Nucleic Acids Research, 2004, Vol. 32, No. 22 **6531–6539**
doi:10.1093/nar/gkh995

Pa-AGOG, the founding member of a new family of archaeal 8-oxoguanine DNA-glycosylases

Alessandro A. Sartori, Gondichatnahalli M. Lingaraju[1], Peter Hunziker[2], Fritz K. Winkler[1] and Josef Jiricny*

Institute of Molecular Cancer Research, University of Zürich, August Forel-Strasse 7, CH-8008 Zürich, Switzerland, [1]Biomolecular Research, Paul Scherrer-Institut, CH-5232 Villigen, Switzerland and [2]Institute of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received October 17, 2004; Revised and Accepted November 23, 2004

ABSTRACT

Oxidative damage represents a major threat to genomic stability, as the major product of DNA oxidation, 8-oxoguanine (GO), frequently mispairs with adenine during replication. In order to prevent these mutagenic events, organisms have evolved GO-DNA glycosylases that remove this oxidized base from DNA. We were interested to find out how GO is processed in the hyperthermophilic archaeon *Pyrobaculum aerophilum*, which lives at temperatures around 100°C. To this end, we searched its genome for open reading frames (ORFs) bearing the principal hallmark of GO-DNA glycosylases: a helix–hairpin–helix motif and a glycine/proline-rich sequence followed by an absolutely conserved aspartate (HhH-GPD motif). Interestingly, although the *P.aerophilum* genome encodes three such ORFs, none of these encodes the potent GO-processing activity detected in *P.aerophilum* extracts. Fractionation of the extracts, followed by analysis of the active fractions by denaturing polyacrylamide gel electrophoresis, showed that the GO-processing enzyme has a molecular size of ~30 kDa. Mass spectrometric analysis of proteins in this size range identified several peptides originating from *P.aerophilum* ORF PAE2237. We now show that PAE2237 encodes AGOG (Archaeal GO-Glycosylase), the founding member of a new family of DNA glycosylases, which can remove GO from single- and double-stranded substrates with great efficiency.

INTRODUCTION

Reactive oxygen species (ROS) arising as by-products of normal metabolism and through oxidative stress pose a considerable threat to genomic integrity. Of the numerous base modifications identified to date, 8-oxo-7,8-dihydro-2′-deoxyguanosine (GO) is the most abundant. In *anti*-conformation, GO forms a stable Watson–Crick base pair with C, but in *syn* it can form a stable Hoogsteen base pair with A. Thus, if unrepaired prior to replication, erroneous incorporation of dAMP opposite template GO or of 8-oxodGMP opposite template A, will give rise to G:C to T:A transversion mutations (1).

In all organisms studied to date, GO is removed from DNA predominantly by the base excision repair (BER) pathway [reviewed in (2)]. This process is initiated by 8-oxoguanine-DNA glycosylases, which cleave the N-glycosidic bond between the aberrant base and the sugar-phosphate backbone to generate an apurinic (AP) site. Some DNA glycosylases possess also an intrinsic AP lyase activity, which cleaves the phosphodiester bond 3′ from the AP site by β- or β,δ-elimination.

In *Escherichia coli*, the mutagenic effects of guanine oxidation are countered by the GO-system, which consists of three enzymes: (i) MutM (also known as Fpg), a DNA glycosylase/lyase that excises GO and its ring-opened form, formamidopyrimidine (Fapy), from base pairs with C; (ii) MutY, a monofunctional DNA glycosylase that excises adenines from A/GO mispairs; and (iii) MutT, a 8-oxo-dGTPase that prevents the incorporation of 8-oxo-dGMP into nascent DNA (1). MutY, MutT and, most recently, also MutM (3) orthologues have been identified also in the human genome. However, processing of GO in eukaryotes is mediated primarily by GO-glycosylases such as yOGG1 in *Saccharomyces cerevisiae* (4,5) and hOGG1 in humans (6), proteins with an associated AP-lyase activity, which belong to a superfamily of repair enzymes that share a common helix–hairpin–helix DNA-binding domain followed by a glycine/proline-rich stretch and an invariant aspartate (HhH-GPD motif) (4). The eukaryotic OGG1 proteins display high selectivity for GO/C pairs (7–9). Although they can also remove GO residues paired with other bases, efficient strand nicking *via* β-elimination was observed only with GO/C (4,10). A second OGG activity, OGG2, was also described; this protein has so far been identified only in yeast and acts preferentially on GO residues paired with G or A. It may have evolved to process GO/A mispairs arising through misincorporation of

*To whom correspondence should be addressed. Tel: +41 1 634 8910; Fax: +41 1 634 8904; Email: jiricny@imr.unizh.ch
Present address:
Alessandro A. Sartori, The Wellcome Trust/Cancer Research UK Gurdon Institute, Tennis Court Road, Cambridge CB2 1QR, UK

8-oxo-dGMP during replication, as *S.cerevisiae* has no MutT homologue (4,11,12).

The rates of spontaneous hydrolysis and oxidation are substantially increased at higher temperatures. We were therefore interested to find out how DNA bases damaged by these processes are repaired in organisms such as *Pyrobaculum aerophilum*, which grows optimally at 100°C, and which was shown in our earlier studies to counteract the mutagenic threat of hydrolytic DNA damage with great efficiency (13–15). Given that it is one of the most thermophilic organisms known that is capable of aerobic respiration (16), we expected it to possess also a highly efficient defense against oxidative damage. We now show that *P.aerophilum* expresses a GO-glycosylase/lyase, which is the founding member of a new family of archaeal DNA glycosylases and which is capable of removing the aberrant base from both single- and double-stranded DNA substrates.

MATERIALS AND METHODS

P.aerophilum whole-cell extracts (WCE) and purified proteins

The *P.aerophilum* WCE were described previously (17). *P.aerophilum* AP Endonuclease IV (Pa-EndoIV, nfo) was expressed and purified as described in (18), and the purified recombinant wild-type human GO-glycosylase (hOGG1) and *E.coli* Fpg proteins were a kind gift of Dr Murat Saparbaev.

Bacterial strains and expression plasmids

The *E.coli* strain XL1Blue was used in all cloning experiments and for plasmid amplifications, and the strain B834(DE3) (Novagen) was used for protein expressions. The plasmid pET28c(+) (Novagen) was used for bacterial expression of N-terminal His_6-tagged proteins.

DNA glycosylase and lyase assays

Glycosylase activity was monitored using duplexes consisting of the fluorescein-labeled (^{F}) 60mer oligo (all sequences written from 5′ to 3′) ^{F}CGGAATTCGTCTAGGTTTGAGGT**GO**-GACATCGGATCCATGGTACCTCGAGGGCAATGTCTA annealed to TAGACATTGCCCTCGAGGTACCATGGAT-CCGATGTC**X**ACCTCAAACCTAGACGAATTCCG (**X** = C, A, G or T) as described in (14). Double-stranded competitor DNA (GO/C or GO/G) consisted of unlabeled 60mer oligos of the same sequence. The assay mixtures (20 µl) contained 50 mM Tris–HCl (pH 8.0), 50 mM KCl, 1 mM EDTA, 1 mM DTT, 1 pmol of labeled DNA duplex and *P.aerophilum* WCE, chromatography fractions, or purified *P.aerophilum* proteins. Recombinant human OGG1 and *E.coli* Fpg proteins were used in a reaction buffer containing 50 mM HEPES–KOH pH 8.0, 100 mM KCl, 0.1 mg/ml BSA, 1 mM EDTA and 5 mM β-mercaptoethanol. Incubations were for 15 min at 60°C (or at 37°C for the mesophilic proteins). Assays involving extracts were terminated by the addition of 1× stop solution (0.5 mg/ml Proteinase K, 5 mM EDTA, 0.5% SDS) and incubated for a further 30 min at 37°C. To measure glycosylase activity (base release and production of AP sites) of the purified recombinant *Pa*-AGOG protein, samples were treated with 100 mM NaOH and incubated at 90°C for 10 min to cleave the DNA at the AP-site prior to precipitation. To measure lyase activity (production of backbone-cleaved DNA), the samples were ethanol-precipitated and the DNA pellets were dissolved in 90% formamide loading buffer supplemented with 50 mM $NaBH_4$ in order to prevent spontaneous hydrolytic cleavage of the labile AP sites prior to analyzing the samples by electrophoresis as described in (14). The fraction of cleaved DNA was calculated from the relative intensities of the substrate and product bands. In glycosylase assays, a correction was made for the amount of substrate cleaved in the absence of enzyme (typically 5–10% after 15 min incubation at 60°C, see e.g. lane 1 in Figure 4C).

