Supporting Information for:

Structure and Stereochemistry of the Base Excision Repair Glycosylase MutY Reveal a Mechanism Similar to Retaining Glycosidases

Ryan D. Woods[1,†] , Valerie L. O'Shea[1,2,†], Aurea Chu[1], Sheng Cao[1], Jody L. Richards[1], Martin P. Horvath[3,*], Sheila S. David[1,*]

[1] Department of Chemistry, University of California, Davis, Davis CA 95616, USA
[2] Department of Chemistry, University of Utah, Salt Lake City, UT 84112, USA
[3] Department of Biology, University of Utah, Salt Lake City, UT 84112, USA

[*] To whom correspondence should be addressed. Sheila S. David Tel: +1 530 752 4280; Email: david@chem.ucdavis.edu
Correspondence may also be addressed to Martin P. Horvath Tel: +1 801 891 3477, Email: martin.horvath@biology.utah.edu

[†] These authors contributed equally to the paper as first authors.

Supplementary Methods

Synthesis and Purification of Oligonucleotides

The 1N transition-state analog, (3*R*,4*R*)-3-hydroxy-4-(hydroxymethyl) pyrrolidin-1-ium, and its phosphoramidite were synthesized according to published procedures (1). 8-oxo-dG-cyanoethyl (CE), dSpacer CE phosphoramidites and abasic II phosphoramidite were purchased from Glen Research. The 1N, OG and THF-containing oligonucleotides were synthesized using the appropriate phosphoramidite monomers at the University of Utah core facility. OG-containing DNA oligonucleotides were cleaved and deprotected with ammonium hydroxide containing 0.25M β-mercapto-ethanol (β-ME). for 17 hours at 55 °C to avoid oxidation of the OG site. All DNA oligonucleotides were purified by 20 % polyacrylamide denaturing gel or by HPLC using standard protocols (2) and desalted with a Sep-Pak C18 desalting cartridge (Waters). The integrity of each DNA was confirmed by electrospray ionization mass spectrometry or MALDI at UC Davis Campus Mass Spectrometry Facilities (CMSF). The following sequences were used:

Duplex 1 (5'-TGTCCA**X**GTCT-3':3'-CAGGT**Y**CAGAA-5'), Duplex 2 (5'-AGCTC**X**TGACT-3':3'-GCTCGAG**Y** ACTGACG-5'), Duplex 3 (5'-CTGTAACGGGAGCT**X**GTGGCTCCATGATCG-3':3'-GACATTGCCCTCGA**Y** CACCGAGGTACTAGC-5'), and Duplex 4 (5'-GCGTCCA**X**GTCTACC-3':3'-FAM-CGCAGGT**Y**CAGATG G-5'), where X is either A, FA, AP, 1N, AP-TBDMS or THF, Y is either OG or G, and FAM is fluorescein. Duplex 4 oligonucleotides containing FA, THF, OG, and G central nucleotides were synthesized by IDT.

Oligonucleotides containing synthetic AP sites were cleaved from the solid support and all protecting groups were removed except the tert-butyldimethylsilyl (TBDMS) group at the AP site. After HPLC purification, the abasic DNA (AP-TBDMS) strands were 5'-end labeled with γ-^{32}P-ATP using T4 Polynucleotide Kinase (NEB). Radiolabeled AP-TBDMS oligomers were dried and then deprotected using 25 µL of 80% acetic acid. The reaction was incubated at room temperature for 30 minutes, followed by the addition of one equivalent of water. After four hours at room temperature, reactions were quenched with two equivalents of 2 M triethylammonium acetate (TEAA) buffer, pH 7. The oligonucleotides were purified with Microspin G-25 columns that had been pre-equilibrated with 20 mM Tris buffer, pH 7.6 and then annealed to complementary OG-containing sequences at 150 mM NaCl for immediate use.

Preparation of enzymes

WT *Ec* MutY was prepared as previously described (3) and stored at -80 ºC for use in binding and kinetic assays. N-terminal His_6-tagged WT and Y126F *Gs* MutY in pET28a expression vectors were overexpressed at 30 ºC in BL21(DE3) Rosetta II *E. coli* (Novagen) along with the proteins involved in iron-sulfur cluster assembly encoded by the pRKISC plasmid (4). *Gs* MutY was purified according to the previously reported procedure (5), which includes Ni^{2+}-NTA (Qiagen) and MonoQ (GE Healthcare) chromatography steps. The His-tag was removed by treatment with Thrombin (Novagen) and passage over Ni^{2+}-NTA resin, followed by gel-filtration on a Superdex 75 or 200 column (GE Healthcare) pre-equilibrated with 20 mM Tris pH 8.0, 150 mM NaCl, and 5 mM β-ME. Pure *Gs* MutY fractions were

combined and stored at 4 ºC for crystallization and methanolysis reactions or mixed with glycerol to 50 % and stored at -80 ºC for use in binding and kinetic assays. Site-directed mutagenesis utilized the Quik-change mutagenesis kit and protocol from Stratagene. Enzyme concentrations used in binding and kinetic assays are corrected for active fraction determined from active site titrations (6).

