Alkylated DNA damage flipping bridges base and nucleotide excision repair

Julie L. Tubbs¹, Vitaly Latypov², Sreenivas Kanugula³, Amna Butt⁴, Manana Melikishvili⁴, Rolf Krachenbuehl⁵†, Oliver Fleck⁶‡, Andrew Marriott², Amanda J. Watson², Barbara Verbeek⁷§, Gail McGown⁷, Mary Thorncroft², Mauro F. Santibanez-Koref⁶, Christopher Millington⁷, Andrew S. Arvai¹, Matthew D. Kroeger¹, Lisa A. Peterson⁸, David M. Williams⁷, Michael G. Fried⁴, Geoffrey P. Margison⁵, Anthony E. Pegg⁷, and John A. Tainer¹,⁹

¹Skaggs Institute for Chemical Biology and Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

²Cancer Research UK Carcinogenesis Group, Paterson Institute for Cancer Research, University of Manchester, Manchester, M20 4BX, UK

³Department of Cellular and Molecular Physiology, Milton S. Hershey Medical Center, Pennsylvania State University College of Medicine, Hershey, PA 17033, USA

⁴Center for Structural Biology, Department of Molecular and Cellular Biochemistry, University of Kentucky, Lexington, KY 40536, USA

⁵NWCRF Institute, Bangor University, Gwynedd LL57 2UW, UK

⁶Institute of Human Genetics, University of Newcastle-upon-Tyne, UK

⁷Centre for Chemical Biology, Department of Chemistry, University of Sheffield, UK

⁸Division of Environmental Health Sciences and the Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455, USA

⁹Life Sciences Division, Department of Molecular Biology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

†Present address: Cancer Research UK DNA Damage Response Group, Paterson Institute for Cancer Research, University of Manchester, Manchester, UK

‡Present address: Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark

§Present address: Department of Toxicology, University of Mainz, D-55131 Mainz, Germany.
Methods

Expression Plasmid Construction. The Atl1 gene was subcloned into the modified pQE30 vector, which contains a C-terminal 6x-His tag. The NvATL coding sequence (nucleotide 2-319 of NZ ABAV01051098) with flanking EcoRI and BamHI sites was synthesised at GenScript Corporation (Piscataway, NJ, USA). The NvATL gene and hAGT C145S mutant (from pQE30-hAGT C145S) were subcloned (using EcoRI and BamHI sites) into a modified pQE80 vector (Qiagen, USA) that has a C-terminal 6x-His tag.

The eAtl, UvrA, UvrB and UvrC genes for ligation into pQE80 expression vector with and without amino-terminal flag tag were separately amplified by PCR with *E.coli* MG1655 genomic DNA and the primer sets shown below. The pQE80-eAtl plasmid with N-terminal flag tag was constructed first and its vector backbone was then used for the construction of N-terminal flag tagged UvrA, UvrB and UvrC plasmids. The PCR products digested with suitable restriction enzymes were ligated into modified pQE80 vector that has C-terminal 6x-His tag (for without flag constructs) and modified pQE80 vector that has both N-terminal flag tag and C-terminal 6x-His tag for the UvrA, UvrB and UvrC constructs with flag tag. All plasmid constructs were sequenced to ensure no mutations were introduced during PCR.

The nucleotide sequence of the restriction enzyme sites in the primers are indicated in italics, the start codon is underlined and the nucleotide sequence coding for the flag tag is in small letters.

Sense primer with EcoRI and SpeI site for eAtl with N-terminal flag tag:

CACACAGAATTCCATTAAAGAGGAGAAATTAACGTATGgactacaaggacgacgatgacaagACTAGTATGCTGGTTTCTTGCGC
Sense primer with *EcoRI* site for eAtl (without flag tag):

CACACAGAATTCAATTAAGAGGAGAAATTAACTATGCTGGTTTCTTGCGC

Anti-sense primer with Bam HI site for both eAtl constructs:

GTGGTGGGATCCGTAGTTCAGCG

Sense primer with *SpeI* site for flag tagged UvrA construct:

GACAAGACTAGTATGGATAAGATCGAAGTTC

Sense primer with BseRI site for UvrA (without flag tag):

ATTAAAGAGGAATTAATGGAGATCGAAGTTC

Anti-sense primer with BamHI site for both UvrA constructs:

GTGGTGGGATCCAGCATCGTTCACAGGA

Sense primer with *SpeI* site for flag tagged UvrB construct:

GACAAGACTAGTATGGAGATTAACCCTCAAACCTG

Sense primer with MfeI site for UvrB (without flag tag):

TAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTTGAAGAGGA

GAAATTAATGAGTAAACCCTTCAAACCTG

Anti-sense primer with BamHI site for both UvrB constructs:

GTGGTGGGATCCCGATGGCCGATAAAACGC

Sense primer with *SpeI* site for flag tagged UvrC construct:

GACAAGACTAGTATGGAGATCGTTCAGTTGACGC
Sense primer with BseRI site for UvrC (without flag tag):

ATTAAAGAGGAGAAATTAACTATGAGTGATCAGTTTGACGC

Anti-sense primer with BamHI site for both UvrC constructs:

