

Supplementary Information

Structures of Endonuclease V with DNA Reveal Initiation of Deaminated Adenine Repair

Bjorn Dalhus^{1,2,3}, Andrew S. Arvai⁴, Ida Rosnes^{1,3}, Oyvind E. Olsen^{1,3}, Paul H. Backe^{1,3},
Ingrun Alseth^{1,2}, Honghai Gao⁵, Weiguo Cao⁵, John A. Tainer⁴ & Magnar Bjoras^{1,2,3}

¹*Centre for Molecular Biology and Neuroscience (CMBN), Rikshospitalet University
Hospital, Sognsvannsveien 20, N-0027 Oslo, Norway.*

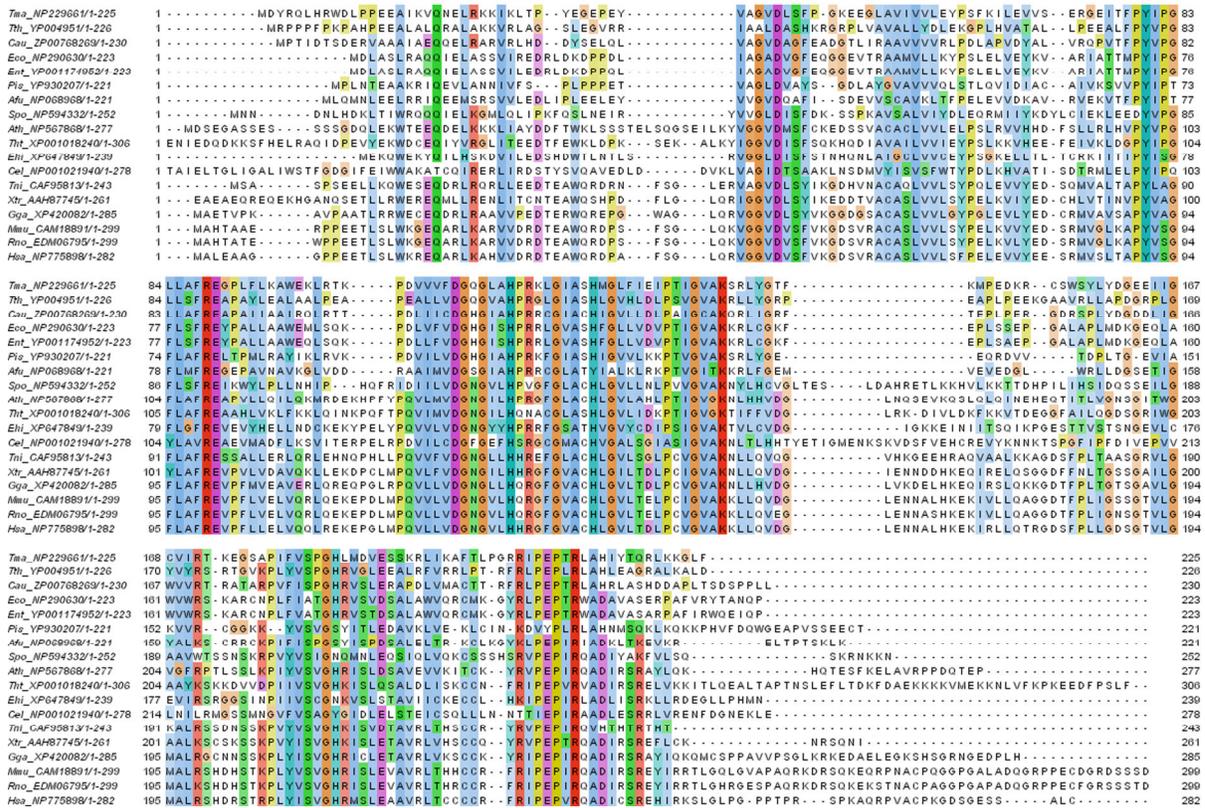
²*Institute of Medical Microbiology, Rikshospitalet University Hospital, Sognsvannsveien
20, N-0027 Oslo, Norway.*

³*Institute of Clinical Biochemistry, University of Oslo, N-0027 Oslo, Norway.*

⁴*Department of Molecular Biology and The Skaggs Institute for Chemical Biology, The
Scripps Research Institute, 10550 North Torrey Pine Road, MB4 La Jolla, CA 92037,
USA.*

⁵*Department of Genetics and Biochemistry, South Carolina Experiment Station, Clemson
University, 51 New Cherry Street, Clemson, SC 29634, USA.*

Correspondence should be addressed to M.B. (magnar.bjoras@rr-research.no) or J.A.T.
(jat@scripps.edu)

