

## **Supporting Information**

### **Analysis of substrate specificity of *Schizosaccharomyces pombe* Mag1 alkylpurine DNA glycosylase**

Suraj Adhikary and Brandt F. Eichman

Tables S1-S4

Extended Methods

Figures S1-S8

References

**Table S1.** Crystallographic data collection and refinement statistics

	<u>Native</u>	<u>SeMet</u>
<b>Data collection</b>		
Wavelength (Å)	0.9787	0.9792
Space group	P2 <sub>1</sub>	P2 <sub>1</sub>
Cell dimensions		
<i>a, b, c</i> (Å)	54.3, 112.3, 61.1	54.1, 111.3, 61.2
$\alpha, \beta, \gamma$ (°)	90, 99.6, 90	90, 99.1, 90
Resolution (Å)	29.10-2.28 (2.38-2.28)	50.00-2.80 (2.90-2.80)
<i>R</i> <sub>sym</sub>	0.109 (0.404)	0.103 (0.326)
<i>I</i> / $\sigma$ <sub>1</sub>	11.50 (2.50)	12.60 (3.20)
Completeness (%)	99.6 (98.1)	98.9 (93.0)
Redundancy	4.8 (3.4)	3.8 (3.0)
<b>Refinement</b>		
Resolution	2.28	
No. reflections	32,188	
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.185 / 0.225	
No. of atoms		
Protein	3184	
DNA	845	
Solvent	114	
B-factors (Å <sup>2</sup> )		
Protein	38.3	
DNA	56.2	
Solvent	40.4	
R.m.s. deviations		
Bond lengths (Å)	0.008	
Bond angles (°)	1.185	

Values in parentheses refer highest resolution shell.

<b>Table S2.</b> Sequence and structural similarity between Mag orthologs						
		<b>Identity (%)</b>	<b>Similarity (%)</b>	<b>RMSD (Å)</b>	<b>Q-score</b>	<b>PDB ID</b>
spMag1	bhMag	26.5	65.2	1.78	0.56	2H56
spMag1	scMag	22.2	48.5			
spMag1	ecAlkA	17.4	46.1	1.81	0.37	1MPG
scMag	bhMag	22.0	49.3			
scMag	ecAlkA	19.9	41.7			
bhMag	ecAlkA	17.0	40.7			

Sequence identity and overall similarity were obtained by CLUSTALW from the PBIL server (<http://npsa-pbil.ibcp.fr>). RMSD and Q-scores for C $\alpha$  atoms only were calculated using the PDBeFold server at EMBL-EBI (<http://www.ebi.ac.uk>). Q-score (Quality of Alignment) is defined by  $Q = (N_{\text{align}})^2 / [(1 + (\text{RMSD}/3\text{\AA})^2) * N_{\text{res1}} * N_{\text{res2}}]$ , where  $N_{\text{align}}$  is the number of aligned residues and  $N_{\text{res}}$  is the number of total residues ([http://www.ebi.ac.uk/msd-srv/ssm/r1\\_qscore.html](http://www.ebi.ac.uk/msd-srv/ssm/r1_qscore.html)). Abbreviations: sp, *Schizosaccharomyces pombe*; sc, *Saccharomyces cerevisiae*; bh, *Bacillus halodurans*; ec, *Escherichia coli*.

<b>Table S3.</b> Base excision activity of wild-type and mutant Mag orthologs				
	$\epsilon$ A		7mG	
	$k_{\text{cat}}$ ( $\times 10^{-5} \text{ sec}^{-1}$ )	Relative activity	$k_{\text{cat}}$ ( $\times 10^{-3} \text{ sec}^{-1}$ )	Relative activity
<b>spMag1</b>				
WT	5.3 ± 0.6	1.0	2.6 ± 0.5	1.0
Q62A	0.6 ± 0.05	0.1	0.4 ± 0.09	0.2
L63A	0.01 ± 0.002	0.003	0.02 ± 0.003	0.01
H64S	21.1 ± 3.2	4.0	2.1 ± 0.4	0.8
FS158SG	5.7 ± 1.1	1.1	3.8 ± 0.6	1.5
D170N	1.1 ± 0.2	0.2	0.09 ± 0.02	0.04
S172G	0.7 ± 0.2	0.1		
<b>scMag</b>				
WT	12.8 ± 1.7	1.0	3.6 ± 0.7	1.0
S97H	2.0 ± 0.3	0.2	1.8 ± 0.4	0.5
SG197FS	14.2 ± 2.1	1.1	3.5 ± 0.7	1.0
D209N	0.1 ± 0.01	0.005	0.01 ± 0.002	0.003
G211S	8.3 ± 0.9	0.6		
<b>bhMag</b>				
WT	9.3 ± 0.4	1.0	3.4 ± 0.7	1.0
S53H	4.5 ± 0.7	0.5	1.7 ± 0.3	0.5
Single-turnover rate constants ( $k_{\text{cat}}$ ) for excision of $\epsilon$ A and 7mG opposite cytosine from a 25mer oligonucleotide were measured under saturating enzyme concentrations at pH 6.0 ( $\epsilon$ A) or 7.5 (7mG), 150 mM ionic strength, and 25°C. Values represent the average from three independent measurements ± standard deviations. Non-enzymatic rate constants for spontaneous depurination ( $k_{\text{non}}$ ) under the same conditions were $7.2 \times 10^{-8} \text{ sec}^{-1}$ ( $\epsilon$ A) and $1.6 \times 10^{-6} \text{ sec}^{-1}$ (7mG).				