Adenine glycosylase assays

Glycosylase assay buffer conditions were 20 mM Tris-HCl pH 7.6, 10 mM EDTA, 0.1 mg/mL BSA, and 30 mM NaCl at 60°C or 25 °C. The rates under single-turnover conditions for WT *Gs* MutY with an OG:A substrate at 60°C were too fast to be measured manually and were determined using a rapid quenched flow instrument (Kintek) (7,8). In experiments monitored by gel eletrophoresis, the methanolysis products was inferred based on reduction of the amount of hydrolysis product in the presence of methanol. These reactions were performed using 20 nM (5% radiolabeled) OG:A-containing Duplex 2 and 100 nM active *Gs* MutY with or without added methanol. Specifially, methanolysis reactions (and relevant controls) were performed in 20% methanol, with 20 mM Tris HCl pH 7.6, 10 mM EDTA, 0.1 mg/mL BSA, and 30 mM NaCl, and at 25°C, to prevent loss of methanol during reaction conditions. Reactions were quenched to a final concentration of 0.2 M NaOH and heated for 5 minutes at 90°C. The amount of hydrolysis product (AP site) was detected by the amount of strand scission observed. Gels quantitation and appropriate fitting of the data to extract rate constants k_3 and k_2 (Fig. S3A), as well as correlation between amplitude of the burst phase to active site concentration are identical to that reported previously for *Ec* MutY (6). Gels were quantified using storage phosphor autoradiography with a Typhoon scanner and ImageQuant software (GE Healthcare). Data from the quantification were fit to the appropriate equations using Grafit 5 (Erithicus Software). In all cases, data are reported as the average of at least three experiments and the error reported as the standard deviation of the sample set.

Electrophoretic mobility shift assays (EMSA)

Experiments were performed similar to previous reports using DNA duplexes that were 5'-end labeled using γ-^{32}P-ATP on the A-, 1N-, or THF-containing strand as appropriate (9). Final reaction buffer components were 20 mM Tris-HCl pH 7.6, 100 mM NaCl, 20 % glycerol, 5 mM EDTA, 0.1 mg/mL BSA, and 0.5 mM DTT. Reactions were performed at 25 °C, with DNA duplex concentrations of 1 or 5 pM, and enzyme concentrations ranging from 10 pM to 300 nM (depending on relevant K_d). After electrophoresis, the gel was imaged and quantitated in a manner similar to that described for the glycosylase assays. The data obtained were fit to the one-site binding equation, $F_{DNA} = Cap * E / (K_d + E)$, where F_{DNA} represents the fraction of bound DNA, E the total enzyme concentration, K_d is the dissociate constant for the DNA-MutY complex, and Cap represents the overall capacity.

Fluorescence polarization binding assays

Fluorescence polarization experiments were performed with DNA duplexes (Duplex 4) harbouring a 3'-FAM label on the OG- or G-containing strand. Reactions contained 1 nM DNA and 0 to 200 nM GsMutY in assay buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 5% glycerol, 0.1 mg/mL BSA, 1 mM DTT, and 1 mM EDTA). Data were collected in a 384-well plate (Greiner Bio-One #784900) at room temperature with a BMG Labtech Clariostar multimode plate reader. Data were analyzed in Prism (Graphpad) using a one-site binding model.

Crystallization and x-ray diffraction data collection

Wild-type *Gs* MutY bound to Duplex 1 containing OG:1N base pairs was crystallized in hanging drops at room temperature with use of a microseeding technique (10). Crystallization well solutions contained 100 mM Tris pH 8.5, 14 % (w/v) PEG 4000, 500 mM calcium acetate, 5 % (v/v) ethylene glycol, and 5 mM β-ME, with some reactions supplemented with 2.5 mM adenine. Crystals appeared overnight as hexagonal rods with a dark golden color as expected for proteins with an iron-sulfur cluster. Crystals were briefly rinsed in cryoprotectant solutions, which were matched in composition with the well solution except for increased 15% (v/v) ethylene glycol, and flash frozen in liquid nitrogen. X-ray diffraction data were collected at 100 K at the Advanced Light Source SIBYLS 12.3.1 beamline with synchrotron radiation tuned to 1.11583 Å (11).

Data processing and model refinement

Data processing and model refinement were accomplished in two stages. In the first stage, diffraction data were indexed, scaled and merged using HKL2000 (12). Phases were determined by molecular replacement with the LRC structure as a search model (PDB 1RRQ) using CNS and PHENIX software (13,14). Rounds of refinement (energy minimization, torsion-angle simulated annealing and restrained individual temperature-factor optimization) performed using CNS (13) and PHENIX (14) were alternated with rounds of model adjustment guided by sigma-weighted $2|F_O| - |F_C|$, $|F_O| - |F_C|$, and simulated annealing composite omit maps in O (15). During this stage of refinement, stereochemical restraints describing the 1N residue had imposed a flat ring structure. The structural model refined with use of data to the 2.3-Å resolution limit obtained at this stage was deposited with the Protein Data Bank with PDB ID 3fsq.

In an effort to provide a more realistic structural model for the 1N residue we initiated a second stage of data processing and model building. Realizing that the most valuable data may have been excluded during the first stage of data processing, the original diffraction data were reprocessed and scaled with XDS and XSCALE (16,17), which allowed inclusion of data to the 2.2-Å resolution limit (Table S3). Intensities were converted to structure factor amplitudes by the method of French & Wilson (18). Of the ~4000 new reflections, 5% were added to the test set. Special care was taken to ensure that reflections reserved for the cross-validation test set in this second stage of model refinement had not previously been included in the working set during the first stage of model refinement. Model refinement with phase extension proceeded along three parallel paths distinguished by different stereochemical restraints describing the 1N group so as to maintain a sugar pucker with 2'-endo, 3'-endo or 1'-exo character. Examples of 1N residues in DNA can be found in the protein data bank and these have either 2'-endo or 1'-exo conformations; the Cambridge Structural Database (CSD) has more examples of 1N groups and some of these show a 3'-endo conformation.

