Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors

Karl A. Haushalter*, P. Todd Stukenberg†, Marc W. Kirschner†‡ and Gregory L. Verdine*‡

Background: The cellular environment exposes DNA to a wide variety of endogenous and exogenous reactive species that can damage DNA, thereby leading to genetic mutations. DNA glycosylases protect the integrity of the genome by catalyzing the first step in the base excision–repair of lesions in DNA.

Results: Here, we report a strategy to conduct genome-wide screening for expressed DNA glycosylases, based on their ability to bind to a library of four synthetic inhibitors that target the enzyme's active site. These inhibitors, used in conjunction with the *in vitro* expression cloning procedure, led to the identification of novel *Xenopus* and human proteins, xSMUG1 and hSMUG1, respectively, that efficiently excise uracil residues from DNA. Despite a lack of statistically significant overall sequence similarity to the two established classes of uracil-DNA glycosylases, the SMUG1 enzymes contain motifs that are hallmarks of a shared active-site structure and overall protein architecture. The unusual preference of SMUG1 for single-stranded rather than double-stranded DNA suggests a unique biological function in ridding the genome of uracil residues, which are potent endogenous mutagens.

Conclusions: The 'proteomics' approach described here has led to the isolation of a new family of uracil-DNA glycosylases. The three classes of uracil-excising enzymes (SMUG1 being the most recently discovered) represent a striking example of structural and functional conservation in the almost complete absence of sequence conservation.

Addresses: *Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. †Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. ‡Harvard Institute for Chemistry and Cell Biology, Boston, Massachusetts 02115, USA.

Correspondence: Gregory L. Verdine
E-mail: verdine@chemistry.harvard.edu

Received: 14 December 1998
Revised: 21 January 1999
Accepted: 21 January 1999

Published: 11 February 1999

Current Biology 1999, 9:174–185
http://biomednet.com/elecref/0960982200900174

 ISSN 0960-9822

Background

The covalent structure of the cellular genome is subject to spontaneous alteration by a wide variety of endogenous and exogenous reactants. The resulting lesions can produce mutations by base-mispairing during DNA replication, or can decrease cell viability by interfering with the template functions of DNA [1,2]. To counter these deleterious effects, all cells and even some viruses express proteins that are responsible for surveillance of the genome and repair of structural aberrations. At least five mechanistically distinct pathways are known to be employed in the eradication of DNA lesions: nucleotide excision–repair [3], base excision–repair [4], recombinational repair [5], mismatch repair [6] and direct reversion [7,8]. These pathways are highly conserved throughout evolution, from bacteria to mammals. Extensive efforts have led to a sophisticated molecular-level understanding of the DNA repair pathways in bacteria and an emerging picture in lower eukaryotes, but the corresponding pathways in higher eukaryotes are now just coming into focus [9]. Progress on the latter front would be accelerated by the isolation and cloning of all the protein components that make up each pathway. Here, we report a potentially general strategy that employs designed synthetic inhibitors to conduct a whole-genome screen for the encoded DNA repair proteins.

As the initial focus of these investigations, we chose the base-excision DNA repair (BER) pathway, which is responsible for the removal of damaged bases from DNA and restoration of the original sequence. The key components of the BER pathway are DNA glycosylases, lesion-specific enzymes that locate aberrant bases and catalyze their ejection through cleavage of the N-glycosidic bond [10,11]. The resulting abasic sites in DNA or cleavage products thereof — common intermediates generated by all known DNA glycosylases — converge upon a series of downstream repair proteins that excise the remainder of the lesion from DNA, replace the missing nucleotide and reseal the DNA backbone [12–14]. Detailed analyses of the catalytic mechanism employed by DNA glycosylases has led to the design of potent inhibitors that bind tightly to the enzymes but cannot be processed by them [15]. These molecules have previously proven valuable for mechanistic and structural studies on known DNA glycosylases. As shown here, the inhibitors can also be used as powerful tools in the isolation and cloning of previously unknown enzymes.

In the past, DNA repair proteins have typically been isolated on the basis of their ability to process a particular substrate of interest, or by virtue of amino-acid sequence similarity to another protein of known function. By contrast, the goal of this study was to screen as broadly as possible for DNA repair activities, irrespective of the lesions they process. The possibility that entirely new classes of DNA repair proteins might be present in vertebrate cells ruled out screens based on sequence similarity to known proteins. Furthermore, the low throughput of bioassay-directed fractionation procedures made this strategy unattractive for broad screening. Instead, a procedure that could rapidly sort through libraries of expressed proteins to detect DNA repair activity was most desirable. Such a method, termed *in vitro* expression cloning (IVEC), has recently been reported [16]. This method involves the screening of proteins transcribed and translated *in vitro* — first as pools, then as single proteins — for an activity of interest. Previous uses of the IVEC strategy include the identification of *Xenopus* proteins that undergo mitosis-specific phosphorylation [17] or degradation [18], identification of substrates for apoptotic proteases [19,20] and isolation of sequence-specific DNA-binding proteins [21]. In this work, we have devised an IVEC screen that aims to identify proteins having enzymatic activity. To overcome the potential liabilities of screening directly on the basis of enzymatic activity, however, we used a surrogate assay, namely binding to mechanism-based inhibitors that target the active-site chemistry of particular enzymes.

As detailed below, IVEC screening of a *Xenopus* cDNA library using a cocktail of four generic inhibitors of DNA repair led to the identification of a cDNA clone encoding a 31 kDa inhibitor-binding polypeptide that is not similar in sequence to any known protein. Biochemical characterization revealed that this protein is kinetically competent as a monofunctional DNA glycosylase specific for uracil residues, and has appreciable selectivity for single-stranded rather than double-stranded DNA substrates, hence the designation xSMUG1 (single-strand-selective monofunctional uracil-DNA glycosylase). On the basis of cloning and preliminary biochemical characterization, a sequence-related human protein was established as the human ortholog of xSMUG1, namely hSMUG1. Despite having less than 10% sequence identity to the two other established classes of uracil-DNA glycosylases (UDGs) — classical UDGs and mismatch-specific uracil-DNA glycosylases (MUGs), which themselves share less than 10% sequence similarity — both xSMUG1 and hSMUG1 contain motifs that strongly suggest a conserved active-site structure and overall protein architecture. These three classes of uracil-excising enzymes thus represent a striking example of structural and functional conservation in the almost complete absence of sequence conservation.

Results

Design of DNA probes for IVEC screening

In considering various *in vitro* assays to detect the presence of DNA repair activities in IVEC protein pools, we were attracted to the electrophoretic mobility shift assay (EMSA) using oligonucleotide probes containing known substrates for DNA repair enzymes. In the present case, however, this approach would suffer from two drawbacks: firstly, a repair enzyme would ordinarily be expected to process the substrate and then dissociate from it rapidly, leaving no stable complex to be detected by EMSA; and secondly, the use of a defined repair substrate would bias the screen toward cognate enzymes and against non-cognate enzymes — this would run counter to our goal of screening as broadly as possible for proteins that recognize damaged DNA. A more powerful screen would instead make use of probes that could form long-lived complexes

Figure 1

Design of mechanism-based inhibitors used to screen for DNA repair proteins by IVEC. (a) The general mechanism of monofunctional DNA glycosylases. The structure indicated in brackets represents the transition state, in which substantial partial positive charge accumulates on the 2′-deoxyribose ring, especially at the positions denoted by δ+. (b) The structure of the four components comprising the inhibitor library used in IVEC screening.