Purification of *P.aerophilum* GO-glycosylase

A total of 200 g of *P.aerophilum* cells were resuspended in 20 mM sodium phosphate (pH 7.0), lysed by sonication and the extract was cleared by centrifugation for 30 min at 4°C using a Sorvall SS-34 rotor at 18 000 r.p.m. The WCE was supplemented with 1× 'Complete^{TM}, protease inhibitor mixture (Roche Applied Science) and 1 mM phenylmethylsulfonyl fluoride (PMSF, Sigma), and extensively dialyzed overnight at 4°C against 3 l of dialysis buffer (DB) containing 20 mM sodium phosphate (pH 7.4), 0.1 M NaCl, 1 mM EDTA, 1 mM PMSF and 2 mM DTT. The protein concentration of the WCE (Fraction I, 160 ml) was 2 mg/ml. All following purification steps were carried out at room temperature using the ÄKTA-purifier^{TM} 10 Chromatography System (Amersham Biosciences). After each chromatography step, the fractions were assayed using the GO/G 60mer as a substrate. Those with GO-glycosylase activity were pooled and used for further analysis.

Fraction I was loaded on a 5-ml HiTrap Heparin HP column (Amersham Biosciences) equilibrated with DB. The column was extensively washed with DB, and the proteins were eluted with a 30-ml linear gradient from 0.1 to 1 M NaCl. Fractions containing the major glycosylase activity, eluting from 0.5 to 0.75 M NaCl, were pooled and dialyzed overnight at 4°C against 2 l of DB (Fraction II, 25 ml @ 0.26 mg/ml). Fraction II was loaded on a 1.3-ml UNO S cation exchange column (Bio-Rad). The column was washed with DB and the proteins were eluted with a 20-ml linear gradient from 0.1 to 0.5 M NaCl. The glycosylase activity eluted between 0.3 and 0.43 M NaCl. The pooled fractions were dialyzed against 2 l of DB (Fraction III, 7 ml, 0.25 mg/ml). Fraction III was loaded on a 1-ml HiTrap Heparin HP column (Amersham Biosciences). The column was washed with DB and the proteins were eluted with a 20-ml linear gradient from 0.1 to 0.5 M NaCl. The glycosylase activity eluted between 0.32 and 0.42 M NaCl. The pooled fractions were dialyzed against 1 l of DB (Fraction IV, 5 ml, 0.2 mg/ml). Fraction IV was loaded on a 0.8-ml Mini S PE 4.6/50 column (Amersham Biosciences). The column was washed with DB and the proteins were eluted with a 30-ml linear gradient from 0.1 to 0.6 M NaCl. The glycosylase activity was in the flow-through, in the wash and in fractions 1–17. All active fractions were pooled and dialyzed overnight at 4°C against 2 l of DB containing 40 mM NaCl (Fraction V). Fraction V was loaded on a 1-ml HiTrap SP Sepharose High Performance (HP) column (Amersham Biosciences) equilibrated with DB containing 40 mM NaCl. The column was washed with DB containing 40 mM NaCl and the bound

proteins were eluted with a 20-ml linear gradient from 0.04 to 0.6 M NaCl. The major glycosylase activity was found in fractions 7–11 containing 0.23–0.37 M NaCl, with fractions 9, 10 and 11 exhibiting the highest activity (see Figure 2A). Fraction 11 was used in various test experiments, and fractions 9 and 10 were pooled (2 ml) and dialyzed overnight against 1 l of DB containing 10 mM NaCl (Fraction VI).

Renaturation of the GO-glycosylase activity

The proteins in Fraction VI were separated on a 15% SDS–polyacrylamide gel loaded as follows: lane 1, 0.5 µl of Bio-Rad SDS–PAGE molecular weight standards; lanes 2, 4, 25 µl of fraction VI; lane 3, 10 µl of Bio-Rad pre-stained SDS–PAGE molecular weight standards. After electrophoresis, lanes 1 and 2 were silver-stained and lane 4 was cut into 11 slices as indicated in Figure 2B (left panel). The glycosylase activity was de/renatured as described in (19) and the renatured samples were concentrated to ~100 µl by centrifuging at 4°C for 35 min and 7500 *g* (Sorvall SS-34 rotor) in Ultrafree Centrifugal Filter Units with Biomax membranes of 10 kDa exclusion limit (NMWL, Millipore), which were pre-washed with 0.5 ml of renaturation buffer containing 0.5 mg/ml BSA; 12 µl of the 11 concentrated fractions were tested in the nicking assay as shown in Figure 2B (right panel).

Matrix-Assisted Laser Desorption-Ionization (MALDI) mass spectrometric analysis

The proteins in fraction VI were precipitated with 20% trichloroacetic acid on ice (45 min). After centrifugation for 30 min at 4°C in an Eppendorf centrifuge (14 000 r.p.m.), the supernatant was removed and the protein pellet was washed two times with 600 µl of ice-cold acetone. The pellet was dried at room temperature and resuspended in 20 µl of 2× alkaline SDS–PAGE loading buffer (120 mM Tris–HCl, pH 8.0, 4% SDS, 200 mM DTT, 20% glycerol). After adding 20 µl of water, the sample was heated for 5 min at 95°C and the proteins were separated on a 12.5% SDS–PAGE gel. Two protein bands in the range between 32 and 26 kDa (see Figure 3A) were cut out from the Coomassie-stained gel and the fragmented slices were destained twice with 100 µl of 100 mM ammonium bicarbonate/50% acetonitrile at 32°C for a total of 1 h and then rehydrated in 30 µl of trypsin solution at 37°C overnight as described in (20). The supernatants were transferred to a new tube and the gel pieces were extracted with 20 µl of 5% formic acid/5% methanol. Of the combined supernatants, 10 µl were loaded onto a ZipTip with a C_{18} resin (Millipore) and the peptides were directly eluted onto the MALDI target with 1.3 µl of a saturated solution of α-cyano-4-hydroxycinnamic acid in 0.05% trifluoroacetic acid/50% acetonitrile. MALDI spectra were recorded on a Bruker Biflex MALDI-TOF mass spectrometer. The Mascot (http://www.matrixscience.com) and ProFound (http://129.85.19.192/profound_bin/WebProFound.exe) programs were used for protein sequence identification.

Cloning of *Pa*-AGOG

The DNA fragment encoding the candidate protein-coding region, PAE2237, was amplified by PCR using *P.aerophilum* genomic DNA as template, and the primers (Ps) GGAATTCCATATGGCCGCAGAGTCGCAATTAAAAAG and (Pas) TATCCCAAGCTTCTATTCTGAGGGAGGCCTACAACC carrying NdeI and HindIII restriction sites, respectively, for subsequent cloning into the pET28c(+) vector (Novagen). The ligation mixture was used to transform competent *E.coli* XL1 Blue cells. The integrity of the insert in the plasmid pET28c(+)paagog was confirmed by DNA sequence analysis.