The first “discovery” maps showed more prominent negative and positive electron density features in the vicinity of the 1N group modelled with 3'-endo sugar pucker, meaning that 1N probably adopts this 3'-endo conformation less frequently than 2'-endo or 1'-exo conformations. After several rounds of refinement in PHENIX (14) and model building in COOT (19,20), electron density maps and structures were again compared. The structural model with 3'-endo was clearly deficient with a lower cross correlation coefficient and stronger positive and negative features. The 2'-endo and 1'-exo conformations both fit electron density nearly equally well; however, the 1'-exo conformation was judged to fit maps slightly better. Model bias was a concern, and for this reason maps included simulated annealing composite omit maps (21). The 1'-exo and 2'-endo models were cross-examined with maps calculated for the other structure (1'-exo structure against maps calculated using phases from the 2'-endo model and vice-versa), and the 1'-exo conformation out-performed the 2'-endo conformation, if only slightly, with this test too. The 1'-exo conformation also featured more ideal hydrogen-bonding geometry for interaction with nearby residues. Rfree and Rwork values along with other validation measures are reported in Table S3. The structural model refined with data to the 2.2-Å resolution limit with 1'-exo conformation for 1N was deposited with the Protein Data Bank with the PDB ID 5dpk, which now supersedes the previously deposited structure. It should be noted that NRI (the residue matching 1N found in other PDB entries) is in the deprotonated state with trigonal planar geometry at the N1' atom. The 1N residue in our TSAC structure is for this reason specified with a different three-letter code: NR1. NR1 is in the protonated state with tetrahedral geometry at the N1' atom as expected for a secondary amine at pH 8.5. Stereochemical restraints describing the 1'-exo, 2'-endo, and 3'-endo conformations for NR1 are available upon request from M.P.H.

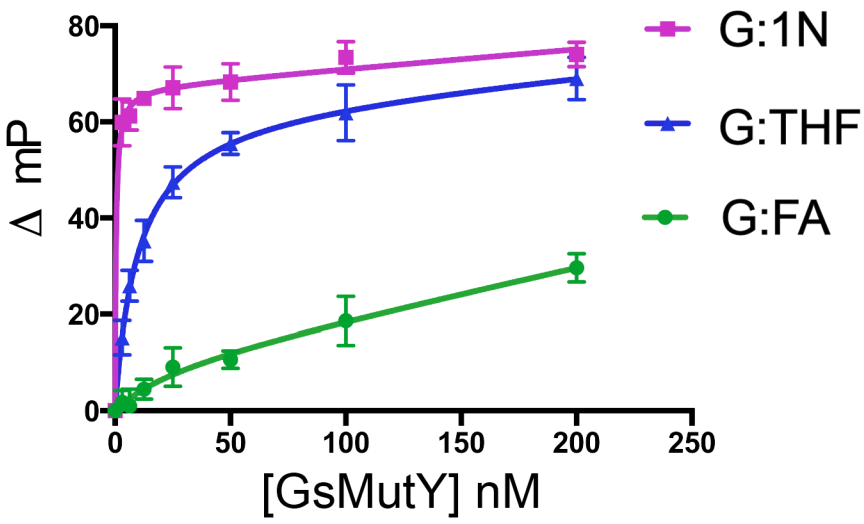
Structure figures were prepared using Pymol (22).

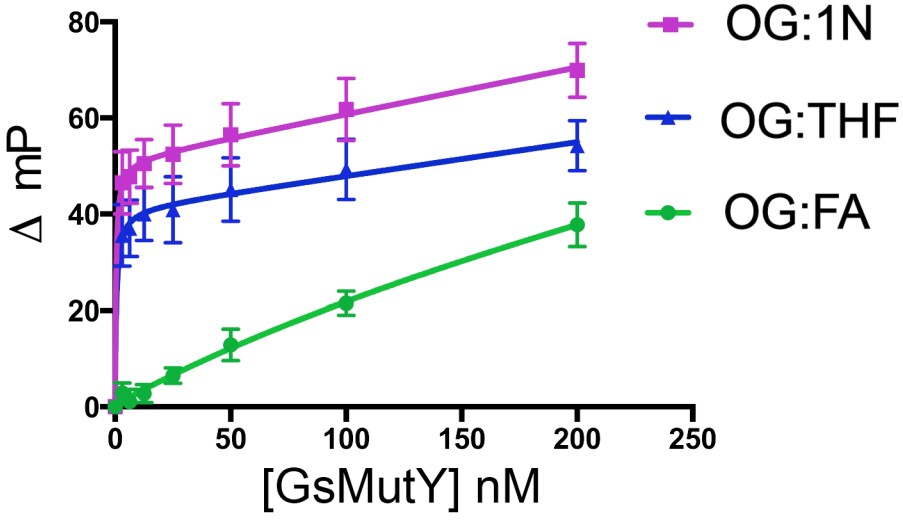
NMR sample preparation, pulse-sequences and assignments

HPLC-purified dROMe-containing oligomers were desalted using Waters Sep-Pak C18 cartridges and pooled from multiple *Gs* MutY methanolysis reactions for NMR analysis. Dried samples were dissolved using 100% D_2O with 50 mM sodium phosphate buffer, pH 7.2. All product spectra were acquired at 50 µM sample concentration using a 5 mm Shigemi Tube Set (Wilmad-LabGlass), with an 8 mm bottom and susceptibility matched to D_2O for Bruker spectrophotometers. ^1H spectra were recorded on an 18.8 T Bruker Avance spectrometer at 303 K with a 5 mm CPTCI probe operating at a proton frequency of 800.15 MHz and data were processed using TopSpin and MestReNova software.