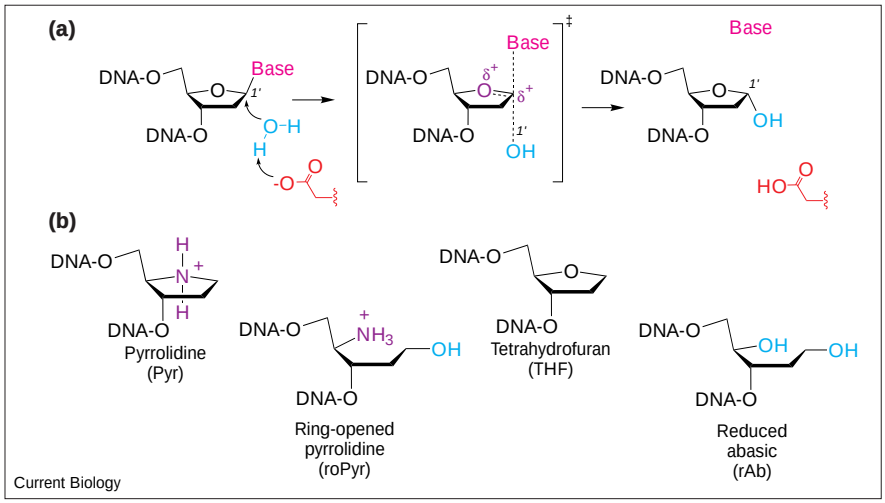
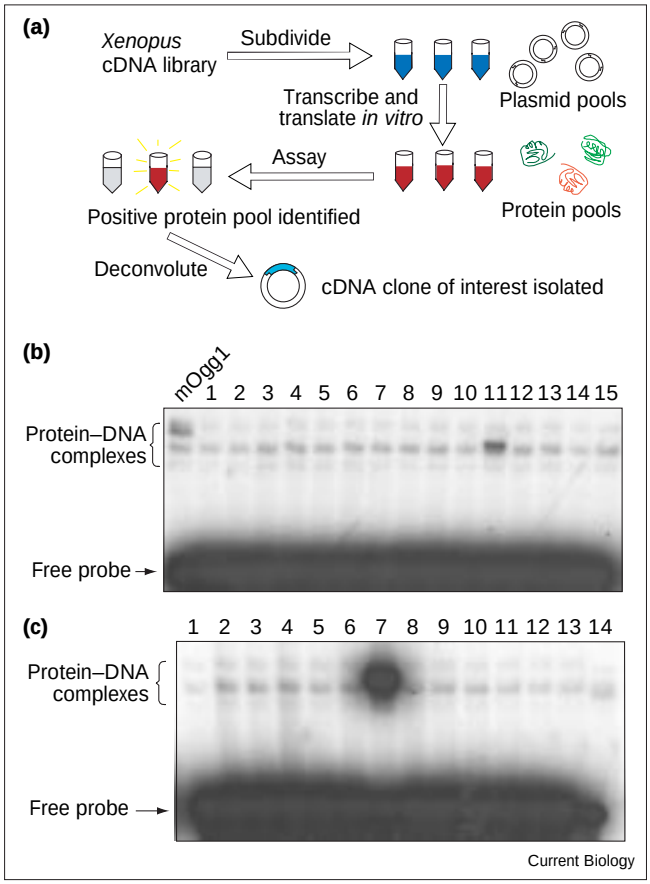


Figure 2



Identification of the TD12 clone by IVEC screening of a *Xenopus* cDNA library. (a) Schematic overview of the IVEC screening procedure [16]. A *Xenopus* cDNA library is subdivided into several hundred pools, each comprising about 100 clones. These cDNA pools are converted to corresponding protein pools via *in vitro* transcription and translation, and the protein pools are screened for activity, in this case, binding to a library of four ^{32}P-labeled generic DNA repair inhibitors. Positive cDNA pools are 'cloned out' (deconvoluted) to obtain single cDNAs encoding the inhibitor-binding activity. (b) EMSA on native polyacrylamide gels of the protein pools generated as shown in (a). Whereas lanes 1–10 and 12–15 have a nearly identical appearance, an extra band is evident in lane 11; the pool corresponding to lane 11 was therefore scored positive for inhibitor-binding activity. The mOgg1 lane is a positive control and size standard containing the inhibitor-bound form of the 39 kDa murine 8-oxoguanine DNA glycosylase enzyme, generated by *in vitro* transcription and translation of an *mOgg1* cDNA [36]. (c) EMSA of single proteins generated by *in vitro* transcription and translation of monoclonal cDNAs obtained through deconvolution of the positive pool in (b). The assay giving a positive result (lane 7) contained the TD12 clone.

with a potentially wide variety of DNA repair proteins; for example, earlier work in these laboratories led to the mechanism-based design of abasic site analogs that bind tightly to numerous base-excision DNA repair proteins [15,22]. As illustrated in Figure 1, two of these analogs (pyrrolidine (Pyr) and ring-opened pyrrolidine (roPyr)) were designed to mimic the positive charge developed in the transition state of the glycosylase reaction. In addition to these, we included in this study two abasic site analogs (tetrahydrofuran (THF) and reduced abasic (rAb)) developed by others to resemble the product of the glycosylase reaction [15]. A library comprising an equimolar mixture of these four inhibitors incorporated into 25-residue double-stranded DNA (duplex 25-mers) was generated and radiolabeled for EMSA screening of IVEC protein pools.

IVEC screening

The availability of a subdivided *Xenopus* embryo cDNA library constructed specifically for expression cloning [23] led us to carry out initial screening for DNA repair proteins expressed by this organism. To our knowledge, no DNA glycosylases have previously been isolated from *Xenopus*, despite the characterization of several enzymes which process the abasic site product of the glycosylase reaction [24–27]. A schematic diagram of the IVEC screening protocol [16] is shown in Figure 2a. Briefly, 120 pools, each containing roughly 100 cDNAs, were transcribed and translated *in vitro* to produce corresponding pools of proteins. These pools were then incubated with the library of four ^{32}P-labeled inhibitors and the resulting mixtures were subjected to native polyacrylamide gel electrophoresis. As shown in Figure 2b, most of the protein pools exhibited an identical pattern of bands, presumably arising from binding activities present in the rabbit reticulocyte lysate used to program *in vitro* transcription and translation. One pool (Figure 2b, lane 11), however, clearly contained a distinct probe-binding activity that was not present in the other pools. The cDNA encoding this binding activity was obtained by deconvolution as described [16]. The final stage of IVEC screening, in which single cDNAs were analyzed, yielded a clone (referred to as TD12 based on its position in the screening grid) encoding a strong binding activity for the library of DNA probes (Figure 2c, lane 7).

Sequencing of the TD12 cDNA clone revealed a single open reading frame encoding a protein of 281 amino acids with a calculated molecular mass of 31 kDa (Figure 3). This mass is in the range expected on the basis of the difference in position of EMSA bands for the TD12-encoded protein and that of the 39 kDa murine Ogg1 protein (compare the mOgg1 lane with lane 11 in Figure 2b). BLAST searching revealed that the TD12-encoded protein showed no significant sequence similarity to any known protein, but did show extensive homology to several human expressed sequence tags (ESTs). Three of these human ESTs were sequenced and two appeared to be full-length, allowing a complete composite sequence to be generated (Figure 3). This composite sequence contains an open reading frame encoding a predicted protein of 270 amino acids (30 kDa). The amino-acid sequences of the *Xenopus* TD12-encoded protein and its human counterpart are 60% (163 out of 270) identical and 71% (191 out

Figure 3

Deduced amino-acid sequence of the *Xenopus* protein encoded by the TD12 clone (xSMUG1) and its human ortholog (hSMUG1). The two active-site motifs that are characteristic of uracil-DNA glycosylases are boxed.

of 270) similar, allowing for conservative substitutions. The complete human nucleotide sequence was used to search the human sequence tag site (STS) database. Three STSs were identified which were identical in sequence to three overlapping segments of the 3′ untranslated region of our sequenced clone. These STSs (WI-6928, Cda0ub02 and stSG3258) have been mapped to the long arm of chromosome 12 in a region bounded by the D12S325 and D12S1691 loci [28,29].

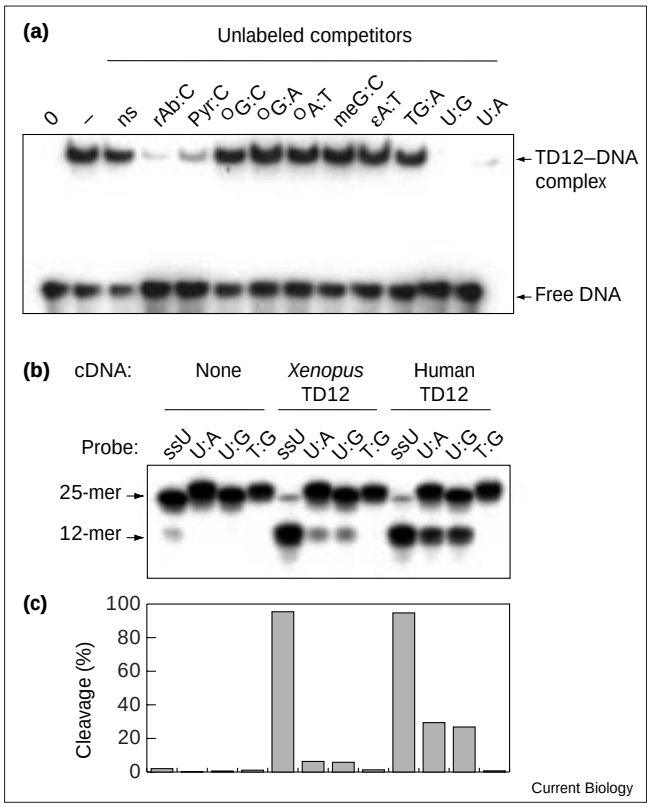
The TD12-encoded protein is a uracil-DNA glycosylase

