPYPKDGT</u> -75	84 -FTK <u>KSIK</u> -90	117-I <u>PWN</u> -120
	Region 4	Region 5	Region 6
Human [ISSP]	210-R <u>AHQANS</u> -216	245-W <u>GSYAQK</u> -251	267-A <u>HPSPLSVYR</u> GFF-279
VACV [4YIG]	126-KL <u>GETKS</u> -132	159-G <u>KTD</u> FSN-165	180-Y <u>H</u> <u>PAAR</u> -DRQFE-190
VACV [4QCB]	126-KL <u>GETKS</u> -132	159-G <u>KTD</u> FSN-165	180-Y <u>H</u> <u>PAAR</u> -DRQFE-190

Motifs included in regions 1–6 (numbering based on hUNG).

a) Region 1: “Catalytic water-activating loop” (hUNG numbering: 143-GQDPYH-148).

b) Region 2: “Pro-rich loop” (hUNG numbering: 165-PPPPS-169).

c) Region 3: Uracil specificity β-strand (hUNG numbering: 201-LLL-N-204).

d) Region 4: “Extended DNA-binding loop” (hUNG numbering: 212-HQAN-215).

e) Region 5: “Gly-Ser loop” (hUNG numbering: 246-GS-247).

f) Region 6: “Leu-intercalation loop” (hUNG numbering: 268-HPSPLSVYR-276).

Shown are regions 1–6 that include the motifs for DNA binding and catalysis for human UNG (ISSP: specific uracil-containing dsDNA) and vaccinia virus UNG (4YIG: specific uracil-containing dsDNA; 4QCB: non-specific dsDNA*). Only 12 of 48 residues total in regions 1–6 are identical (25%) between hUNG and VACV UNG (D4).

Residues that interact with the abasic ORP site:

4YIG: 67-71 (IDPYP), 86-88 (KKS), 130 (T), 181 (H), 183 (A), 185 (R).

ISSP: 144-148 (QDPYH), 152 (Q), 158 (F), 167-169 (PPS), 214 (A), 268 (H), 270-272 (SPL).

Residues that interact with the cleaved uracil (URA):

4YIG: 66-71 (IDPYP), 78 (P), 79 (F), 88 (S), 120 (N), 181 (H).

ISSP: 144-147 (QDPY), 153 (A), 156-158 (LCF), 204 (N), 268 (H).

DNA interface residues with hydrogen bonds are highlighted in bold (of these residues with side chain hydrogen bonds are underlined), and DNA interface residues with non-bonded contacts are shaded.

*Residues 184–188 were missing from the PDB entry 4QCB due to disorder.

(subunit D). Of these, the side chain NE of Arg187 participates in hydrogen bonding with the deoxyribose of nucleotide 28, and Arg187 NH1 forms a hydrogen bond with the phosphate backbone of nucleotide 29. The buried surface of this D4:DNA interface (subunits A and D) is less than 50% (487 Å²) of the main interface.

Only nine (Asp68, Pro69, Tyr70, Phe79, Ser88, Asn120, Gly159, His181 and Arg187) of the 29 residues that form the interface with DNA (including binding of ORP and uracil) are conserved in D4.

An overview of the D4 complex with specific uracil-containing DNA (subunits ACD) is shown in Figure 6. Figure 6(A) provides a schematic diagram of the specific double-stranded DNA, and shows D4:DNA interface residues with hydrogen bonds (< 3.9 Å) and non-bonded contacts (< 3.5 Å). Figure 6(B) displays a cartoon representation of the structure highlighting the interacting loops and key interface residues of D4.

Differences between specific DNA complexes of D4 and hUNG

Comparison of free hUNG [IAKZ] with its specific DNA complex [ISSP] showed an “open-to-close” conformational change, which is characteristic for productive DNA binding.⁴⁰ On the other hand, a similar conformational transition was not observed in D4 upon DNA binding.⁹ Except for the movement of the intercalation loop and a few changes in the orientations of the side chains of the interface residues the conformation of D4 in the DNA complex [4YIG] and

in the DNA-free state [4OD8; 4YGM] is very similar.⁹ An overview of the hUNG complex with specific dsDNA is shown in Figures 7(A) (schematic diagram of protein-DNA interactions) and 7(B) (cartoon representation highlighting the interacting loops). A detailed analysis of the hUNG:DNA interface is provided in the Supporting Information (Supporting Information Table VII).

The conformation of the specific DNA in the D4 and the hUNG complexes is very similar. Burmeister *et al.* suggested that the conservation of the DNA conformation in these two complexes is indicative of a dominant role of the DNA conformation for UNG function.⁹ The same nucleotides are involved in binding to both proteins but orientation of the protein on the DNA differed by 15° in the two complexes (see Fig. 8).⁹

The contact surface between D4 and DNA [in 4YIG] may be more positively charged than in the hUNG:DNA complex [ISSP].⁹ Overall, D4 interface residues in the specific complex are more charged, and there are more electrostatic interactions than in the hUNG complex.^{9,40} In hUNG, the Pro and Ser residues of the “Pro-rich loop” and the “Gly-Ser loop” are involved in DNA backbone pinching. In D4, three basic residues (Lys86, Lys87, Lys160) replace these interactions. Moreover, in D4 the catalytic residue Arg185 of the “Leu-intercalation” loop replaces the conserved Leu residue in hUNG [Fig. 9(A); Tables IV and V]. This latter replacement renders the protein-DNA interface more charged in the D4