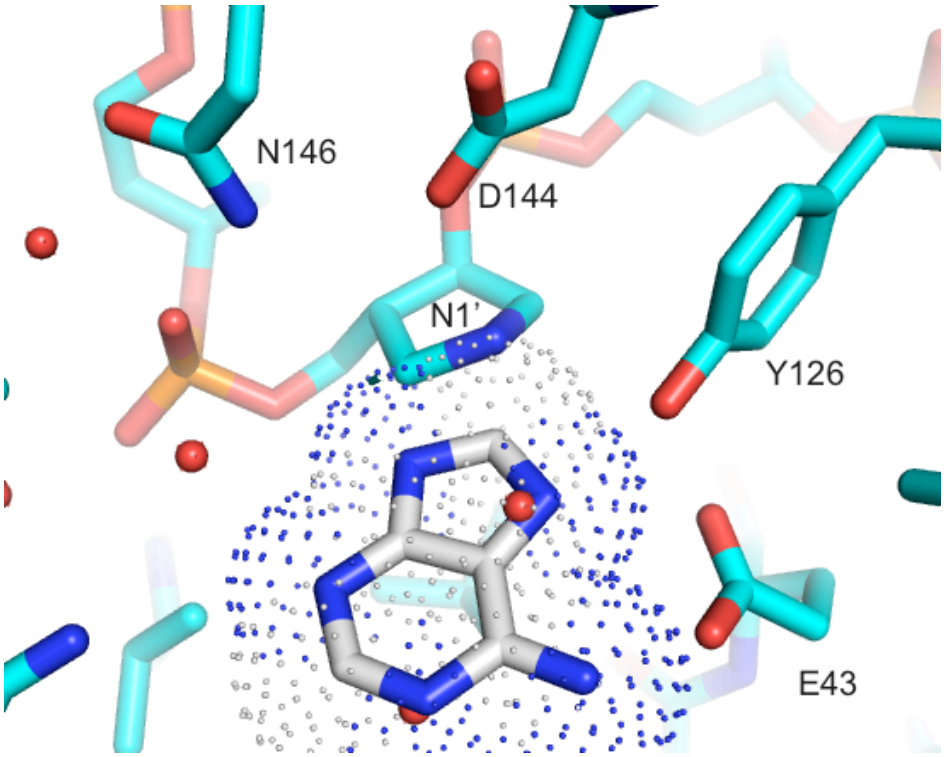
1D NMR spectral data were collected using a *zgpr* pulse sequence, with 64 scans 16k acquisition size, and 2.5 s relaxation delay. Data were processed with 0.3 Hz exponential apodization, zero filed to 64k data points, and a third-order polynomial baseline correction. The 1D selective TOCSY spectrum was acquired using a *selmlgp* pulse sequence, with 1024 scans, 16k acquisition size, and 3.0 s relaxation delay. Data were processed with 5 Hz exponential apodization, zero filed to 64k data points, and a third-order polynomial baseline correction. The 2D NOESY spectrum was collected using a *noesyphpr* pulse sequence with 128 scans, 2048 and 256 acquisition size in the t2 and t1 dimensions, and 2.5 s relaxation delay. NOESY data were processed with sine square 90 apodization, zero filed t1 dimension to 1024 data points, and noise was reduced with a third-order polynomial baseline correction and Savitzky-Golay smoothing applied to both dimensions. The 2D COSY spectrum was collected using a *cosygpprqf* pulse sequence with 128 scans, 2048 and 128 acquisition size in the t2 and t1 dimensions, and 2.5 s relaxation delay. COSY data were processed with sine bell 0 apodization, zero filed t1 dimension to 1024 data points, and noise was reduced with a third-order polynomial baseline correction and Savitzky-Golay smoothing applied to both dimensions.

Supplementary Figure 1: Fluorescence polarization binding assays with Gs MutY and 15 bp Duplex 4. Experiments were performed by titration of Gs MutY with DNA duplexes harbouring the indicated centrally located base pair and a 3'-FAM label on the OG- or G-containing strand. Relative dissociation constants are as follows: G:THF, K_d = 9.8 ± 1.6 nM; G:1N, K_d < 0.5 nM, G:FA > 200 nM, OG:THF, < 0. 5 nM, OG:1N, K_d < 0.5 nM; OG:FA, K_d > 200 nM. With the requirement for 1 nM concentration of DNA in order to detect polarization, these relative dissociation constants are in many cases best estimates of the upper limit of the true dissociation constants. Nevertheless, binding reactions followed by polarization changes indicate enhanced binding for DNAs containing the TS analog 1N compared with product (THF) and substrate (FA) analogs.