The TD12-encoded protein was isolated as a result of its binding to a library of four radiolabeled probes containing generic forms of DNA damage, rather than as a result of its ability to catalyze a reaction. Thus, it was unclear at first whether the protein had enzymatic activity, and indeed whether it had any relevance to DNA repair. In an attempt to identify a function for this 'orphan' protein, we took into account the fact that it bound tightly to molecules that had been designed to inhibit DNA glycosylases. We therefore tested its ability to interact with several DNA lesions that are known substrates for DNA glycosylases. Specifically, a series of competition EMSA experiments were carried out in which the gel-shift complex — comprising a radiolabeled inhibitor and the TD12-encoded protein — was challenged by coincubation with an excess of unlabeled oligonucleotide competitors containing various alkylated, oxidized and deaminated DNA lesions (Figure 4a). As positive controls, two of the four inhibitors present in the original IVEC screening library paired to cytosine residues (rAb:C and Pyr:C) were also tested and, as expected, these were found to compete efficiently with the radiolabeled inhibitor for binding to the TD12-encoded protein. Of the base lesions tested, neither 8-oxoguanine paired to cytosine (^{O}G:C) or adenine (^{O}G:A), 8-oxoadenine paired to thymine (^{O}A:T), N7-methylguanine paired to cytosine (meG:C), N1,N^{6}-ethenoadenine paired to thymine (εA:T), nor thymine glycol paired to adenine (TG:A) showed an appreciable ability to compete with the radiolabeled inhibitor. But uracil paired to either guanine (U:G) or adenine (U:A) nearly abolished binding of the TD12-encoded protein to the radiolabeled inhibitor probe. Thus we concluded that the TD12-encoded protein is capable of recognizing uracil residues in DNA.

To determine whether the TD12-encoded protein catalyzes endonucleolytic cleavage at uracil residues in DNA, we performed DNA strand scission assays (Figure 4b). A radiolabeled oligonucleotide containing a single uracil residue was used in these assays, either alone (ssU) or base-paired to a complementary strand containing either adenine (U:A) or guanine (U:G) at the corresponding position on the opposite strand. One class of uracil-DNA glycosylases, typified by the human G:T glycosylase, is capable of processing both uracil and thymine residues mispaired to guanine; to test for such an activity in the TD12-encoded protein, we included a substrate containing a G:T mismatch that was radiolabeled on the T-containing strand. These substrates were incubated with either the *Xenopus in vitro* transcribed and translated TD12-encoded protein or its human counterpart, after which the DNA was treated with aqueous sodium hydroxide to effect strand scission at abasic sites. Both the human and the *Xenopus* proteins clearly processed uracil residues in the DNA, as evidenced by DNA strand scission at these positions (Figure 4b). The proteins acted upon U:A and U:G with no preference for either, but, interestingly, they processed uracil residues in single-stranded DNA much more efficiently than in double-stranded DNA. Neither the *Xenopus* nor human protein showed any enzymatic activity toward G:T mismatches.

To further characterize TD12, the *Xenopus* TD12-encoded protein was overexpressed in *Escherichia coli* with a carboxy-terminal hexahistidine tag and purified in a two-step chromatography procedure. The overexpressed protein contained significant uracil-excision activity, which could not be inhibited by the PBS1 bacteriophage uracil-glycosylase inhibitor protein, UGI (data not shown). As the endogenous *E. coli* uracil-DNA glycosylase (UDG)

Figure 4



Substrate specificity of the TD12-encoded protein. (**a**) Competition EMSA analyzing the effect of excess unlabeled competitors on the binding of a radiolabeled rAb-containing duplex to the TD12-encoded protein. Reaction products from *in vitro* transcription and translation of the TD12 clone were added to all lanes with the indicated competitors except 0, to which no protein was added; no competitor was added to the – lane. Abbreviations: ns, undamaged duplex DNA; rAb, reduced abasic; Pyr, pyrrolidine; ^{O}G, 8-oxoguanine; ^{O}A, 8-oxoadenine; meG, N7-methylguanine; εA, N1,N^{6}-ethenoadenine; TG, thymine glycol; U, uracil. (**b**) DNA cleavage assays using *in vitro* transcribed and translated reaction products generated from the *Xenopus* or human TD12 clone. The 25-mer band corresponds to the full-length DNA substrate, and the 12-mer represents the cleavage product generated by enzymatic processing followed by hydroxide-catalyzed strand scission. The four lanes on the left were from control *in vitro* transcription and translation reactions in which no cDNA was present. The radiolabel was present on the uracil-containing (ssU, U:A, U:G lanes) or thymine-containing (T:G lanes) strand. (**c**) The data from (b) are represented as the percentage of each DNA substrate processed by the TD12-encoded proteins.

is potently inhibited by UGI, the single-stranded DNA glycosylase activity observed in the presence of UGI must derive from the *Xenopus* protein. All assays using the recombinant *Xenopus* enzyme were therefore run in the presence of UGI to abolish any contaminating *E. coli* UDG activity.

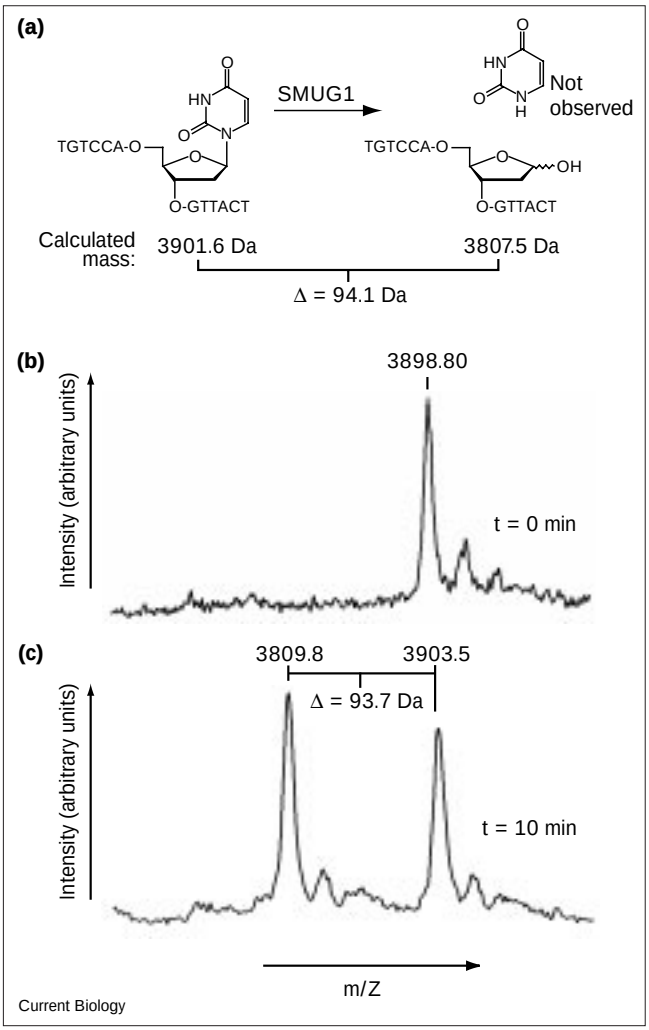
The DNA strand scission assay shown in Figure 4b does not differentiate between the possible modes of enzyme action. One possibility is that the enzyme simply catalyzes phosphodiester bond cleavage at the uracil-containing site. A second, more likely, possibility is that the enzyme acts as a DNA glycosylase, catalyzing scission of the glycosidic bond linking the uracil base to its sugar moiety. Two different processing mechanisms are employed by the two known classes of DNA glycosylases: monofunctional DNA glycosylases use an activated water molecule to displace the aberrant base, leaving behind an abasic site; whereas bifunctional DNA glycosylases/lyases use an activated amine moiety on the enzyme for the initial displacement, and then catalyze subsequent degradations of the enzyme-linked sugar, ultimately leading to DNA-strand cleavage [30–32]. To distinguish unambiguously between these modes of action, we used the recombinant *Xenopus* TD12-encoded enzyme to process a single-stranded 13-mer substrate, the products of which were directly analyzed by MALDI-TOF mass spectrometry. The use of a single-stranded substrate and the UGI inhibitor in the reaction buffer eliminated any activity arising from possible contaminant *E. coli* MUG and UDG, respectively. In a time-dependent manner, the approximately 3,900 Da peak arising from the starting 13-mer was replaced by a single product peak at 3,809.8 Da (Figure 5b,c), the expected mass for an abasic site product. The observed decrease in mass of approximately 94 Da corresponds to the loss of uracil (111 Da) and the gain of a hydroxyl group (17 Da). Identical results were obtained with the known monofunctional glycosylase *E. coli* UDG (data not shown). We thus conclude that the TD12-encoded protein acts as a monofunctional DNA glycosylase specific for uracil residues in DNA and has a preference for single-stranded DNA substrates. On the basis of these functional characteristics, we designated the enzyme as single-strand-selective monofunctional uracil-DNA glycosylase, SMUG1.

