SUPPLEMENTARY INFORMATION

TO

Structural basis for enzymatic excision of N^1-methyladenine and N^3-methylcytosine from DNA

Ingar Leiros[1,5], Marivi P. Nabong[2,3,5], Kristin Grøsvik[3], Jeanette Ringvoll[2], Gyri T. Haugland[4], Lene Uldal[2], Karen Reite[2], Inger K. Olsbu[3], Ingeborg Knævelsrud[3,4], Elin Moe[1], Ole A. Andersen[1], Nils-Kåre Birkeland[4], Peter Ruoff[3], Arne Klungland[2] and Svein Bjelland[3,*]

[1]The Norwegian Structural Biology Centre, University of Tromsø, N-9037 Tromsø, Norway. [2]Centre for Molecular Biology and Neuroscience and Institute of Medical Microbiology, Rikshospitalet-Radiumhospitalet HF, University of Oslo, N-0027 Oslo, Norway. [3]Faculty of Science and Technology, Department of Mathematics and Natural Sciences, University of Stavanger, N-4036 Stavanger, Norway. [4]Department of Biology, University of Bergen, PO Box 7800, N-5020 Bergen, Norway. [5]These authors contributed equally to this work.

*Corresponding author. Faculty of Science and Technology, Department of Mathematics and Natural Sciences, University of Stavanger, N-4036 Stavanger, Norway. Tel.: + 47 51831884; Fax: + 47 51831750; E-mail: svein.bjelland@uis.no

Derivation of kinetic equations for *Af*AlkA

The reaction catalyzed by *Af*AlkA (E) can be summarized as

$$\mathrm{DNA} \rightarrow \mathrm{P} \qquad \text{(R1)}$$

where P represents products (*i.e.* DNA and free base lesion). Reaction R1 is a first-order process with reaction rate

$$v = d[\mathrm{P}]/dt = k\,[\mathrm{DNA}] \qquad (1)$$

Equation 1 is readily solved and the increase of product concentration [P] as a function of time t is given by the equation

$$[\mathrm{P}] = [\mathrm{DNA}]_{tot}(1 - e^{-k \cdot t}) \qquad (2)$$

where $[\mathrm{DNA}]_{tot}$ denotes the total (initial) DNA substrate concentration and k is the (first-order) rate constant.

The experimental results can be interpreted by considering a non-specific glycosylase binding or adsorption to DNA and looking specifically at the site where catalysis (excision) is taking place, *i.e.*, the specific position where m^1A, εA or m^3C are inserted (Figure 1A). If DNA_i is the location i on the DNA where catalysis occurs (when E binds to this site) we can write

$$\mathrm{E} + \mathrm{DNA}_i \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{E} \bullet \mathrm{DNA}_i \xrightarrow{k_2} \mathrm{E} + \mathrm{DNA} + \text{free base lesion} \qquad \text{(R2)}$$

Considering this process, we can write the following mass balance at DNA_i:

$$[\mathrm{DNA}]_{tot} = [\mathrm{DNA}_i] + [\mathrm{E}\cdot\mathrm{DNA}_i] \qquad (3)$$

where $[\mathrm{DNA}]_{tot}$ is the total DNA concentration, $E \cdot DNA_i$ denotes any DNA molecule where E occupies the catalytic site i, and DNA_i is any DNA molecule where E has not bound to the excision site i. Calculating $d[E \cdot DNA_i]/dt$ and assuming that $E \cdot DNA_i$ is in a steady state (leading to $d[E \cdot DNA_i]/dt = 0$), it results in

$$d[\mathrm{E}\cdot\mathrm{DNA}_i]/dt = k_1[\mathrm{E}][\mathrm{DNA}_i] - (k_{-1} + k_2)[\mathrm{E}\cdot\mathrm{DNA}_i] = 0 \qquad (4)$$

By inserting $[DNA_i]$ from Equation 3 into Equation 4, the steady state value of $[E \cdot DNA_i]$, $[E \cdot DNA_i]_{ss}$, can be expressed as

$$[\mathrm{E}\cdot\mathrm{DNA}_i]_{ss} = \frac{[\mathrm{E}][\mathrm{DNA}]_{tot}}{K_D + [\mathrm{E}]} \quad (5)$$

with

$$K_D = \frac{k_{-1} + k_2}{k_1}$$

Because the concentration of E is in excess of the DNA template (single-turnover conditions; $[\mathrm{E}]_{tot} >> [\mathrm{DNA}]_{tot}$), [E] in Equation 5 can be replaced by $[\mathrm{E}]_{tot}$:

$$[\mathrm{E}\cdot\mathrm{DNA}_i]_{ss} = \frac{[\mathrm{E}]_{tot}[\mathrm{DNA}]_{tot}}{K_D + [\mathrm{E}]_{tot}} \quad (6)$$