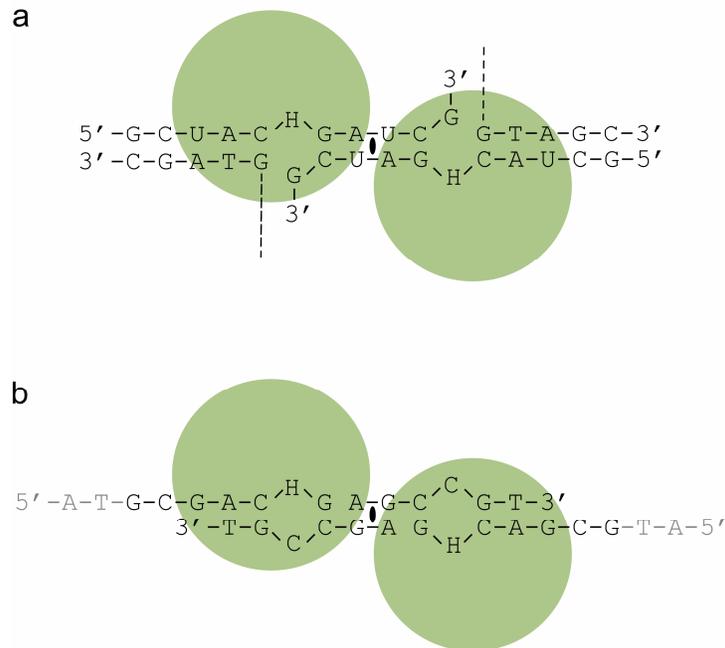


Supplementary Figure 1. Multiple sequence alignment of endonuclease V homologs from a selection of organisms spanning all three domains of life.

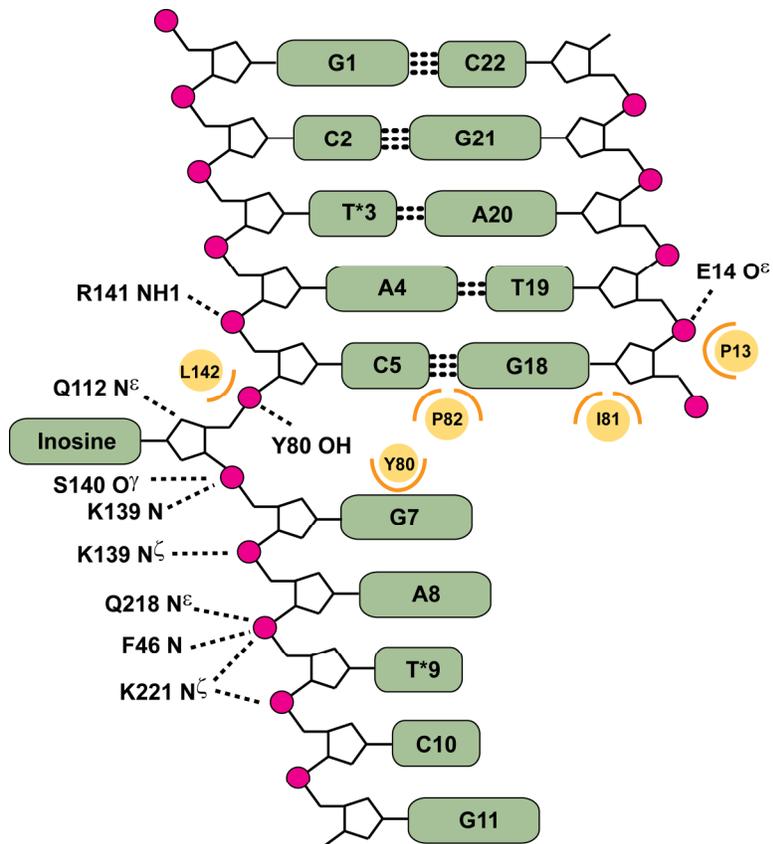
Eubacteria: Tma = *Thermotoga maritima*; Tth = *Thermus thermophilus*; Cau = *Chloroflexus aurantiacus*; Eco = *Escherichia coli*; Ent = *Enterobacter* sp.;

Archaea: Pis = *Pyrobaculum islandicum*; Afu = *Archaeoglobus fulgidus*;