The preference of SMUG1 for single-stranded DNA was further characterized through measurement of Michaelis–Menten kinetics using recombinant *Xenopus* SMUG1 (xSMUG1) with a single-stranded uracil-containing oligonucleotide substrate (ssU), or with the same strand base-paired to its complement (U:G mismatched duplex) (Table 1). The value of k_{cat} for the processing of the single-stranded substrate by xSMUG1 was more than 1,500-fold greater than the k_{cat} for the processing of the duplex DNA substrate by xSMUG1. The K_m for the duplex substrate was about 20-fold lower than the K_m for the single-stranded DNA substrate, however, suggesting that the protein binds double-stranded DNA more tightly than single-stranded. Thus, the overall catalytic efficiency (k_{cat}/K_m) of SMUG1 is about 60-fold greater for a single-stranded substrate than for a duplex substrate.

SMUG1 is localized to the nucleus

In order to function as a DNA repair protein, SMUG1 should co-localize to the same intracellular compartments

Figure 5



Products of the reaction catalyzed by the TD12-encoded enzyme SMUG1. (a) Calculated masses of the 13-mer single-stranded DNA substrate containing a single, centrally located uracil residue, and of the DNA product generated by hydrolysis of the N-glycosidic bond to this uracil residue. Representative MALDI-TOF spectra of (b) the starting 13-mer and (c) the product mixture obtained by incubation with the recombinant TD12-encoded protein for 10 min. Relative masses in MALDI-TOF spectra are more accurately determined than absolute masses.

in which DNA is located. In the case of mammalian UDG, differential RNA splicing directs two variant forms to either the nucleus or mitochondria [33]. To ascertain the sub-cellular localization of the human SMUG1 (hSMUG1), the cDNA was fused in-frame to green fluorescent protein (GFP) and transiently transfected into HeLa cells, which were then analyzed by confocal microscopy. As shown in Figure 6, the hSMUG1–GFP fusion protein localizes predominantly to the nucleus, whereas GFP alone is distributed throughout the cell. These initial experiments could not rule out the possibility that a small fraction of hSMUG1 also localizes to the mitochondria.

Discussion

In this work, we have used IVEC to isolate a protein that binds a library of four synthetic DNA repair inhibitors. Biochemical characterization of this protein revealed that it catalyzes excision of uracil residues from DNA through hydrolysis of the N-glycosidic bond. Given the appreciable preference of the protein for uracil residues in single-stranded DNA rather than double-stranded DNA, we have designated it as single-strand-selective monofunctional uracil-DNA glycosylase, or SMUG1.

The proteomics-based approach used to isolate SMUG1 offers an alternative to the genetic and biochemical strategies employed in the past. Although of proven value, gene-based approaches relying on the reversion of mutator phenotypes or on homology to known DNA repair activities sometimes fail for subtle phenotypes or in cases of insufficient sequence similarity. Protein-based strategies employing bioassay-directed purification have also figured prominently in the isolation of DNA repair proteins, but these can often be labor-intensive and expensive, especially for mammalian proteins present in low abundance. IVEC, the proteomics procedure employed here, offers a rapid, systematic screen of all expressed proteins on the basis of their biochemical activity. Because this functional property is linked throughout the assay to its genetic encoding, IVEC enables regeneration and amplification of the protein through successive rounds of screening and directly yields a cDNA clone encoding the activity of interest. Whereas traditional isolation methods may involve screening mixtures comprising tens of thousands of proteins or genes, IVEC reduces the number being evaluated simultaneously in any one test tube, because the cDNA pools being transcribed and translated *in vitro* have been pre-sorted to contain approximately 100 clones or less. A general concern regarding IVEC might be that random cDNA sorting could lead to separation of protein subunits that are required together for function; in the present case, however, we were aware that most, if not all, known base-excision DNA repair proteins are monomeric.

IVEC with mechanism-based inhibitors

In previous studies, IVEC has been used to isolate substrates for mitotic kinases [17], for mitotic protein degradation pathways [18] and for apoptotic proteases [19,20]. This method has also been used to isolate Mix-related sequence-specific DNA-binding proteins, using an assay method much like the one employed here [21]. The present report, however, is the first in which IVEC has been used to isolate an enzyme on the basis of its activity. This application presented a unique challenge, namely how to detect the interaction of an enzyme with a substrate, when the two are ordinarily bound for only the

Table 1

Michaelis–Menten kinetic parameters for uracil removal by recombinant *Xenopus* SMUG1, human UDG and human TDG.

Enzyme	Substrate	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} M^{-1})	Reference
xSMUG1	ssU	1.09	2.5	2.3×10^6	This work
xSMUG1	U:G	0.035	0.0014	3.9×10^4	This work
hUDG	ssU	0.40	9.1	2.3×10^7	[66]*
hUDG	U:G	10016	4.6	2.8×10^6	[66]*
hTDG	U:G	0.012	0.00035	2.9×10^4	[50]†

*The values listed here (units converted) are for 10 mM NaCl, the concentration of salt that gave the maximal k_{cat}/K_m. †See also [67].

fleeting instant during which catalysis takes place. To solve this problem, we turned to mechanism-based enzyme inhibitors, which can be recognized specifically but not processed by DNA repair enzymes; these characteristics enabled a screen based on the formation of long-lived enzyme–inhibitor complexes to be performed. Such inhibitors had not previously been employed in expression-based cloning of DNA repair enzymes, but had found extensive use in affinity purification, mechanistic studies and high-resolution structural analysis [15,34].

Although one class of such inhibitors contains an attached base, which confers exquisite selectivity for cognate DNA repair enzymes, the four inhibitors that comprised our screening library belong to a second class, which lacks a base and is therefore less restricted in specificity for target proteins. Notwithstanding this broad substrate scope, the highly stringent conditions employed in our screen led to the isolation of a single positive clone, TD12, through screening 120 pools (approximately 1% coverage of the total library). Thus, while it remains to be demonstrated experimentally, one would expect that further screening of the *Xenopus* cDNA library with the abasic inhibitors would identify several other DNA glycosylases. These might include well-characterized enzymes that are known to be widely distributed throughout evolution, such as those that process 8-oxoguanine (xOgg1) and 3-methyladenine (xAag); the orthologous human glycosylases bind tightly to inhibitors present in our screening panel [35,36].

SMUG1 is a uracil-DNA glycosylase enzyme

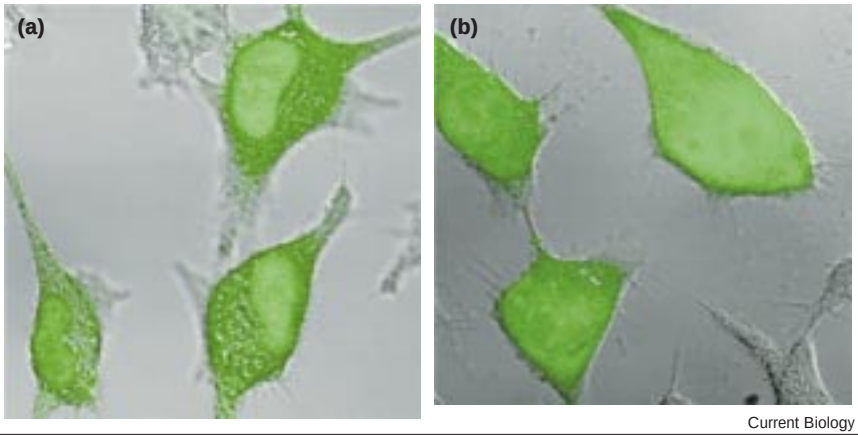
Deamination of cytosine residues generates uracil, which can cause mutations. The biochemical data presented here demonstrate that SMUG1 can initiate the repair of these mutagenic lesions by catalyzing excision of uracil from DNA. Several other vertebrate proteins have been shown or suggested to have uracil-DNA glycosylase activity: the predominant enzyme in vertebrates, UDG, is similar to SMUG1 in its exquisite specificity for uracil over thymine and its preference for single-stranded DNA (Table 1). A more recently described enzyme, thymine-DNA glycosylase (TDG), shows less preference for uracil than for thymine, but does exhibit apparently absolute specificity for double-stranded substrates, especially those containing U:G and T:G mismatches (Table 1). The catalytic efficiency (k_{cat}/K_m) of SMUG1 is 10-fold lower than UDG for single-stranded substrates and about 70-fold lower for double-stranded substrates; on the other hand, SMUG1 is nearly identical to TDG in its efficiency for repairing double-stranded DNA (Table 1). Reports of uracil-DNA glycosylase activity in two other polypeptides [37,38] remain unconfirmed and the proteins have not been sufficiently characterized to allow detailed comparison with SMUG1.