Since the reaction rate is expressed by $\nu = k_2[\mathrm{E}\cdot\mathrm{DNA}_i]_{ss}$, ν can be written as

$$\nu = k_2 \frac{[\mathrm{E}]_{tot}[\mathrm{DNA}]_{tot}}{K_D + [\mathrm{E}]_{tot}} \quad (7)$$

showing that the rate constant k of Equation 1 is specified by the following relationship:

$$k = \frac{k_2[\mathrm{E}]_{tot}}{K_D + [\mathrm{E}]_{tot}} \quad (8)$$

An analogous expression to Equation 7 can be derived by assuming a rapid equilibrium between E, DNA_i and $\mathrm{E}\cdot\mathrm{DNA}_i$. In this case K_D is represented by k_{-1}/k_1. In the case reaction R2 is irreversible ($k_{-1} = 0$), $K_D = k_2/k_1$.

Due to these situations it is difficult to make a precise interpretation about the physical nature of K_D based on k_2 and K_D values alone. The observation that for $\mathrm{m}^1\mathrm{A}$, k_2 is decreased while K_D is increased for most of the mutant proteins (Table I), may suggest that the binding between the DNA and the enzyme has become less effective with a somewhat slower turnover. For εA, both k_2 and K_D are decreased for the mutant compared to the wild-type proteins (Table I). This may indicate the presence of a steady state with still a relatively strong binding between DNA and the enzyme, but with a less effective turnover.

Supplementary Table I

Data collection, structure solution and refinement statistics

Dataset	Peak	Native
Data collection statistics		
Beamline	ID14-4	BL14.1
Wavelength (Å)	1.009	0.919
Resolution range (Å)	20.0–1.80(1.90–1.80)	50.0-1.90(1.95-1.90)
R_{sym} (%)	8.4(49.9)	6.7(43.4)
Multiplicity	8.0(7.6)	2.9(2.9)
Mean $I/\sigma I$	18.0(3.5)	11.4(2.5)
Completeness (%)	98.7(97.9)	99.9(100.0)
Anom. completeness (%)	98.5(97.5)	-
Space group	$P2_1$	$P2_1$
Unit cell parameters:		
a(Å)	69.5	69.9
b(Å)	50.0	48.85
c(Å)	105.8	104.5
β(°)	107.4	106.1
Phasing statistics		
No. Heavy atoms/a.s.u.	2 Hg ∗	-
R_{cullis} **§**	0.675	-
Phasing power **§**	1.656	-
FOM SHARP **§**	0.430	-
FOM DM **†**	0.792	-
Refinement statistics		
No. Atoms ♣	5493/4900/495/4/4/4/4	5425/4826/571/0/4/4/0
B-factors ♣	20/19/29/33♠/27♠/18/33♠	21/19/31/0/34/20/0
R_{free} (%)	22.3	23.9
R_{work} (%)	18.5	18.1
Geometrical deviation		
Bonds (Å)	0.015	0.015
Angles (°)	1.360	1.366
ESU (Å) **‡**	0.078	0.100

∗: Two additional Mercury atoms were identified at the refinement stage through phased anomalous fourier maps, but were not used at the preceding phasing stage. **§**: Figure-of-merit (FOM) from SHARP-phasing for acentric data to 2.5 Å. **†**: FOM from DM after extending phases to 1.8 Å. ♣: Total/Protein/Water/Mercury/Glycerol molecules/Sodium/MES molecules. ♠: Modeled with reduced occupancy. **‡**: Estimated overall coordinate error from REFMAC5 based on maximum likelihood.