GTGGTGGGATCCATGTTTCAACGACCAGAGATC

**Protein Expression and Purification.** Hexa-histidine tagged proteins were purified over metal affinity columns (Ni-NTA, Qiagen or Talon, Clontech) from JM109 cells (At1l) induced with 50μM isopropyl-β-D-thiogalactopyranoside (IPTG) or epicurian XL1 blue E.coli cells (all other 6x-His-tagged proteins) induced with 100 μM IPTG. Cell pellets were resuspended in a minimum of Lysis Buffer (50 mM Tris-HCl, 300 mM NaCl, 10mM Imidazole, pH 8.0, 2 mM β-mercaptoethanol, pH 8.0), then lysed by sonication. The lysate supernatant was incubated with a 50% slurry of Ni-NTA agarose with rocking for 1 hour, 4°C, then poured into a column for the wash and elution steps. The column was washed twice with 50mM Tris 8.0, 300mM NaCl, 30mM Imidazole, 2mM β-mercaptoethanol and eluted with high salt buffer (50mM Tris 8.0, 1.5M NaCl, 10mM Imidazole, 5mM β-mercaptoethanol). The eluted protein was concentrated and run over a Superdex 75 16/60 gel filtration column in 20mM Tris 8.0, 300mM NaCl, 5mM β-mercaptoethanol. The pure protein was concentrated to ~8mg/mL, flash frozen in liquid nitrogen and stored at -80°C.

**At1l-DNA Complex Preparation for Crystallization.** Single-strand oligodeoxiribonucleotides of sequence 5’-GCCATGCTAGTA-3’, where X = O6-methylguanine or O6-4-(3-pyridyl)-4-oxobutylguanine, were annealed with equimolar complementary oligodeoxiribonucleotides of sequence 5’-CTACTAGCCATGG-3’ to produce 13-mer double-strand DNA with an overhang on either end. The resulting double-strand oligodeoxyribonucleotide was then mixed with purified At1l at a DNA:protein molar ratio of 1.5:1.
Crystallization and X-ray Diffraction Data Collection. Crystals were grown by the hanging drop vapor diffusion method mixing 1 μL of protein with 1 μL well solution. Crystals of Atl1 were grown from 100 mM CHES, pH 9.5. 30% sucrose in mother liquor was used as a cryoprotectant. Diffraction data were collected at ALS beamline 12.3.1 at a wavelength of 1.1158 on an ADSC Q315 detector. Initial crystallization conditions of Atl1:Oβ-methylguanine-DNA were obtained by liquid diffusion using a Fluidigm Topaz FID Crystallizer and a Fluidigm Topaz AutoInspeX Workstation to visualize crystals. Initial crystals were adapted for growth by vapor diffusion and refined for quality and size. Refined crystals of Atl1:Oβ-methylguanine-DNA were grown from well solution containing 10% mPEG 2000, 200 mM imidazole-malate, pH 5.8, 5% sodium formate, and 4mM cobalt hexamine. Crystals were briefly transferred to 30% ethylene glycol in mother liquor solution for cryoprotection and frozen in liquid nitrogen. Atl1: Oβ-4-(3-pyridyl)-4-oxobutylguanine-DNA crystals were grown from well solution containing 20% mPEG 2000, 5% sodium formate, 30% xylose, 200 mM imidazole-malate, pH 6.2. Diffraction data for Atl1:Oβ-mG and Atl1:Oβ-pobG DNA complexes were collected at SSRL beamline 11-1 at a wavelength of 0.97945 on a MAR325 detector. Diffraction data were processed with HKL2000. Structures were solved by molecular replacement with Phaser. A modified version of wild-type Ada (PDB code 1SFE) in which all side chains were replaced with the Atl1 sequence by the Fold & Function Assignment System server (http://ffas.ljcrf.edu/ffas-cgi/cgi/ffas.pl) was made and used as a molecular replacement search model for Atl1. The refined Atl1 structure was then used as a molecular replacement search model for Atl1:DNA complexes.

Crystallographic Structure Refinement and Analysis. Crystallographic refinement was done with Crystallography & NMR System (CNS). Xfit was used for manual model building into 2Fo - Fc and composite omit and simulated annealing omit 2Fo - Fc and Fo - Fc electron density maps. DNA was built into the model after two rounds of
refinement. Water molecules were added to regions of greater than 3σ positive density in a $F_o - F_c$ difference map after two rounds of refinement. Ramachandran plot statistics were as follows (most favored, additionally allowed, generously allowed, and disallowed regions): 168 (91.3%), 14 (7.6%), 2 (1.1%), 0 (0.0%) for Atl1; 76 (81.7%), 12 (12.9%), 3 (3.2%), 2 (2.2%) for Atl1:$O^6$-mG-DNA complex; and 62 (66.7%), 27 (29.0%), 3 (3.2%), 1 (1.1%) for Atl1:$O^6$-pobG-DNA complex. Structural superpositions were done with Sequoia\textsuperscript{7}. Binding pocket analysis was performed with the CASTp (http://sts-fw.bioengr.uic.edu/castp/index.php)\textsuperscript{8} and MOLEOnline (http://mole.chemi.muni.cz/index.php)\textsuperscript{9} servers. Sequence conservation was mapped onto the Atl1 primary sequence with the ConSeq server (http://conseq.tau.ac.il/)\textsuperscript{10} and onto the Atl1 structure surface with the ConSurf server (http://consurf.tau.ac.il/)\textsuperscript{11}. Sequence alignments were made with ClustalW2 (http://www.clustal.org/)\textsuperscript{12}, sequence conservation, quality, and consensus were calculated with JalView\textsuperscript{13}, and sequence logos were made with Weblogo (http://weblogo.berkeley.edu/)\textsuperscript{14}. Structure figures were made with PyMol (http://www.pymol.org).