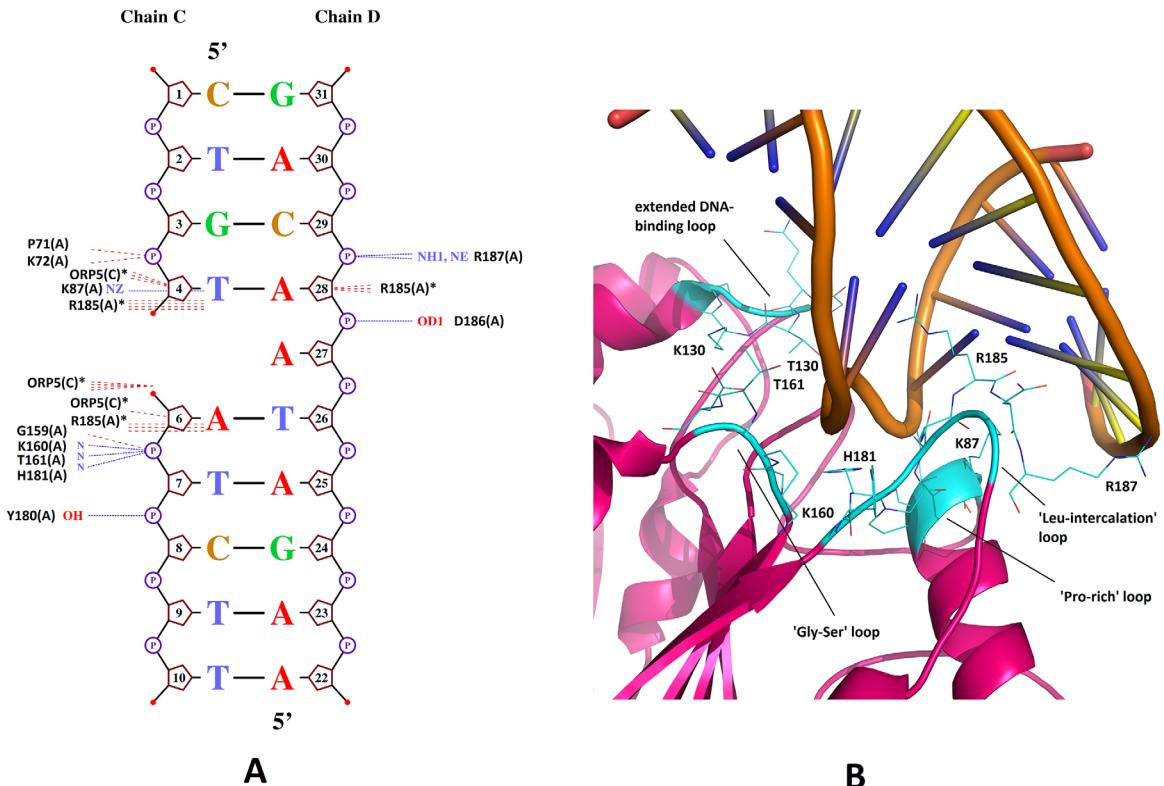


Figure 6. DNA binding and DNA repair: specific DNA complex of D4 [4YIG]. A. Schematic diagram of D4:DNA interactions. Blue (dashed) lines represent hydrogen-bonding contacts ($<3.9\text{ \AA}$) while non-bonded contacts ($<3.5\text{ \AA}$) are depicted by brown (dashed) lines. The relative low cut-off for non-bonded contacts is chosen to avoid clutter in this Figure. For generation of the schematic diagram the program NUCPLOT (<http://www.ebi.ac.uk/thornton-srv/software/NUCPLOT/>) recognizes only regular nucleotides. Therefore, the interactions of the baseless nucleotide (ORP5) in this figure are highlighted instead as multiple non-bonded contacts, although ORP is of course still bonded to nucleotides 4 and 6 after cleavage of the uracil base. The asterisk “**” indicates that the corresponding protein residue (R185) and the nucleotide (ORP5) are shown more than once on the plot. Notice that the intercalating charged residue R185 displays multiple non-bonded contacts in the vicinity of the abasic site with both DNA strands but no hydrogen bonds are observed. D4 residues D186 and R187 make hydrogen bonds with the phosphate backbone of the DNA strand opposite to the lesion, which may have a stabilizing effect. B. Overview of specific D4:DNA Complex [4YIG]. The figure depicts D4 in complex with specific dsDNA [subunits ACD only] in cartoon representation (D4: pink; DNA: orange). D4 interface residues from four DNA binding loops (“Pro-rich loop”: 86-88; “extended DNA-binding loop”: 128-132; “Gly-Ser loop”: 158-162; “Leu-intercalation loop”: 180-187) are shown as line models (color code: C cyan, O red, N blue). The motifs and several residues in each motif are labeled.

complex.⁹ The Arg185 substitution is especially notable since in UDGs in which DNA distortion results from side-chain intercalation into the minor groove, the intercalating side chain is usually hydrophobic.⁴¹ The predominance of positively charged residues in motifs involved in interactions with DNA may account for the increased sensitivity of D4 to higher ionic strength.^{14,39}

Differences between the non-specific and the specific DNA complex of D4

Like in the specific DNA complex there was no major conformational change in D4 upon binding to non-specific DNA.²⁰ Moreover, D4 subunits in the non-specific and the specific DNA complexes superimpose well (*rmsd* of 0.5 \AA), and major differences are confined to the residues in the “Leu-intercalation” loop and residues that interact with

A20₁₋₅₀. In contrast to the non-specific DNA complex where no kinking and bending of the DNA was observed, in the specific DNA complex the DNA is bent and kinked to the same degree as in the similar complex of hUNG [1SSP].^{9,20,40} Thus, a combination of the pinching action of residues in the “Pro-rich loop” and the “Gly-Ser loop” and the intercalation of Arg185 in the “Leu-intercalation loop” is sufficient for generating the distortion of the DNA helix in the specific DNA complex of D4 without the need for an “open-to-close” conformational change in the protein. On the other hand, since the same residues in the “Pro-rich loop” and the “Gly-Ser loop” interact with the phosphate backbone in the non-specific DNA complex but these interactions do not lead to distortion of the DNA helix, it may imply a more important role for the “Leu-intercalation loop” in distorting the DNA. Since the residues 184–188 of