Although we have not undertaken an exhaustive search for potential SMUG1 substrates, the following properties of the enzyme are fully consistent with uracil being a physiological substrate of this enzyme. First, the catalytic efficiency of SMUG1 for uracil in single-stranded DNA compares favorably with that of most DNA glycosylases operating on their cognate substrates [39–42] and is within an order of magnitude of the exceptionally efficient human UDG. Second, the catalytic efficiency of xSMUG1 for double-stranded DNA, its less preferred substrate, is still above the well-characterized TDG (Table 1) on its optimal substrates. Third, the preference to process single-stranded DNA is a unique and characteristic property of UDGs from many organisms, and SMUG1 is the only other DNA glycosylase known to exhibit this property. Finally, SMUG1 activity is virtually abrogated by changing uracil to thymine — the addition of a single methyl group. This property is essential for an enzyme that can process single-stranded DNA.

Mouse cells bearing a homozygous targeted disruption of the gene encoding UDG retain residual uracil-DNA glycosylase activity attributable to the product of another gene (T. Lindahl, personal communication). SMUG1 may be responsible for this activity, as may TDG. The differences in biochemical activity of these two candidates should make it straightforward to differentiate between the two.

Figure 6

Intracellular localization of SMUG1. Fluorescence data from confocal microscopy were superimposed on differential interference contrast images to produce the composite images shown. (a) HeLa cells transfected with a hSMUG1–GFP expression construct. (b) Control HeLa cells transfected with a GFP expression construct.



Two critical motifs are present in SMUG1

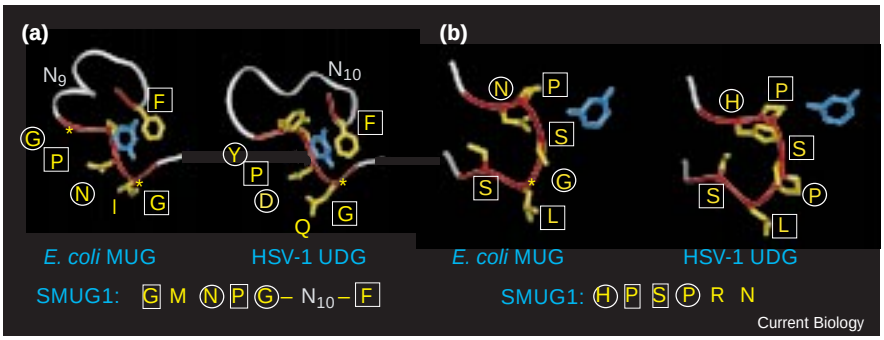
Of all the base-excision DNA repair proteins known, UDGs stand alone as the most highly conserved, being the only DNA glycosylases encoded by the primitive bacterium *Mycoplasma genitalium* [43] and even being encoded by certain viruses [44]. Despite the evolutionary space separating these enzymes, their protein sequences are sufficiently homologous to permit ready alignment and to allow one to infer a high degree of structural similarity. Indeed, the X-ray structure of the human UDG is nearly identical to that of the herpes simplex virus-1 (HSV-1) enzyme, as would be expected on the basis of the 39% sequence identity between the two. Given the striking parallels between the biochemical properties of these UDGs and SMUG1, it is perplexing to note that there is a virtually complete absence of statistically significant sequence similarity between these two classes of uracil-excising enzymes. For example, the sequences of hSMUG1 and hUDG are less than 8% identical to one another, far below the level of statistical significance.

A similar but perhaps less compelling paradox had been raised by the discovery of a class of uracil-DNA glycosylases, exemplified by TDG, members of which are specific for U:G mismatches but are less than 10% homologous to UDGs. The dramatic resolution to this TDG–UDG paradox came with the X-ray structure of a bacterial U:G-mismatch-specific uracil-DNA glycosylase, MUG, bound to DNA. Comparison of the MUG–DNA structure [45] with that of HSV-1 UDG [46] bound to DNA revealed that the two protein folds are virtually superimposable on one another (root-mean-square deviation, 3.4 Å), in a striking example of structural conservation in the absence of sequence conservation. Furthermore, an unpublished structure of MUG bound to an inhibitor-containing duplex (R. Savva, O.D. Schärer, G.L.V. and L. Pearl, unpublished observations) reveals that the two proteins interact in fundamentally similar ways with the uracil base and DNA backbone. Comparison of these structures led to the identification of two active-site motifs (Figure 7), in which residues that are particularly important for substrate recognition and catalysis are located. Inspection of the SMUG1 sequence reveals that it too contains these two motifs (see Figure 3).

Motif I in *E. coli* MUG has the sequence GINPG-N_9-F (single-letter amino-acid code). Crystallographic evidence strongly implicates the asparagine at the third position as the residue that positions and perhaps activates the nucleophilic water molecule for attack on the N-glycosidic bond, and the terminal phenylalanine stacks up against the uracil base (Figure 7a). In HSV-1 UDG, motif I has the sequence GQDPY-N_{10}-F. In this case, it is an aspartic acid rather than an asparagine residue at the third position that activates the water molecule. Furthermore, the presence of a tyrosine at position 5 of the motif sterically blocks the thymine methyl group from binding in the enzyme active site; MUG, which contains a glycine at this position, binds both uracil and thymine. Motif I of SMUG1 has the sequence GMNPG-N_{10}-F, which is closer to that of MUG than that of UDG. Interestingly, SMUG1 is able to discriminate strongly against thymine, despite the presence of a glycine at position 5 of motif I.

The structure of motif II is shown in Figure 7b. In the sequence of MUG motif II, NPSGLS, noteworthy residues are: the asparagine at position 1, which forms hydrogen bonds, either directly or through a tightly bound water molecule, with the uracil base; and the glycine at position 4, which forms hydrogen bonds with the guanine residue of the U:G mismatch. In motif II of HSV-1 UDG, the asparagine is replaced by a histidine having the same presumed function, and the glycine residue is replaced by a proline, which is consistent with the lack of specificity in UDG for the base opposite uracil. Motif II of SMUG1 has the sequence HPSPRN, which is more closely related to

Figure 7



SMUG1 contains two active-site motifs that are characteristic of uracil-DNA glycosylases. (a) Motif I; (b) Motif II. Red and gray tubes represent the main-chain backbone of *E. coli* MUG (R. Savva, O.D. Schärer, G.L.V. and L. Pearl, unpublished observations) and HSV-1 UDG [46]. Certain key side-chains are shown in yellow licorice, and the uracil base of the substrate is shown in blue. Identical residues are boxed. Circled residues denote functionally important residues that differ between the two motifs; see the text for specific details.

UDG than to MUG. Thus, with regard to active-site functionality, SMUG1 can be viewed as a chimera of MUG and UDG, having a motif I that is MUG-like and a motif II that is UDG-like.

A unique role for SMUG1?

The demonstration that SMUG1 is a kinetically competent uracil-DNA glycosylase raises the question as to why vertebrate cells contain at least three such enzymes. Because each of the three enzymes exhibits unique biochemical properties, it seems unlikely that they serve redundant functions. It is more likely that each uracil-DNA glycosylase has evolved to fulfil a specific role in protecting vertebrate cells from the deleterious effects of cytosine deamination. Owing to its abundance and high kinetic efficiency on both single- and double-stranded DNA, UDG is probably responsible for initiating the repair of most uracil residues, at least in cycling cells [47]. The strong preference of TDG for U:G and T:G mismatches, wherein the substrate base is flanked on the 3′ side by a guanine [48], has led to the proposal that this enzyme is primarily dedicated to the repair of CpG sites that have undergone methylation followed by deamination or methyltransferase-catalyzed deamination [49]. The non-redundant function of uracil-DNA glycosylases may extend to the processing of lesions other than uracil; for example, human TDG has been found to process N3,N^4-ethenocytosine (εC) almost as well as U:G and better than T:G [50].