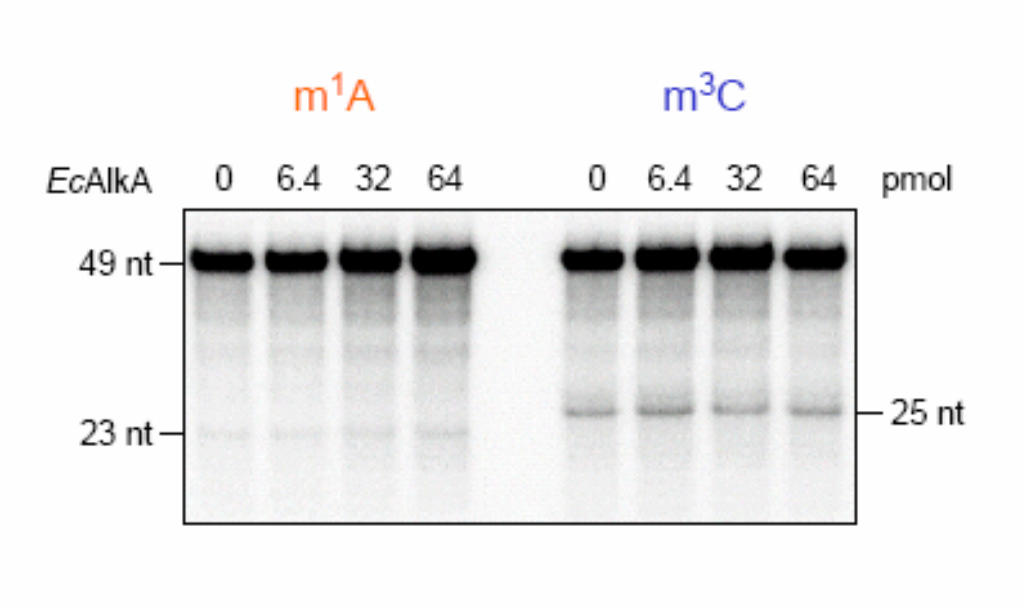
Supplementary Table II

Structural alignment of domains of *Af*AlkA and *Ec*AlkA

	*Ec*AlkA (#aligned/total)	rmsd (Å)	Domain
*Af*AlkA	(84/111)	1.55	N-term
	(33/48)	1.14	Central
	(114/130)	1.38	C-term
	(159/289)	1.58	All residues

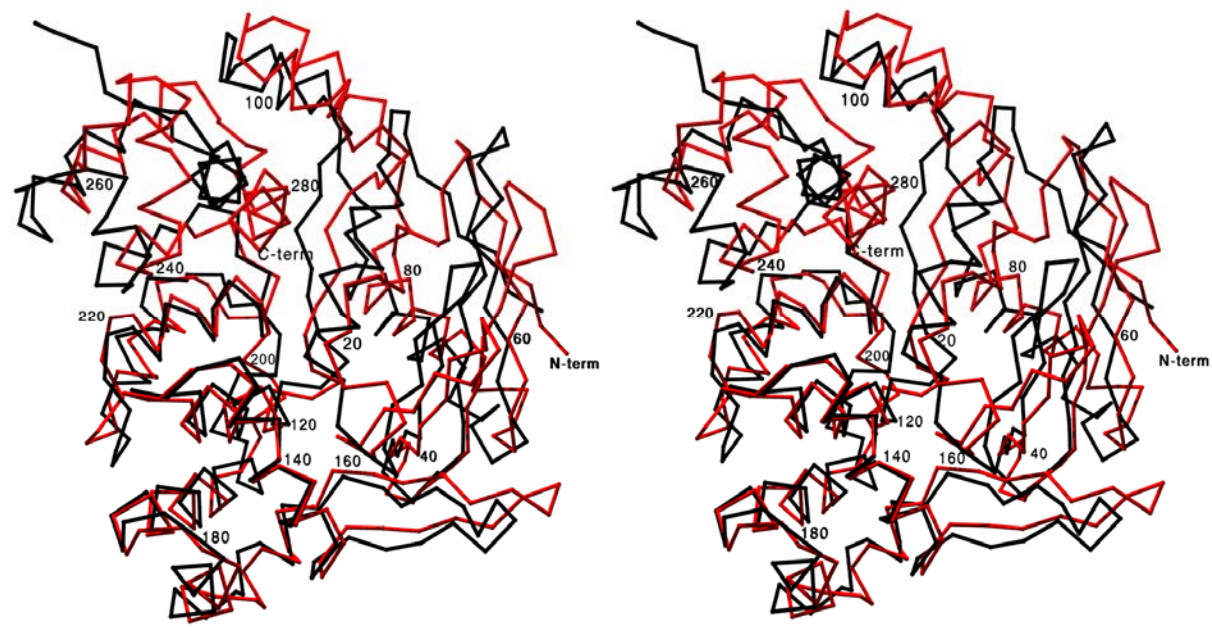
The table illustrates structural conservation between individual domains of *Af*AlkA and *Ec*AlkA, where the numbers in parenthesis are number of residues aligned out of the total number of residues in that domain and the last number lists the root mean square deviation (rmsd) between the aligned residues in *Af*AlkA and *Ec*AlkA.

