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}

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

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

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

⁴Department 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.

⁵Department 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*)

Tma_NP229661/1-225	1 · · · · · · · · · MDYRQLHRWDLPPEEAIKV <mark>D</mark> NEL <mark>R</mark> KKIKLTP · · YE <mark>GEP</mark> EY · · · · · · · · · · VA <mark>G</mark> VDLSFP · GKEEGLAVIVVLEYPSFKILEVVS · · ERGEITF	PYIPG 83
Tth YP004951/1-226	1 · · · · · · · · MRPPPF PKPAHPEEALAL GRALAKKVRLAG · · SLEGVRR · · · · · · · · IAAL DASHKRGRPL VAVALLYDLEKGPL HVATAL · · LPEEAL	PYVPG 83
Gau ZP00768269/1-230	2 1 · · · · · · · · · · · · · · · · · ·	PYVPG 82
Eco NP290630/1-223		APYIPG 76
Ent VP001174952/1-222		APVIPG 78
Pie VP930207/1-221		PYIPT 73
A4: ND00000004 224		DVIDT 77
AJU_NF0683687-227		
Spo_NP594332/1-252	1	DAALAG 82
Ath_NP567868/1-277	1 · · MDSEGASSES · · · · SSSGDDLEKWTEEDDELKKKLTAYDDFTWKLSSSTELSDGSETLKYVGGVDMSFCKEDSSVACACLVVLELPSLRVVHD.FSLLRLH	PYVPG 103
Tht_XP001018240/1-306	6 1 ENIEDQDKKSFHELRAQID <mark>P</mark> EVYEKWDCE <mark>D</mark> IYVRGLITEEDTFEWKLDPK+++SEK++ALKYIGGVDISFSQKHQDIAVAILVVIEYPSLKKVHEE+FEIVKLDC	PYIPG 104
Ehi_XP647849/1-239	1RVGGLDISF STNHQNLAIGCLVVCEYPSGKELLTL.TCRKIII	PYISG 78
Cel_NP001021940/1-278	8 I TAIELTGLIGALIWSTF <mark>GDGIFEIWW</mark> AKATCQI <mark>R</mark> ERLIRDSTYSVQAVEDLD·····DVKLVA <mark>GIDTS</mark> AAKLNSDMVYISVSFWTYPDLKHVATI·SDTRMLEI	. PYIPQ 103
Tni_CAF95813/1-243	1 · · · · · · MSA · · · · · SPSEELLKQWESEQDRLRQRLLEEDTEAWQRDRN · · · FSG · · · LERVAGVDLSYIKODHVNACAQLVVLSYPQLEVVYED · SQMVALTA	APYLAG 90
Xtr_AAH87745/1-261	1 · · · EAEAEQREQEKHGANQSETLLRWEREQMLLRENLITCNTEAWQSHPD · · · FLG · · · LQRIGGVDLSYIKEDDTVACASLVVLSYPELKVIYED · CHLVTIN	/ P Y V A G 100
Gga_XP420082/1-285	1 · · · MAETVPK · · · · · AVPAATLRRWECE <mark>O</mark> DRL <mark>R</mark> AAVVPEDTERWQRE <mark>PG</mark> · · · WAG · · · LQR <mark>VGGVDLS</mark> YVK <mark>GODO</mark> SACASLVVLGYPGLEVLYED · CRWVAVSA	PYVAG 94
Mmu_CAM18891/1-299	1 · · · MAHTAAE · · · · · · RPPEETLSLWKGEQARLKARVVDRDTEAWQRDPS · · · FSG · · · LQKVGGVDVSFVKGDSVRACASLVVLSVPELKVVYED · SRMVGLKA	APYVSG 94