**Surface plasmon resonance interaction analysis.** Oligodeoxyribonucleotides of sequence 5'-GAA CTX CAG CTC CGT GCT GGC CC-3', where X = $O^6$-methyl- or $O^6$-4-(3-pyridyl)-4-oxobutylguanine, were obtained by employing either methanol or 4-oxo-4-(3-pyridyl)butanol (synthesized by modification of the literature procedure\textsuperscript{15}) in post-DNA synthesis chemistry as described\textsuperscript{16}. An abasic site oligonucleotide of the same sequence where X was a spacer;AP site was obtained from DNA Technology A/S Denmark. Oligonucleotide-Atl1 interactions were analyzed using a ProteOn XPR36 (Bio Rad Laboratories Ltd.) at 25°C. Lesion-containing or control oligonucleotides (10µM) were annealed to biotinylated complement (C opposite X), and the resultant double-stranded oligonucleotides were diluted 1:1000 in running buffer (ProteOn PBS/Tween, pH 7.4). An aliquot (70ul) was immobilized on the streptavidin-coated surface of a Biacore SA chip at a flow rate of 30 µl/min to achieve a loading level of
approximately 100 RU (150RU for the AP oligonucleotide). Serial dilutions of Atl1 in running buffer (0-4nM for methyl and poh, 0-5nM for AP) were applied to the cell at a flow rate of 60 µl/min for 200 secs contact time and 900 secs dissociation time and signals recorded. Data were fitted using a 1:1 Langmuir model considering mass transfer limitation. No kinetic data could be obtained for the AP site oligonucleotide.

**Electrophoretic Mobility Shift Assays.** The binding reactions were carried out for 60 min at 20±1°C in 10 mM Tris (pH 8.3 at 20°C), 1 mM EDTA, 50 mM NaCl, 1 mM DTT, 5% glycerol, 0.1 mg/mL BSA. Electrophoresis was performed using 15% polyacrylamide gels (30:0.4 Acrylamide:N,N'-methylene bis-acrylamide) cast and run in 40 mM Tris-acetate (pH 7.6), 2.5 mM EDTA, 5% ethylene glycol.

Self-consistent estimates of binding stoichiometry (n) and the association constant (K_a) were obtained by the method of Fried and Crothers^{17,18}. For a single binding step in which n protein molecules associate with DNA the association constant is

\[ K_a = \frac{[P_nD]}{[D][P]_{free}^n} \]

Separating variables and taking logarithms gives

\[ \ln\frac{[P_nD]}{[D]} = n\ln[P]_{free} + \ln K_a \]

For these experiments, [D]total << [P]total, so [P]total is an acceptable estimate of [P]free. A graph of \( \ln([P_nD]/[D]) \) as a function of \( \ln[P]_{free} \) has a slope equal to the stoichiometry of the binding step, n. The equilibrium constant is most simply estimated at the midpoint of the reaction, where \( \ln([P_nD]/[D]) = 0 \) and \( \ln K_a = -\ln[P]_{free} \). A summary of the analysis for each binding reaction is below:

1. Binding to double stranded normal and \( O^6 \)-mG-containing13mers:

   Binding to double-stranded 13mer:
We tried to measure binding with double-stranded normal 13mer, but the complexes were too unstable to resolve by gel electrophoresis (note smears instead of resolved complexes in Fig. 3b top).

Binding to double-stranded $O^6$-mG-containing 13mer:

\[ \begin{align*}
5'\text{-CCC ATG XCT AGT A-3'} & \quad [\text{DNA}] = 0.17 \mu\text{M} \\
3'\text{-GG TAC CGA TCA TC-5'} & \quad 0 \leq [\text{ATL}] \leq 14.5 \ \mu\text{M}
\end{align*} \]

$X = O^6$-mG

Two independent experiments were performed (symbols $\bullet$, $\blacksquare$ in Fig. 2F bottom right). The slope of the graph is a measure of stoichiometry. This value was confirmed by sedimentation equilibrium analysis. Values of $n$ and $K$ are summarized in table S2.

2. Binding to double stranded normal and $O^6$-mG-containing 16mers:

Binding to double stranded 16mer:

\[ \begin{align*}
5'\text{-GAC TGA CTG ACT GAC T-3'} & \quad [\text{DNA}] = 0.248 \ \mu\text{M} \\
3'\text{-CTG ACT GAC TGA CTG A-5'} & \quad 0 \leq [\text{ATL}] \leq 13.7 \ \mu\text{M}
\end{align*} \]

Binding to double stranded normal 16mer appears to take place in a single-step. Complexes can be resolved, although significant dissociation occurs during the gel run.
In the quantitative analysis (fig. S2A), the binding stoichiometry appears to be 2 protein:1DNA (slope of the ln-ln plot is $1.9 \pm 0.15$). We estimate that the equilibrium constant for the overall reaction, $K_n = (6.74 \pm 0.83) \times 10^{10} \text{ M}^{-2}$, corresponding to a monomer-equivalent association constant of $K = (5.0 \pm 0.3) \times 10^5 \text{ M}^{-1}$. The data in fig. S2A were obtained in two independent experiments (symbols ●, ▲) the solid line is the fit to the data closest to the mid-point of the reaction (filled symbols).

Binding to double stranded $O^6$-mG-containing 16mer:

$5'$-GAC TGA CTG ACT XAC T-3'  
$3'$-CTG ACT GAC TGA CTG A-5'  
$[\text{DNA}] = 0.15 \mu\text{M}$

$0 \leq [\text{ATL}] \leq 13.7 \mu\text{M}$

$X = O^6$-mG

Binding to double-stranded 16mer containing $O^6$-mG takes place in two steps. The resulting complexes are comparatively stable to electrophoresis. As shown by the slopes of the ln–ln plots (fig. S2A), each binding step appears to add one protein molecule. For the first binding step, $K = (9.09 \pm 0.56) \times 10^5 \text{ M}^{-1}$ and for the second binding step, $K = (8.64 \pm 1.22) \times 10^5 \text{ M}^{-1}$.