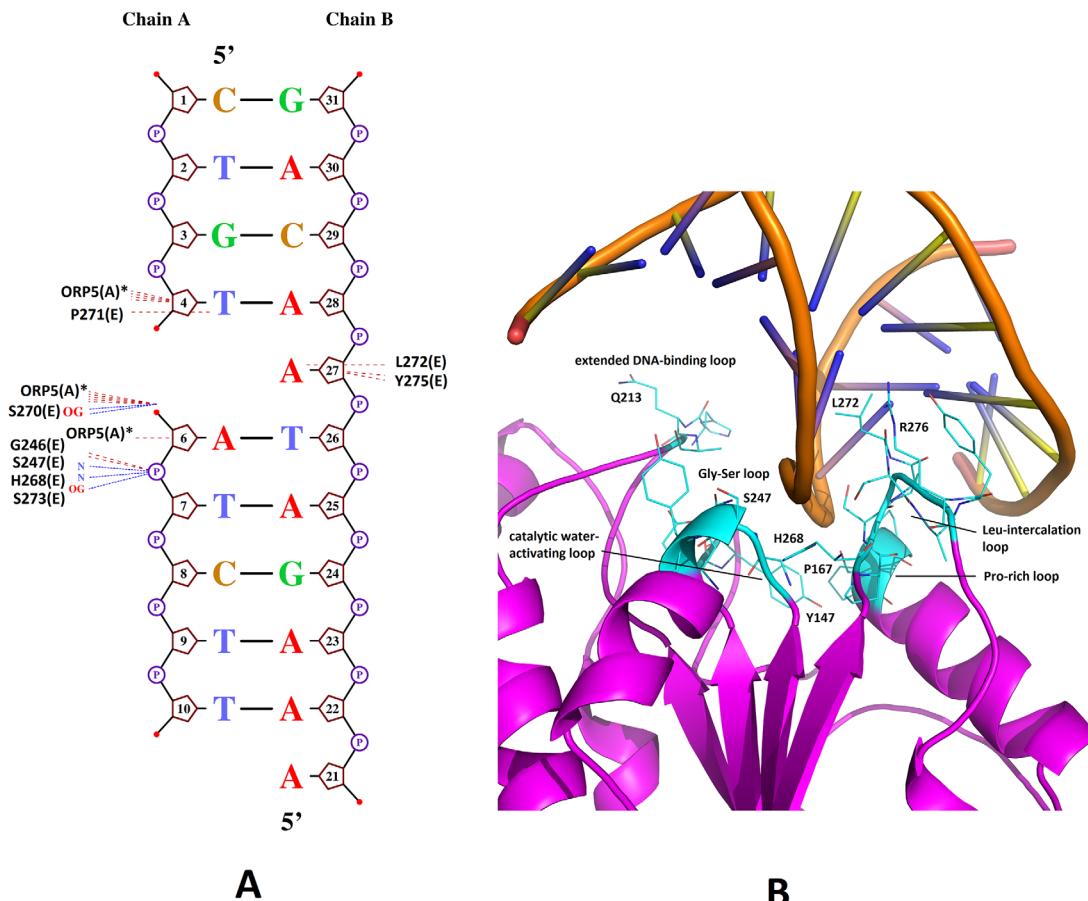


Figure 7. DNA binding and DNA repair: specific DNA complex [1SSP]. A. Schematic diagram of hUNG:DNA interactions. The overall description for this schematic diagram is the same as described in Figure 4(A). Blue (dashed) lines represent hydrogen-bonding contacts ($<3.9 \text{ \AA}$) while non-bonded contacts ($<3.5 \text{ \AA}$) are depicted by brown (dashed) lines. The cut-off for non-bonded contacts is chosen to avoid clutter. As described for Figure 6(A), ORP is bonded to nucleotides 4 and 6 but it is not recognized as a regular nucleotide by the program NUCPLOT (<http://www.ebi.ac.uk/thornton-srv/software/NUCPLOT/>). Therefore, instead of bonds interactions of ORP are highlighted as multiple non-bonded contacts. The asterisk indicates that the corresponding nucleotide (ORP5) is shown more than once on the plot. Hydrogen bonds with the phosphate backbone are provided by Ser residues in the “Gly-Ser loop” (S247) and the “Leu-intercalation loop” (S270 and S273). Leu272 and Tyr275 provide non-bonded contacts with nucleotide 27 opposite to the lesion (abasic site). B. Overview of specific hUNG:DNA Complex [1SSP]. The figure shows hUNG in complex with specific dsDNA [subunits ECD] in cartoon representation (hUNG: magenta; DNA: orange). Interface residues of hUNG from five DNA binding loops (“catalytic wateractivating loop”: 144-147; “Pro-rich loop”: 167-169; “extended DNA-binding loop”: 212-214; “Gly-Ser loop”: 246-248; “Leu-intercalation loop”: 268-276) are highlighted as line models (color code: C cyan, O red, N blue). Several residues in each motif and the motifs are labeled.

the “Leu-intercalation” loop in 4QCB could not be modelled due to structural flexibility, we cannot discuss their role during the transition from non-specific to specific DNA binding by D4.

An “open-to-close” conformational change is not observed for D4 when transitioning between free form, non-specific and specific DNA complex, and most of the interface residues in the motifs for DNA-binding and catalysis are already in position in the free enzyme for interaction with the DNA phosphate backbone [Fig. 9(B); Tables IV and V; Supporting Information Table VII]. Larger movement of residues is confined only to the “Leu-intercalation” loop. The other DNA binding loops show minor changes mostly in the orientation of side-chains.

Some differences are noticed in the nature of interactions between the protein and DNA in the two complexes. In the non-specific DNA complex the side chain of Lys87 forms a hydrogen bond with the phosphate backbone of the opposite strand, while in the specific DNA complex the side chain of Lys87 is hydrogen bonded to the thymine base of the nucleotide adjacent to the baseless nucleotide ORP. In addition, Thr130 OH interacts with the deoxyribose sugar of the baseless nucleotide ORP in the specific DNA complex, while in the non-specific complex the main chain nitrogen atom of Thr130 and the side chain NZ atom of the adjacent Lys131 form hydrogen bonds with the phosphate backbone. These two residues are part of the “extended DNA binding” loop. Comparison of the two DNA complexes reveals