Supplementary Figure 1



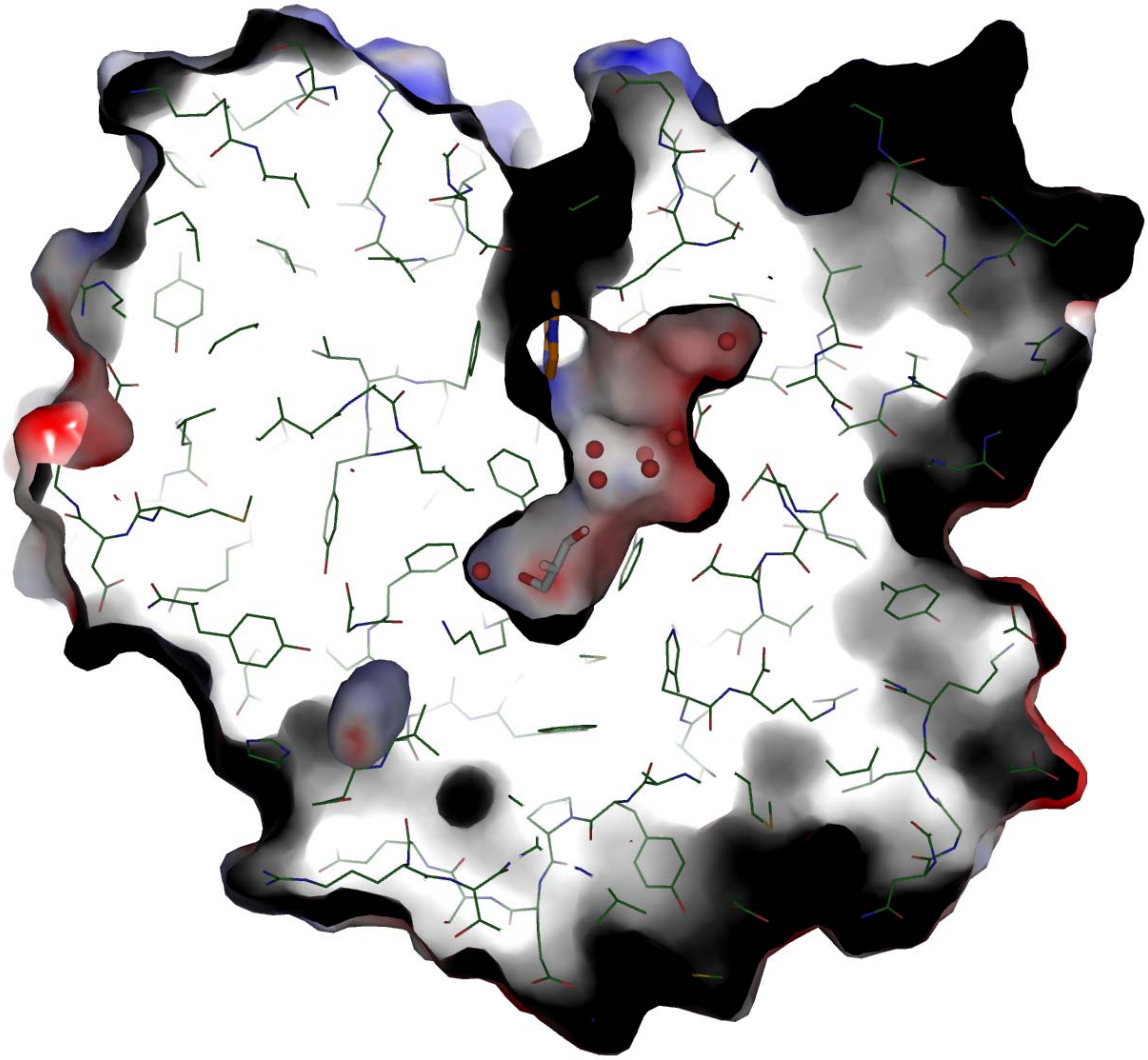
Legend to Supplementary Figure 1 Exposure of *Ec*AlkA to m^1A- and m^3C-containing DNA. Assay for cleavage of 5´[^{32}P]labeled 49-nt DNA into repair product (23 and 25 nt) is described in Figure1A,B. *Ec*AlkA was incubated with DNA (10 fmol) in 15 mM Hepes [4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid]-KOH, pH 7.5, 0.3 mM EDTA, 1.5 mM 2-mercaptoethanol at 37°C for 30 min (The same result was obtained when incubation was carried out at 37°C under exactly the same conditions as employed for *Af*AlkA).