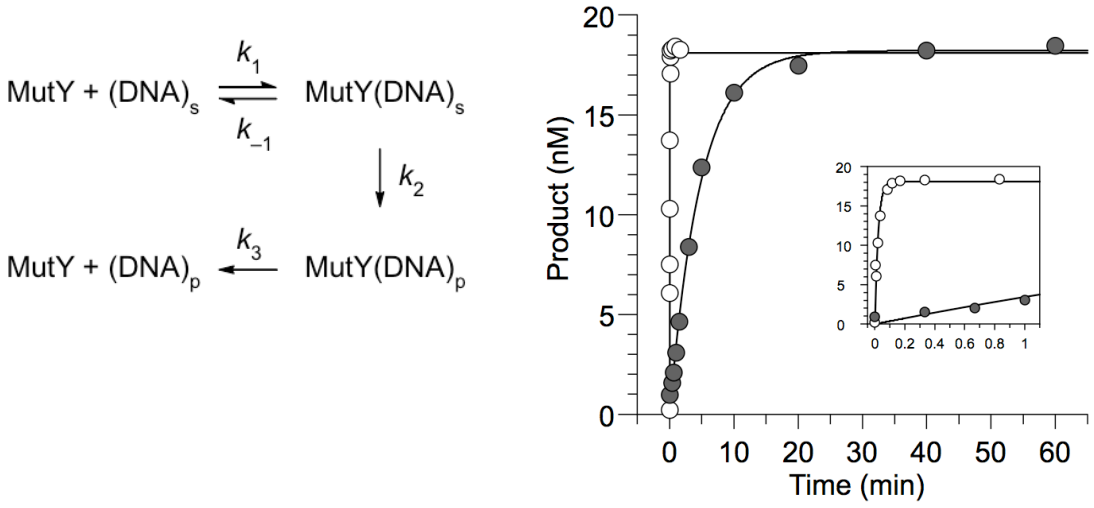




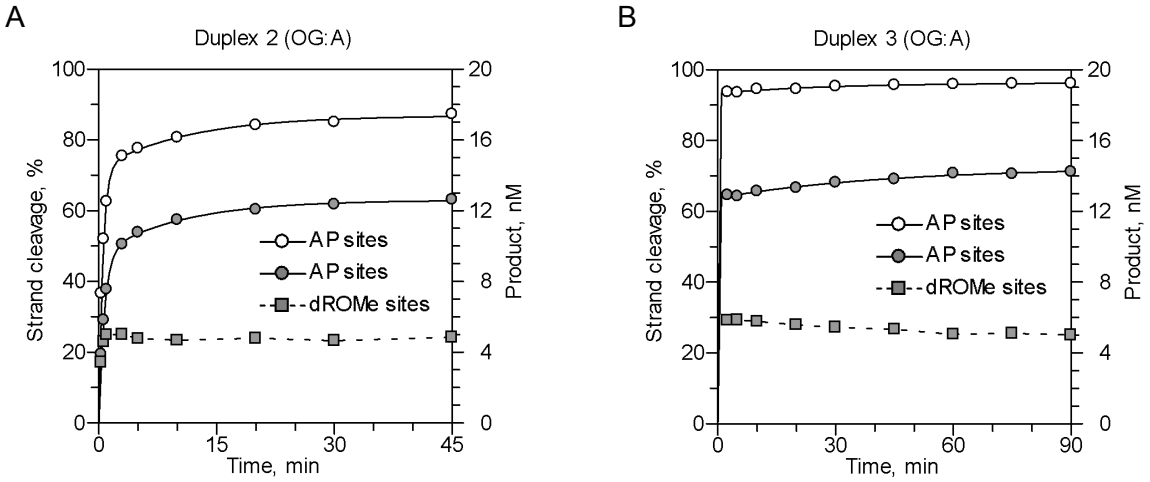
Supplementary Figure 2: Superposition of adenine from rAP structure (PDB ID IVRL) onto TSAC illustrates steric clashes between 1N and adenine. Distances: from C6' (which is analogous to C4' in a normal nucleotide) of 1N to N7 of adenine (1.6 Å), from N1' of 1N to C8 of adenine (1.7 Å)



Supplementary Figure 3. Y126F *Gs* MutY mutant is catalytically compromised. A) Minimal kinetic scheme used for analysis of adenine removal activity of *Gs* MutY. B) Representative data from single-turnover experiments at 60 °C with 120 nM active enzyme and 20 nM OG:A containing DNA for WT *Gs* MutY (open circles ○) and Y126F *Gs* MutY (black filled circles ●). The inset highlights the data for the first minute to allow visualization of the dramatic difference in rate of processing of substrate between WT and Y126F *Gs* MutY. Rate constants derived from single-exponential curve fitting are as follows: WT k_2 = 54 ± 4 (min^{-1}), Y126F k_2 = 0.21 ± 0.01 (min^{-1}).



Supplementary Figure 4. Influence of methanol in Gs MutY-catalyzed hydrolysis of OG:A duplex substrates monitored using gel-based AP-site cleavage assays. Formation of base-labile AP sites in solutions in the absence or presence of methanol with solutions containing 20 nM Duplex 2 (A) or Duplex 3 (B) and 100 nM *Gs* MutY. The dROMe-containing product (gray squares) is indirectly measured by the difference between observed strand-scission for reactions without methanol (white circles) and with 20% methanol (gray circles). This difference indicated that up to 5 nM of methanolysis product was formed within the first minute of the reaction. Results averaged from several experiments with both duplexes in the presence of 20% methanol show a decrease of 25-30% of AP product, indirectly providing the amount of dROMe-containing product. In representative data shown in panels A and B, the data fit best with a two exponential equation: $y = Ae^{-k_{fast}t} + Be^{-k_{slow}t}$. Parameters from the curve fitting are compiled in panel C. The k_{fast} rate constant represents initial product formation with a large capacity, *A*. With the 30-bp Duplex 3, the major component of the observed rate is attributed to k_{fast}. The larger contribution of k_{slow} (but still considerably less than k_{fast}) in the reaction with Duplex 2 (15:11 nts) suggest that this slower rate is related to the stability of the duplex.