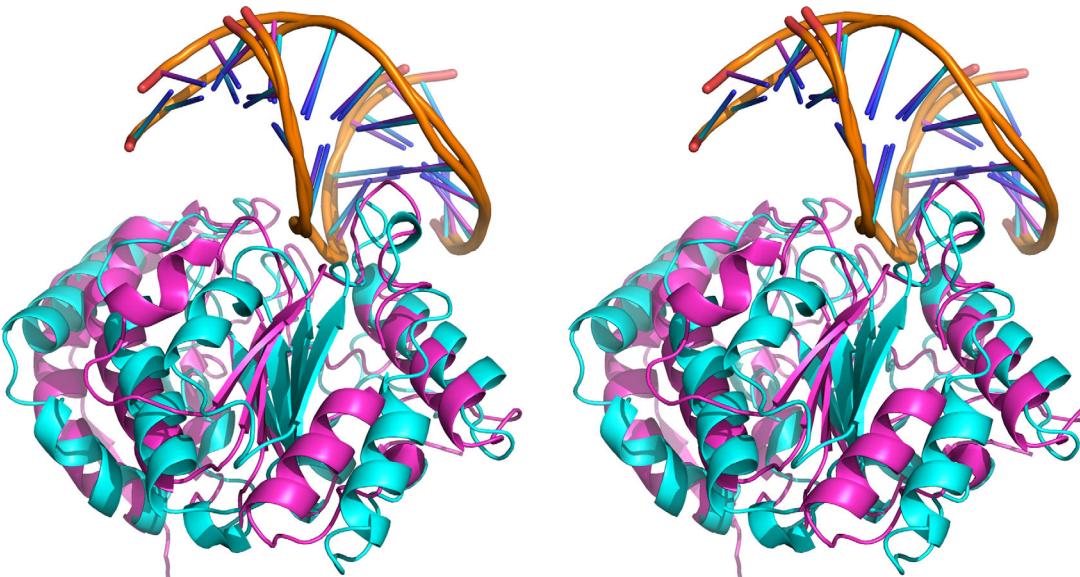


Figure 8. Superimposition of hUNG and D4 in complex with specific DNA [1SSP and 4YIG]. The stereo figure shows the superimposition of 1SSP (subunits ECD) and 4YIG (subunits ACD only). As compared to hUNG D4 is rotated by ~15°. Protein and DNA are in cartoon representation (D4: light magenta; hUNG: cyan).

that this loop superimposes quite well [Fig. 9(B)]. Therefore, the observed differences in their interactions may be related to the bending and kinking of the specific DNA in the vicinity of these residues after uracil excision.

The other difference observed between the non-specific and the specific DNA complexes is a consequence of A20 binding in the latter complex. In the non-specific DNA complex Arg167 is involved in hydrogen bonding with the main chain oxygen atoms of Ser172 and Val174, and the side chain of Thr176.²⁰ However, in the specific DNA complex (as in the A20₁₋₅₀:D4 complex) Arg167 and Pro173 of D4 are engaged in stacking interactions with A20 residue Trp43 at the A20₁₋₅₀:D4 interface.^{9,19}

D4 Homodimer Interface Mutants

Before any structural data on the complexes of D4 with A20 or DNA were available, in order to investigate the functional significance of the D4 homodimer interface observed in the crystals we generated three mutants by altering one or more amino acid residues on a loop (residues 167–173) at the interface (see Fig. 5 for their location with respect to A20₁₋₅₀ and dsDNA binding sites).¹⁸ This loop represents one of the most flexible parts in the D4 structure.^{15,17,20} Using site-directed mutagenesis, recombinant D4 mutants were prepared by (i) replacing Arg167 with Ala, (ii) deleting residues 171 and 172 (Δ 171-172) and (iii) by substituting residues 171–173 with Gly (3G171-173).¹⁸ Comparison of circular dichroism (CD) spectra of purified protein samples showed that the overall folds of the mutants and *wt*D4 were similar (Supporting Information Fig. S3; method described in Supporting Information).

Crystal structure analyses showed that the overall structures of these mutants were very similar to that of *wt*D4 and the main conformational differences were in the interface loop areas.

Subsequently, the crystal structure of the A20₁₋₅₀:D4 complex revealed the importance of the interface loop area in D4 for binding to A20.¹⁹ In the complex side chains of Arg167 and Pro173 of D4 are directly involved in cationic π-stacking interactions with Trp43 of A20. Mutation of Arg167 to alanine and Pro173 to glycine also adversely affected the stability of the complex of each D4 mutant with A20.¹⁹ The experimental melting temperature (T_m) of the A20:D4 Arg167Ala complex and the A20:D4 Pro173-Gly complex was lower by 3°C and 4°C, respectively, as compared to that for the A20:*wt*D4 complex.¹⁹ Structures of the three loop mutants are shown in Figure 10. For the Arg167Ala mutant [4QCA] it can be seen that the substitution of Arg167 by alanine removes the crucial stacking interaction of Arg167 with A20 residue Trp43 [Fig. 10(A)]. For the Δ 171-172 mutant [4IRB] and the 3G mutant [4QC9] the homodimer interfaces highlight an additional short two-stranded β-sheet not seen in *wt*D4 [Fig. 10(B,C)].

Recently crystal structures of A20₁₋₅₀ in complex with D4 Arg167Ala and Pro173Gly mutants have been reported.²⁸ Although these structures revealed only local changes at the site of the mutation, the mutations result in significant reduction in the interface surface area. In the crystal structure of the A20₁₋₅₀ Trp43Ala mutant in complex with *wt*D4 the Arg167 side chain remains disordered. Thus these structures reiterate the importance of this region in D4 for interaction with A20, although Arg167 seems to be not essential for binding.