Rno_EDM06795/1-299	1 · · · MAHTATE · · · · · · WPPEETLSLWKGEDARLKARVVDRD TEAWORDPS · · · FSG · · · LOKVGGVDVSFVKGDSVRACASLVVLSYPELKVLYED · SRMVGLKA	APYVSG 94
Hsa NP775898/1-282	1 · · · MALEAAG · · · · · · · · · · · · · · · · · ·	APYVSG 94
-		
Tma NP229661/1-225		
7th VP004954/4 226		GPPIG 180
Car 700700200/4 220		DDL 10 188
Call_2P00768269/1-230		
EC0_NF 250630 1-223		CEQLA 100
Ent_1P001114952/1-223	3 7/ FLSFRETPALLAAWEULSUKPDLLFVD0H0TSHPRKL0VASHF0LLVDVFTT6VARKKL00KFEPLSAEPGALAFLMDF	GEGLA 100
Pis_YP930207/1-221	74 FLAFRELTPMLRAYIKLRVKPOVILVDGHGIAHPRKFGIASHIGVVLKKPTVGVAKSRLYGEEQRDVVTDPLTC	- EVIA 151
Afu_NP068968/1-221	78 FLMFREGEPAVNAVKGLVDDRAAIMVDGSGIAHPRCGLATYIALKLRKPTVGITKKRLFGEMVEVEDGLWRLLDG	SETIG 158
Spo_NP594332/1-252	86 FLSFREIKWYLPLLNHIPHQFRIDIILVDGNGVLHPVGFGLACHLGVLLNLPVVGVAKNYLHCVGLTESLDAHRETLKKHVLKKTTDHPILIHSIDQS	3 S E I L G 188
Ath_NP567868/1-277	104 FLAFREAPVLLOILOKMRDEKHPFYPOVLMVDGNGILHPRGFGLACHLGVLAHLPTIGVGKNLHHVDGLNOSEVKOSLOLOINEHEOTITLVGNS	3G I TWG 203
Tht_XP001018240/1-306	6 105 FLAF <mark>RE</mark> AAHLVKLFKKLQINKPQFT <mark>PQVIMVDGNGILHQNACGLASHLGVLIDKPTIGVGK</mark> TIFFVDG · · · · · · · · LRK · DIVLDKF <u>KKV</u> TDEG <u>G</u> FAILQ <u>G</u> DS	3 G R I WG 203
Ehi_XP647849/1-239	79 FLGFREVEVYHELLNDCKEKYPELYPQVVLVDGNGYYHPRRFGSATHVGVYCDIPSIGVAKTVLCVDGIGKKEINIITSQIKPGESTTVSTSI	IGEVLC 176
Cel_NP001021940/1-278	8 104 YLAV <mark>REAEVMADFLKSVITERPELRPDVILCDGFGEFHSRGCGMACHVG</mark> ALSGIA <mark>SIGVAK</mark> NLTLHHTYETIGMENKSKVDSFVEHCREVYKNNKTS <mark>PGFIP</mark> FDI	I VE <mark>PV</mark> V 213
Tni_CAF95813/1-243	91 FLAFRESSALLERLORLEHNOPHLLPOVVFVDGNGLLHHREFOLACHLGVLSGLPCVOVARNLLOVOGVHKGEEHROVAALKKAGDSFPLTAAS	3 G R V L G 190
Xtr_AAH87745/1-261	101 YLAFREVPVLVDAVQKLLEKDPCLMPQVLFVDGNGILHHRGFOVACHLGILTDLPCIOVAKNLLQVDGIENNDDHKEQIRELQSGGDFFNLTGS	3 G A I L G 200
Gga_XP420082/1-285	95 FLAFREVPFMVEAVERLQREQPGLRPQVLLVDGNGLLHQRGFGVACHLGVLTDLPCVGVAKNLLHVDG······LVKDELHKEQIRSLQKKGDTFPLTGT	3 <mark>6 A V L 6</mark> 194
Mmu_CAM18891/1-299	95 F L A F REVPF L VELVQ R L Q E K E P D L M PQ V V L V DG NG V L H Q R G F O V A C H L G V L T E L P C I O V A K K L L Q V D G · · · · · · · L E N N A L H K E K I V L L Q A G G D T F P L I G S S	3 G T V L G 194
Rno_EDM06795/1-299	95 FLAF REVPFLVELVQRLQEKEPDLMPQVLLVDGNGVLHQRGFGVACHLGVLTELPCVGVAKKLLQVEGLENNASHKEKIVLLQAGGDTFPLIGNS	3 G T VL G 194
Hsa NP775898/1-282	95 FLAF REVPFLLELVQQLREKEPGLMPQVLLVDGNGVLHHRGFGVACHLGVLTDLPCVGVAKKLLQVDGLENNALHKEKIRLLQTRGDSFPLLGDS	3 G T VL G 194