What then is the unique role of SMUG1 in the repair of uracil residues? It is not possible at present to answer this question conclusively, but the appreciable selectivity of SMUG1 for single-stranded DNA may provide a clue. It is possible that the enzyme might be involved in surveillance of transiently single-stranded intermediates in the cells, such as those that arise during transcription, recombination and replication; in this regard, it is noteworthy that the rate of cytosine deamination in single-stranded DNA is more than 100-fold greater than that for the duplex form of DNA [51,52]. It is also attractive to speculate that SMUG1 might act in concert with an ATP-dependent motor protein that unpairs the DNA, thereby presenting single-stranded DNA to the SMUG1 active site. Whether the endogenous substrate of SMUG1 is single-stranded or duplex DNA, it is possible that the enzyme functions as a dedicated repair subunit of a larger multiprotein complex, as has been suggested for UDG and the replication machinery [53]. The fact that SMUG1 was not isolated in earlier biochemical studies of uracil-DNA glycosylase activity in extracts might be taken to suggest that SMUG1 has a highly restricted pattern of expression, but this is inconsistent with the identification of hSMUG1 ESTs in cDNA libraries derived from a wide variety of tissues, including breast, ovary, lung, placenta, prostate, brain, liver, heart, kidney, colon and whole embryos.

Conclusions

We have described the isolation and cloning of a unique uracil-DNA glycosylase enzyme, SMUG1, on the basis of its biochemical activity. Kinetic measurements and product analysis by mass spectrometry reveal SMUG1 to be a monofunctional DNA glycosylase, specific for uracil residues and selective for single-stranded DNA. The discovery of this enzyme presents an especially compelling argument for proteomics-based gene discovery, because the level of sequence similarity between SMUG1 and its functional relatives is too low (less than 8%) to be detected by current computer algorithms that detect similarity in either sequence or structure. The observation that three classes of uracil-DNA glycosylases — UDG, TDG/MUG and now SMUG — have arrived at a similar function and (presumably, in the case of SMUG1) structure, despite such remarkable dissimilarity in sequence, highlights the relevance of function-based genome-wide screening technology, in addition to raising fascinating questions about convergency versus divergency in the evolution of these enzymes.

Materials and methods

Oligonucleotide probes and substrates

Oligonucleotides were synthesized and purified by standard methods, and radiolabeled using T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]ATP (New England Nuclear). The phosphoramidite required for incorporation of uracil into DNA was purchased from Glen

Table 2

Oligonucleotide sequences.

ss dU	5´ GGATAGTGTCCA U GTTACTCGAAGC 3´
Duplex	5´ GGATAGTGTCCA X GTTACTCGAAGC 3´ 3´ CCTATCACAGGT Y CAATGAGCTTCG 5´
U:G	X=U Y=G
U:A	X=U Y=A
THF:C	X=THF Y=C
roPyr:C	X=roPyr Y=C
^{O}G:C	X=^{O}G Y=C
^{O}G:A	X=^{O}G Y=A
^{O}A:T	X=^{O}A Y=T
εA:T	X=εA Y=T
ns	X=A Y=T
rAb:C	5´ GTGATCCTGAGC rAb TAGCTCAGTAAC 3´ 3´ CACTAGGACTCG C ATCGAGTCATTG 5´
dG:dT	5´ GGATAGTGTCCAA G TTACTCGAAGC 3´ 3´ CCTATCACAGGTT T AATGAGCTTCG 5´
meG:C	5´ CGC meG AATT C GCG 3´ 3´ GCG C TTAA meG CGC 5´
TG:A	5´ CGCAG TG CAGCC 3´ 3´ GCGTC A GTCGG 5´

Wherever relevant, the radiolabeled strand is the upper one in the sequences shown.

Research. The phosphoramidites for THF [54] and Pyr [55] were prepared by published methods. The rAb phosphoramidite was a gift of Huw M. Nash (H.M.N. and G.L.V., unpublished) and the roPyr phosphoramidite was a gift from Steve D. Bruner (S.D.B. and G.L.V., unpublished). Oligonucleotides containing the base lesions ^{O}G [56,57], ^{O}A [58], meG [59], TG [60] and εA [61] were prepared by published methods. Radiolabeled duplex oligonucleotides (Table 2) were prepared by annealing the radiolabeled strand with a 10-fold molar excess of the complementary strand in annealing buffer (1 × TE, 100 mM NaCl). Non-radiolabeled duplex DNA competitors were made by annealing equimolar amounts of the two complementary strands.

Electrophoretic mobility shift assays

Radiolabeled inhibitor-containing duplex DNA substrates (5 fmol in 1 µl annealing buffer each) were mixed with 5 µl of *in vitro* transcription/translation reaction mixture (TNT Coupled Reticulocyte Lysate System, Promega) programmed with either pooled or single cDNA clones as described [16]. The mixture was adjusted to a final volume of 20 µl in 1 × EMSA buffer (50 mM Tris, 250 mM NaCl, 3 mM EDTA, 500 µM DTT, 5% glycerol, 100 µg/ml BSA, pH 7.4) by appropriate additions of 5 × EMSA buffer and deionized, distilled water. The mixture was allowed to incubate for 20 min at room temperature and was then loaded onto a pre-run 10% native polyacrylamide gel (Protogel, National Diagnostics). The gel was run at 250 V for 1.5 h, then dried and exposed to a phosphorimaging plate (Fuji BAS 1000 s). Competition EMSA assays followed a similar protocol, except that a 100-fold molar excess of unlabeled competitor was mixed together with the radiolabeled inhibitor (10 fmol, rAb) immediately prior to incubation with the *in vitro* transcription/translation products.

DNA cleavage assays

Radiolabeled substrates were incubated at 37°C in 20 µl of 1 × UDG reaction buffer (20 mM Tris, 1 mM EDTA, 1 mM DTT, 50 mM NaCl, 0.5 U uracil-glycosylase inhibitor (New England Biolabs), 100 mg/ml BSA, pH 8.0) for 10 min with varying amounts of either *in vitro* transcription/translation reaction mixture (programmed with either the xSMUG1 or hSMUG1 cDNA) or purified recombinant protein (xSMUG1). The reaction was stopped, cleavage of abasic sites was effected by the addition of 15 µl stop solution (70% formamide, 0.3 M NaOH, 1 × TBE), and the mixture was then heated at 95°C for 15 min. The reaction mixtures were then loaded onto a pre-run 20% denaturing polyacrylamide gel (Sequagel, National Diagnostics). The gel was run at 250 V for 4 h and imaged (wet) using a phosphorimaging plate (Fuji BAS 1000 s). Band intensities were quantified using the MacBas software package (Fuji). Michaelis–Menten parameters were determined by measurement of initial rates (< 30% conversion) at various concentrations of substrate (250–3000 nM, ssU; 12.5–500 nM, U:G), followed by non-linear regression analysis using the software program EnzymeKinetics (Trinity Software).

Bacterial overexpression and purification of xSMUG1

The xSMUG1 coding sequence, PCR-amplified from the xTD12 cDNA, was cloned into pET30b (Novagen) as a *Nde*I–*Kpn*I fragment so as to express the protein as a fusion containing a carboxy-terminal hexahistidine tag. The resulting expression construct (pET30b-xSMUG1) was transformed into *E. coli* BL21-DE3 cells (Novagen) and expression was induced in mid-log phase by the addition of ispropyl-β-D thiogalactopyranoside to 1 mM. The cells were induced for 4 h at 30°C and then harvested by centrifugation. The induced cell pellet was resuspended in buffer A (30 mM Tris, 250 mM ammonium acetate, 10 mM 2-mercaptoethanol, pH 8.0) containing a cocktail of standard protease inhibitors. The cells were lysed in a French press and the resulting lysate was clarified by centrifugation. To the supernatant was added 2 ml Talon resin (Clontech), preequilibrated in buffer A and Triton X-100 (1% v/v). The cell lysate and Talon resin were allowed to mix for 30 min using a mechanical rotary platform and then transferred to a disposable gravity-flow column. The resin was washed with three column volumes of buffer A and one volume of 10 mM imidazole in buffer A, and then eluted with 200 mM imidazole in buffer A. The eluate was exchanged into buffer B (30 mM Tris, 50 mM ammonium acetate, 10 mM 2-mercaptoethanol, pH 8.0) by centrifugal dialysis (Centriprep-10, Amicon) and then loaded onto a MonoQ FPLC column (Pharmacia), washed with buffer B, and then eluted with a linear gradient of elution buffer (30 mM Tris, 10 mM 2-mercaptoethanol, pH 8.0) from 0 to 1 M ammonium acetate over 20 ml. Fractions containing xSMUG1 eluted at ~300 mM ammonium acetate. After concentration by centrifugal dialysis (Centriprep-10), the combined xSMUG1-containing fractions were supplemented with DTT (1 mM final concentration) and stored at 4°C. xSMUG1 can be stored under these conditions for up to one month without any significant loss of activity.