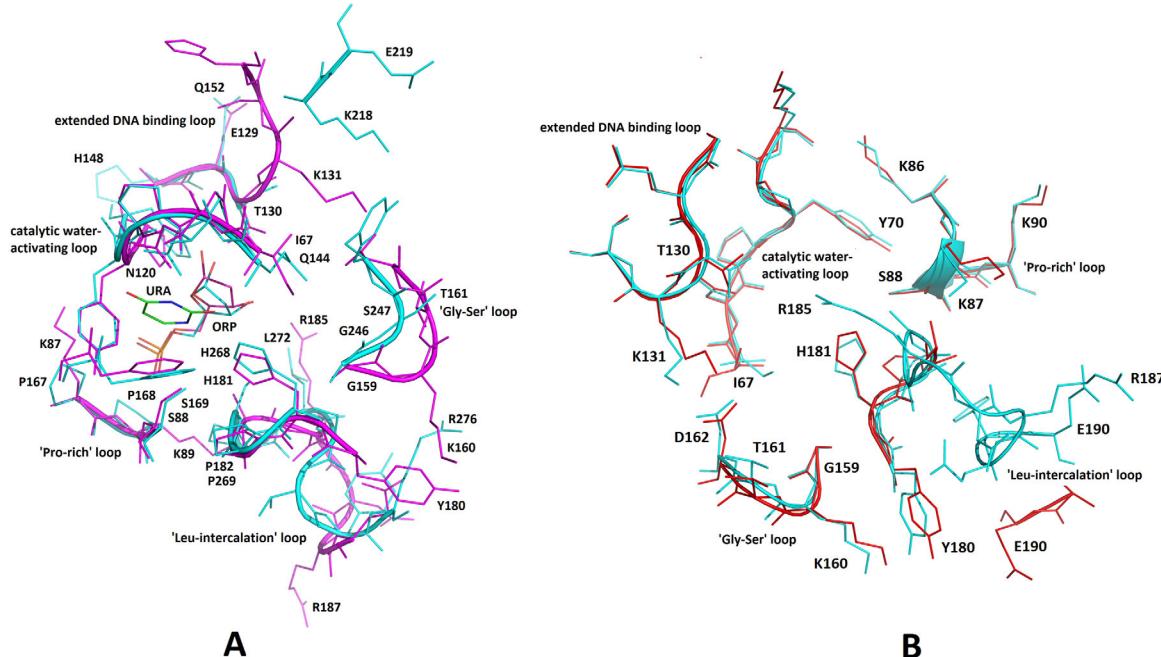


Figure 9. DNA binding and DNA repair: D4-DNA complexes - motifs and interface residues. A. Superimposition of motifs and interface residues in specific dsDNA complexes of D4 [4YIG] and hUNG [1SSP]. The figure shows the superimposition of motifs and interface residues (cartoon and line representation) for the two complexes of D4 [4YIG] and hUNG [1SSP] with specific dsDNA (4YIG: magenta; 1SSP: cyan). Although some motifs (“catalytic water-activating loop”; “Pro-rich loop”; part of the “extended DNA-binding loop”) superimpose quite well, a large number of interface residues in these loops in D4 differ from otherwise reasonably conserved (in UNG enzymes) residues. In addition, larger conformational differences are observed for the “Gly-Ser loop” and the “Leu-intercalation loop”. Several important DNA interface residues are labeled for both enzymes. The motifs, uracil and the baseless nucleotide ORP are also labeled. B. The figure shows the superimposition of motifs and interface residues (cartoon and line representation) for the two D4 complexes with non-specific (4QCB: red) and specific (4YIG: cyan) dsDNA. Several important interface residues and the motifs are labeled. As can be seen motifs and interface residues superimpose quite well. The “Leu-intercalation loop” is incomplete in 4QCB due to disorder.

Processivity Mutants

In an earlier study, Druck Shudofsky *et al.* generated a series of D4 mutants and showed that four D4 point mutants (Lys126Val, Lys131Val, Lys160Val, and Arg187Val) were deficient in DNA processivity function but each retained the ability to bind DNA and A20 (see Fig. 5 for their location in D4 with respect to A20₁₋₅₀ and dsDNA binding sites).¹⁶ However, a quantitative comparison of DNA-binding affinity of these mutants and *wt*D4 was not made. With the exception of Lys131Val, all mutants also retained glycosylase activity. The crystal structure of one of these mutants (Arg187Val) was described [3NT7].¹⁶ The overall structure of this mutant was similar to *wt*D4 but showed some changes in the electrostatic properties near the site of the mutation. Arg187 is important for binding of UNGs to dsDNA since the corresponding residue Arg276 in hUNG was found to bind to dsDNA and a mutation of this residue led to a reduced affinity for dsDNA.⁴² In the crystal structure of A20₁₋₅₀:D4:dsDNA [4YIG] the side-chain of D4 residue Arg187 interacts with a deoxyribose sugar of one of the DNA strands. The loss of this interaction between the DNA and the

Arg187Val mutant may be responsible for the reduction in D4’s ability to support DNA processivity.

Lys160 and Thr161 are part of the “Gly-Ser” loop in D4. In the DNA complexes these residues form hydrogen bonds through the peptide nitrogen atoms with the DNA phosphate backbone.^{9,20} Only the side chain hydroxyl group of Thr161 interacts with a deoxyribose sugar in the specific DNA complex. The side chain nitrogen atom of Lys160, on the other hand, makes a hydrogen bond with the peptide oxygen atom of Val178 in the C-terminal β-strand of D4 in both DNA complexes. The same interaction is also observed in the free *wt*D4 structures, and therefore it may play a role for the overall stability of the protein structure.²⁰

The side chain NZ atom of Lys131 is hydrogen bonded to the carboxyl oxygen atoms of Asp162 in the free and in complexed D4 structures, which may be important in stabilization of the local structure.²⁰ Although Lys131 makes no contact with the strand containing the lesion in the specific DNA complex, the adjacent side chain of Thr130 is hydrogen bonded to the deoxyribose sugar of the baseless nucleotide (ORP). In contrast, in the non-specific DNA complex the side chain of Lys131 shows a hydrogen