Site-directed mutagenesis of *Pa*-AGOG

In vitro mutagenesis of the *P.aerophilum* GO-glycosylase was performed using the QuikChange™ site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions. pET28c(+)-paagog served as template for mutagenesis, using the following oligonucleotide primers (sense-strand sequences shown only; mutation sites are underlined):

Pa-AGOG K140Q, CGCGAGCAACAGACCTTGGTTTTC
Pa-AGOG K147Q, GGTTTTCACGATACAAATACTGAACTATGCC
Pa-AGOG D172N, CATACCCGTTAACTACAGAGTCGCC

The resulting plasmids were named pET28c(+)paagogK140Q, pET28c(+)paagogK147Q and pET28c(+)paagogD172N, respectively.

Expression and purification of recombinant *Pa*-AGOG proteins

The plasmids expressing the His-tagged fusion proteins pET28c-paagog, pET28c-paagog K140Q, pET28c-paagog K147Q and pET28c-paagog D172N were electroporated into competent *E.coli* B834(DE3) cells for large-scale protein expression. The transformed cells were used to inoculate 100 ml of LB medium containing 50 µg/ml kanamycin. The cells were allowed to grow overnight at 37°C. The saturated cultures were diluted 1:100 in 1 l of LB medium containing 100 µg/ml kanamycin and grown with shaking at 37°C until the OD_{600} reached 1.5. Expression of the proteins was then induced with 1 mM IPTG. After 5 h incubation at 30°C, the cells were harvested by centrifugation at 4000 *g* for 15 min at 4°C. The cell pellet was washed once with 200 ml of ice-cold washing buffer (20 mM Tris–HCl pH 7.5, 10 mM NaCl and 0.1 mM PMSF) and then resuspended in 50 ml of sonication buffer (50 mM Tris–HCl pH 7.5, 300 mM NaCl and 1 mM PMSF). The cells were lysed by sonication and the WCE was clarified by centrifugation at 15 000 *g* for 45 min at 4°C. The supernatant was incubated with gentle shaking for 30 min at 4°C with 5 ml of Ni-NTA-agarose (Qiagen), pre-equilibrated in sonication buffer. The suspension was then packed into a disposable column and, after washing with 50 ml of sonication buffer, the bound proteins were eluted with sonication buffer containing increasing concentrations of imidazole (100 ml of 5 mM imidazole and 50 ml of 40 mM imidazole). Finally, the column was washed with sonication buffer containing 1 M NaCl to remove any nucleic acids. The His-tagged proteins were eluted with 50 ml of sonication buffer containing 500 mM imidazole. After adding 1 mM EDTA, the recombinant *P.aerophilum* proteins, having a theoretical pI of 9.2, were dialyzed overnight at 4°C against 2 l of ion exchange buffer A (50 mM Tris–HCl, pH 7.5, 50 mM NaCl, 0.1 mM PMSF and 1 mM DTT). The dialyzed fractions were spun down and

loaded onto a 1 ml Resource™-S pre-packed column (Amersham Biosciences). Elution was performed with a 200 ml linear gradient of 50–650 mM NaCl. The *Pa*-AGOG proteins, which eluted around 180 mM NaCl, were pooled (30 ml) and concentrated to 5 ml using 10 000 MWCO Vivaspin concentrators (Sartorius AG, Germany). The proteins were further purified by gel filtration using a HiLoad Superdex™75-26/60 column (Amersham Biosciences) and buffer containing 20 mM HEPES–NaOH, pH 7.5, 100 mM NaCl and 1 mM DTT. Fractions containing the homogeneous Pa-AGOG proteins were pooled (20 ml) and concentrated to 5 ml as above. The pure proteins were stored at −80°C at a concentration of 1.6 mg/ml as determined by UV-absorption.

RESULTS

P.aerophilum extracts efficiently process oligonucleotide duplexes carrying GO residues, irrespective of whether they are in single-stranded DNA, or in duplexes opposite T, C, G or A (Figure 1). As shown, the GO/G duplex was cleaved with the greatest efficiency. This was rather unexpected, as GO glycosylases of the OGG1 family preferentially remove GO from base pairs with C (7–9).

In an attempt to identify the enzyme responsible for the observed GO/G processing, we searched the *P.aerophilum* genome for ORFs carrying the HhH-GPD motif. Two of the three potential candidates were previously shown to encode *Pa*-Nth (21), which excises dihydrothymine, and *Pa*-MIG (13), which removes uracil or thymine from U/G or T/G mispairs, respectively. We postulated that the third ORF carrying an easily identifiable HhH-GPD motif, PAE3092, might encode a GO-processing enzyme, as it displayed homology to OGG glycosylases from other organisms (data not shown). Interestingly, this was not the case; the recombinant protein expressed in *E.coli* possessed an AP-lyase activity, but was unable to process GO-containing substrates (data not shown).

When the *P.aerophilum* extracts were fractionated (see Materials and Methods), the GO/G processing activity eluted from the final column, a High-Trap SP Sepharose HP, predominantly in fractions 9–11 (Figure 2A), which displayed no activity towards the MutY substrates A/GO and A/G, hydroxymethyluracil, hypoxanthine, uracil and G/G, C/C or T/G mispairs (data not shown). However, SDS–PAGE analysis of the pooled fractions 9 and 10 (fraction VI) showed it to contain at least seven polypeptides (Figure 2B, left panel). In order to establish the identity of the GO-processing activity, an unstained lane was cut into 11 slices according to the bands visualized in the stained half of the gel (Figure 2B). The proteins were then extracted and renatured by dialysis (Materials and Methods). The activity was primarily in slice 9, which corresponded to a molecular size of ~30 kDa (Figure 2B). The remainder of fraction VI was therefore loaded on a preparative gel, the two bands in the region of 30 kDa (Figure 3A, left panel) were cut out, and subjected to tryptic digests. Mass spectrometric analysis of peptide products generated from band 2 contained a protein predicted to function in cobalamine biosynthesis and was not analyzed further. Band 1 yielded several peptide masses (Figure 3A, right panel) that could be assigned to ORF PAE2237 (Figure 3B), encoding a hypothetical protein of theoretical molecular weight of 29.5 kDa, a pI of 9.2 and of unknown function.

Analysis of the primary structure of PAE2237 revealed no homology to any known GO-processing enzymes. However, prediction of the secondary structure (22) of PAE2237 indicated that it might be structurally related to the HhH superfamily of DNA repair glycosylases. This evidence suggested that the product of ORF PAE2237 is likely to be involved in DNA damage processing. In an attempt to identify a close relative of PAE2237 the function of which is known, we carried out extensive BLAST searches. We identified several archaeal ORFs that contained regions of identity with

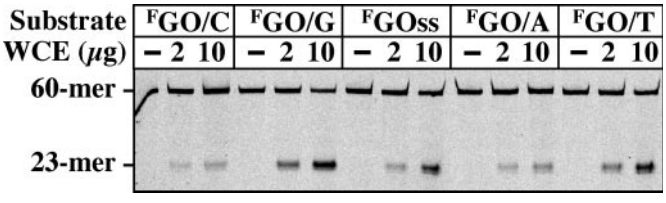


Figure 1. Substrate specificity of GO-processing activity in crude extracts of *P.aerophilum*: 1 pmol of the 60mer substrates was incubated with 0, 2 or 10 μg of WCE as described in Materials and Methods. The positions of the full-length 60mer substrate and of the 23mer product are indicated; F indicates the fluorescein-labeled strand; ss, single-stranded GO 60mer.