Supplementary Figure 2



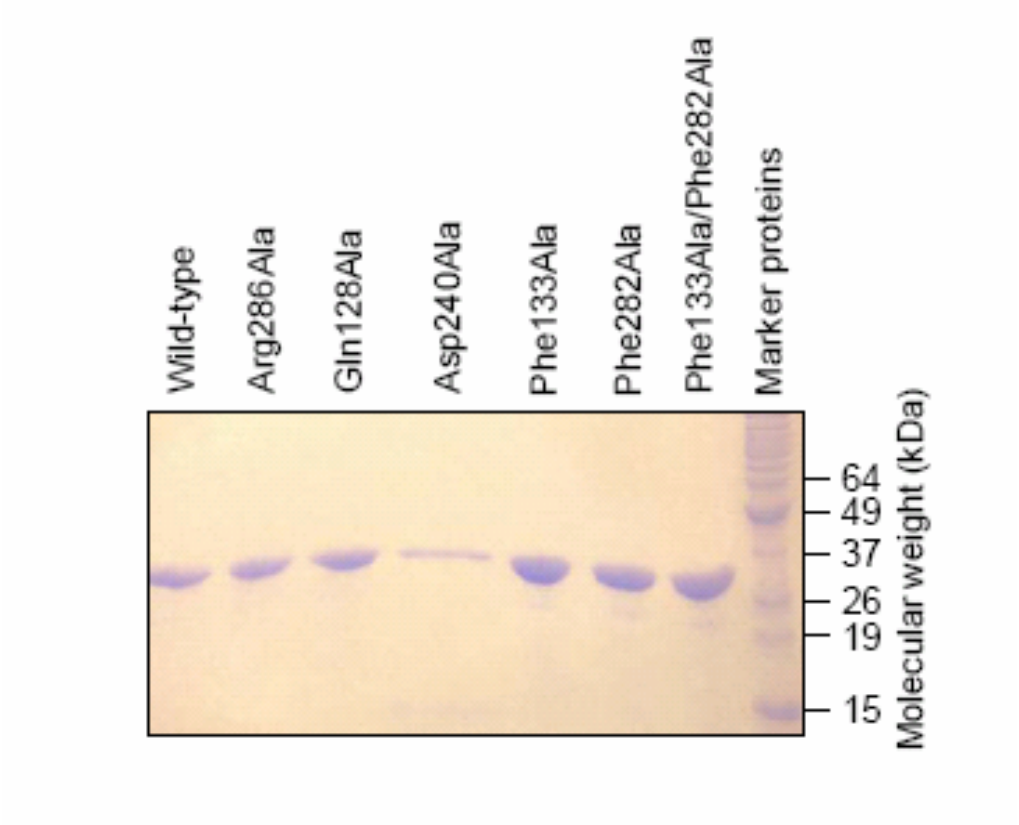
Legend to Supplementary Figure 2 Stereo representation showing the superpositioning of the Cα-traces of *Af*AlkA (red) and *Ec*AlkA (black). *Af*AlkA residue numbers are labeled.

Supplementary Figure 3



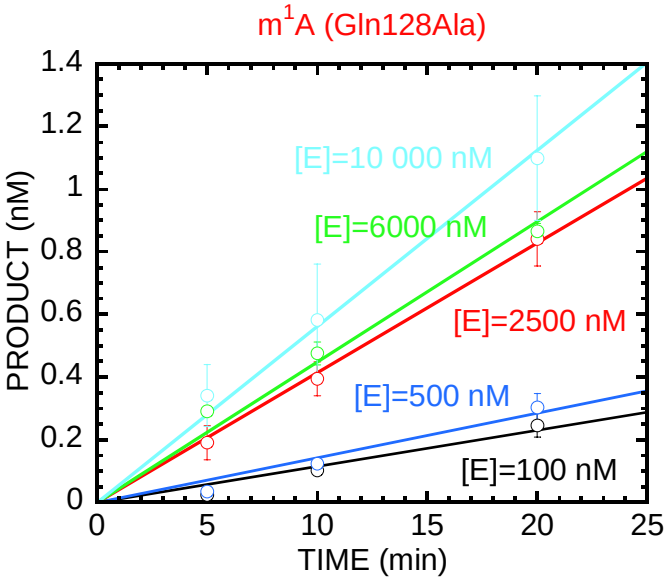
Legend to Supplementary Figure 3 Sliced view of the molecular surface of *Af*AlkA enabling visualization of the water-accessible channel extending from the base of the substrate-binding pocket into the core of the protein. The modeled εA moiety is shown with orange carbon atoms, while protein carbon atoms are shown in dark green. Other atoms are shown in atom colors (oxygen, red; nitrogen blue; sulfur, yellow). The cavity contains a total of one glycerol and 13 water molecules. The electrostatic surface potential is calculated as for Figure 4.