3. Binding to double stranded normal and $O^6$-mG-containing 26mers:

Binding to double stranded 26mer:

$5'$-AGT CAG TCA GTC AGT CAG TCA GTC AG-3'  
$3'$-TCA GTC AGT CAG TCA GTC AGT CAG TC-5'  
$[\text{DNA}] = 0.17 \mu\text{M}$

$0 \leq [\text{ATL}] \leq 11.6 \mu\text{M}$
Binding to double stranded normal 26mer appears to take place in a single-step. Complexes can be resolved, although significant dissociation occurs during the gel run. In the quantitative analysis (fig. S2B), the binding stoichiometry appears to be 3 protein:1 DNA (the slope of the ln-ln plot is $2.77 \pm 0.12$). We estimate that the equilibrium constant for the overall reaction, $K_a = (2.58 \pm 0.35) \times 10^{16} \text{ M}^{-3}$, corresponding to a monomer-equivalent association constant of $K = (8.40 \pm 0.39) \times 10^{5} \text{ M}^{-1}$. The data shown were obtained in two independent experiments (symbols ●, ■).

Binding to double stranded $O^6$-mG-containing 26mer:

5'-AGT CAG TCA GTC AXT CAG TCA GTC AG-3'  [DNA] = 0.15 μM

3'-TCA GTC AGT CAG TCA GTC AGT CAG TC-5'  $0 \leq [ATL] \leq 11.6 \mu M$

$X = O^6$-mG

Binding to double stranded 26mer containing $O^6$-mG takes place in two steps. The first complex to form (labeled 1 in fig. S2B) is comparatively stable to electrophoresis. The second complex to form (labeled 2 in fig. S2B) dissociates significantly during the gel run. As shown in the ln–ln plot (fig. S2B), the first binding step appears to add one protein molecule (slope = $1.08 \pm 0.05$). For the this binding step, $K = (2.41 \pm 0.33) \times 10^7 \text{ M}^{-1}$. In the second binding step, two equivalents of protein bind the DNA (slope = $1.89 \pm 0.09$). The apparent equilibrium constant for this step is $K_a = (9.83 \pm 0.94) \times 10^{10} \text{ M}^{-2}$. This corresponds to a monomer-equivalent association constant $K = (6.54 \pm 0.32) \times 10^5 \text{ M}^{-1}$.

**Analytical ultracentrifugation.** Sedimentation equilibrium analysis was used to establish the stoichiometries of the DNA complexes formed under conditions of protein saturation. Samples (1μM DNA duplex and 5-25μM protein) were prepared in buffer
containing 10 mM Tris (pH 8.3 at 20°C), 200 mM NaCl, 1 mM DTT, 5% glycerol. Centrifugation was performed in a Beckman-Coulter XL-A centrifuge using an AN60Ti rotor. Scans were obtained at 252 nm<sup>19</sup>. Typically, sedimentation times ≥24h were required for attainment of equilibrium. Data were acquired at 25,000, 30,000 and 35,000 rpm. Five scans were averaged for each sample and rotor speed. Smooth curves are global fits of Eq. 1 to nine data sets (three protein concentrations at three rotor speeds).

At sedimentation equilibrium, the radial distribution of absorbance for an equilibrium mixture containing protein, DNA and a protein-DNA complex is given by equation 1.

\[
A(r) = \alpha_p \exp \left[ \sigma_p (r^2 - r_0^2) \right] + \alpha_D \exp \left[ \sigma_D (r^2 - r_0^2) \right] + \alpha_{pD} \exp \left[ \sigma_{pD} (r^2 - r_0^2) \right] + \varepsilon \quad (1)
\]

Here \(A(r)\) is the absorbance at radial position \(r\) and \(\alpha_p\), \(\alpha_D\) and \(\alpha_{pD}\) are absorbances of protein, DNA and protein-DNA complex at the reference position, \(r_0\), and \(\varepsilon\) is a baseline offset that accounts for radial position-independent differences in the absorbances of different cell assemblies. The reduced molecular weights of ATL protein, DNA and protein-DNA complexes are given by \(\sigma_p = M_p \left(1 - \frac{v_p}{v_D} \right) \omega^2 / (2RT)\), \(\sigma_D = M_D \left(1 - \frac{v_D}{v_p} \right) \omega^2 / (2RT)\) and \(\sigma_{pD} = (nM_p + M_D) \left(1 - \frac{v_{pD}}{v_D} \right) \omega^2 / (2RT)\). Here \(M_p\) and \(M_D\) are the molecular weights of protein and DNA, \(n\) is the protein:DNA ratio of the complex; \(\rho\) is the solvent density, \(\omega\), the rotor angular velocity, \(R\) is the gas constant and \(T\) the temperature (Kelvin). The partial specific volume of ATL \((v_p = 0.716 \, \text{mL/g})\) was calculated by the method of Cohn and Edsall<sup>20</sup>, using partial specific volumes of amino acids tabulated by Laue and co-workers<sup>21</sup>. The partial specific volume of double-stranded NaDNA at 0.1M NaCl (0.540 ml/g) was estimated by interpolation of the data of Cohen and Eisenberg<sup>22</sup>. 
Partial specific volumes of protein-DNA complexes were estimated using Eq. 2.

\[
\bar{v}_{p,nD} = \frac{(nM_p \bar{v}_p + M_D \bar{v}_D)}{(nM_p + M_D)}
\]

This relationship is based on the assumption that there is no significant change in partial specific volumes of the components upon association\textsuperscript{23-25}. Equation 1 was used in global analysis of multiple data sets obtained at different macromolecular concentrations and rotor speeds\textsuperscript{26}. In this method, the values of \( \alpha_p \), \( \alpha_D \), \( \alpha_{PN} \) and \( \epsilon \) are unique to each sample but the value of \( n \) must be common to all data sets. Terms accounting for non-ideal effects were not included because there was no evidence of nonideality (results not shown).