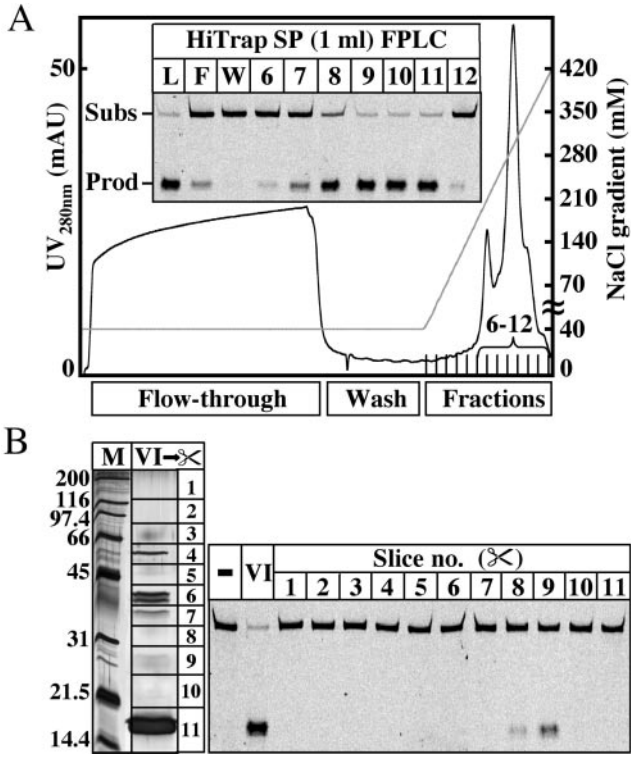
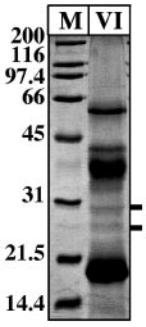


Figure 2. Fractionation of *P.aerophilum* extracts and estimation of the molecular weight of the GO-DNA glycosylase. (**A**) Profile of Fraction V eluted from a HiTrap SP column with a linear gradient from 0.04 to 0.6 M NaCl. Inset shows the GO/G activity in 8 μl of the load (L), flow-through (F), wash (W) and fractions 6–12. Incubations were carried out for 30 min at 60°C, using 1 pmol of the GO/G substrate. The positions of the full-length 60mer substrate and of the 23mer product are indicated. (**B**) Recovery and renaturation of the protein eluted from 11 slices of the 15% SDS–PAGE gel (left panel; M, molecular weight standards; VI, 25 μl of fraction VI). The GO/G-processing activity was found to reside predominantly in slice 9 (right panel).

A



Measured Mass (monoisomers)	ΔM (Da)[a]	Sequence	AA residues
834.49	0.08	RDPQYR	28-33
842.59	0.19	HGETVGSR	42-49
886.69	0.15	RVIETLR	9-15
886.69	0.15	VIETLRR	10-16
980.69	0.13	QLSHIVGAR	128-136
1097.69	0.11	YIETSPFLK	91-99
1232.69	0.05	LERRDPQYR	25-33
1454.89	0.03	RLGIEEVLKLER	16-27
1454.89	0.03	LGIEEVLKLERR	17-28

B

MAAESQLK**RVIETLRRLGIEEVLKLERRDPQYR**AVCNV
VKR**HGETVGSR**LAMLNALISYRLTGKGEEHWEYFGKYF
SQLEVIDLCRDFLK**YIETSPFLK**IGVEARKKRALKACD
YVPNLEDLGLTLR**QLSHIVGAR**REQKTLVFTIKILNYA
YMCSRGVNRVLPFDIPIPVDYRVARLTWCAGLIDFPPE
EALRRYEAVQKIWDAVARETGIPPLHLDTLLWLAGRAV
LYGENLHGVPKEVIALFQWRGGCRPPSE

Figure 3. Identification of the *P.aerophilum* GO-glycosylase (PAE2237) by mass spectrometry. (**A**) Left panel: TCA precipitate from 2 ml of Fraction VI (Materials and Methods) was resolved on a 12.5% SDS–PAGE gel and stained with Coomassie blue. M, Bio-Rad SDS–PAGE molecular weight standards. The two bands (1 and 2) detected at ~30 kDa were cut out and subjected to a tryptic digest (Materials and Methods) followed by MALDI-TOF Mass Spectrometry (MS) analysis of the peptide fragments. Right panel: MALDI-TOF MS analysis of band 1. Measured masses and amino acid sequences of the tryptic peptides are listed in the table. 'a' indicates measured minus calculated mass. (**B**) The peptides identified by MS (shown in boldface) could be assigned to ORF PAE2237 in the *P.aerophilum* genome that encodes a hypothetical protein of 256 amino acids (29.5 kDa).

PAE2237, most notably in the putative HhH-motif containing the absolutely conserved lysine (K140, K147) and aspartate (D172) residues—characteristic features of the active sites of most HhH-GPD glycosylases/lyases (Figure 4A), but none of the proteins encoded by these sequences has been biochemically characterized to date.

The polypeptide encoded by PAE2237 was therefore overexpressed in *E.coli* and purified to apparent homogeneity (Materials and Methods). As shown in Figure 4B, the recombinant protein cleaved all duplex substrates in the GO-containing strand, irrespective of the base opposite the modification (T/GO and single-stranded substrates are not shown), and the cleavage occurred by β-elimination, as in other OGG-like enzymes. The human OGG1 protein generated an identical product, but was very inefficient on GO/G and GO/A. The *E.coli* MutM protein cleaved both GO/C and GO/G, but the products migrated faster through the gel, as this polypeptide catalyzes a β,δ-elimination, such that the 5′-labeled fragment is terminated with a 3′-phosphate group, rather than with a 3′-α,β unsaturated aldehyde group arising from cleavage with hOGG1 or the product of ORF PAE2237. The β,δ-elimination products such as those generated by MutM could be obtained if the substrates incubated with hOGG1 or PAE2237 were subsequently treated with hot alkali (Figure 4B, lanes OH). In contrast, cleavage of the AP-sites with the *P.aerophilum* endonuclease IV, a 5′AP-endonuclease (Figure 4B, lanes *nfo*), generated 5′-labeled 23mers terminated with 3′-hydroxyl groups; these products migrated between the two species generated by the elimination reactions.

Given that the *P.aerophilum* ORF PAE2237 appears to have homologues only in the archaeal world and that it processes GO residues in any substrate, we named the enzyme archaeal GO-glycosylase (*Pa*-AGOG). This enzyme has most likely the most potent GO-glycosylase activity described to date. Due to complete denaturation of the double-stranded oligonucleotide substrates at 100°C (calculated T_m of ~87°C), we were unable to assay the activity of *Pa*-AGOG at the optimal growth temperature of *P.aerophilum*. However, at 60°C, *Pa*-AGOG was able to process more than 90% of the GO/G and GO/C substrates in 15 min, at a substrate to enzyme ratio of 10:1 (Figure 4C, lanes 2 and 5). The activity on the labeled GO/G substrate was not noticeably reduced even in the presence of a 100-fold molar excess of the unlabeled GO/C substrate, and was only slightly decreased upon the addition of a similar excess of the GO/G competitor (Figure 4C, lanes 2–4). In contrast, a 100-fold molar excess of unlabeled GO/G duplex inhibited the processing of the labeled GO/C substrate by more than 50% (Figure 4C, lane 6). The substrate specificity of the recombinant *Pa*-AGOG (GO/G>GO/C) thus largely reflected the relative efficiencies of GO-processing seen in the *P.aerophilum* extracts (Figure 1).