<b>Table S4. DNA binding by spMag1</b>			
	<b>K<sub>d</sub> (μM)</b>		<b>DNA sequences</b>
	<u>Wild-type</u>	<u>H64S D170N</u>	
THF	0.8 ± 0.1	0.8 ± 0.04	5' TGACTACTACATG <b>X</b> TTGCCTACCAT 3' *ACTGATGATGTACCAACGGATGGTA
εA	0.6 ± 0.1	0.8 ± 0.1	5' TGACTACTACATG <b>X</b> TTGCCTACCAT 3' *ACTGATGATGTACCAACGGATGGTA
Gua	1.3 ± 0.3		5' TGACTACTACATG <b>G</b> TTGCCTACCAT 3' *ACTGATGATGTACCAACGGATGGTA
Gua, 5'-oh	0.9 ± 0.1		5' <b>T</b> TGACTACTACATG <b>G</b> TTGCCTACCA 3' *ACTGATGATGTACCAACGGATGGTA

Dissociation constants (K<sub>d</sub>) were determined by fluorescence anisotropy changes upon adding protein to 6-carboxyfluorescein(\*)-DNA as described in the Supplementary Methods. Values shown are averages ± standard deviations from three independent measurements.

## Methods

### Protein purification

The spMag1, scMag, and bhMag genes were PCR amplified from *S. pombe*, *S. cerevisiae*, and *B. halodurans* (ATCC BAA-125D-5) genomic DNA and cloned into expression vector pBG100 (Vanderbilt University Center for Structural Biology) that produces an N-terminal His<sub>6</sub>-tagged protein. Wild-type and mutant yeast proteins were overexpressed in *E. coli* C41 (spMag1) or BL21 (scMag) cells for 4 hr at 25° C in LB media. BhMag was overexpressed in *E. coli* HMS174 (wild-type) or BL21 (S53H mutant) for 16 hr at 16° C. Cells were lysed in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, and 10% glycerol and proteins isolated using Ni-NTA (Qiagen) affinity chromatography. Following cleavage of the His<sub>6</sub> tag, proteins were purified by heparin and gel filtration chromatography in 20mM Tris-HCl (pH 7.5), 150 mM NaCl, 2mM DTT and 0.05 mM EDTA. Mutant protein constructs were generated using a Quik-Change kit (Stratagene), purified the same as wild-type, and their structural integrity verified using circular dichroism spectroscopy (Fig. S6). SeMet-spMag1 was overexpressed in C41 cells for 16 hr at 16° C in minimal media supplemented with 70 mg/L selenomethionine (Acros Organics) under conditions that suppress normal methionine biosynthesis (Van Duyne et al, 1993) and was purified the same as wild-type, except that 5 mM methionine and 5 mM DTT were added to all purification buffers after the Ni-NTA step.

### X-ray crystallography

Protein-DNA complexes were assembled by incubating 0.20 mM SpMag1 with 0.24 mM DNA (d(TGTCCA(THF)GTCT)/d(AAGACTTGGAC) at 4° C for 20 min. Crystals were obtained by vapor diffusion at 21°C against reservoir solution containing 100 mM MES (pH 6.5), 20% (w/v) PEG 8K and 2.4% (v/v) glycerol and were flash frozen in mother liquor containing 15% (v/v) glycerol prior to data collection. X-ray diffraction data (Table I) were collected at the Advanced Photon Source (Argonne, IL) beamlines 21-ID (native) and 22-BM (SeMet) and processed using HKL 2000 (Otwinowski & Minor, 1997). SAD data were collected at the selenium absorption peak. Positions of 10 Se atoms were identified and phases calculated using the program SHARP (Vonrhein et al, 2007). A crystallographic model corresponding to

amino acids 16-221 and nucleotides 1-22 for each of 2 protein/DNA complexes in the asymmetric unit was built into 2.8 Å Se-SAD electron density maps using Coot (Emsley & Cowtan, 2004). The two crystallographically distinct protein/DNA complexes were virtually identical with an r.m.s. deviation of 0.33 Å for main chain atoms.

The crystallographic model was refined against native diffraction data extending to 2.2 Å using SAD phase combination and a maximum likelihood target as implemented in PHENIX (Adams et al, 2002). Improvements to the model were made by manual inspection of  $\sigma_A$ -weighted  $2mF_o - F_c$  and  $mF_o - DF_c$  electron density maps. Translation/libration/screw-rotation (TLS) refinement was used to model anisotropic motion of each protein/DNA complex, and individual anisotropic B-factors derived from the refined TLS parameters were held fixed during subsequent rounds of refinement. The final model was validated using PROCHECK (Laskowski et al, 1996) and deposited in the Protein Data Bank under accession number 3S6I.

### **Enzymatic activity**

Base excision activities were measured by following the alkaline cleavage of the abasic DNA product of alkylbase excision from a 25-mer oligonucleotide duplex containing a centrally positioned  $\epsilon A \bullet C$  or  $7mG \bullet C$  base pair. Oligonucleotides were purchased from Integrated DNA Technologies (IDT, USA), and 7mG-DNA was prepared enzymatically as previously described (Asaeda et al, 2000; Rubinson et al, 2008). Nucleotide sequences used were 5'-<sup>32</sup>P-d(GACCA CTACA CCXTT TCCTA ACAAC) annealed to 5'-d(GTTGT TAGGA AACGG TG TAG TGGTC). Reaction mixtures (75  $\mu$ L) contained 10  $\mu$ M enzyme, 100 nM radiolabeled DNA duplex, 100 mM KCl, 2 mM DTT, and either 50mM sodium acetate (pH 6.0) for  $\epsilon A$ -DNA and HEPES (pH 7.5) for 7mG-DNA. We verified that under these conditions, the enzyme concentration is saturating. Reactions were initiated by addition of enzyme and incubated at 25 °C. As a control to verify that the reaction rates were not affected by changes in the protein as a result of the long reaction times, we performed control reactions in which the protein was pre-incubated for 4 hours under the reaction conditions prior to initiating the enzymatic reaction (Fig. S8). Aliquots (8  $\mu$ L) were stopped at various time points by addition of 0.2 M NaOH and heated at 70 °C for 2 min. The cleaved 12-mer product and unreacted 25-mer substrate oligonucleotides were separated by 15% polyacrylamide/7M urea gel electrophoresis and quantified by

autoradiography. Fraction product ( $F_P$ ) at each time point was calculated by  $F_P = I_P/(I_S+I_P)$ , where  $I_S$  and  $I_P$  are the integrated intensities of substrate and product bands. Rate constants ( $k_{cat}$ ) were determined from single-exponential fits to the data from three separate experiments. Rate differences between wild-type and mutant proteins were judged to be significant based on  $p$ -values derived paired t-test analysis.