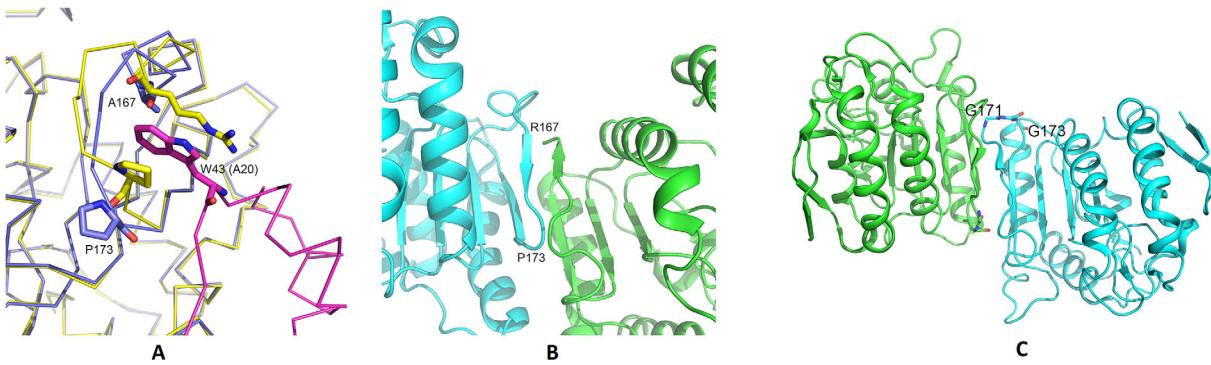


Figure 10. Structures of three D4 interface mutants. A. D4 Arg167Ala mutant [4QCA]. The figure shows a close-up view of the site of mutation for the Arg167Ala mutant. D4 subunit A from the crystal structure of the mutant (blue) was superimposed on the D4 subunit A (yellow) from the A20₁₋₅₀:D4 complex [4OD8]. A20₁₋₅₀ is shown in magenta. In the complex D4 Arg167 is involved in a stacking interaction with Trp43 of A20. Shown in stick model are Arg167 and Pro173 of D4 in the complex (C atoms: yellow), Ala167 of the mutant D4 subunit (C atoms: blue) and Trp43 of A20 in the complex (C atoms: magenta). B. Homodimer interface in the crystal structure of D4 Δ171-172 mutant [4IRB]. This cartoon drawing provides a close-up view of the homodimer interface in the crystal structure of the mutant. The homodimer interface exhibits an additional 2-stranded short β-sheet not seen in wild type D4. C. Homodimer interface in the crystal structure of D4 3G mutant [4QC9]. The cartoon drawing shows the close-up view of the homodimer interface in the crystal structure of the mutant. Subunits are colored by chain. The homodimer interface exhibits an additional two-stranded short β-sheet not seen in the wild type protein. Mutated residues are shown in stick and labeled.

bond with the DNA phosphate backbone.^{9,20} The loss of these interaction(s) could likely affect processivity activity of the Lys131Val mutant.²⁰ Consistent with its role in processivity is the fact that Lys131 is strictly conserved in poxviruses but not in UNG enzymes. However, for this mutant major alteration of the protein folding cannot be ruled out since it did not exhibit glycosylase activity.

Lys126 in D4 is located on the opposite side of the A20 binding site. In the crystal structures of the D4:DNA complexes, this residue is not in contact distance with DNA. However, in D4 crystal structures the NZ atom of Lys126 usually forms a hydrogen bond with one of the carboxyl oxygen atoms of Glu129 located in the “extended DNA-binding loop” (residues 128–132). Similar interactions are noted in the DNA complexes of D4. Interaction of this Lys side chain may be important for the stability of the “extended DNA-binding loop.” This stabilizing effect will be absent in the Lys126Val mutant.

Temperature-Sensitive D4 Mutants

A role of D4 in viral replication became apparent from studies on a temperature-sensitive (*ts*) mutant *ts*149, which was later renamed *Dts*30.^{7,8} This *ts* mutant contained a single amino acid substitution in D4 (Gly179Arg) and showed severe impairment of DNA replication when infections were performed at the non-permissive temperature (39.7°C). Later, a second *ts* mutant *Dts*27, which showed a similar phenotype, was described. A lesion in the D4 gene, which generated the Leu110Phe replacement, was shown to be responsible for the temperature sensitivity of this mutant.^{7,8} Stanitsa *et al.* reported that

the Gly179Arg mutant, but not the Leu110Phe mutant, exhibited reduced binding to A20 as compared to *wt*D4.⁷ The Gly179Arg mutant was partially impaired in virus production and DNA synthesis at 31.5°C, while both were completely blocked at 39.7°C. Extracts infected with this mutant *in vitro* were unable to sustain processive DNA synthesis.⁷ Molecular modelling using the known structure of D4 with the N-terminal domain of A20 indicates that the replacement of a Gly by an Arg residue at the end of β-strand 9 weakens the interaction of D4 with A20 since this residue is located at the A20:D4 interface.¹⁹ The crystal structure of the Leu110Phe mutant shows that the bulkier side chain of Phe110 extends into the hydrophobic core formed by a number of hydrophobic amino acids including several aromatic residues (Fig. 11). However, the structural basis for the temperature sensitivity is not clear.

Protein-Protein Interaction Inhibitors

As discussed earlier, for processive DNA synthesis in the DNA polymerase complex E9 binds to A20, which must also bind to D4. Therefore, inhibitors that disrupt interactions between D4 and A20 can inhibit viral DNA replication by E9. Several groups used different binding assays to screen libraries of compounds that interfere with D4 binding to A20 or A20 fragments.^{21,23–27} Silverman *et al.* developed a rapid plate assay to screen for compounds that block vaccinia virus DNA synthesis.²¹ In their approach, they differentiated between inhibitors specific for blocking DNA polymerase activity or processivity. Further screening campaigns combining the rapid plate assay with functional validation led to the

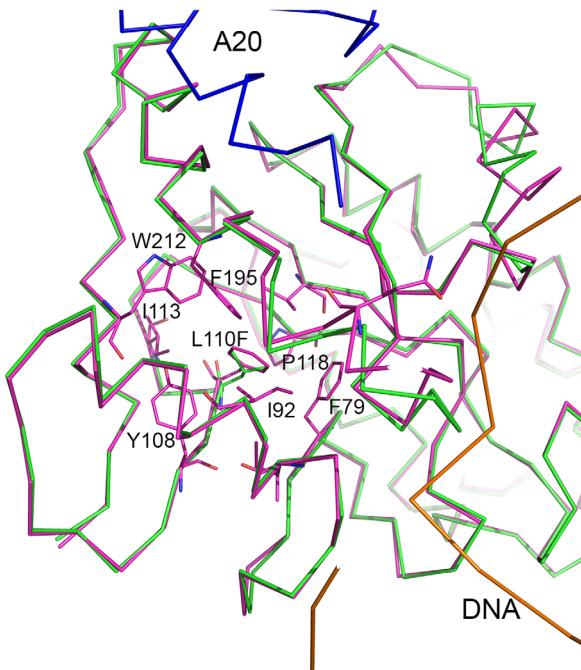


Figure 11. Structure of the temperature sensitive mutant Leu110Phe [5JX0]. The cartoon drawing highlights the site of the mutation. Subunit A of the mutant protein structure (magenta) is superimposed on the D4 subunit (green) of the A20₁₋₅₀:D4 complex [4YIG]. A20₁₋₅₀ is shown in blue and DNA in orange. The side chains of Leu110 in wtD4 and Phe110 in the mutant are inserted in the hydrophobic cleft. These residues along with some of the hydrophobic residues in the cavity are shown in stick model. The overall structure of the mutant was nearly identical to that of wtD4.

discovery of non-nucleoside inhibitors of DNA synthesis that do not target E9.^{22,25,27} IC₅₀ values (50% inhibitory concentration) for DNA synthesis inhibition of the hit compounds ranged from 26 to 239 μM and their cytotoxicity values were in range of > 100 to > 200 μM in a cytotoxicity assay.²² Medicinal chemistry efforts resulted in the synthesis of two derivatives that were ~2000 times more potent (IC₅₀ of 42 and 46 nM) than the parent scaffold.²⁷ These compounds were capable of inhibiting vaccinia virus DNA synthesis, which prevented late gene expression. The binding site for these inhibitors was predicted to be in the vicinity of D4 residue Arg193.²⁷