Interestingly, the DNA glycosylase and AP-lyase activities of *Pa*-AGOG could be uncoupled. By treating samples taken at different time points with hot alkali, more than 80% of the GO/G substrate could be hydrolyzed after only 5 min reaction (Figure 4D), which indicates that most GO residues have been removed at this time. Yet, AP-lyase activity of this enzyme cleaved only ~20% of these AP-sites at the same time point. Notably, during the processing of the GO/C substrate, the AP-lyase activity appeared to be somewhat more efficient.

Further analysis revealed *Pa*-AGOG to be extremely heat stable, as pre-treatment of the purified protein at 80°C for 15 min had no notable effect on its catalytic activity (Figure 5). In an attempt to test which residues were involved in catalysis, we carried out side-directed mutagenesis experiments. Alignments with other members of the HhH protein superfamily suggested that aspartate 172 might be involved in catalysis, and we mutated this amino acid to asparagine. Prediction of the catalytic site lysine was not as unambiguous. Secondary structure alignment (22) of *Pa*-AGOG with the C-terminus of hOGG1 implicated K147 of *Pa*-AGOG as a likely candidate, but we also had to consider K140, given that it is also absolutely conserved among the AGOG family members (Figure 4A). We therefore mutated both these sites to glutamines. We also mutated aspartate 166 to glutamine, even though this position is not absolutely conserved in the AGOG proteins (Figure 4A). The D166N mutant had wild-type activity (data not shown), but mutants K140Q and D172N were essentially inactive and the enzymatic activity of the K147Q mutant was severely attenuated (Figure 5). This evidence implicated K140 and D172 in catalysis and suggested that K147 has an important, but not essential function in catalysis. Interestingly, the residual activity of the K147Q mutant was lost upon heating, which implied that this residue was required also for thermal stability of the polypeptide.

DISCUSSION

Hydrolytic and oxidative damage represent major threats to genomic integrity. As rates of these spontaneous reactions rise with temperature, it may be anticipated that hyperthermophilic

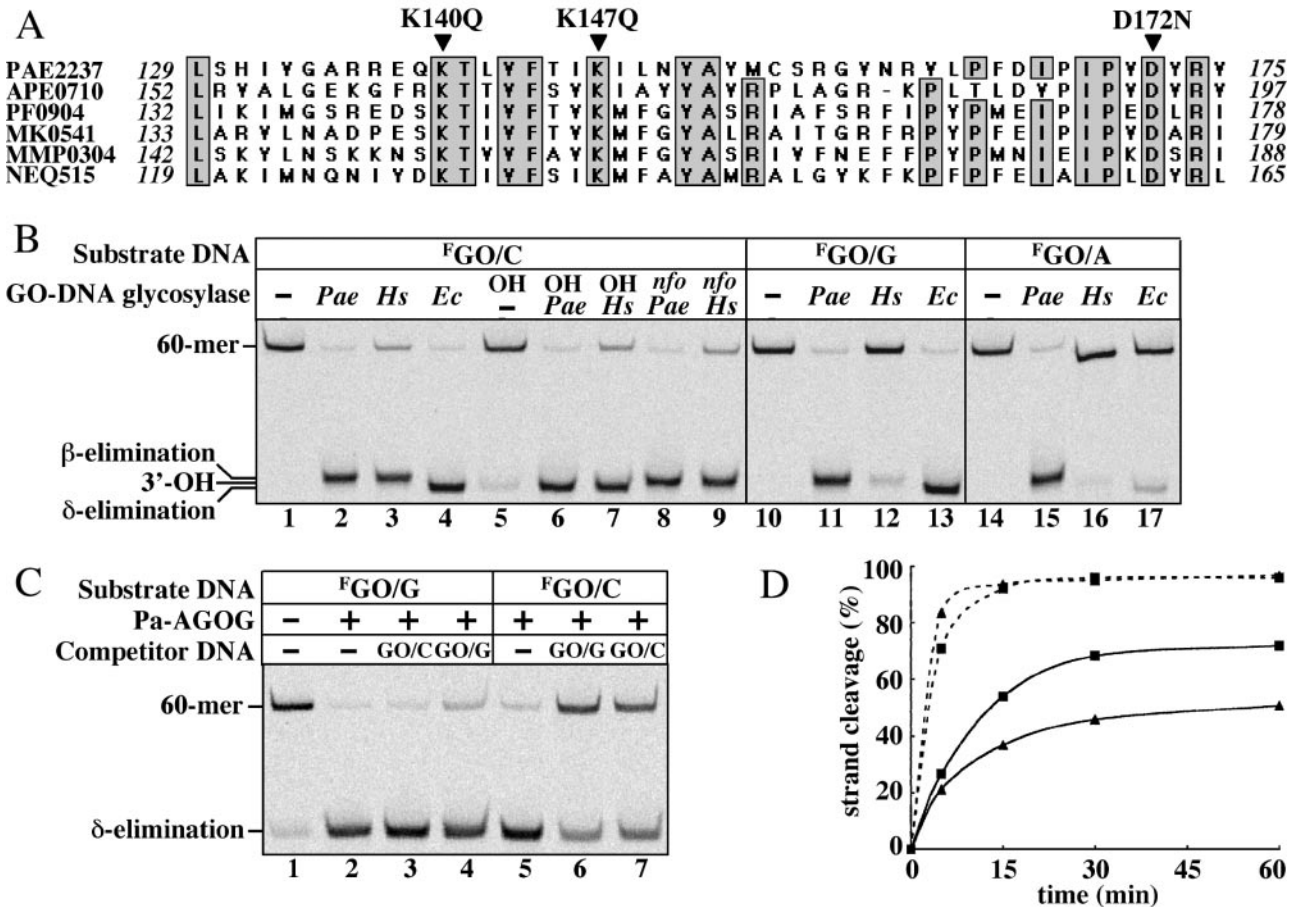


Figure 4. *Pa*-AGOG is a member of a new family of DNA glycosylases. (**A**) Partial sequence alignment of the putative 'extended' helix–hairpin–helix motif of AGOG homologues. PAE2237, *P.aerophilum* (TrEMBL:Q8ZVK6); APE0710, *Aeropyrum pernix* (TrEMBL: Q9YE60); PF0904, *Pyrococcus furiosus* (TrEMBL: Q8U2D, for simplicity, the sequence of only one of the three *Pyrococcus* species is shown); MK0541, *Methanopyrus kandleri* (TrEMBL: Q8TXW8); MMP0304, *Methanococcus maripaludis* (Trnew: CAF29860); NEQ515, *Nanoarchaeum equitans* (Trnew: AAR39356). Identical residues are shaded and the putative active site residues (K140Q, K147Q and D172N) are indicated by arrowheads. The sequence alignment was generated using the MultAlin software (23), available at www.toulouse.infra.fr. (**B**) Substrate specificity comparison of *Pa*-AGOG, human OGG1 and *E.coli* MutM. The GO/C, GO/G and GO/A substrates (1 pmol) were incubated for 15 min at 60°C in the absence of enzyme (lanes 1, 5, 10 and 14) or with 1 pmol of *Pa*-AGOG (*Pae*, lanes 2, 6, 8, 11 and 15) or at 37°C with 10 pmol of hOGG1 (*Hs*, lanes 3, 7, 9, 12 and 16) or 1 pmol of MutM (*Ec*, lanes 4, 13 and 17). Samples were then either directly precipitated (lanes 1–4), treated with 100 mM NaOH (OH) for 10 min at 90°C (lanes 5–7) or incubated with 0.5 pmol of *Pa*-EndoIV (*nfo*) for 10 min at 60°C (lanes 8 and 9). The 60mer substrate and the positions of the three possible 23mer products are indicated on the left. NB: Some GO is lost spontaneously at 60°C and the resulting AP-sites (~10%) are cleaved by the hot alkali treatment (lane 5). (**C**) The highly efficient DNA glycosylase activity of *Pa*-AGOG is modulated by the base opposite GO; 10 pmol of the labeled GO/G and GO/C substrates were incubated for 15 min at 60°C with 1 pmol of *Pa*-AGOG in the presence or absence of 100 pmol of unlabeled 60mer competitor duplexes GO/C (lanes 3 and 7) or GO/G (lanes 4 and 6). The reactions were immediately quenched by the addition of 100 mM NaOH (10 min at 90°C) to measure DNA glycosylase activity. Lane 1, no enzyme; lanes 1, 2 and 5, no competitor DNA added. (**D**) Comparison of DNA-glycosylase (dashed lines) and DNA-glycosylase/lyase (solid lines) activities of *Pa*-AGOG on GO/G (squares) and GO/C (triangles) substrates. *Pa*–AGOG (1 pmol) was incubated with 10 pmol of the labeled substrates. After 5, 15, 30 and 60 min, aliquots of the reaction mixtures were removed and either immediately treated with hot alkali to measure glycosylase activity or directly ethanol-precipitated to measure combined glycosylase/AP-lyase activity.