## **DNA binding**

DNA binding was measured by the change in fluorescence anisotropy as spMag1 was added to oligonucleotide duplexes containing a centrally located THF or Gua in one strand and a 6-carboxyfluorescein on the 3'-end of the other (Table S2). Increasing concentrations of protein (0-30  $\mu$ M) were added to a 50 nM DNA in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM DTT, and 0.01 mM EDTA. Polarized fluorescence intensities using excitation and emission wavelengths of 485 and 538 were measured at 25° C using a SpectraMax M5 microplate reader (Molecular Devices). Dissociation constants were derived by fitting the data using the equation,  $A = A_{max}[\text{protein}]/(K_d+[\text{protein}])$ , in which  $A$  is the anisotropy value at a given protein concentration and  $A_{max}$  is the anisotropy value at maximal binding.



## Figure Legends

**Fig. S1.** SpMag1-DNA structural data. A cross-section of the final protein-DNA model is shown superimposed onto experimental (solvent-flattened) SAD (A) and refined 2Fo-Fc (B) electron density contoured at  $1.5\sigma$ . Protein and DNA carbon atoms are colored cyan and gold, respectively.

**Fig. S2.** Helix-hairpin-helix mediated protein-DNA contact in spMag1. (A) Contacts between the HhH motif (helices  $\alpha$ I- $\alpha$ J) and the DNA (yellow) backbone are mediated by the hairpin loop (blue sticks), and are largely conserved among AlkA, EndoIII and hOgg1 protein-DNA structures (Bruner et al, 2000; Hollis et al, 2000; Kuo et al, 1992). The hairpin turn ligates a metal ion, which we have modeled as a sodium ion (purple sphere) on the basis of octahedral coordination and bonds lengths, shown as dashed lines and distances in Angstroms. The sodium ion is coordinated by main-chain carbonyl oxygen atoms of Thr142, Ile144 and Ile147, the O1P phosphate oxygen of Cyt10, and two water molecules (Fig. S7). In addition to metal ion mediated contacts, the Cyt10 phosphate is anchored to the protein through a hydrogen bond from main-chain Gly146 nitrogen. The adjacent phosphate group from Thy9, immediately 3' to the THF abasic site, is hydrogen bonded to Gly148 nitrogen and the side-chain  $O_{\gamma}$  oxygen from Thr151. (B) Structure based sequence alignment between Mag proteins with AlkA. Hairpin atoms are enclosed by a dashed box, and base binding pocket and catalytic aspartate residues are highlighted orange and yellow, respectively.

**Fig. S3.** Comparison of Mag orthologs. (A) Structure-based sequence alignment of 3-methyladenine DNA glycosylases from *Schizosaccharomyces pombe* (spMag1), *Bacillus halodurans* (bhMag), *Saccharomyces cerevisiae* (scMag), and *Escherichia coli* (ecAlkA). The alignment of spMag1, bhMag and ecAlkA was generated by structural superposition using PDB ID 3S6I, 2H56, and 1DIZ. ScMag was added by multiple sequence alignment using CLUSTALW. Secondary structure of spMag1 (blue, cyan) and AlkA (residues 1-89, grey) are shown schematically above the alignment. Residues lining the extrahelical base binding pocket are highlighted orange, and conserved aspartic acid residues important for catalysis in HhH

glycosylases are highlighted yellow. DNA intercalating plug and wedge residues are highlighted green and light green, respectively, and the novel minor-groove interrogating histidine residue observed in spMag1 is highlighted pink. Residues contacting DNA are boxed. (B) Crystallographic models of ecAlkA (Hollis et al, 2000), spMag1 and bhMag (PDB, 2006) are shown in the same orientation. HhH motifs are shown in cyan and AlkA residues 1-89 are grey. DNA is shown as an orange backbone trace.

**Fig. S4.** Structural alignment of spMag1 with ecAlkA (A) and bhMag (B). In the superpositions at left, the structures are in the same orientation as in Fig. S2, with SpMag colored blue and AlkA/bhMag colored grey. AlkA residues 1-89 are colored gold. At the right are stereo views of the nucleobase binding pockets, in which spMag1 is blue with gold DNA and AlkA/bhMag are grey. Structures were taken from PDB ID 1DIZ (ecAlkA) and 2H56 (bhMag).

**Fig. S5.** Packing of spMag1-DNA crystals. (A) Two spMag1-DNA complexes in the asymmetric unit are colored blue and orange. Protein molecules are shown as van der Waals surfaces and DNA is shown as sticks. (B) A 90° rotation along the vertical axis of the view shown in A. (C) Close-up view of the base binding pocket of spMag1 from complex #1 (blue), with Thy1 nucleotide from complex #2 (orange) inserted into the cavity. (D) Stereo view of the Thy1 interactions in the active site. Protein and DNA residues from complex #1 are colored cyan and blue, and the DNA from complex #2 is orange. (E) The position of the invading DNA (orange) from the adjacent molecule in spMag1 is in the same location as the N-terminal  $\beta$ -sheet domain in AlkA (gold).