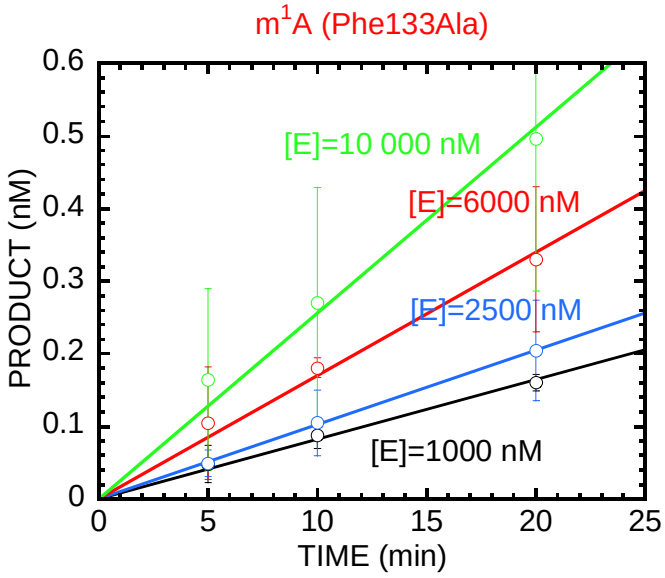
Supplementary Figure 4

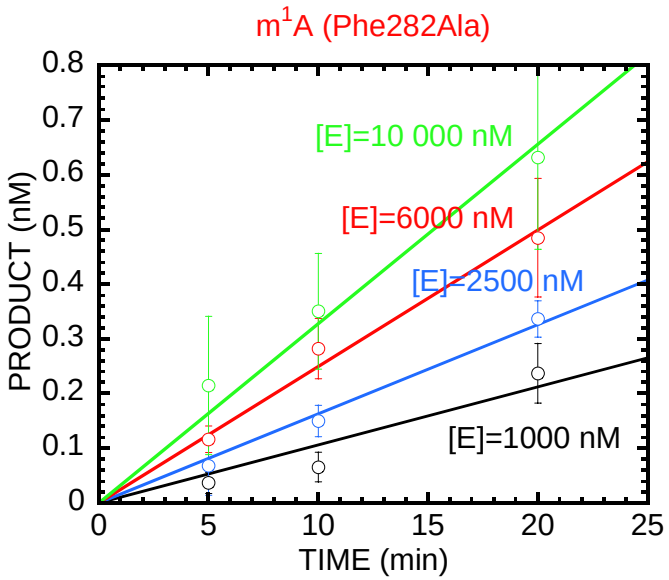


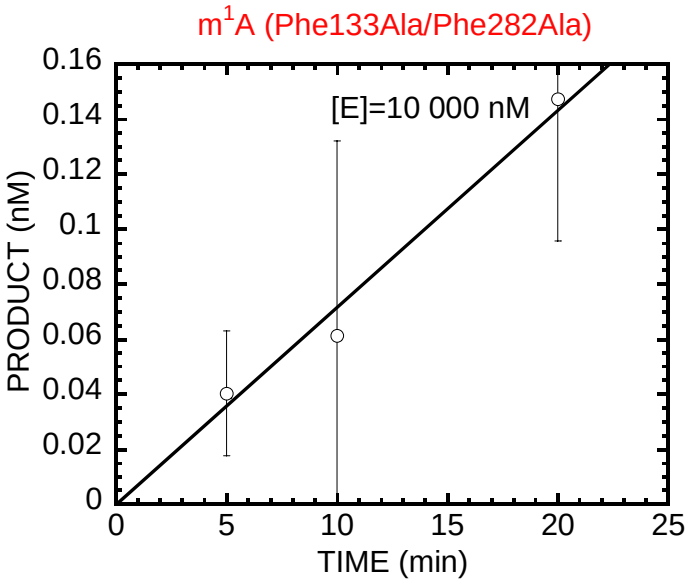
Legend to Supplementary Figure 4 SDS–PAGE (12 %) of purified *Af*AlkA mutant and wild type proteins stained with Coomassie Blue. Samples (2 μg; 10 μl) and protein markers (10 μl; BenchMark™ Pre-Stained Protein Ladder, Cat. No. 10748-010, Invitrogen) were loaded onto the gel (Bio-Rad, Ready Gel® Tris–HCl Gels, 12 % Resolving Gel, Cat. # 161-1156) and run for 2 h at 120 V.