organisms require highly efficient repair pathways that remove deaminated and oxidized bases from their DNA. The crenarchaeon *P.aerophilum* lives at temperatures around 100°C and is well protected against hydrolytic deamination of cytosines: it carries at least three uracil-processing enzymes (13–15). As far as the repair of oxidative DNA damage is concerned, *P.aerophilum* possesses no detectable MutY-like activity to process A/GO mispairs. Moreover, as the organism encodes no homologues of mismatch repair proteins and as its extracts possess no mismatch binding or processing activity (data not shown), it is unlikely that A/GO mismatches are processed by the latter system, as is the case in *S.cerevisiae* (24). In contrast, it expresses a robust GO-DNA glycosylase (Figure 1), *Pa*-AGOG, encoded by PAE2237 (Figure 3). Like other GO-processing DNA glycosylases, *Pa*-AGOG has an associated AP-lyase activity that can cleave the sugar-phosphate backbone 3′ from AP-sites by β-elimination (Figure 4B). Interestingly, the AP-lyase activity was more efficient on GO/C substrates (Figure 4D), while the glycosylase activity of *Pa*-AGOG removed GO with the highest efficiency from GO/G duplexes (Figure 4C). The differential efficiencies of base removal and enzyme-catalyzed cleavage of the sugar-phosphate backbone might be best explained by the model suggested by Verdine and colleagues (25). The initial nucleophilic attack of the active site lysine on the C1′ position of the sugar bearing the aberrant base results

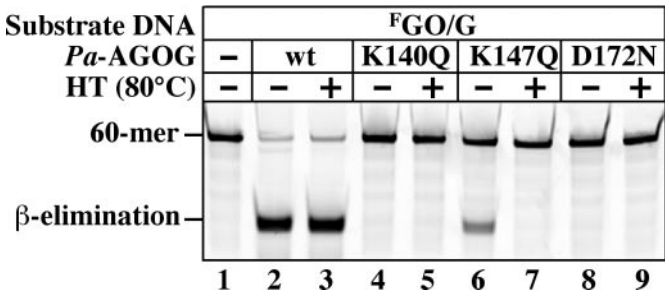


Figure 5. Effect of mutations of the conserved lysines 140 and 147 and aspartate 172 on the glycosylase/lyase activity of *Pa*-AGOG. The labeled GO/G substrate (1 pmol) was incubated for 15 min at 60°C alone (lane 1), or with 1 pmol of either wild-type (wt) *Pa*-AGOG (lanes 2 and 3), *Pa*-AGOG K140Q (lanes 4 and 5), *Pa*-AGOG K147Q (lanes 6 and 7) or *Pa*-AGOG D172N (lanes 8 and 9). HT indicates that the proteins were heat-treated for 15 min at 80°C prior to addition of the substrate (lanes 3, 5, 7 and 9). Heat treatment did not affect the glycosylase/lyase activity of the wild-type enzyme (lane 3). The K140Q (lanes 4 and 5) and D172N (lanes 8 and 9) mutations severely attenuated the enzymatic activity of *Pa*-AGOG on the GO/G substrate, while substitution of K147 with glutamine resulted in a substantial reduction of the catalytic activity of the protein, while making the polypeptide also thermolabile (lanes 6 and 7).

in the scission of the glycosidic bond, and in the formation of a Schiff's base intermediate, in which the enzyme and the DNA substrate are covalently linked. Under normal circumstances, the Schiff's base can be either hydrolyzed to release the enzyme and generate the abasic site, or, alternatively, it might decompose through abstraction of the 2′ proton, which would result in a β-elimination reaction and would release the enzyme, while forming an α,β-unsaturated aldehyde—a process that also leads to a scission of the sugar-phosphate backbone of the DNA substrate. The relative efficiency of these two pathways will depend on the geometry of the complex; successful β-elimination requires that the 2′ proton and the leaving group, the phosphodiester moiety, be in *trans*. Like other DNA glycosylases/lyases, *Pa*-AGOG is likely to make specific contacts with the widowed base in the strand opposite the baseless sugar-phosphate. Its interaction with cytosine might induce a conformation favoring efficient β-elimination, whereas interaction with the larger guanine may force the enzyme into a conformation that favors the simple hydrolysis of the Schiff's base intermediate. This would explain why GO is removed with high efficiency from both GO/G and GO/C substrates, but why the rates of cleavage of the AP-site through β-elimination differ (Figure 4D).

Is the *in vitro* substrate preference of *Pa*-AGOG relevant to its biological function? In order to answer this question, we need to consider the entire range of pathways by which GO can be introduced into DNA and compare this with the palette of *P.aerophilum* detoxifying enzymes. During oxidation of DNA, GO residues arise opposite C. In this case, an antimutator enzyme has to initiate the repair process by removing the GO and thus allowing its replacement with a G. This function is highly conserved through evolution, in the form of the MutM and OGG enzymes (2). If GO persists in the DNA until the next round of replication, it will pair preferentially with A. These GO/A mispairs are addressed by the MutY protein family, which remove the mispaired adenine. BER then excises the baseless sugar-phosphate and inserts a C to generate a GO/C pair (26) that can now be repaired to a G/C with the help of the MutM/OGG enzymes. In the latter process, the co-operation of MutY and MutM is again antimutagenic. But MutY can also act as a mutator, by removing adenines from GO/A mispairs that arose during DNA replication through the incorporation of 8-oxo-dGMP residues opposite template adenosines. Because the BER-mediated A→C substitution takes place in the template strand, this process fixes rather than repairs the mutation. In most organisms, however, incorporation of 8-oxo-dGMP into newly synthesized DNA is kept to a minimum by the MutT family of enzymes, which hydrolyze 8-oxo-dGTP to the monophosphate. A *P.aerophilum* MutT homologue has not been identified to date, but it is likely to be present, as this organism encodes several Nudix family hydrolases. In contrast, we failed to detect a MutY-like activity in the extracts of *P.aerophilum* (data not shown), nor did we find an ORF that might encode such a polypeptide. Is it possible that *Pa*-AGOG is the only enzyme in *P.aerophilum* capable of processing GO-containing lesions in the DNA? If so, how can it fulfill the role of both MutM and MutY? The answer to this question may lie in its catalytic efficiency. If *Pa*-AGOG removes all GO residues from GO/C pairs prior to replication, then the only way that GO could arise in *P.aerophilum* DNA is through incorporation of 8-oxo-dGMP residues into the nascent strand during replication. At ambient temperatures, GO is incorporated preferentially opposite A. However, recent evidence suggests that GO can form stable Hoogsteen base pairs also with G, and it was postulated that formation of GO/G mispairs is favored over GO/A at elevated temperatures (27). Thus, by catalyzing the removal of GO from GO/C pairs from oxidized resting DNA and from GO/G or GO/A pairs from newly replicated DNA, *Pa*-AGOG could effectively protect the genome of *P.aerophilum* against mutagenesis caused by ROS, even in the absence of a MutT protein. In this scenario, the absence of a MutY homologue would lower the mutation frequency.