**Fig. S6.** Excision of  $\epsilon$ A and 7mG by Mag1 H64S and Mag S97H. (A) Denaturing polyacrylamide gels showing the disappearance of radiolabeled  $\epsilon$ A-containing 25-mer DNA substrate (S) and appearance of alkaline-cleaved 12-mer abasic-DNA product (P) as a function of time after addition of wild-type (WT) or mutant enzymes. For each enzyme, lanes correspond to 0, 30, 60, 180, 300, 480, 1440, 2880 min reaction times. (B) Same as (A) but using a 7mG-DNA substrate. For each enzyme, lanes correspond to 0, 2, 5, 15, 30, 60, 240, 1440 min reaction times.

(C,D) Quantification of the data shown in (A) and (B), respectively. Black curves correspond to spMag1 and blue curves correspond to scMag. Circles denote wild-type and squares denote mutant proteins. Non-enzymatic  $\epsilon$ A and 7mG depurination is shown as crosses and a dotted line curve fit.

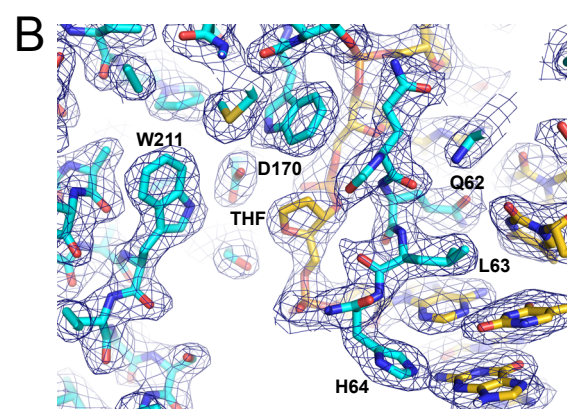
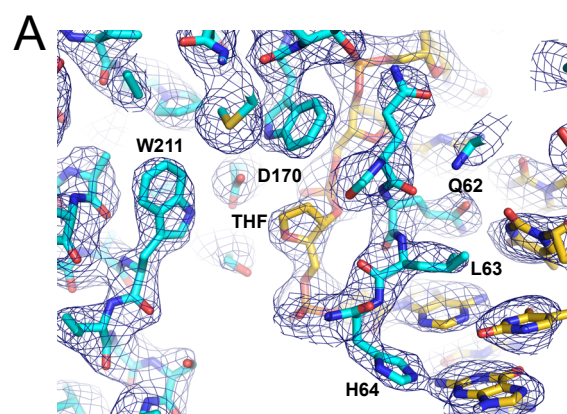
**Fig. S7.** Thermal denaturation of wild-type and mutant spMag1 (A) and scMag (B) proteins. Protein unfolding was monitored using CD spectroscopy, measuring the change in molar ellipticity at 222 nm as a function of temperature. (A,B) Molar ellipticity values were normalized to reflect the fraction of unfolded protein and plotted against temperature. The melting temperatures were determined by fitting the data using the equation  $f_u = [(A_f - A_u)/(1 + e^{-(T - T_m)/w})]$ , in which  $f_u$  is the fraction of unfolded protein,  $A_f$  and  $A_u$  are the CD values for folded and unfolded protein respectively,  $T_m$  is the melting temperature, and  $w$  is the cooperativity coefficient for the transition. For curves showing bi-phasic transitions, melting temperatures for each transition were determined by fits to the following equation:  $f_u = [(\Delta A_1)/(1 + e^{-(T - T_{m1})/w_1})] + [(\Delta A_2)/(1 + e^{-(T - T_{m2})/w_2})]$ , in which  $\Delta A$  refers to the change in CD signal for a particular transition.

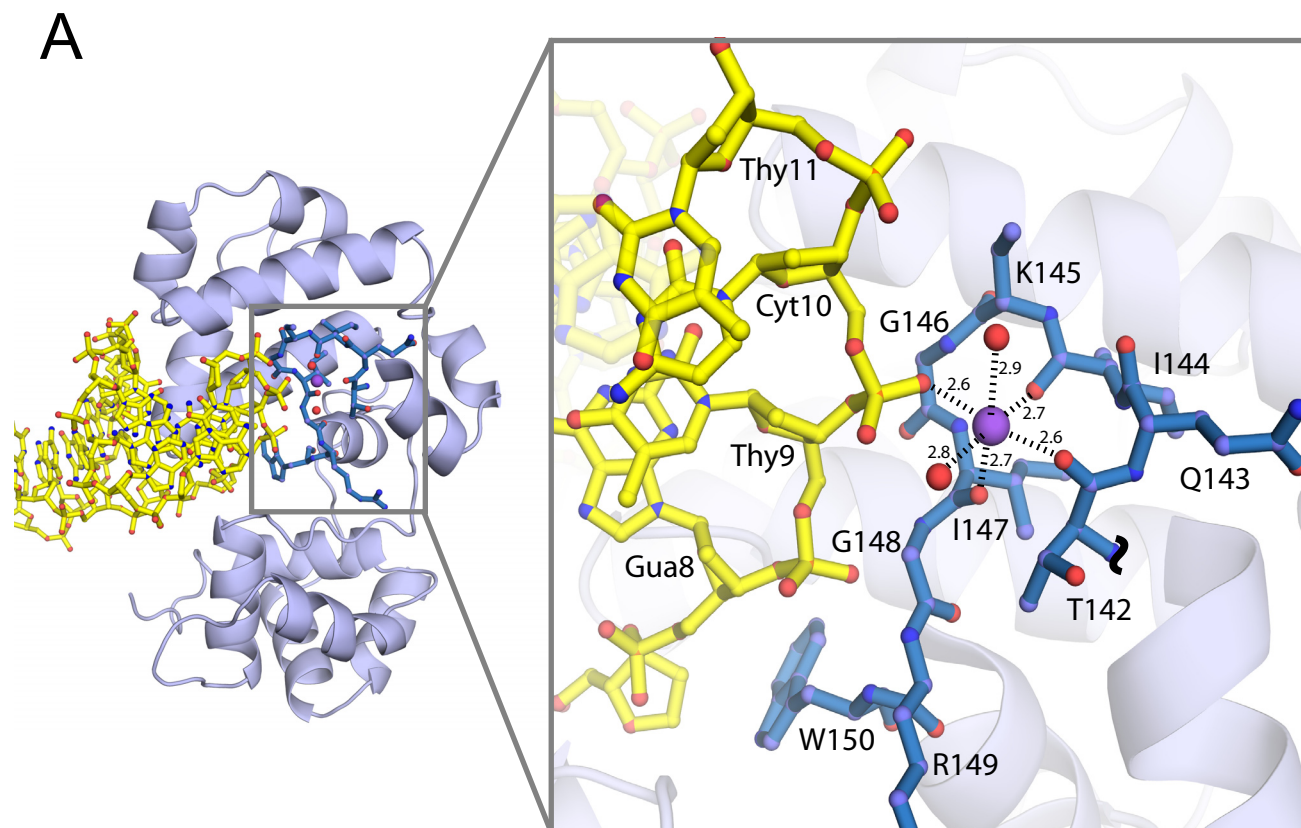
**Fig. S8.** Protein stability control for spMag1 7mG activity. Denaturing polyacrylamide gels showing the disappearance of radiolabeled 7mG-containing 25-mer DNA substrate and appearance of alkaline-cleaved 12-mer abasic-DNA product as a function of time after addition of spMag1 that had been either pre-incubated for 4h in the reaction buffer or not pre-incubated. The plot at the bottom shows the quantification of the data shown in the gel strips.