Using an AlphaScreen based high throughput assay we screened a library of 28,000 small molecules for inhibitors of D4 and A20 interactions.²⁴ This screening effort resulted in identification of several compounds that exhibited antiviral activity against vaccinia and cowpox virus in a cytopathic assay. In this assay the EC₅₀ value (half maximal effective concentration) for a number of compounds was below 10 μM . Further analysis using the rapid plate assay and a vaccinia virus plaque reduction assay confirmed that at least five compounds effectively reduced viral plaques (EC₅₀ = 7.1–21.5 μM) and also inhibited processive DNA synthesis

(IC₅₀ = 5.1–137.6 μM).²⁵ Calorimetric analysis was used to verify that these compounds targeted D4 (ΔT_m = 4.6–14.9°C).²⁵ Based on docking results using the crystal structure of D4 with A20₁₋₅₀ Contesto-Richefeu *et al.* predicted that at least some of these compounds may interfere with the stacking interaction of A20 residue Trp43 and D4 residues Arg167 and Pro173.¹⁹ It is suggested that the hydrophobic phenyl group in these compounds can possibly substitute for Trp43 of A20 and mimic the key stacking interactions with D4 residues Arg167 and Pro173.

Conclusions

D4 is the most diverse member of family I UNGs in terms of sequence and structure. The conserved sequence motifs of UNG enzymes that are responsible for DNA binding and catalysis are substantially altered in D4. D4 plays an essential role in viral replication as a component of the DNA processivity factor. This unique function of D4 does not require enzymatic activity but involves its physical integration into the replication machinery where the DNA scanning activity and the glycosylase activity are effectively coupled during DNA synthesis. Sequence and structure of D4 therefore have evolved to accommodate the novel activity within the evolutionarily conserved framework of UNGs.

To address how DNA processivity and DNA repair are coupled in poxvirus it is important to know the orientation of the trimeric DNA polymerase holoenzyme at the replication fork during DNA replication. Recent small-angle X-ray scattering studies of the D5 helicase domain⁴³ allowed the refinement of the model of the poxvirus replication fork that was first proposed by Sèle *et al.*¹² These authors favor a model in which D5 opens the replication fork, followed by the E9:A20:D4 holoenzyme. D4 interacts with the newly synthesized DNA after DNA replication of the leading strand. Although A20 has a D5 binding region, in the proposed model D5 is not bound to A20. In addition, the model only shows the helicase domain of D5 and currently does not allow positioning of the primase domain. Boyle *et al.* demonstrated that *in vitro* E9 can incorporate dUTP in DNA, and that the holoenzyme can excise uracil bases incorporated during DNA synthesis using the catalytically active UNG in the holoenzyme complex.⁸ The proposed orientation of D4 in the holoenzyme allows DNA repair of generated U:A mismatches during DNA replication by the enzyme. However, this model cannot explain how the holoenzyme deals with uracil sites on the leading DNA strand that is to be replicated. It has been suggested that in these cases uracil sites are excised by free viral UNG alone, independently of the DNA polymerase activity.¹² One of the most important unanswered question in the field arises from the fact that the polymerase complex stalls upon encountering an

abasic site. A protein (or an encoding gene), which can repair the abasic site, has not been identified in poxviruses. New research should be directed to discovering the mechanism of base excision repair in poxviruses.

Acknowledgments

The authors thank Drs. E Eugenia Kharlampieva and Veronika A Kozlovskaya for assistance with DLS measurement and analysis, and Darika Sartmatova and Taishayla Nash for assistance with crystallization of mutant D4.