Using the amino acid sequence of *Pa*-AGOG, we searched for possible structural homologues of this protein. The 3D-PSSM program (22) identified EndoIII of *E.coli* and the C-terminal domain of hOGG1 as bearing similar structural motifs. Both these enzymes are DNA glycosylases with HhH motifs. Our site-directed mutagenesis experiments (Figure 5) strongly suggested that *Pa*-AGOG does indeed possess such a structural motif, because mutation of the evolutionarily highly conserved lysine and aspartate residues characteristic of the HhH-GPD glycosylases inactivated our enzyme. However, even though *Pa*-AGOG may be structurally related to EndoIII and hOGG1, its primary amino acid sequence bears no significant homology to these enzymes, nor does it resemble other bacterial or eukaryotic GO-glycosylases. This was substantiated by our phylogenetic analysis, which showed that the AGOG proteins do not group with any of the known DNA glycosylase families of the HhH superfamily (Figure 6). *Pa*-AGOG is thus the founding member of a new family of proteins, which currently comprises eight members (including the three *Pyrococcus* species; Figure 4A). Interestingly, this family contains the hyperthermophile *N.equitans* (28), an organism with the smallest microbial genome sequenced to date (<0.5 Mb containing 552 coding sequences). An AGOG homologue appears to be present also in *M.maripaludis*, a mesophilic relative of

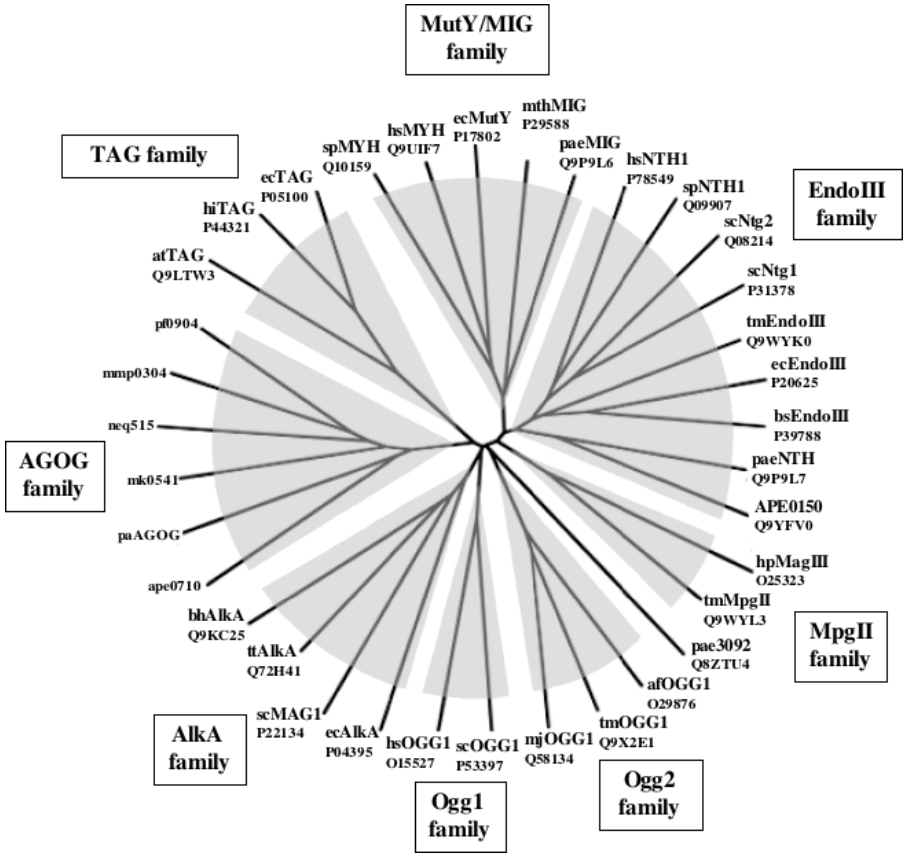


Figure 6. Unrooted Phylogenetic Tree of the HhH-Superfamily of DNA glycosylases. All 35 protein sequences were aligned using the ClustalW program (http://www.ebi.ac.uk/clustalw). Distance analysis of the phylogenetic tree was performed using the program TreeView 1.6.6 (31). The abbreviations used in the figure are as follows: Bacteria: bh, *Bacillus halodurans*; bs, *Bacillus subtilis*; ec, *E.coli*; hi, *Haemophilus influenzae*; hp, *Helicobacter pylori*; tm, *Thermotoga maritima*; tt, *Thermus thermophilus*; Archaea: af, *Archaeoglobus fulgidus*; mth, *Methanobacterium thermoformicicum*; mj, *M.jannaschii*; mk, *Methanopyrus kandleri*; Eukaryotes: at, *Arabidopsis thaliana*; hs, *Homo sapiens*; sc, *S.cerevisiae*; sp, *S.pombe*. Members of the AGOG family are indicated in Figure 4A. Swiss-Prot/TrEMBL accession numbers are also indicated.

Methanococcus jannaschii, which implies that AGOG proteins are not restricted to extreme thermophiles and thus that the enzyme has not evolved specifically to counteract the enhanced mutagenic threat of DNA oxidation in species living in hot environments.

Most recently, we succeeded in solving the crystal structure of *Pa*-AGOG at 1.0 Å resolution. Our findings confirm that, as predicted on the basis of our biochemical data, the enzyme does indeed have an HhH-GPD motif. Interestingly, in spite of its primary sequence divergence, the 3D structure of this motif is largely superimposable with those of the other HhH DNA glycosylases. It therefore appears that nature has independently evolved a similar protein fold from completely divergent primary sequences, due to the necessity to effectively protect DNA from the mutagenic threat of oxidative damage. However, *Pa*-AGOG differs from the other HhH family members in two aspects: the hairpin is much longer as predicted from the sequence alignments shown in Figure 4A (29) and the mode of GO recognition is substantially different too, inasmuch as the enzyme stabilizes the aberrant base in its active site through several GO-specific hydrogen bonds that include also the O^8 atom. This differs from other GO-recognizing enzymes, which differentiate between GO and G merely by sensing the protonation status of the N^7 position (30). These differences confirm that *Pa*-AGOG is indeed the founding member of a new family of GO-glycosylases that are an integral part of the HhH glycosylase superfamily.

It is interesting to note that the unrooted phylogenetic tree (Figure 6) places ORF PAE3092 on its own, separate from the eight HhH families. The fact that its branch lies very close to the OGG2 family supports our initial expectation that this protein could be responsible for the GO-processing activity detected in *P.aerophilum* extracts. This was not the case (see above), however, the phylogenetic evidence suggests that PAE3092, which could to date be shown to possess only an AP-lyase activity (data not shown), might be involved in the processing of substrates containing other, as yet uncharacterized, form of oxidative DNA damage.