## References

- Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**: 1948-1954
- Asaeda A, Ide H, Asagoshi K, Matsuyama S, Tano K, Murakami A, Takamori Y, Kubo K (2000) Substrate specificity of human methylpurine DNA N-glycosylase. *Biochemistry* **39**: 1959-1965
- Bruner SD, Norman DP, Verdine GL (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**: 859-866
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126-2132
- Hollis T, Ichikawa Y, Ellenberger T (2000) DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, Escherichia coli AlkA. *Embo J* **19**: 758-766
- Kuo CF, McRee DE, Fisher CL, O'Handley SF, Cunningham RP, Tainer JA (1992) Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III. *Science* **258**: 434-440
- Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8**: 477-486
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymology* **276**: 307-326
- PDB (2006) Crystal structure of DNA-3-methyladenine glycosidase (10174367) from Bacillus halodurans at 2.55 Å resolution. *Joint Center for Structural Genomics, PDB ID 2H56*
- Rubinson EH, Metz AH, O'Quin J, Eichman BF (2008) A new protein architecture for processing alkylation damaged DNA: the crystal structure of DNA glycosylase AlkD. *J Mol Biol* **381**: 13-23
- Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* **229**: 105-124
- Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods in molecular biology (Clifton, NJ)* **364**: 215-230

Figure S1



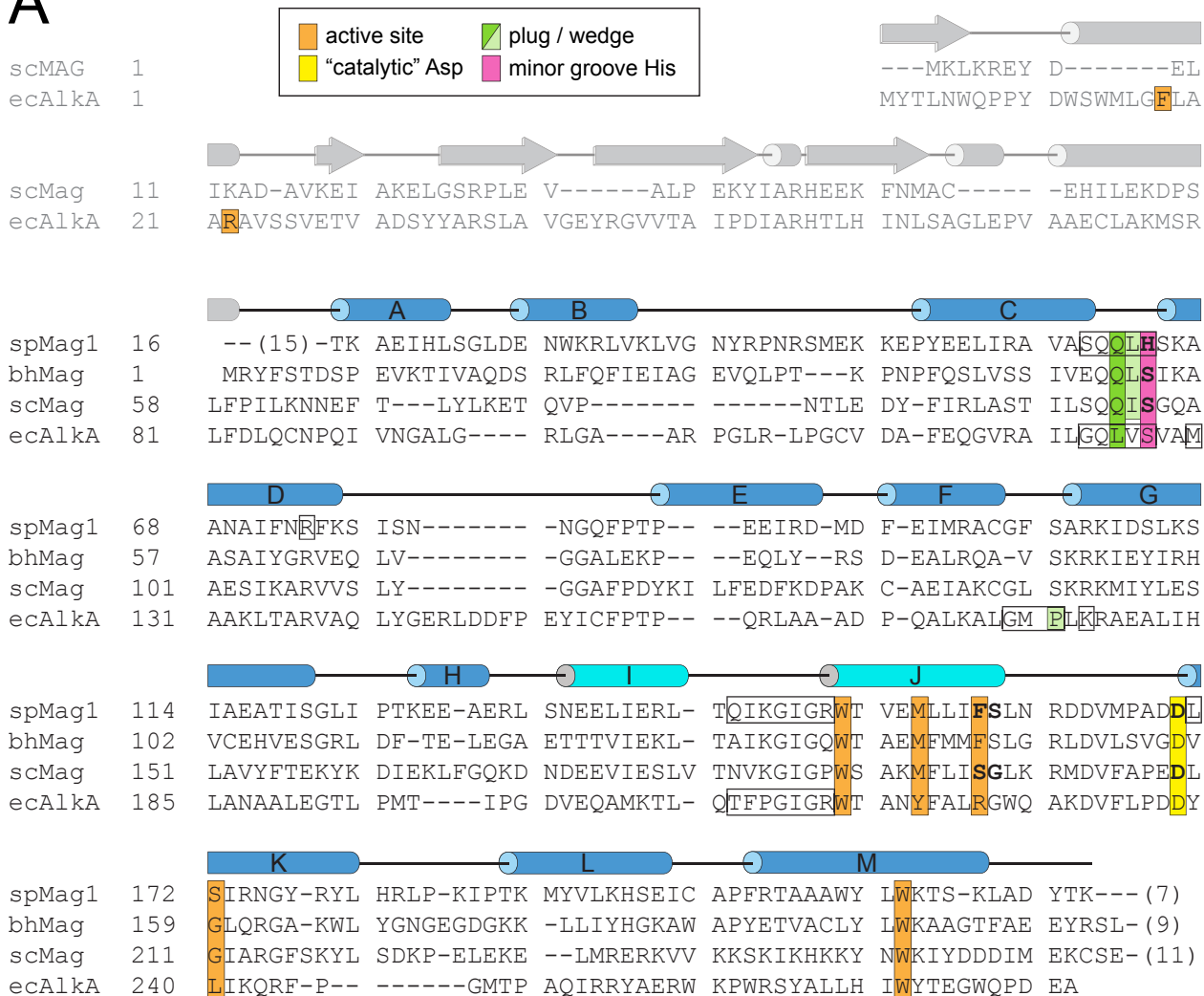


**B**

			I		J		
spMag1	133	SNEELIERL-	TQIKGIGRWT	VEMLLI	<b>F</b> SLN	RDDVMPAD	<b>D</b> L
bhMag	120	ETTTVIEKL-	TAIKGIGQWT	AEMFMM	FSLG	RLDVLSVG	DV
scMag	171	NDEEVIESLV	TNVKGIGPWS	AKMFLI	<b>S</b> GLK	RMDVFAPED	<b>D</b> L
ecAlkA	201	DVEQAMKTL-	QTFPGIGRWT	ANYFAL	RGWQ	AKDVFLPDD	<b>Y</b>