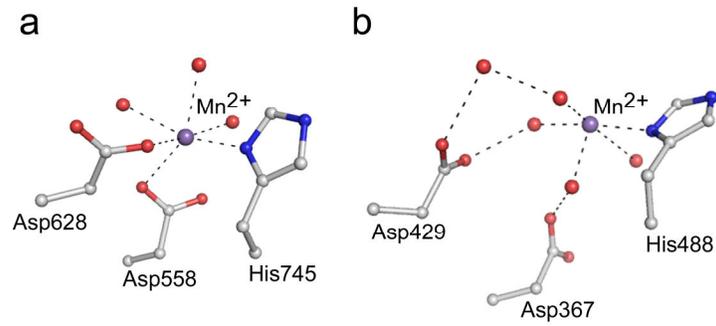
Eukaryotes: Spo = *Schizosaccharomyces pombe*; Ath = *Arabidopsis thaliana*; Tht = *Tetrahymena thermophila*; Ehi = *Entamoeba histolytica*; Cel = *Caenorhabditis elegans*; Tni = *Tetraodon nigroviridis*; Xtr = *Xenopus tropicalis*; Gga = *Gallus gallus*; Mmu = *Mus musculus*; Rno = *Rattus norvegicus*; Hsa = *Homo sapiens*. Only the C-terminal part of Cel and Tht and the N-terminal part of Mmu and Rno are shown in this diagram. Sequences were aligned by ClustaW and the figure was made with the JalView software.



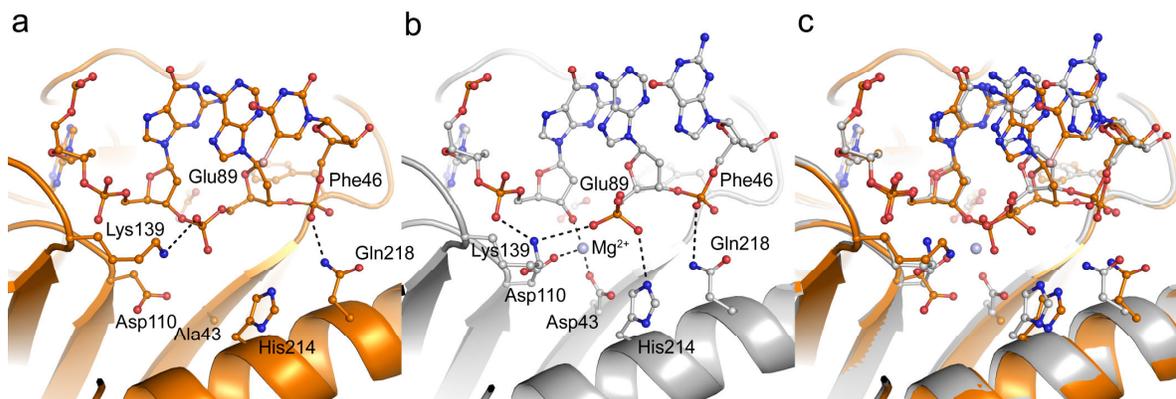
Supplementary Figure 2. Homo-dimerization in the two structures is mediated by partial DNA duplex formation of the 3' tail of the lesion containing strand. Green spheres represent EndoV. **(a)** Diagram showing the situation in the structure of D43A EndoV (lesion recognition complex). U denotes 5-bromo-2'-deoxyuridine, and H denotes hypoxanthine. Note that 5 nucleotides from the original complementary strand anneal with the lesion strand. The dashed line represents the disordered 5' tail of the complementary strand. **(b)** Diagram showing the situation in the structure of wild-type EndoV (product complex). Note that the longer 15-mer extends around the PYIP wedge and displaces the original complementary strand. Bases in grey are not included in the crystallographic model as they are not visible in the difference electron density maps.



Supplementary Figure 3. Diagram showing all protein-DNA contacts in the D43A mutant damage recognition complex. Hydrogen bond/ionic interactions are shown with dashed lines. Symbol N denotes protein main chain amide nitrogen. Steric interactions involving amino acid side chains (symbolized by yellow circles) are shown as orange arcs. T* denotes 5-bromouracil used for phasing.



Supplementary Figure 4. Active site metal coordination in related structures. The positions of the co-crystallized Mn²⁺ ions are shown as violet spheres. **(a)** *P. furiosus* Argonaute. **(b)** C-terminal domain of *T. maritima* UvrC.



Supplementary Figure 5. Comparison of protein fold, side-chain conformation and DNA conformation close to the active site in the Lesion Recognition Complex (LRC, gold) and Product Complex (PC, silver) of *Tma* EndoV. Dashed lines show direct protein-DNA or protein/DNA-metal interactions. **(a)** Structure of the D43A EndoV LRC structure with continuous DNA. **(b)** Structure of the wild-type EndoV PC structure with nicked DNA. **(c)** Overlay of the LRC and PC structures. Superposition of all atoms in the displayed amino acids gives an rmsd of 0.55Å for 50 atoms. Superposition of C^α-atoms only for the same six residues gives an rmsd of 0.38Å.