**Inhibition of human AGT activity by different ATLs.** The inhibition of hAGT activity by ATLs was measured by adding purified hAGT to a preformed mixture of \(^3\text{H}\)methylated DNA and different ATLs or hAGT-C145S and then assaying the mixture for alkyltransferase activity. The alkyltransferase activity was measured by determining the transfer of \(^3\text{H}\)methyl groups to the purified human AGT protein from \( O^6-[^3\text{H}] \text{methylguanine} \) in DNA\textsuperscript{27}. The assay mixture (1.0 ml), incubated at 37°C for 15 min., contained 50 mM Tris-HCl (pH 7.6), 5 mM dithiothreitol (DTT), 50 \( \mu \)g hemocyanin, 0.1mM EDTA, 15 \( \mu \)g of the \(^3\text{H}\)methylated calf thymus DNA and different amounts of purified ATLs or hAGT-C145S mutant. Ten microliters of purified human AGT protein (0.5 pmoles) was added to the above reaction mixture, incubation was continued for 60 min at 37°C and assayed for alkyltransferase activity.
Far western analysis. Purified proteins were separated on SDS-15% polyacrylamide gels, transferred to a PVDF membrane and renatured with guanidine-HCl as described by Jayaraman, et al. The membrane was blocked with 5% non-fat dry milk in 50 mM Tris-HCl (pH 7.6), 150 mM NaCl, 0.5% Tween 20 for 1 hour at 24°C. The membrane was then incubated for 6 hrs at 4°C with 0.5μg/ml Flag-tagged UvrA or Flag-tagged eAtl in 0.25% non-fat dry milk in 50 mM Tris-HCl (pH 7.6), 150 mM NaCl and 0.5% Tween 20. The blot was then probed with Anti-Flag M2 monoclonal antibody (Sigma).

Yeast media. S. pombe media of malt extract agar, YEA (yeast extract agar), and YEL (yeast extract liquid) were as described. Minimal medium agar (MMA) consists of 0.17% Yeast nitrogen base without amino acids without ammonium sulphate, 0.5% ammonium sulphate, 1% glucose, and 1.8% agar. The pH of MMA was adjusted to 6.5 with 1 N NaOH. Media were supplemented with adenine, histidine, leucine, lysine, and/or uracil (each 0.01%) where required.

S. pombe strains. The S. pombe strains used in this study originated from GM4 (h− att1::ura4 ura4-D18 leu1-32 his7-366 ade6-M210), RO131 (h+ rad13::kanMX ura4-D18 his3-D1; RK and OF, unpublished).

Agar plate (“spot”) assays. Plate assays were a modified version of a method described previously. Plates were prepared by adding the appropriate volume of MNNG stock solution (20mM in dry DMSO) to cooled molten agar. Dilutions of S. pombe cultures containing 1.2 x 10^7 cells/ml were made (100, 10^-1, 10^-2, 10^-3, and 10^-4) and 5 µl aliquots were spotted onto the plates. Photographs were taken after 2-3 days of incubation.

Determination of MNNG sensitivity by clonogenic assay. To determine MNNG sensitivity, exponentially growing cell suspensions (10^7 cell/ml) were exposed to increasing concentrations of MNNG for 10 mins and the reaction stopped by 100-fold
dilution in distilled water. Serial cell dilutions were then plated on YEPD medium and cell titer was determined by plating the original cell suspension without MNNG treatment. After 4 days of growth at 30°C, cell survival was calculated as ratio of the number of colonies and the total amount of cells plated. All experiments were repeated two to four times. Mean, standard deviation, and standard error of the mean were calculated by the following formulas:

Mean:

$$\bar{x} = \frac{\sum x_i}{n}$$

Standard deviation

$$\sigma = \sqrt{\frac{\sum (x - x_i)^2}{n - 1}}$$

Standard error of the mean

$$m_x = \frac{\sigma}{\sqrt{n}}$$

$$X_{real} = \bar{x} \pm m_x$$

**Mutation rate determination in *S. pombe.*** Mutation rates were determined as reversions of ade6-485 to Ade⁺. *ade6-485* is a C to G transversion that reverts via various base substitutions³¹,³². Each fluctuation test included nine cultures, grown for approximately 16 hours in 10 ml YEL medium to a final density of 0.5-2.0 x 10⁷ cells/ml. Cells were washed with 5 ml H₂O, suspended in 200 μl H₂O and plated on minimal medium to select for adenine prototrophs. Appropriate dilutions were plated on complex medium for determination of the cell titre. Plates were incubated for seven days at 30°C. For each background reversion assays were performed at least three times. Mutation rates were determined by the method of the median or from the proportion of
cultures without Ade$^{+33,34}$. The number of experiments was in all cases 3, except for atl1-MNNG (5) and rad13+MNNG (4). Results are shown as mean ± standard deviation, which was calculated by the following formula:

$$SD = \sqrt{\frac{\sum(X - \bar{x})^2}{n - 1}}$$

where $\bar{X}$ = mean value (average), $x$ = individual value of a given experiment, and $n$ = number of experiments ($\geq 3$).