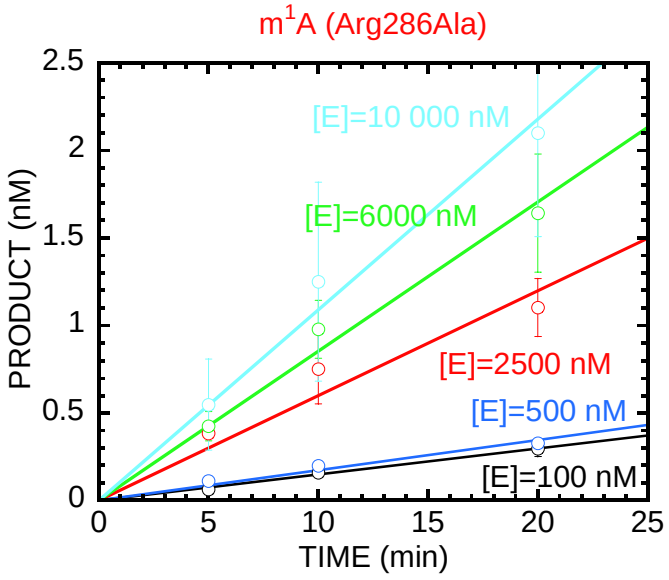
Supplementary Figure 5





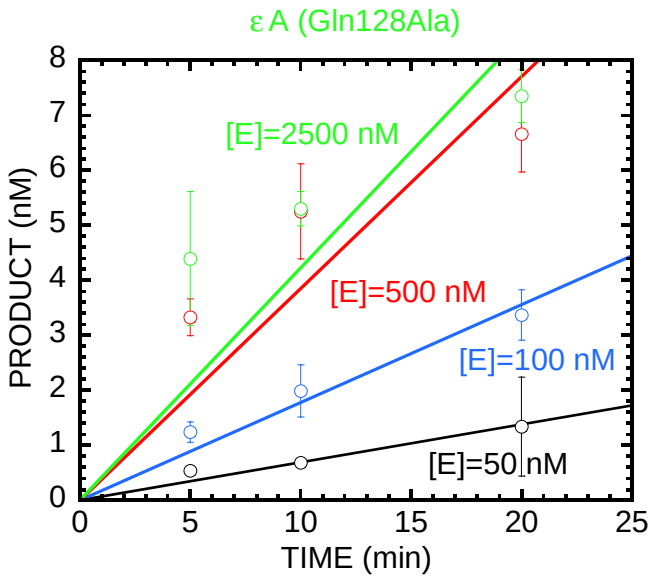


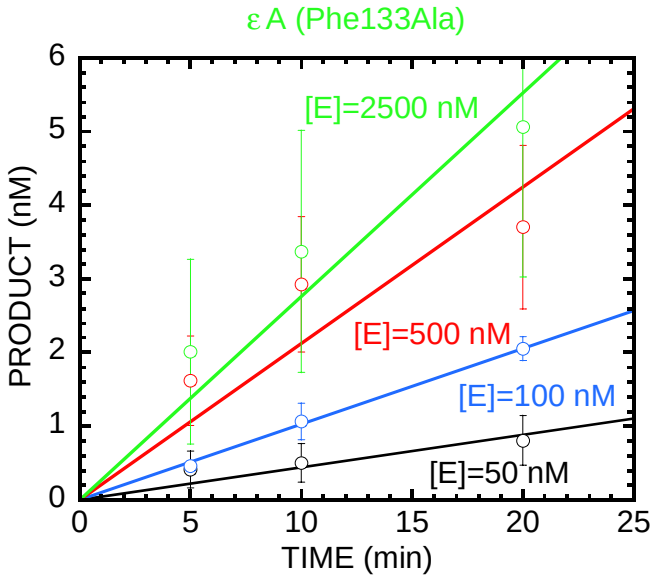


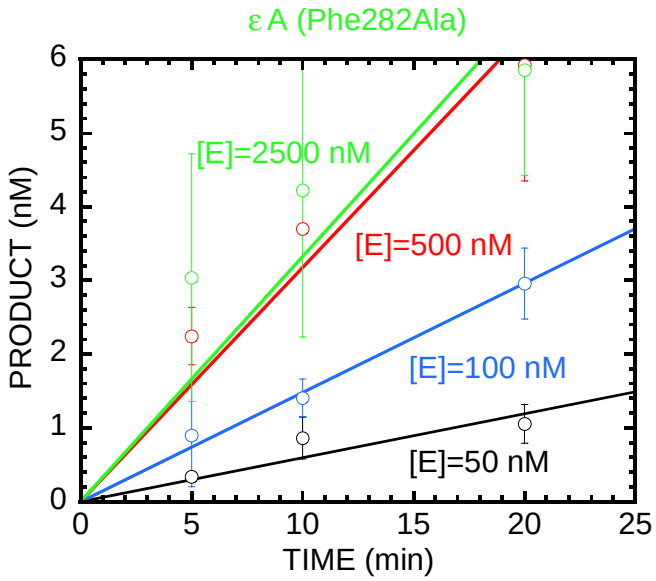


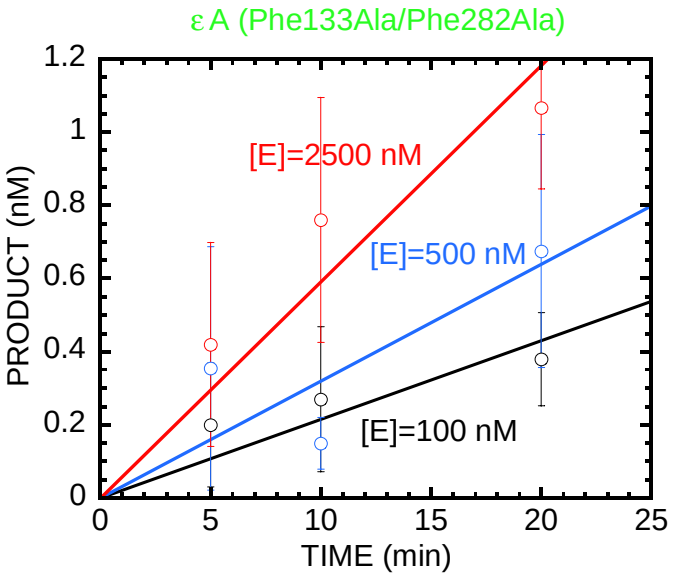
Legend to Supplementary Figure 5 Single-turnover kinetics for excision of m^1A, where 10 nM of substrate was incubated with different concentrations of *Af*AlkA at 70°C for increasing