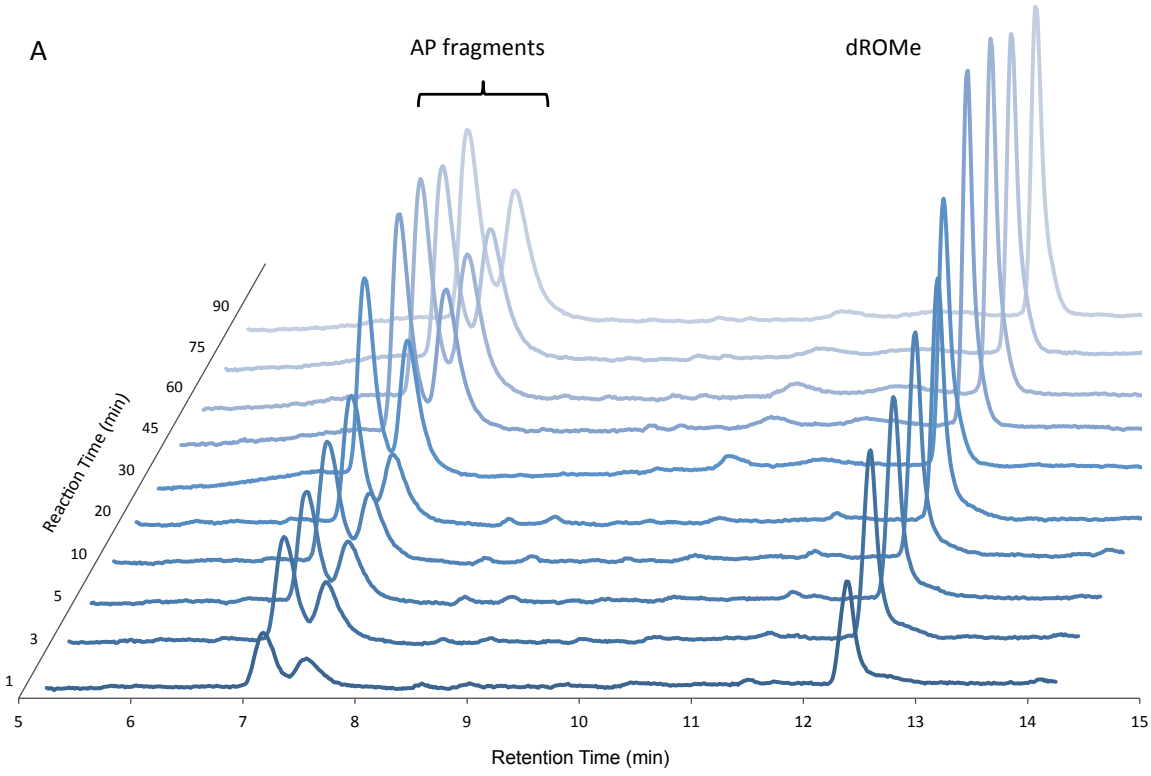


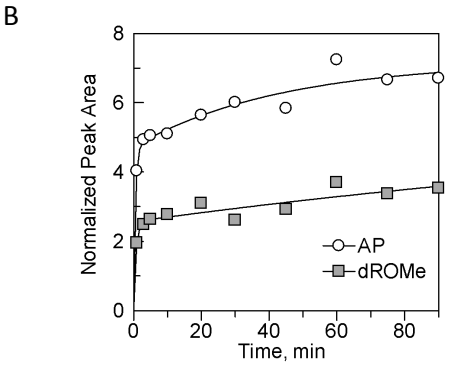
C

	Duplex 2 (OG:A)		Duplex 3 (OG:A)	
	No MeOH (white)	20% MeOH (gray)	No MeOH (white)	20% MeOH (gray)
A (nM)	14.3	9.5	18.7	12.7
k_{fast} (min^{-1})	2.0	1.4	> 2.3[a]	> 2.3[a]
B (nM)	3.0	3.1	0.6	1.7
k_{slow} (min^{-1})	0.09	0.10	0.03	0.02

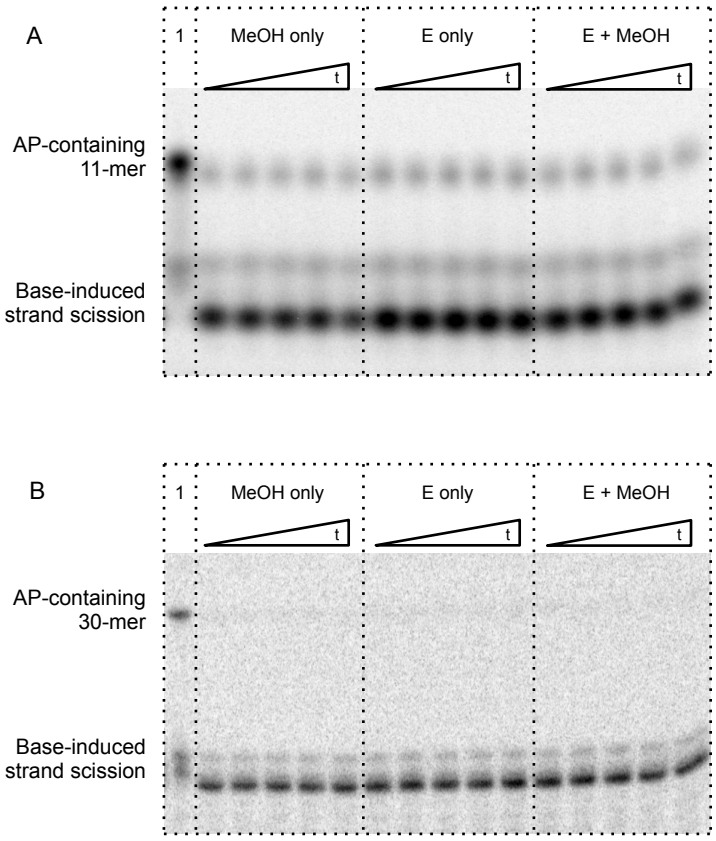
[a] k_{fast} was estimated from the first time point, which surpassed 90% completion within 1 minute.