**Atl1 expression in *E. coli*.** Expression of Atl1 in *E. coli* reduces MNNG-induced mutations: *E. coli* GWR109 ada$^{-}$ ogt$^{-}$ cells$^{35}$ containing the pQE-30 empty vector and *E. coli* GWR109 ada$^{-}$ ogt$^{-}$ atl$^{-}$ cells or *E. coli* GWR109 his$^{-}$ ada$^{-}$ ogt$^{-}$ cells containing the pQE-30 empty vector or pQE-Atl1 were grown overnight in one ml LB broth containing 50 μg/ml ampicillin and 50 μg/ml kanamycin. Two ml of fresh LB medium (containing 50 μg/ml ampicillin and 50 μg/ml kanamycin) were then inoculated with 20 μl of the overnight culture and grown with agitation at 200 rpm at 37°C.

For mutation frequencies expressed as the number of MNNG-induced Rifampicin resistant mutants: When the OD$_{600}$ reached 0.5, the cultures were exposed to MNNG (from 0 to 4 μg/ml by addition of a 0.4 mg/ml stock solution in 10% DMSO). After 30 min at 37°C, the cells were washed in M9 minimal medium and resuspended in LB medium and plated on either LB plates (containing 50 μg/ml ampicillin and 50 μg/ml kanamycin) to estimate the number of surviving cells or LB Rifampicin (100μg/ml) plates to estimate the number of Rifampicin resistant mutants, respectively. The plates were incubated for 24-36 hours at 37°C. Mutation frequencies were expressed as the number of MNNG-induced Rifampicin resistant mutants (Rif$^{\beta}$) per $10^8$ surviving cells. The eAtl gene in *E.coli* BW25141 was replaced by eAtl:cat cassette as described by Datsenko and Wanner$^{36}$. The *E. coli* GWR109 ada$^{-}$ ogt$^{-}$ atl$^{-}$ triple mutant was then
constructed by P1 transduction of the eAtl:cat cassette into the E. coli GWR109 ada\(^{-}\) ogt\(^{-}\) strain.

For mutation frequencies expressed as the number of MNNG-induced his\(^{+}\) revertants: When the OD\(_{600}\) reached 0.5, the cultures were exposed to MNNG (from 0 to 4 µg/ml by addition of a 0.5 mg/ml stock solution in 100 mM Na acetate buffer, (pH 5.0). After 30 min at 37\(^{\circ}\)C, the cells were washed in M9 minimal medium and plated on either minimal plates containing histidine to estimate the number of surviving cells or minimal plates lacking histidine to estimate the number of his\(^{+}\) revertants, respectively. The plates were incubated for 24-36 hours at 37\(^{\circ}\)C. Mutation frequencies were expressed as the number of MNNG-induced his\(^{+}\) revertants per 10\(^{8}\) surviving cells.

References

1. Wang, L., Spratt, T. E., Pegg, A. E. & Peterson, L. A. Synthesis of DNA oligonucleotides containing site-specifically incorporated \(O^{6}\)-[4-oxo-4-(3-pyridyl)butyl]guanine and their reaction with \(O^{6}\)-alkylguanine-DNA alkyltransferase. *Chem. Res. Toxicol.* **12**, 127-131 (1999).
2. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* **276**, 307-326 (1997).
3. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658-674 (2007).
4. Moore, M. H. et al. Crystal structure of a suicidal DNA repair protein: the Ada \(O^{6}\)-methylguanine-DNA methyltransferase from *E. coli*. *EMBO J.* **13**, 1495-1501 (1994).
5. Brunger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905-921 (1998).
6. McRee, D. M. XtalView/Xfit-A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156-165 (1999).
7. Bruns, C. M. et al. Human glutathione transferase A4-4 crystal structures and mutagenesis reveal the basis of high catalytic efficiency with toxic lipid peroxidation products. *J. Mol. Biol.* **288**, 427-439 (1999).
8. Dundas, J. et al. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **34**, W116-118 (2006).
Petrek, M., Kosinova, P., Koca, J. & Otyepka, M. MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure* **15**, 1357-1363 (2007).

Berezin, C. et al. ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics* **20**, 1322-1324 (2004).

Landau, M. et al. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299-302 (2005).

Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).

Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. The Jalview Java alignment editor. *Bioinformatics* **20**, 426-427 (2004).

Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188-1190 (2004).

Hecht, S. S., Chen, C.-h. B. & Hoffman, D. Synthesis of N-nitrosamino aldehydes. *Tetrahedron Lett.* **17**, 593-596 (1976).

Shibata, T. et al. Novel synthesis of $O^6$-alkylguanine containing oligodeoxyribonucleotides as substrates for the human DNA repair protein, $O^6$-methylguanine DNA methyltransferase (MGMT). *Nucleic Acids Research* **34**, 1884-1891 (2006).

Fried, M. & Crothers, D. M. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**, 6505-6525 (1981).

Fried, M. G. & Daugherty, M. A. Electrophoretic analysis of multiple protein-DNA interactions. *Electrophoresis* **19**, 1247-1253 (1998).

Daugherty, M. A. & Fried, M. G., in *Modern Analytical Ultracentrifugation: Techniques and Method*, (eds D. Scott) 195-209 (Royal Society of Chemistry, 2005).

Cohn, E. J. & Edsall, J. T., in *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*, (eds E.J. Cohn and J.T. Edsall) 370-381 and 428-431 (Reinhold, 1943).

Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L., in *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, (eds S.E. Harding, A.J. Rowe, and J.C. Harding) 90-125 (The Royal Society of Chemistry, 1992).