ACKNOWLEDGEMENTS

The authors wish to thank Primo Schär and Orlando Schärer for many helpful discussions and for critical reading of the manuscript. This project was supported in part by grants from the UBS AG (A.S. and J.J.) and the Swiss National Science Foundation (L.G. and F.K.W.).

REFERENCES

1. Michaels,M.L. and Miller,J.H. (1992) The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J. Bacteriol.*, **174**, 6321–6325.
2. Scharer,O.D. and Jiricny,J. (2001) Recent progress in the biology, chemistry and structural biology of DNA glycosylases. *Bioessays*, **23**, 270–281.
3. Hazra,T.K., Izumi,T., Kow,Y.W. and Mitra,S. (2003) The discovery of a new family of mammalian enzymes for repair of oxidatively damaged DNA, and its physiological implications. *Carcinogenesis*, **24**, 155–157.
4. Nash,H.M., Bruner,S.D., Scharer,O.D., Kawate,T., Addona,T.A., Spooner,E., Lane,W.S. and Verdine,G.L. (1996) Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily. *Curr. Biol.*, **6**, 968–980.
5. van der Kemp,P.A., Thomas,D., Barbey,R., de Oliveira,R. and Boiteux,S. (1996) Cloning and expression in *Escherichia coli* of the OGG1 gene of *Saccharomyces cerevisiae*, which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine. *Proc. Natl Acad. Sci. USA*, **93**, 5197–5202.
6. Boiteux,S. and Radicella,J.P. (1999) Base excision repair of 8-hydroxyguanine protects DNA from endogenous oxidative stress. *Biochimie*, **81**, 59–67.
7. Radicella,J.P., Dherin,C., Desmaze,C., Fox,M.S. and Boiteux,S. (1997) Cloning and characterization of hOGG1, a human homolog of the OGG1 gene of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **94**, 8010–8015.
8. Roldan-Arjona,T., Wei,Y.F., Carter,K.C., Klungland,A., Anselmino,C., Wang,R.P., Augustus,M. and Lindahl,T. (1997) Molecular cloning and functional expression of a human cDNA encoding the antimutator enzyme 8-hydroxyguanine-DNA glycosylase. *Proc. Natl Acad. Sci. USA*, **94**, 8016–8020.
9. Rosenquist,T.A., Zharkov,D.O. and Grollman,A.P. (1997) Cloning and characterization of a mammalian 8-oxoguanine DNA glycosylase. *Proc. Natl Acad. Sci. USA*, **94**, 7429–7434.
10. Bjoras,M., Luna,L., Johnsen,B., Hoff,E., Haug,T., Rognes,T. and Seeberg,E. (1997) Opposite base-dependent reactions of a human base excision repair enzyme on DNA containing 7,8-dihydro-8-oxoguanine and abasic sites. *EMBO J.*, **16**, 6314–6322.
11. Bruner,S.D., Nash,H.M., Lane,W.S. and Verdine,G.L. (1998) Repair of oxidatively damaged guanine in *Saccharomyces cerevisiae* by an alternative pathway. *Curr. Biol.*, **8**, 393–403.
12. Hazra,T.K., Hill,J.W., Izumi,T. and Mitra,S. (2001) Multiple DNA glycosylases for repair of 8-oxoguanine and their potential *in vivo* functions. *Prog. Nucleic Acid Res. Mol. Biol.*, **68**, 193–205.
13. Yang,H., Fitz-Gibbon,S., Marcotte,E.M., Tai,J.H., Hyman,E.C. and Miller,J.H. (2000) Characterization of a thermostable DNA glycosylase specific for U/G and T/G mismatches from the hyperthermophilic archaeon *Pyrobaculum aerophilum*. *J. Bacteriol.*, **182**, 1272–1279.
14. Sartori,A.A., Schar,P., Fitz-Gibbon,S., Miller,J.H. and Jiricny,J. (2001) Biochemical characterization of uracil processing activities in the hyperthermophilic archaeon *Pyrobaculum aerophilum*. *J. Biol. Chem.*, **276**, 29979–29986.
15. Sartori,A.A., Fitz-Gibbon,S., Yang,H., Miller,J.H. and Jiricny,J. (2002) A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site. *EMBO J.*, **21**, 3182–3191.
16. Volkl,P., Huber,R., Drobner,E., Rachel,R., Burggraf,S., Trincone,A. and Stetter,K.O. (1993) *Pyrobaculum aerophilum* sp. nov., a novel nitrate-reducing hyperthermophilic archaeum. *Appl. Environ. Microbiol.*, **59**, 2918–2926.
17. Hudepohl,U., Reiter,W.D. and Zillig,W. (1990) *In vitro* transcription of two rRNA genes of the archaebacterium Sulfolobus sp. B12 indicates a factor requirement for specific initiation. *Proc. Natl Acad. Sci. USA*, **87**, 5851–5855.
18. Sartori,A.A. and Jiricny,J. (2003) Enzymology of base excision repair in the hyperthermophilic archaeon *Pyrobaculum aerophilum*. *J. Biol. Chem.*, **278**, 24563–24576.
19. Neddermann,P. and Jiricny,J. (1993) The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. *J. Biol. Chem.*, **268**, 21218–21224.
20. Schrimpf,S.P., Langen,H., Gomes,A.V. and Wahlestedt,C. (2001) A two-dimensional protein map of *Caenorhabditis elegans*. *Electrophoresis*, **22**, 1224–1232.
21. Yang,H., Phan,I.T., Fitz-Gibbon,S., Shivji,M.K., Wood,R.D., Clendenin,W.M., Hyman,E.C. and Miller,J.H. (2001) A thermostable endonuclease III homolog from the archaeon *Pyrobaculum aerophilum*. *Nucleic Acids Res.*, **29**, 604–613.
22. Kelley,L.A., MacCallum,R.M. and Sternberg,M.J. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.*, **299**, 499–520.
23. Corpet,F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.*, **16**, 10881–10890.
24. Ni,T.T., Marsischky,G.T. and Kolodner,R.D. (1999) MSH2 and MSH6 are required for removal of adenine misincorporated opposite 8-oxo-guanine in *S. cerevisiae*. *Mol. Cell*, **4**, 439–444.
25. Bruner,S.D., Norman,D.P. and Verdine,G.L. (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*, **403**, 859–866.
26. Hashimoto,K., Tominaga,Y., Nakabeppu,Y. and Moriya,M. (2004) Futile short-patch DNA base excision repair of adenine:8-oxoguanine mispair. *Nucleic Acids Res.*, **32**, 5928–5934.
27. Thiviyanathan,V., Somasunderam,A., Hazra,T.K., Mitra,S. and Gorenstein,D.G. (2003) Solution structure of a DNA duplex containing 8-hydroxy-2-deoxyguanosine opposite deoxyguanosine. *J. Mol. Biol.*, **325**, 433–442.
28. Huber,H., Hohn,M.J., Rachel,R., Fuchs,T., Wimmer,V.C. and Stetter,K.O. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature*, **417**, 63–67.
29. Lingaraju,G.M., Sartori,A.A., Kostrewa,D., Prota,A.E., Jiricny,J. and Winkler,F.K. (2005) Crystal structure of Pa-AGOG, a novel 8-oxoguanine DNA glycosylase from the hyperthermophilic archaeon *Pyrobaculum aerophilum*. *Structure (Camb.)*, in press.
30. Fromme,J.C., Banerjee,A. and Verdine,G.L. (2004) DNA glycosylase recognition and catalysis. *Curr. Opin. Struct. Biol.*, **14**, 43–49.
31. Page,R.D. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci*, **12**, 357–358.