Supplementary Figure 5: **The simultaneous production of Gs MutY catalyzed methanolysis and hydrolysis products monitored using anion-exchange HPLC**. A) *Gs* MutY methanolysis reactions performed with Duplex 2 were quenched at various time points, ranging from 1 minute to 90 minutes, and the reaction products were resolved by anion-exchange HPLC. B) The peaks that correspond to hydrolysis (AP) and methanolysis (dROMe) products were normalized to the initial baseline and plotted against time. Note: that the reaction is >50% complete at the first time point of 1 min which is consistent with data derived from the gel-based assay. The ratio of dROMe to AP products was estimated by the amplitude of the burst phase, and indicate that 33% of the reaction proceeds via methanolysis. This ratio corresponds well with the amount of methanolysis product of 25-30% determined in the gel-based assays (Supplementary Fig. 4).





Supplementary Figure 6. *Gs* MutY does not catalyze esterification of abasic site-containing DNA. Radiolabeled duplex DNA containing a synthetically prepared abasic site, AP, opposite OG was treated with 20% methanol and active *Gs* MutY at 25 °C. Reactions were quenched at 5, 10, 20, 40, and 60 minutes with sodium hydroxide (0.2M) and separated by PAGE to reveal conversion of base-sensitive AP sites to base-resistant dROMe adducts. All reactions of MutY with the AP-containing Duplex 2 (A) or Duplex 3 (B) produced equivalent amounts of length 11-mer strand resulting from cleavage of the AP site. Lane 1 was not treated with sodium as an indicator of intact AP-containing DNA at the beginning of the reaction.



Supplementary Table 1. Gs MutY exhibits high affinity for transition state mimic (1N) containing duplex DNA.[a, b]

	Dissociation constants (K_d, nM)	
X\Y	1N	THF
OG	<0.005[c]	<0.005[c]
G	<0.005[c,d]	0.2 ± 0.1[d]

[a] K_d values were measured with WT Gs MutY with the transition state mimic (1N) and product analog (THF) in 30 bp duplex 3 at described in the methods section. Consult the methods section for details of sequences and duplex synthesis.

[b] We were unable to observe a shifted band in EMSA experiments for the GsMutY/OG:FA-duplex complex. Gs MutY does not behave well in EMSA experiments and therefore only extrememly tight complexes persist during electrophoresis. Due to this issue, we also include K_d values for Ec MutY (Table S2).

[c] With these duplexes, the DNA was completely bound at all enzyme concentrations (0.010-300 nM), thus the upper limit for the K_d reported is based on the DNA duplex concentration.

[d] The data fits better to a two-site binding isotherm. The K_d value reported is the one with the higher capacity. The second K_d value obtained with a two-site fitting is ~ 5 nM and corresponds to the K_d observed with a 30 nt ss DNA containing the TS/product mimic (23).

Supplementary Table 2. Ec MutY exhibits high affinity for transition state mimic (1N) containing duplex DNA.[a]

	Dissociation constants (K_d, nM)			
X\Y	C	FA	1N	THF
OG	26 ± 10[b]	0.8 ± 0.4[c]	< 0.005[d]	< 0.005[d]
G	150 ± 60[b]	7 ± 3[c]	< 0.005[d]	2 ± 1

[a] K_d values were measured with WT Ec MutY with the transition state mimic (1N) and product analog (THF) in 30 bp duplex 3. Consult the methods section for details of sequences and duplex synthesis.
[b] We previously reported K_d values with the same duplex containing C at the position of 1N (24).
[c] We previously reported K_d values with substrate analog (FA) in same duplex (8,24).
[d] With these duplexes, the DNA was completely bound at all enzyme (0.010-300 nM), thus the upper limit for the K_d reported is based on the DNA duplex concentration which was 5 pM.

Supplementary Information Table 3. Data collection and model refinement statistics

Data collection	
Wavelength (Å)	1.1158
Resolution range (Å) *	41.3 – 2.20 (2.26 – 2.20)
Space group	$P2_12_12_1$
Unit cell (Å)	37.85, 86.62, 140.77
Total reflections	170560 (12449)
Unique reflections	24104 (1695)
Completeness (%)	99.0 (96.8)
Mean I/sigma(I)	18.5 (2.8)
Wilson B-factor ($Å^2$)	47.9
R-merge	0.071 (0.764)
R-measure ‡	0.076 (0.823)
CC1/2 (%)	99.9 (75.1)

Model refinement	
Resolution range (Å) *	41.3 – 2.20 (2.30 – 2.20)
Reflections in working set	22941 (2752)
Reflections used in test set	1163 (153)
R-work	0.242 (0.369)
R-free	0.257 (0.391)
Number of non-hydrogen atoms (average B value, $Å^2$)	3255 (50.7)
Protein	2754 (52.6)
DNA	421 (40.0)
NR1	11 (24.0)
8OG	23 (27.1)
Metal ions (2 Ca^{2+}, 4Fe-4S)	10 (56.7)
Solvent	70 (38.8)
RMSD for bond lengths (Å)	0.0042
RMSD for bond angles (°)	0.76
Ramachandran favored (%)	96.5
Ramachandran allowed (%)	3.5
Ramachandran outliers (%)	0
Rotamer outliers (%)	7.8
Clash score	8.0

* Values in parentheses are for the highest resolution shell.

‡ R-measure is the redundancy independent metric developed by Diederichs and Karplus (25).

Supplementary Table 4. Sugar chemical shifts for dROMe-containing product DNA (ppm).