References

- Chung JH, Im EK, Park H-Y, Kwon JH, Lee S, Oh J, Hwang K-C, Lee JH, Jang Y (2003) A novel uracil-DNA glycosylase family related to the helix-hairpin-helix DNA glycosylase superfamily. *Nucleic Acids Res* 31:2045–2055.
- Kosaka H, Hoseki J, Nakagawa N, Kuramitsu S, Masui R (2007) Crystal structure of family 5 uracil-DNA glycosylase bound to DNA. *J Mol Biol* 373:839–850.
- Lee H-W, Dominy BN, Cao W (2011) New family of deamination repair enzymes in uracil-DNA glycosylase superfamily. *J Biol Chem* 286:31282–31287.
- Schormann N, Ricciardi R, Chattopadhyay D (2014) Uracil-DNA glycosylases—structural and functional perspectives on an essential family of DNA repair enzymes. *Protein Sci* 23:1667–1685.
- Pearl LH (2000) Structure and function in the uracil-DNA glycosylase superfamily. *Mutat Res* 460:165–181.
- De Silva FS, Moss B (2003) Vaccinia virus uracil DNA glycosylase has an essential role in DNA synthesis that is independent of its glycosylase activity: catalytic site mutations reduce virulence but not virus replication in cultured cells. *J Virology* 77:159–166.
- Stanitsa ES, 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.
- Boyle KA, Stanitsa ES, Greseth MD, Lindgren JK, 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.
- Burmeister WP, Tarbouriech N, Fender P, Contesto-Richefeu C, Peyrefitte CN, Iseni F (2015) Crystal structure of the vaccinia virus uracil-DNA glycosylase in complex with DNA. *J Biol Chem* 290:17923–17934.
- McCrath S, Holtzman T, Moss B, Fields S (2000) Genome-wide analyses of vaccinia virus protein-protein interactions. *Proc Natl Acad Sci USA* 97:4879–4884.
- Ishii K, Moss B (2002) Mapping interaction sites of the A20R protein component of the vaccinia virus DNA replication complex. *Virology* 303:232–239.
- Sèle C, Gabel F, Gutsche I, Ivanov I, Burmeister WP, Iseni F, Tarbouriech N (2013) Low-resolution structure of vaccinia virus DNA replication machinery. *J Virology* 87:1679–1689.
- Ellison KS, Peng W, McFadden G (1996) Mutations in active-site residues of the uracil-DNA glycosylase encoded by vaccinia virus are incompatible with virus viability. *J Virol* 70:7965–7973.
- Scaramozzino N, Sanz G, Crance JM, Saparbaev M, Drillien R, Laval J, Kavli B, Garin D (2003) Characterization of the substrate specificity of homogeneous vaccinia virus uracil-DNA glycosylase. *Nucleic Acids Res* 31:4950–4957.
- Schormann N, Grigorian A, Samal A, Krishnan R, DeLucas L, Chattopadhyay D (2007) Crystal structure of vaccinia virus uracil-DNA glycosylase reveals dimeric assembly. *BMC Struct Biol* 7:45.
- Druck Shudofsky AM, Silverman JEY, Chattopadhyay D, Ricciardi RP (2010) Identification of vaccinia virus D4 point mutants that are defective in processivity yet retain UDG catalytic activity and are functional in A20 and DNA binding. *J Virol* 84:12325–12335.
- Schormann N, Banerjee S, Ricciardi R, Chattopadhyay D (2013) Structure of the uracil complex of vaccinia virus uracil DNA glycosylase. *Acta Cryst F* 69:1328–1334.
- Sartmatova D, Nash T, Schormann N, Nuth M, Ricciardi R, Banerjee S, Chattopadhyay D (2013) Crystallization and preliminary X-ray diffraction analysis of three recombinant mutants of Vaccinia virus uracil DNA glycosylase. *Acta Cryst F* 69:295–301.
- Contesto-Richefeu C, Tarbouriech N, Brazzolotto X, Betzi S, Morelli X, Burmeister WP, Iseni F (2014) Crystal structure of the vaccinia virus DNA polymerase holoenzyme subunit D4 in complex with the A20 N-terminal domain. *PLOS Pathogens* 10:e1003978.
- Schormann N, Banerjee S, Ricciardi R, Chattopadhyay D (2015) Binding of undamaged double stranded DNA to vaccinia virus uracil-DNA glycosylase. *BMC Struct Biol* 15:10.
- Silverman JEY, Ciustea M, Druck Shudofsky AM, Bender F, Shoemaker RH, Ricciardi RP (2008) Identification of polymerase and processivity inhibitors of vaccinia DNA synthesis using a stepwise screening approach. *Antiviral Res* 80:114–123.
- Ciustea M, Silverman JE, Druck SAM, Ricciardi RP (2008) Identification of non-nucleoside DNA synthesis inhibitors of vaccinia virus by high-throughput screening. *J Med Chem* 51:6563–6570.
- Saccucci L, Crance J-M, Colas P, Bickle M, Garin D, Iseni F (2009) Inhibition of vaccinia virus replication by peptide aptamers. *Antiviral Res* 82:134–140.
- Schormann N, Sommers CI, Prichard MN, Keith KA, Noah JW, Nuth M, Ricciardi RP, Chattopadhyay D (2011) Identification of protein-protein interaction inhibitors targeting vaccinia virus processivity factor for development of antiviral agents. *Antimicrob Agents Chemother* 55:5054–5062.
- Nuth M, Huang L, Saw YL, Schormann N, Chattopadhyay D, Ricciardi RP (2011) Identification of inhibitors that block vaccinia virus infection by targeting the DNA synthesis processivity factor D4. *J Med Chem* 54:3260–3267.
- 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.
- Nuth M, Guan H, Zhukovskaya N, Saw YL, Ricciardi RP (2013) Design of potent poxvirus inhibitors of the heterodimeric processivity factor required for viral replication. *J Med Chem* 56:3235–3246.
- Contesto-Richefeu C, Tarbouriech N, Brazzolotto X, Burmeister WP, Peyrefitte CN, Iseni F (2016) Structural analysis of point mutations at the vaccinia virus A20/D4 interface. *Acta Cryst F* 72:687–691.

29. Mol CD, Arvai AS, Sanderson RJ, Slupphaug G, Kavli B, Krokan HE, Mosbaugh DW, Tainer JA (1995) Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* 82:701–708.
30. Jeruzalmi D, O'Donnell M, Kuriyan J (2002) Clamp loaders and sliding clamps. *Curr Opin Struct Biol* 12: 217–224.
31. McDonald WF, Traktman P (1994) Vaccinia virus DNA polymerase. In vitro analysis of parameters affecting processivity. *J Biol Chem* 269:31190–31197.
32. Ishii K, Moss B (2001) Role of vaccinia virus A20R protein in DNA replication: construction and characterization of temperature-sensitive mutants. *J Virol* 75:1656–1663.
33. Parikh SS, Putnam CD, Tainer JA (2000) Lessons learned from structural results on uracil-DNA glycosylase. *Mutation Res* 460:183–199.
34. Parker JP, Bianchet MA, Krosky DJ, Friedman JI, Amzel LM, Stivers JT (2007) Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* 449:433–438.
35. Porecha RH, Stivers JT (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc Natl Acad Sci USA* 105: 10791–10796.
36. Cravens SL, Hobson M, Stivers JT (2014) Electrostatic properties of complexes along a DNA glycosylase damage search pathway. *Biochemistry* 53:7680–7692.
37. Zharkov DO, Mechetin GV, Nevinsky GA (2010) Uracil-DNA glycosylase: Structural, thermodynamic and kinetic aspects of lesion search and recognition. *Mutat Res* 685:11–20.
38. Slupphaug G, Mol CD, Kavli B, Arvai AS, Krokan HE, Tainer JA (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* 384:87–92.
39. Duraffour S, Ishchenko AA, Saparbaev M, Crance J-M, Garin D (2007) Substrate specificity of homogeneous monkeypox virus uracil-DNA glycosylase. *Biochemistry* 46:11874–11881.
40. Parikh SS, Mol CD, Slupphaug G, Bharati S, Krokan HE, Tainer JA (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J* 17: 5214–5226.
41. Dlakic M, Kerppola TK (2001) DNA structure changes coupled to protein binding. *Encyclopedia Life Sci* 1–10. DOI: 10.1038/npg.els.0003009.
42. Chen C-Y, Mosbaugh DW, Bennett SE (2005) Mutations at arginine 276 transforms human uracil-DNA glycosylase into a single-stranded DNA-specific uracil-DNA glycosylase. *DNA Repair* 4:793–805.
43. Hutin S, Ling WL, Round A, Effantin G, Reich S, Iseni F, Tarbouriech N, Schoehn G, Burmeister WP (2016) Domain organization of vaccinia virus helicase-primase D5. *J Virol* 90:4604–4613.