Cohen, G. & Eisenberg, H. Deoxyribonucleate solutions: sedimentation in a density gradient, partial specific volumes, density and refractive index increments, and preferential interactions. *Biopolymers* **6**, 1077-1100 (1968).

Rippe, K., Mucke, N. & Schulz, A. Association states of the transcription activator protein NtrC from *E. coli* determined by analytical ultracentrifugation. *J. Mol. Biol.* **278**, 915-933 (1998).

Fried, M. G., Dyckman, D. & Heyduk, T. The *Escherichia coli* cyclic AMP receptor protein forms a 2 : 2 complex with RNA polymerase holoenzyme, *in vitro*. *Biophys. J.* **82**, 906 (2002).

Kar, S. R. et al. SmtB-DNA and protein-protein interactions in the formation of the cyanobacterial metallothionein repressor complex: $Zn^{2+}$ does not dissociate the protein-DNA complex *in vitro*. *Biochemistry* **40**, 13378-13389 (2001).

Cole, J. L. Analysis of heterogeneous interactions. *Methods Enzymol.* **384**, 212-232 (2004).
Kanugula, S. & Pegg, A. E. Novel DNA repair alkyltransferase from *Caenorhabditis elegans*. *Environ. Mol. Mutagen.* **38**, 235-243 (2001).

Jayaraman, L. et al. High mobility group protein-1 (HMG-1) is a unique activator of p53. *Genes Dev.* **12**, 462-472 (1998).

Gutz, H., Heslot, H., Leupold, U. & Loprieno, N., in *Handbook of Genetics*, (eds R. C. King) 395–446 (Plenum Press, 1974).

Pearson, S. J. et al. A novel DNA damage recognition protein in *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **34**, 2347-2354 (2006).

Schar, P. & Kohli, J. Marker effects of G to C transversions on intragenic recombination and mismatch repair in *Schizosaccharomyces pombe*. *Genetics* **133**, 825-835 (1993).

Fleck, O., Lehmann, E., Schar, P. & Kohli, J. Involvement of nucleotide-excision repair in *msh2 pms1*-independent mismatch repair. *Nat. Genet.* **21**, 314-317 (1999).

Luria, S. E. & Delbruck, M. Mutations of bacteria from virus sensitivity to virus resistance *Genetics* **28**, 491-511 (1943).

Lea, D. E. & Coulson, C. A. The distribution of the number of mutants in bacterial populations. *J. Genet.* **49**, 264-285 (1949).

Rebeck, G. W. & Samson, L. Increased spontaneous mutation and alkylation sensitivity of *Escherichia coli* strains lacking the *ogt* *O*<sup>6</sup>-methylguanine DNA repair methyltransferase. *J. Bacteriol.* **173**, 2068-2076 (1991).

Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640-6645 (2000).
**Supplementary Table 1. Data collection and refinement statistics**

|                      | Atl1                | Atl1:O6-mG-DNA complex | Atl1:O6-pobG-DNA complex |
|----------------------|---------------------|------------------------|--------------------------|
| **Data collection**  |                     |                        |                          |
| Space group          | P3\(_1\)21          | P6\(_1\)22             | P6\(_1\)22               |
| Cell dimensions      |                     |                        |                          |
| \(a, b, c\) (Å)      | 46.3, 46.3, 169.3   | 59.7, 59.7, 238.0      | 60.2, 60.2, 235.5        |
| \(\alpha, \beta, \gamma\) (°) | 90.0, 90.0, 120.0 | 90.0, 90.0, 120.0      | 90.0, 90.0, 120.0        |
| Resolution (Å)       | 50.0-2.00 (2.07-2.00) * | 50.0-2.70 (2.80-2.70)  | 50.0-2.80 (2.90-2.80)    |
| \(R_{sym}\) or \(R_{merge}\) | 7.1 (28.6)   | 8.4 (50.0)             | 5.5 (50.0)               |
| \(I/\sigma I\)       | 25.2 (4.9)          | 25.3 (1.9)             | 72.8 (2.6)               |
| Completeness (%)     | 98.7 (98.5)         | 94.6 (74.6)            | 93.7 (67.1)              |
| Redundancy           | 9.3 (8.8)           | 10.8 (5.1)             | 17.0 (9.5)               |
| **Refinement**       |                     |                        |                          |
| Resolution (Å)       | 50.0-2.00           | 50.0-2.70              | 50.0-2.80                |
| No. reflections      | 13976              | 6716                   | 6069                     |
| \(R_{work}\)/\(R_{free}\) | 22.5 (25.2) | 22.7 (29.6)            | 21.6 (28.3)              |
| No. atoms            |                     |                        |                          |
| Protein              | 1766                | 891                    | 840                      |
| DNA                  | 0                   | 491                    | 538                      |
| Ligand/ion           | 0                   | 14                     | 0                        |
| Water                | 122                 | 22                     | 6                        |
| Average B factor (Å\(^2\)) | 41.4              | 39.9                   | 56.6                     |
| R.m.s deviations    |                     |                        |                          |
| Bond lengths (Å)     | 0.0051              | 0.0068                 | 0.0086                   |
| Bond angles (°)      | 1.1                 | 1.4                    | 1.4                      |

*Highest resolution shell is shown in parenthesis.
Supplementary Table 2. Lesion-binding pocket sizes for Atl1 and AGTs of known structure.