MALDI-TOF mass spectrometry

The 13-mer dU substrate (300 pmol) was incubated with 500 ng recombinant xSMUG1 in 20 µl of 1 × UDG reaction buffer at 37°C. At various time points, 1 µl aliquots were removed and added to 1 µl reaction-quenching matrix (2 parts 2,4,6-trihydroxyacetophenone (300 mM in ethanol), 1 part diammonium citrate (100 mM in water), and 1 part acetonitrile). MALDI-TOF mass spectra were acquired as described [62] using a Bruker BIFLEX spectrometer.

Green fluorescent protein fusion

The hSMUG1 coding sequence, PCR-amplified from the hTD12 cDNA, was cloned into pEGFP-N1 (Clontech) as an *Eco*RI–*Bam*HI fragment to express the protein as a fusion containing a carboxy-terminal green fluorescent protein tag. Transient transfection of cultured adherent HeLa cells grown with coverslips was done with TransfectAMINE (GIBCO-BRL). At 36 h post-transfection, samples were fixed and mounted on slides. Images were obtained using a Zeiss LSM410 confocal microscope equipped with a Krypton/Argon laser.

Expressed sequence tags and cloning hSMUG1

The full-length hSMUG1 cDNA clone (ATCC clone number 101881), a product of the EST sequencing effort of the Institute for Genomic

Research [63], was purchased from ATCC. The cDNA clones with IMAGE clone numbers 415608 and 726197 were products of the IMAGE Consortium sequencing effort [64] and were purchased from Research Genetics. The clone 415608 contained the complete coding sequence, but lacked a portion of the 5′ untranslated region. The clone 726197 contained a portion of the 5′ untranslated region and a portion of the coding sequence. The hSMUG1 coding sequence, PCR-amplified from the clone 415608, was cloned into pCS2 [65] as a *Eco*RI–*Xba*I fragment so as to allow expression by *in vitro* transcription/translation.

GenBank accession numbers

The sequence data for xSMUG1 and hSMUG1 have been submitted to the GenBank database under accession numbers AF125181 and AF125182, respectively.

Acknowledgements

We thank Randy King (Harvard Institute of Chemistry and Cell Biology) and the members of the Verdine and Kirschner labs, especially Steve Bruner, Sheldon Park, Huw Nash and Rongzhen Lu for advice and helpful discussions. We thank Kevin Lustig for providing the subdivided *Xenopus* cDNA library, Clemens Richert and David Sarracino (Tufts University) for expert assistance with mass spectral analyses and Tomohiko Kawate for providing specialty DNA synthesis reagents. We thank Renos Savva and Laurence Pearl for sharing the coordinates of the MUG–FdU complex structure prior to publication. K.A.H. was supported by a predoctoral fellowship from the National Science Foundation, P.T.S. by a postdoctoral fellowship from the Fleet Investment Management, Trustee of the Charles A. King Trust. Funding for this research was provided by the National Institute of General Medical Sciences (Grant number 5 R01 GM 26875 to M.W.K. and GM 51330 to G.L.V.).

References

1. Friedberg EC, Walker GC, Siede W: *DNA Repair and Mutagenesis.* Washington, D.C.: ASM Press; 1995.
2. Lindahl T: Instability and decay of the primary structure of DNA. *Nature* 1993, 362:709-715.
3. Sancar A: DNA excision repair. *Annu Rev Biochem* 1996, 65:43-81.
4. Hickson ID: Base excision repair of DNA damage. In *Molecular Biology Intelligence Unit.* Austin, Texas: Landes Bioscience, 1997.
5. Baumann P, West SC: Role of the human RAD51 protein in homologous recombination and double-stranded-break repair. *Trends Biochem Sci* 1998, 23:247-251.
6. Modrich P, Lahue R: Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu Rev Biochem* 1996, 65:101-133.
7. Sancar A: No “End of History” for photolyases. *Science* 1996, 272:48-49.
8. Lindahl T, Sedgwick B, Sekiguchi M, Nakabeppu Y: Regulation and expression of the adaptive response to alkylating agents. *Annu Rev Biochem* 1988, 57:133-157.
9. Wood RD: DNA repair in eukaryotes. *Annu Rev Biochem* 1996, 65:135-167.
10. Cunningham RP: DNA Glycosylases. *Mutat Res* 1997, 383:189-196.
11. Krokan HE, Standal R, Slupphaug G: DNA glycosylases in the base excision repair of DNA. *Biochem J* 1997, 325:1-16.
12. Kubota Y, Nash RA, Klungland A, Schar P, Barnes DE, Lindahl T: Reconstitution of DNA base excision-repair with purified human proteins – interaction between DNA polymerase beta and the Xrcc1 protein. *EMBO J* 1996, 15:6662-6670.
13. Nicholl ID, Nealon K, Kenny MK: Reconstitution of human base excision repair with purified proteins. *Biochemistry* 1997, 36:7557-7566.
14. Klungland A, Lindahl T: Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirements for DNase IV (FEN1). *EMBO J* 1997, 16:3341-3348.
15. Schärer OD, Deng L, Verdine GL: Chemical approaches toward understanding base excision DNA repair. *Curr Opin Chem Biol* 1997, 1:526-531.
16. Lustig KD, Stukenberg PT, McGarry TJ, King RW, Cryns VL, Mead PE, *et al.*: Small pool expression screening: identification of genes involved in cell cycle control, apoptosis, and early development. *Methods Enzymol* 1997, 283:83-99.
17. Stukenberg PT, Lustig KD, McGarry TJ, King RW, Kuang J, Kirschner MW: Systematic identification of mitotic phosphoproteins. *Curr Biol* 1997, 7:338-348.
18. McGarry TJ, Kirschner MW: Geminin, an inhibitor of DNA replication, is degraded during mitosis. *Cell* 1998, 93:1043-1053.
19. Kothakota S, Azuma T, Reinhard C, Klippel A, Tang J, Chu K, *et al.*: Caspase-3-generated fragment of gelsolin: effector of morphological change in apoptosis. *Science* 1997, 278:294-298.
20. Cryns VL, Byun Y, Rana A, Mellor H, Lustig KD, Ghanem L, *et al.*: Specific proteolysis of the kinase protein kinase C-related kinase 2 by caspase-3 during apoptosis. Identification by a novel, small pool expression cloning strategy. *J Biol Chem* 1997, 272:29449-29453.
21. Mead PE, Zhou Y, Lustig KD, Huber TL, Kirschner MW, Zon LI: Cloning of mix-related homeodomain proteins using fast retrieval of gel shift activities, (FROGS), a technique for the isolation of DNA-binding proteins. *Proc Natl Acad Sci USA* 1998, 95:11251-11256.
22. Schärer OD, Nash HM, Jiricny J, Laval J, Verdine GL: Specific binding of a designed pyrrolidine abasic site analog to multiple DNA glycosylases. *J Biol Chem* 1998, 273:8592-8597.
23. Lustig KD, Kirschner MW: Use of an oocyte expression assay to reconstitute inductive signaling. *Proc Natl Acad Sci USA* 1995, 92:6234-6238.
24. Kim K, Biade S, Matsumoto Y: Involvement of flap endonuclease 1 in base excision DNA repair. *J Biol Chem* 1998, 273:8842-8848.
25. Pinz KG, Bogenhagen DF: Efficient repair of abasic sites in DNA by mitochondrial enzymes. *Mol Cell Biol* 1998, 18:1257-1265.
26. Matsumoto Y, Kim K: Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. *Science* 1995, 269:699-702.
27. Matsumoto Y, Kim K, Bogenhagen DF: Proliferating cell nuclear antigen-dependent abasic site repair in ***Xenopus laevis*** oocytes: an alternative pathway of base excision DNA repair. *Mol Cell Biol* 1994, 14:6187-6197.
28. Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, *et al.*: A gene map of the human genome. *Science* 1996, 274:540-546.
29. Deloukas P, Schuler GD, Gyapay G, Beasley EM, Soderlund C, Rodriguez-Tome P, *et al.*: A physical map of 30,000 human genes. *Science* 1998, 282:744-746.
30. Sun B, Latham KA, Dodson ML, Lloyd RS: Studies on the catalytic mechanism of five DNA glycosylases. Probing for enzyme-DNA imino intermediates. *J Biol Chem* 1995, 270:19501-19508.
31. Zharkov DO, Rieger RA, Iden CR, Grollman AP: NH2-terminal proline acts as a nucleophile in the glycosylase/AP-lyase reaction catalyzed by ***Escherichia coli*** formamidopyrimidine-DNA glycosylase (Fpg) protein. *J Biol Chem* 1997, 272:5335-5341.
32. Nash HM, Lu R, Lane WS, Verdine GL: The critical active-site amine of the human 8-oxoguanine DNA glycosylase, hOgg1: direct identification, ablation and chemical reconstitution. *Chem Biol* 1997, 4:693-702.
33. Nilsen H, Otterlei M, Haug T, Solum K, Nagelhus TA, Skorpen F, *et al.*: Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the Ung gene. *Nucleic Acids Res* 1997, 25:750-755.
34. Lau AY, Schärer OD, Samson L, Verdine GL, Ellenberger TE: Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* 1998, 95:249-258.
35. Labahn J, Schärer OD, Long A, Ezaz-Nikpay K, Verdine GL, Ellenberger TE: Structural basis for the excision repair of alkylation-damaged DNA. *Cell* 1996, 86:321-329.
36. Lu R, Nash HM, Verdine GL: A mammalian DNA repair enzyme that excises oxidatively damaged guanines maps to a locus frequently lost in lung cancer. *Curr Biol* 1997, 7:397-407.
37. Caradonna S, Ladner R, Hansbury M, Kosciuk M, Lynch F, Muller S: Affinity purification and comparative analysis of two distinct human uracil-DNA glycosylases. *Exp Cell Res* 1996, 222:345-359.
38. Meyer-Siegler K, Mauro DJ, Seal G, Wurzer J, DeRiel JK, Sirover MA: A human nuclear uracil DNA glycosylase is the 37-kDa subunit of glyceraldehyde-3-phosphate dehydrogenase. *Proc Natl Acad Sci USA* 1991, 88:8460-8464.
39. Wang D, Essigmann JM: Kinetics of oxidized cytosine repair by endonuclease III of ***Escherichia coli***. *Biochemistry* 1997, 36:8628-8633.
40. Roy R, Brooks C, Mitra S: Purification and biochemical characterization of recombinant N-methylpurine-DNA glycosylase of the mouse. *Biochemistry* 1994, 33:15131-15140.