-		
Tma NP229661/1-225	168 CVIRT-KEGSAPIFVSPGHLMDVESSKRLIKAFTLPGRRIPEPTRLAHIYTORLKKGLF	225
Tth YP004951/1-226	170 YVYRS-RTGVKPLYVSPGHRVGLEEALRFVRRLPT-RFRLPEPLELAHLEAGRALKALD	226
Gau ZP00768269/1-230	2 187 WVVRT-RATARPVFISPGHRVSLERAPDLVMACTT-RFRLPEPTRLAHRLASHDDAPLTSDSPPLL	230
Eco NP290630/1-223	181 WWWRS, KARCNRIFIATGHRVSVDSALAWVORCHK, GYRIPEPTRWADAVASERPAFVRYTANOP.	223
Ent VP001174952/1-223		223
Pia VP920207/4-224	10 KUVD CAAVY VEVEN TEDAVY VE VICIN KNYVDIDIANNAGY VOVVDIDAADVSSEECT	221
A4. ND00000011 221		221
See NP504222/4 252		252
Sp0_NF 554 532 1-252		202
Atti_NP567868/1-277		2//
Int_XP001018240/1-306	204 AAYKSKROVOPTTVSVGHKTSLUSALDLISKCKN-FRIPEPVRVADIRSRELVKKTTLUEALTAPTNSLEFLIDKFDAEKKRKVMEKKNLVFKPKEEDFPSL	300
Em_XP647849/1-239	1// EVIRSROOSINPTIVSCONKVSLSTAVIICKECCL. HKIPEPINLADLISKKLLRDEGLLPHMN.	239
Cel_NP001021940/1-278	8 214 LNILRMGSSMNGVFVSAGYGIDLELSTEICSGLLLN.NTTTEPTRAADLESRRLVRENFDGNEKLE.	278
Ini_GAF95813/1-243	191 KALRSSDNSSKYVEVBVSVGHKISVDTAVRLEHSCCR. YRVPEPIROVHTHTRTHI	243
Xtr_AAH87745/1-261	201 AALKSCSKSSKPVYISVGHKISLETAVRLVHSCCQ. · YRVPEPTHQADIRSREFLCK. · · · · · · · · · · · · · · · · · · ·	261
Gga_XP420082/1-285	195 MALROCINNSSKPLYISVOHRICLETAVRLVKSCCK. YRIPEPIROADIRSRAYIQKQMCSPPAVVPSGLKRKEDAELEGKSHSGRNGEDPLH	285
Mmu_CAM18891/1-299	195 MALRSHDHSTKPLYVSVGHRISLEVAVRLTHHCCR··FRIPEPIROADIRSREVIRRTLGQLGVAPAQRKDRSQKEQRPNACPQGGPGALADQGRPPECDGRDS	SD 299
Rno_EDM06795/1-299	195 MALKSHDHSTKPLYVSVGHRISLEVAVRLTHRCCR··FRIPEPIROADIRSREVIRRTLGHRGESPAQRKDRSQKEKSTNACPAGGPGAPADQGRPPECFGRSS	STD 299
Hsa_NP775898/1-282	195 MALRSHDRSTRPLYISVGHRMSLEAAVRLTCCCCCC・・FRIPEPVR0ADICSREHIRKSLGLPG・PPTPR・・・SPKAQRPVACPKGDSGESS・・・・ALC・・・・・	

Supplementary Figure 1. Multiple sequence alignment of endonucleaseV homologs from a selection of organisms spanning all three domains of life. Eubacteria: Tma = Thermotoga maritima, Tth = Thermus thermophilius; Cau = Chloroflexus aurantiacus; Eco = Escherichia coli; Ent = Enterobacter sp.; Archea: 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 norwegicus; Has = 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 ClustalW 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 rmds of 0.55Å for 50 atoms. Superposition of C^{α}-atoms only for the same six residues gives an rmsd of 0.38Å.