| Protein                  | PDB ID | Area ($\text{Å}^2$) | Volume ($\text{Å}^3$) |
|--------------------------|--------|----------------------|------------------------|
| Atl1+DNA                 | 3GX4   | 610.07               | 1075.60                |
| Atl1                     | 3GVA   | 477.45               | 827.09                 |
| Human AGT+DNA            | 1T38   | 250.57               | 319.55                 |
| Human AGT                | 1EH6   | 228.22               | 302.76                 |
| *S. tokodaii* AGT        | 1WRJ   | 174.14               | 174.72                 |
| *P. kodakaraensis* AGT   | 1MGT   | 161.03               | 148.15                 |
| *E. coli* Ada-C          | 1SFE   | 139.97               | 114.23                 |
### Supplementary Table 3. Stoichiometries and equilibrium constants from EMSA and sedimentation equilibrium analyses.

| DNA     | Binding step | $\Delta n$     | $K_{assoc}$                  |
|---------|--------------|----------------|------------------------------|
| ds13    | saturating   | $0.97 \pm 0.21^b$ |                              |
|         |              |                |                              |
| ds13+Me | 1            | $0.94 \pm 0.09^a$ | $(5.20 \pm 0.78) \times 10^5$ M$^{-1}$ |
|         | saturating   | $1.15 \pm 0.08^b$ |                              |
| ds16    | 1            | $1.90 \pm 0.15^a$ | $(6.74 \pm 0.83) \times 10^{10}$ M$^{2}$ |
|         | saturating   | $1.91 \pm 0.13^b$ |                              |
| ds16+Me | 1            | $0.98 \pm 0.04^a$ | $(9.09 \pm 0.56) \times 10^5$ M$^{-1}$ |
|         | 2            | $1.20 \pm 0.09^a$ | $(8.64 \pm 1.22) \times 10^6$ M$^{-1}$ |
|         | saturating   | $1.93 \pm 0.19^b$ |                              |
| ds26    | 1            | $2.77 \pm 0.12^a$ | $(2.58 \pm 0.35) \times 10^{16}$ M$^{-3}$ |
|         | saturating   | $3.03 \pm 0.20^b$ |                              |
| ds26+Me | 1            | $1.08 \pm 0.05^a$ | $(2.41 \pm 0.33) \times 10^7$ M$^{-1}$ |
|         | 2            | $1.89 \pm 0.09^a$ | $(9.83 \pm 0.94) \times 10^{10}$ M$^{2}$ |
|         | saturating   | $2.99 \pm 0.21^b$ |                              |

a. Measured by EMSA

b. Measured by sedimentation equilibrium analysis
Supplementary Table 4: Atl1 Expression in *E. coli* reduces MNNG-induced mutations.

| MNNG (µg/µl) | pQE-30 | pQE-Atl1 |
|--------------|--------|----------|
|              | Cell survival | His⁺ rev/10⁸ cfu | Cell survival | His⁺ rev/10⁸ cfu |
| 0            | 100          | <5        | 100           | <5          |
| 2            | 52.6 (16.0)  | 4484 (991) | 69.3 (8.4)    | 2528 (625) |
| 4            | 19.0 (7.0)   | 5161 (1113) | 25.0 (7.3)    | 3378 (947) |

Values are from 4 experiments with standard deviations shown in parentheses.

| MNNG (µg/µl) | GWR109ada⁻ogf⁻/pQE-30 | GWR109ada⁻ogf⁻ atf⁻/pQE-30 | GWR109ada⁻ogf⁻ atf⁻/pQE- SpATL |
|--------------|------------------------|----------------------------|---------------------------------|
|              | Cell survival | Rif³ mutants/10⁸ cfu | Cell survival | Rif³ mutants/10⁸ cfu | Cell survival | Rif³ mutants/10⁸ cfu |
| 0            | 100          | <2        | 100           | <2          | 100           | <2          |
| 2            | 87.5         | 6054      | 77.4          | 6073        | 98.7          | 3524        |
| 4            | 58.4         | 9358      | 54.8          | 10289       | 85.4          | 5063        |

Values are an average of two experiments.
**Supplementary Figure 1 | Atl1 gel-shift assays.** Binding of Atl1 to double-stranded nonmethylated and methylated 16mer (a, left) and 26mer (b, left) oligonucleotides and associated binding isotherm analysis (a and b, right). For binding isotherm analyses two independent experiments are shown.
Supplementary Figure 2 | Complementation of Δatl1 cells by pREP3X-atl1.

Cell survival (single experiment in duplicate; a) further validated by spot assay (b). Spot assays show six days of growth on selective MMA media containing leu, ura, his but lacking leu (non-inducing conditions for nmt1 promoter that assure basal level of transcription).
Supplementary Figure 3 | Cell viability of untreated (black) and MNNG-treated (gray) cells. There were nine cultures of a given strain for each fluctuation test, and each experiment was repeated several times. For each culture the number of cells was counted and a dilution of the culture was plated on non-selective medium. After seven days incubation at 30°C the number of colonies were counted.
Supplementary Figure 4 | Sequence alignment for select ATLs. Nucleotide flipping residues pink, binding site motif yellow, binding site loop cyan, HTH motif blue box. Atl1 numbering and secondary structure are shown. Sequence logo, conservation, quality, and consensus derived from 197 ATL sequences.
Supplementary Figure 5 | Sequence conservation of ATL proteins (197 sequences). Sequence conservation mapped onto Atl1 molecular surface (a) and amino acid sequence (b). Binding site motif, Tyr25, and Arg39 are underlined in blue in (b).
Supplementary Figure 6 | Archaeal ATLendoV. Sequence alignment of ATL (a) and EndoV (b) domains.
Supplementary Figure 7 | Inhibition of human AGT activity by hAGTC145S and ATLs from *N. vectensis* (NvATL), *S. pombe* (Atl1), and *E. coli* (eAtl).