time periods. Each value represents the average of three independent measurements.

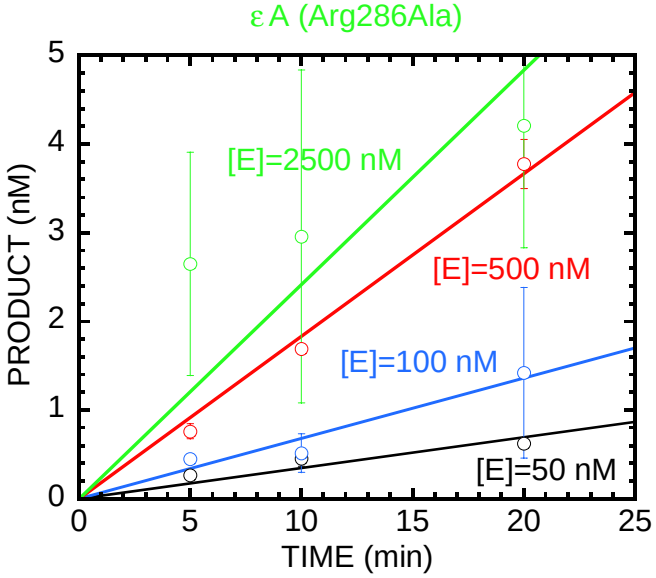
Supplementary Figure 6











Legend to Supplementary Figure 6 Single-turnover kinetics for excision of εA, where 10 nM of substrate was incubated with different concentrations of *Af*AlkA at 70°C for increasing time periods. Each value represents the average of 2–3 independent measurements.