Biochemical reconstitution and genetic characterization of the major oxidative damage base excision DNA repair pathway in *Thermococcus kodakarensis*.

```
Alexandra M. Gehring<sup>1</sup>, Kelly M. Zatopek<sup>1</sup>, Brett W. Burkhart<sup>2</sup>, Vladimir Potapov<sup>1</sup>, Thomas J. Santangelo<sup>2</sup>,
Andrew F. Gardner<sup>1*</sup>
```

SUPPLEMENTARY DATA

¹New England Biolabs, Inc., Ipswich, MA 01938 ²Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, CO 80523



Figure S1: PacBio sequencing confirms the deletion of TK0940 which encodes AGOG. PacBio sequencing reads for libraries generated from *T. kodakarensis* Δ AGOG mapped to the reference genome using SMRT Portal. The reference sequences used were Δ AGOG (top) or TS559 (bottom). The x-axis indicates genomic position and the y-axis indicates the fold coverage at each position.

S1



Figure S2: Trapping of TkoAGOG Schiff Base intermediate. *A*, Formation of the borohydratetrapped TkoAGOG and DNA complex. *B*, Coomassie stained SDS PAGE showing formation of a Schiff base intermediate. Lane 1: 80x0G DNA and sodium cyanoborohydride, Lane 2: 80x0G DNA and TkoAGOG Lane 3: 80x0G DNA, TkoAGOG, and sodium cyanoborohydride, Lane 4: TkoAGOG and sodium cyanoborohydride. Molecular weight standards in kDa are identified in the rightmost lane.



Figure S3: TkoAGOG and TkoEndoIV activity on various DNA Lesion. TkoAGOG has no activity on dsDNA containing dU, dI, dX or 80x0A. TkoEndoIV has no activity on dsDNA containing dU, dI or dX. Both TkoAGOG and TkoEndoIV have activity on AP site containing DNA. A visualization of 60-bp dsDNA substrates containing a centralized DNA lesion incubated with either TkoAGOG or TkoEndoIV for 1 hour and then analyzed with capillary electrophoresis.