# Figure S3

## A



## B

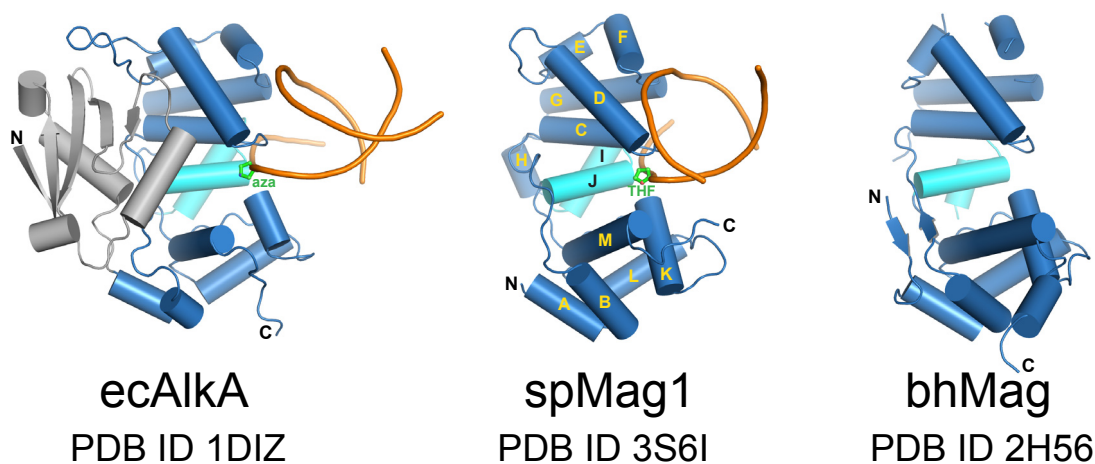


Figure S4

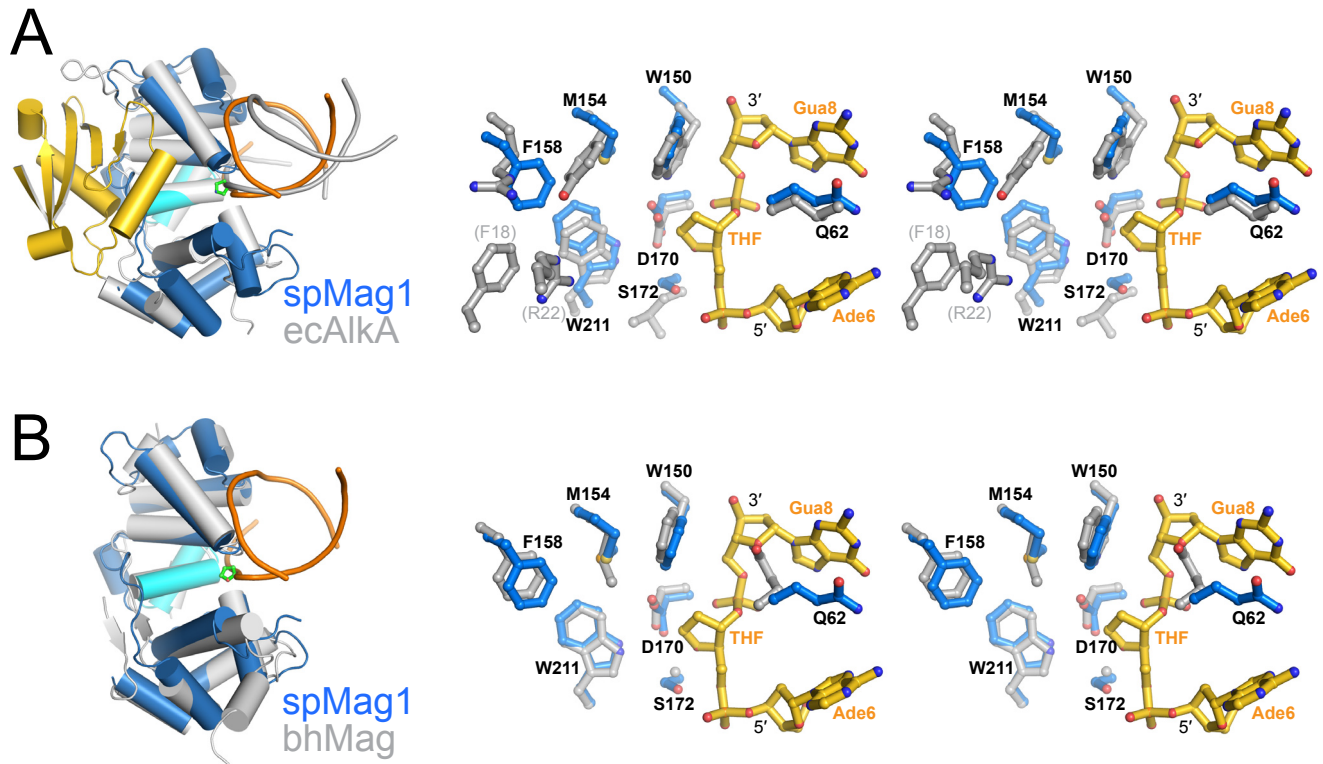




Figure S5