Nucleotide	Proton H1'	H2'	H2"	H3'	H4'	H5'/H5"
A1	6.11	2.32	2.54	-	4.20	-
G2	5.97	2.74	2.65	5.02	4.40	4.12/4.12
C3	6.21	2.21	2.51	-	4.32	-
T4[a]	6.18	2.18	2.47	-	-	-
C5	6.27	2.31	2.51	-	4.31	-
X6	5.15	2.15	2.21	4.57	4.19	-
T7	5.88	1.82	2.16	4.70	4.02	-
G8	5.63	2.61	2.61	4.93	-	3.98/3.98
A9	6.30	2.79	2.73	5.00	4.42	4.16/4.16
C10[a]	6.22	2.27	2.42	-	-	-
T11[a]	6.13	2.27	2.52	-	-	-

Proton signals were identified using 2D COSY and 2D NOESY experiments. Resonances that could not be unambiguously identified are marked with a dash, -. Six H3' resonances were buried under the residual water resonance at 303K. Of those six, only A1, C3 and C5 had NOEs for identifying H4' or H5'/H5" protons.

[a] Three sugar spin systems were identified that are assigned to T4, C10, or T11. A lack of internucleotide NOEs prevented assigning these systems to specific positions on the oligonucleotide. Importantly, all spin systems had clear intranucleotide NOEs used to distinguish the canonical nucleotides A, G, C and T from the O-glycoside lesion of interest. The affected resonances are underlined.

Supplementary Table 5. Nucleobase chemical shifts for dROMe-containing product DNA (ppm).

	Nucleotide										
proton	A1	G2	C3	T4	C5	X6	T7	G8	A9	C10	T11
H8/H6	8.04	7.96	7.67	7.59	7.82		7.36	7.86	8.3	7.62	7.62
H2	8.06								8.45		
H5			5.78		6.01					5.69	
-CH_3[a]				1.80			1.74				1.82
-OCH_3						3.32					

Resonances that do not occur for the particular nucleotide are shaded gray.
[a] T4 and T11 could not be distinguished from each other due to lack of internucleotide NOEs. The affected resonances are underlined.

References

1. Chu, A.M., Fettinger, J.C. and David, S.S. (2011) Profiling base excision repair glycosylases with synthesized transition state analogs. *Bioorganic & Medicinal Chemistry Letters*, **21**, 4969-4972.
2. Sambrook, J. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Second ed. Cold Spring Harbor Press, Cold Spring Harbor.
3. Chmiel, N.H., Golinelli, M.-P., Francis, A.W. and David, S.S. (2001) Efficient recognition of substrates and substrate analogs by the adenine glycosylase MutY requires the C-terminal domain. *Nucleic Acids Res.*, **29**, 553-564.
4. Tokumoto, U. and Takahashi, Y. (2001) Genetic Analysis of the isc operon in Escherichia coli involved in biogenesis of cellular iron-sulfur proteins. *J. Biochemistry*, **130**, 63-71.
5. Fromme, J.C., Banerjee, A., Huang, S.J. and Verdine, G.L. (2004) Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature*, **427**, 652-656.
6. Porello, S.L., Leyes, A.E. and David, S.S. (1998) Single-turnover and pre-steady-state kinetics of the reaction of the adenine glycosylase MutY with mismatch-containing DNA substrates. *Biochemistry*, **37**, 14756-14764.
7. Chepanoske, C.L., Lukianova, O.L., Lombard, M., Golinelli-Cohen, M.-P. and David, S.S. (2004) A residue in MutY important for catalysis indentified by photo-crosslinking and mass spectrometry. *Biochemistry*, **43**, 651-662.
8. Livingston, A.L., Kundu, S., Henderson Pozzi, M., Anderson, D.W. and David, S.S. (2005) Insight into the roles of tyrosine 82 and glycine 253 in the Escherichia coli adenine glycosylase MutY. *Biochemistry*, **44**, 14179-14190.
9. Porello, S.L., Williams, S.D., Kuhn, H., Michaels, M.L. and David, S.S. (1996) Specific recognition of substrate analogs by the DNA mismatch repair enzyme MutY. *J. Am. Chem. Soc.*, **118**, 10684-10692.
10. Bergfors, T. (2003) Seeds to crystals. *J Struct Biol*, **142**, 66-76.
11. Classen, S., Hura, G.L., Holton, J.M., Rambo, R.P., Rodic, I., McGuire, P.J., Dyer, K., Hammel, M., Meigs, G., Frankel, K.A. *et al.* (2013) Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *Journal of applied crystallography*, **46**, 1-13.

12. Otwinowski, Z. and Minor, W. (1997) In Carter, C. W. and Sweet, R. M. (eds.), *Methods in Enzymology, Macromolecular Crystallography Part A*. Academic Press, New York, Vol. 276, pp. 307-325.
13. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, **54**, 905-921.
14. Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K. and Terwilliger, T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr*, **58**, 1948-1954.
15. Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A*, **47 (Pt 2)**, 110-119.
16. Kabsch, W. (2010) Xds. *Acta Crystallogr D Biol Crystallogr*, **66**, 125-132.
17. Diederichs, K., McSweeney, S. and Ravelli, R.B. (2003) Zero-dose extrapolation as part of macromolecular synchrotron data reduction. *Acta Crystallogr D Biol Crystallogr*, **59**, 903-909.
18. French, S. and Wilson, K. (1978) On the treatment of negative intensity observations. *Acta Crystallogr A*, **A34**, 517-525.
19. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, **66**, 486-501.
20. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **60**, 2126-2132.
21. Hodel, A., Kim, S.-H. and Brünger, A.T. (1992) Model bias in macromolecular crystal structures. *Acta Crystallogr*, **A48**, 851-858.
22. DeLano, W.L. (2002) The PyMOL Molecular Graphics System.
23. Chu, A.M. (2011), University of California, Davis, Davis CA.
24. Chepanoske, C.L., Porello, S.L., Fujiwara, T., Sugiyama, H. and David, S.S. (1999) Substrate recognition by Escherichia coli MutY using substrate analogs. *Nucleic Acids Res*, **27**, 3197-3204.
25. Diederichs, K. and Karplus, P.A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol*, **4**, 269-275.