41. Tchou J, Bodepudi V, Shibutani S, Antoshechkin I, Miller J, Grollman AP, *et al.*: Substrate specificity of Fpg protein. Recognition and cleavage of oxidatively damaged DNA. *J Biol Chem* 1994, 269:15318-15324.
42. O'Connor TR: Purification and characterization of human 3-methyladenine-DNA glycosylase. *Nucleic Acids Res* 1993, 21:5561-5569.
43. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, *et al.*: The minimal gene complement of Mycoplasma genitalium. *Science* 1995, 270:397-403.
44. Krokan HE, Standal R, Bharati S, Otterlei M, Haug T, Slupphaug G, *et al.*: Uracil in DNA and the family of conserved uracil DNA glycosylases. In *Base Excision Repair of DNA Damage*. Edited by I.D. Hickson. Austin, Texas: Landes Bioscience; 1997.
45. Barrett TE, Savva R, Panayotou G, Barlow T, Brown T, Jiricny J, *et al.*: Crystal structure of a G-T/U mismatch-specific DNA glycosylase – mismatch recognition by complementary-strand interactions. *Cell* 1998, 92:117-129.
46. Savva R, McAuley-Hecht K, Brown T, Pearl L: The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* 1995, 373:487-493.
47. Slupphaug G, Olsen LC, Helland D, Aasland R, Krokan HE: Cell-cycle regulation and ***in vitro*** hybrid arrest analysis of the major human uracil-DNA glycosylase. *Nucleic Acids Res* 1991, 19:5131-5137.
48. Sibghat U, Gallinari P, Xu YZ, Goodman MF, Bloom LB, Jiricny J, *et al.*: Base analog and neighboring base effects on substrate specificity of recombinant human G:T mismatch-specific thymine DNA-glycosylase. *Biochemistry* 1996, 35:12926-12932.
49. Lettieri T, Jiricny J: Mammalian mismatch-specific DNA-glycosylases. In *Base Excision Repair of DNA Damage*. Edited by I.D. Hickson. Austin, Texas: Landes Bioscience; 1997.
50. Saparbaev M, Laval J: 3,N4-ethenocytosine, a highly mutagenic adduct, is a primary substrate for the ***Escherichia coli*** double-stranded uracil-DNA glycosylase and the human mismatch-specific thymine-DNA glycosylase. *Proc Natl Acad Sci USA* 1998, 95:8508-8513.
51. Frederico LA, Kunkel TA, Shaw BR: A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry* 1990, 29:2532-2537.
52. Lindahl T, Nyberg B: Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry* 1974, 13:3405-3410.
53. Nagelhus TA, Haug T, Singh KK, Keshav KF, Skorpen F, Otterlei M, *et al.*: A sequence in the N-terminal region of human uracil-DNA glycosylase with homology to Xpa interacts with the C-terminal part of the 34-KDa subunit of replication protein A. *J Biol Chem* 1997, 272:6561-6566.
54. Takeshita M, Chang CN, Johnson F, Will S, Grollman AP: Oligodeoxynucleotides containing synthetic abasic sites. Model substrates for DNA polymerases and apurinic/apyrimidinic endonucleases. *J Biol Chem* 1987, 262:10171-10179.
55. Schärer OD, Ortholand J-Y, Ganeson A, Ezaz-Nikpay K, Verdine GL: Specific binding of the DNA repair enzyme AlkA to a pyrrolidine-based inhibitor. *J Am Chem Soc* 1995, 117:6623-6624.
56. Roelen HC, Saris CP, Brugghe HF, van den Elst H, Westra JG, van der Marel GA, *et al.*: Solid-phase synthesis of DNA fragments containing the modified base 7-hydro-8-oxo-2′-deoxyguanosine. *Nucleic Acids Res* 1991, 19:4361-4369.
57. Bodepudi V, Iden CR, Johnson F: An improved method for the preparation of the phosphoramidites of modified 2′-deoxynucleotides: Incorporation of 8-oxo-2′-deoxy-7H-guanosine into synthetic oligomers. *Nucleosides Nucleotides* 1991, 10:755-761.
58. Guy A, Duplaa A-M, Harel P, Teoule R: Synthesis and characterization of DNA fragments bearing an adenine radiation product: 7,8-dihydroadenin-8-one. *Helv Chim Acta* 1988, 71:1566-1572.
59. Ezaz-Nikpay K, Verdine GL: The effects of N7-methylguanine on duplex DNA structure. *Chem Biol* 1994, 1:235-240.
60. Kao JY, Goljer I, Phan TA, Bolton PH: Characterization of the effects of a thymine glycol residue on the structure, dynamics, and stability of duplex DNA by NMR. *J Biol Chem* 1993, 268:17787-17793.
61. Srivastava SC, Raza SK, Misra R: 1,N6-etheno deoxy and ribo adenosine and 3,N4-etheno deoxy and ribo cytidine phosphoramidites. Strongly fluorescent structures for selective introduction in defined sequence DNA and RNA molecules. *Nucleic Acids Res* 1994, 22:1296-1304.
62. Sarracino D, Richert C: Quantitative MALDI-TOF MS of oligonucleotides and a nuclease assay. *Biorg Med Chem Lett* 1996, 6:2543-2548.
63. Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, *et al.*: Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* 1995, 377:3-174.
64. Lennon G, Auffray C, Polymeropoulos M, Soares MB: The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* 1996, 33:151-152.
65. Turner DL, Weintraub H: Expression of achaete-scute homolog 3 in ***Xenopus*** embryos converts ectodermal cells to a neural fate. *Genes Dev* 1994, 8:1434-1447.
66. Slupphaug G, Eftedal I, Kavli B, Bharati S, Helle NM, Haug T, *et al.*: Properties of a recombinant human uracil-DNA glycosylase from the Ung gene and evidence that Ung encodes the major uracil-DNA glycosylase. *Biochemistry* 1995, 34:128-138.
67. Waters TR, Swann PF: Kinetics of the action of thymine DNA glycosylase. *J Biol Chem* 1998, 273:20007-20014.

Because *Current Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cub – for further information, see the explanation on the contents page.