	1	10	20		30	4	0	5	0	6	0	7	0 I	80		90		100
Eco EndolV	M-MYIGAH	VSAAG	-GLANAA	AIRAAEII	DATAFA	LFTKNQ	WRA.	APLTTO	TID	FKAACE	KYHY	TSAQUL	PBD S	YLINLG	PVTEA	LEKSR	DAF	DEMOR
Mth EndolV	WVLFL-IF	VGPAG	-PVGYRG	GSUVN	VFSKIR	AMGLDA	TEYQA	TYGLRI	KKEN	AIRIGE	NSRK	NDILVS	MEGP	YMINLS	SARKET	MEKSI	ERUF	DCAVA
Pae EndolV	MANVY	LIGPAG	PQFVV	AKSULD	AVKVVR	ELGLNA		VQGVRM	ISREI	AKQVG	AAQD	YQVKUS	VEAP	YEINLC	SEEADK	VEKSR	QRIV	DSLDE
Pfu EndolV	MFKIDRLF	FIGTAG	PLSTP	- PSUIA	GIERVR	ELGLDA		RGVN	RPEI	AKKIKY	VAKK		ALAP	YMINLN	KEKEK	VESSK	RRIII	QSAER
Sis EndolV	MVKII Y	LIGPAG	PHSSK		GIRTVK	ELGLNA		VQGVRM	ISRE	ALETGO	VAKE		VEAP	YEINLC	EEKDK	TEASK	QRIL	DTADE
Teu EndolV	FEVDRLF	FIGTAG	PLSAP	-RSIID	GIIHVR	NLGLDA		RGVN	KPEI	AKKIKY	VAKK	HDVLUT	ALAP	YMINLN	SEKAK	VEASK	RRIII	QSAER
Tko EndolV	FEVDRLF	FIGTAG	PLSTP	- PSUIT	GIEHVR	NLGLDA		RGVN	SPEI	AKKIKY	VAKK	NDVLUT	ABAP	YMINLN		VEASK	KR	QSAER
Tth EndolV	MPRYG-FH	LSIAG	KGVAGA	EEAMAL	GUTAFQ	IFAKSPI	SWRP	RALSPA	EVE	FRAURE	ASGG	LPAVE-	-BAS	YLVNLG	-EGEL	WEKSV	ASLA	DDLEK
		110	120)	130	1.	40	14	0	16	50	17	0	18)			200
Eco EndolV	CEOURISI		SEL MOLS			NIALDK.	0 G	TA	LEN	TAGOGSN	ILMER	FUHLAA	UIDG	VENKSR		TCHAF	AGY	DI -RT
Mth EndolV	AFINGANG		ENGEH-0	SPSGALGI	CIKA	DELLERI	HGAG	DEET	ADE	TGKDS			- 1 KI	SELEDEDE	MDIT		ARGG	CIPD
Pae EndolV	AVVNCAWA		MMGKI - (PERCVE		FKAVKE	GLGS	GWVI	GVE	TARTNO		F B	AFRI	GKEIISE		WGELL	ARINNI	active
Pfu EndolV	LVEARGWS	WAVERAN		PEKVYO	TEST	KDIEREI	KDRIG			TGKPTO		KB		SONEEL			MONK	RECNIS
Sis EndolV	AFLMCADA		ENGRM-9	PECVO		AEVIDE	KEMG		GVE	TMAKET/			VISI	SKELKG		MALTE	AROG	GE
Tou EndolV	LHOARGWS			DAKVVNI		KEIEKEI	MDPIG		DDEI	TCKDT		KB	INKI	SFELLENS			MDINK	RECNIS
Tko EndolV	LVEARCWO			DESVVOI		KBLOKE	MDKG			TCKPT			VKI	CEPTICM			MOND	REVENS
Tth EndolV	AALURVEY			DEDVKE		DIAGVD				LAGGER	VEAD		- 414/1	VANTDI		TCHAN	MAGY	DV
TUTLINUOIV	AALIENVLI	210	10 700		220	RLAGVR.	JKF	20	O	20	O	27	O.	200		200	AGI	200
F F L B (210	220	•	230	+	+	23	0	20	50	21	0	28	,	290		298
Eco EndolV	PALECEKTE	ADFAR	TVICEFKY LEF	RGMELND	AKSTIEI-	-GSRVD	HSL	GEG-N	GHD	AFRW	DDRF	DGIPL	LETI	NPDIWA	IAW	KAQQT	EKAV	A
Mth Endolv	ASSMRGI	ERESE	REGSHHE	BCEF	FG∎E¥I-	TDAGER	HSL	SEGYGE	PIEF	PLIEMEN	DGGW	DA-TI	SET-	-PM-KD	DARKI	KGIIK	EYLE	R
Pae EndolV	- INMGEI	DLWARE	FGNAHM	BTHF	TSVR Y R	NGKFVD	EEP	ERN-ME	PFEF	AREL	(NRD	TI-TL	CES-	- PL - LE	DALL	KEILE	QVGV	NLA
Pfu EndolV	EEEWREM	ALENE			SGIEY-	TEKGEK		EIES-DI	KWE		EFKV	KG-VV	SES -		BDALL	KKKWE	ELKI	
Sis EndolV	- IDYGKI	DRLIKE	I GL THI	NSHFI	ESLVSR	KGKYVD	IP	DAN-AF	PFEF	AKELL	. KRD	SI-TL	CES -	- RE - LEI	RDALK	KEVLE	RLGY	RLE
Teu EndolV	AEEWREM	SFMEDF	REGREALD	DNMETHV	SGIEY-	TDKGEK	RHLNL	QIE S - DA	INWE	LGVL	EFRV	RG-VV	SES -	-PN-IE	BDALLL	KKKYE	EIKV	
Tko EndolV	VEEWREM	SFLEDF	REGREALD		SG∎E Y -	TDKGEK	RHLPL	QIES-DA	INWE		EFRV	KG-VV	SES -	-PN-IIE	DALLIN	KKKYE	EIK-	A
Tth EndolV	AFDPLGV	DALDRA	VIELERVI	VVELND	SVGGL -	-GSRVDI	H A H	0G+	LGEO	KRVEL	DPRI	KDRVF	1 B T -	- PRGPE	DAWN	RVIRA	WEFE	A

Figure S4: Amino acid sequence alignment of characterized EndolV enzymes. The nine metalbinding active site residues are indicated by arrows. The species abbreviation and GenBank ascension numbers used are as follows: Bacteria: Eco, *Escherichia coli* (NP_416664.1); Tth, *Thermus thermophilus* (AAS80830.1); Archaea: Tko, *Thermococcus kodakarensis* (BAD84359); Teu, *Thermococcus eurythermalis* (WP_050002723.1); Pfu, *Pyrococcus furiosus* (AAL80382.1); Mth, *Methanothermobacter thermautotrophicus* (AAB85506.1); Pae, *Pyrobaculum aerophilum* (AAL64792.1); Sis, *Sulfolobus islandicus* (ADX86710.1).



Figure S5. TkoEndolV does not stimulate TkoAGOG activity. An 80xoG-containing substrate was incubated with AGOG (5 nM) alone (black filled circles), with BSA (5 nM) (open circles) or TkoEndolV (10 nM) (grey filled circles) at 65°C and quenched with formamide and EDTA at various times (as described in the Material and Methods section). The rate of AGOG cleavage (5.3 min⁻¹) was not significantly effected by addition of TkoEndolV (6.3 min⁻¹) or BSA (6.2 min⁻¹). As controls, TkoEndolV (10 nM) (open squares) or BSA (5 nM) were incubated with 80xoG-containing substrate and showed no activity.



Figure S6: DNA substrates were pre-incubated with TkoAGOG and TkoEndolV leaving a 3'-OH with a FAM (blue) label. DNA replication proteins were then added as shown on the left and incubated at 65°C for 15 minutes. The repair was monitored by the appearance of the 60-nt dual-labeled FAM/ROX product (repaired) and the 60-nt FAM only labeled product (strand displaced).