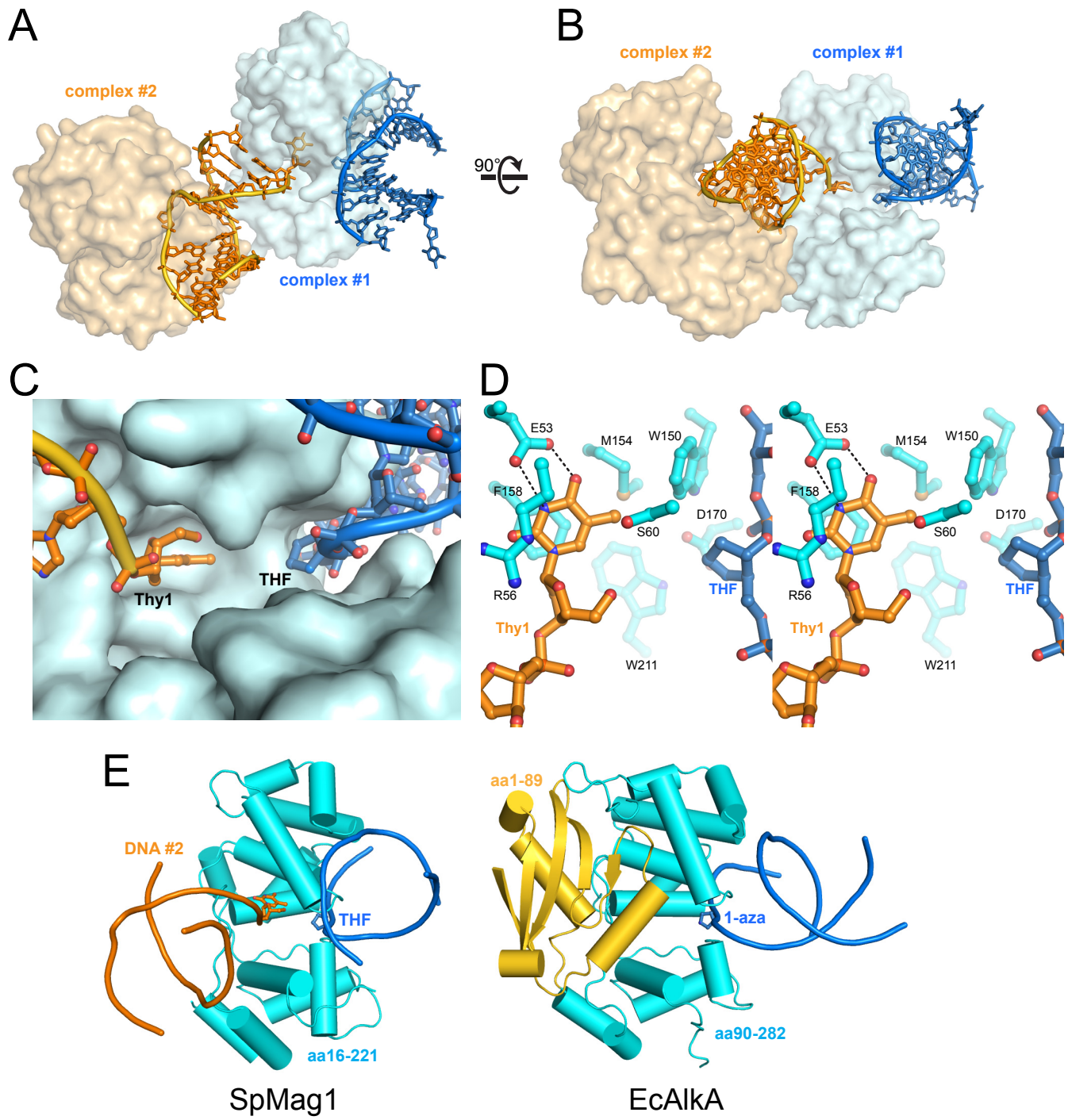


Figure S6

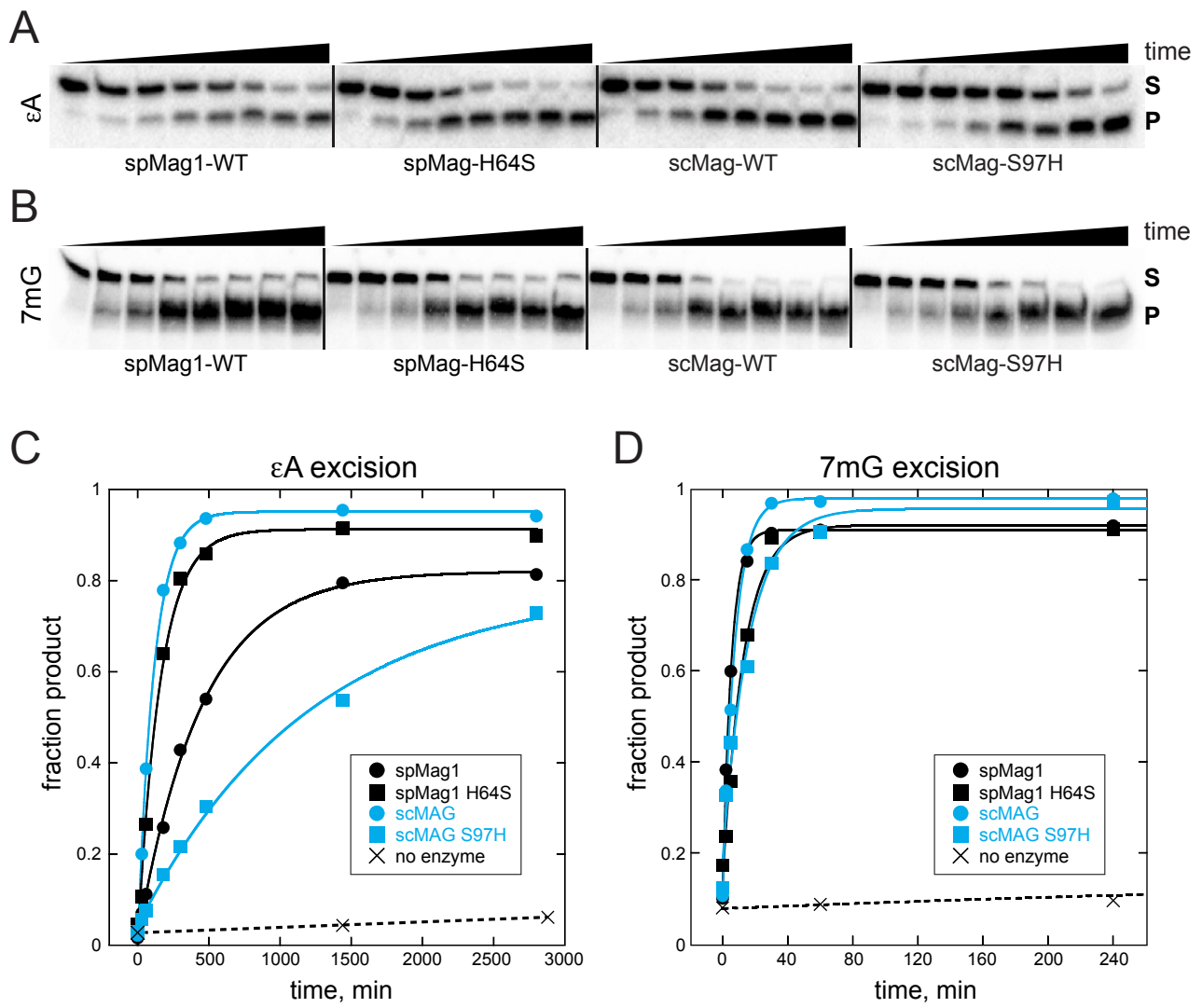
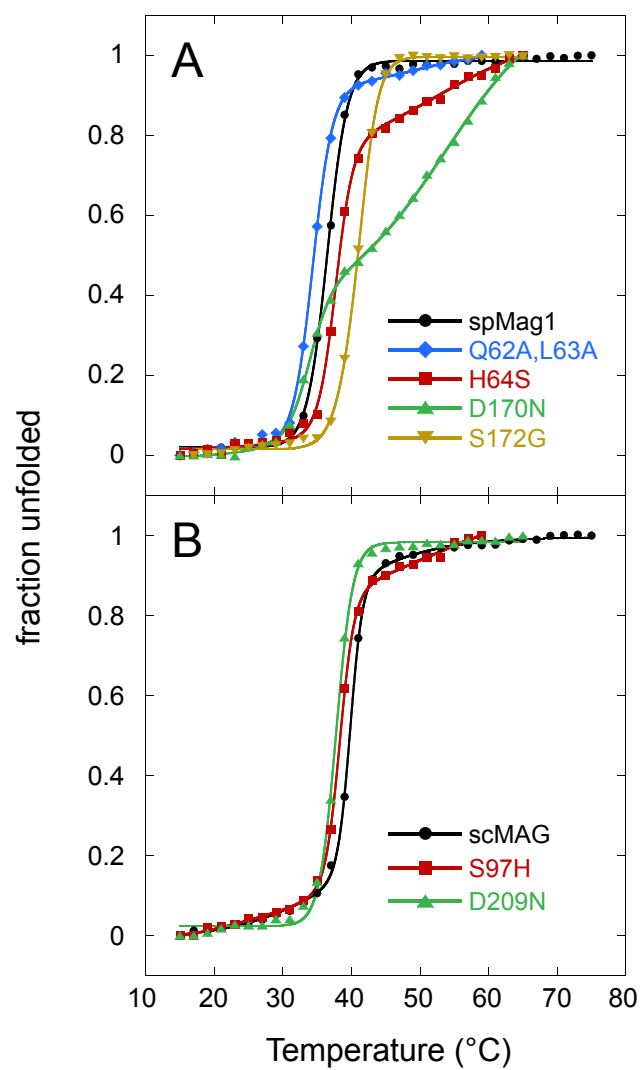


Figure S7



C

	$T_{m1}$ (°C)	$T_{m2}$ (°C)
spMag1	36.5	
Q62A,L63A	34.6	
H64S	37.7	52.4
D170N	33.9	53.7
S172G	40.9	
scMAG	39.7	
S97H	38.3	
D209N	37.7	

Figure S8

