 © 1997 Oxford University Press

Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the *UNG* gene

Hilde Nilsen, Marit Otterlei, Terje Haug, Kristin Solum, Toril A. Nagelhus, Frank Skorpen and Hans E. Krokan*

UNIGEN Center for Molecular Biology, The Medical Faculty, Norwegian University of Science and Technology, N-7005 Trondheim, Norway

Received November 6, 1996; Revised and Accepted December 23, 1996

DDBJ/EMBL/GenBank accession nos+

ABSTRACT

A distinct nuclear form of human uracil-DNA glycosylase [UNG2, open reading frame (ORF) 313 amino acid residues] from the *UNG* gene has been identified. UNG2 differs from the previously known form (UNG1, ORF 304 amino acid residues) in the 44 amino acids of the N-terminal sequence, which is not necessary for catalytic activity. The rest of the sequence and the catalytic domain, altogether 269 amino acids, are identical. The alternative N-terminal sequence in UNG2 arises by splicing of a previously unrecognized exon (exon 1A) into a consensus splice site after codon 35 in exon 1B (previously designated exon 1). The UNG1 sequence starts at codon 1 in exon 1B and thus has 35 amino acids not present in UNG2. Coupled transcription/translation in rabbit reticulocyte lysates demonstrated that both proteins are catalytically active. Similar forms of UNG1 and UNG2 are expressed in mouse which has an identical organization of the homologous gene. Constructs that express fusion products of UNG1 or UNG2 and green fluorescent protein (EGFP) were used to study the significance of the N-terminal sequences in UNG1 and UNG2 for subcellular targeting. After transient transfection of HeLa cells, the pUNG1-EGFP-N1 product colocalizes with mitochondria, whereas the pUNG2-EGFP-N1 product is targeted exclusively to nuclei.

INTRODUCTION

Uracil-DNA glycosylase is the first enzyme in base excision repair (BER) for removal of uracil from DNA and its main function is probably to remove mutagenic uracil residues resulting from deamination of cytosine in DNA (1). A catalytically fully active recombinant form of human uracil-DNA glycosylase lacking an N-terminal sequence encoded by the open reading frame (ORF) (2) has been used to study structure–function relationships as determined by site directed mutagenesis and X-ray crystallography (3). These studies identified this form of human uracil-DNA glycosylase as a one domain structure with a positively charged DNA-binding groove and a deep uracil-binding pocket (3,4) which is accessed by flipping of uracil with its deoxyribose and 5′-phosphate out of the helix (5). The complete sequence of the *UNG* gene, which encodes human uracil-DNA glycosylase (UNG) (6), and comparison with *UNG* cDNA from human placenta (7) indicated that the gene contained six exons and a TATA-less promoter. This cDNA contains an ORF of 304 amino acids (7), but a homogeneous form of uracil-DNA glycosylase purified from human placenta (8) lacked 77 N-terminal amino acids predicted from the ORF (7). Subsequent work demonstrated that the *UNG* gene encoded both the mitochondrial and nuclear forms of UNG-proteins and that this N-terminal sequence was required for transport of the enzyme to mitochondria, but not for nuclear import as determined by transfection studies (9). The similarity of the mitochondrial and nuclear forms was also demonstrated by subcellular fractionation and Western blotting, as well as by immunostaining (9,10). It was thus assumed that proteolytic removal of the N-terminal sequence could explain the differential transport of UNG either to mitochondria or nuclei, although artificial cleavage during purification could not be excluded.

Here we report a distinct form of human nuclear UNG. We designate this form UNG2 and the previously known human form (7) UNG1. Both forms are encoded by the *UNG* gene, but have different N-terminal sequences. In addition we have isolated cDNAs for mouse homologs of UNG1 and UNG2. Transient transfection experiments with constructs that directed expression of fusion products of either UNG1 or UNG2 and a variant of green fluorescent protein indicated that the N-terminal sequences of UNG1 and UNG2 direct targeting to mitochondria and nuclei, respectively.

MATERIALS AND METHODS

Materials

Mouse embryonic carcinoma cDNA library, human liver cDNA library and NT2 neuronal precursor cell cDNA library were from Stratagene (La Jolla, CA, USA). All libraries were propagated in

*To whom correspondence should be addressed. Tel: +47 73 598680; Fax: +47 73 598705; Email: hans.krokan@unigen.unit.no

+X08975, X15653, X89398, X99018 and Y09008.

the Uni-ZAP™XR vector using XL-1 blue as host. [α-^{32}P]dCTP, [^{35}S]methionine, Rediprime random labelling kit and Hybond N+ filters were all from Amersham (UK). All sequencing primers were from MedProbe (Oslo, Norway). Dye terminator cycle sequencing ready reaction kit was from Applied Biosystems (Foster City, CA). The Dynazyme PCR kit was purchased from Finnzymes Oy (Espoo, Finland). TNT *in vitro* transcription/translation rabbit reticulocyte lysate system kit, pGEM-T TA cloning kit, Alter Sites II *in vitro* Mutagenesis System, primers for sequencing from T3 and T7 promoters and T3 RNA polymerase were from Promega (Madison, WI). The plasmid encoding the red-shifted variant of green fluorescent protein (pEGFP-N1) was from Clontech (Palo Alto, CA, USA). Restriction enzymes were from New England Biolabs Inc. (Beverly, MA, USA).

Screening of cDNA libraries

All libraries were screened as recommended by the manufacturer, using ^{32}P-labeled *UNG40* cDNA (7) as probe. Hybridization was carried out at 65°C overnight in 6× SSC, 5× Denhardt's solution and 0.1% SDS. Filters were washed in 0.1× SSC/0.5% SDS at 65°C and autoradiographed. Three rounds of screening were done. *In vivo* excision of pBluescript phagemids from the Uni-ZAP™XR vector was performed as recommended by the manufacturer.

Sequence analysis of clones

Sequencing was performed on an Applied Biosystems Model 373A DNA Sequencing System using the Dye terminator cycle sequencing ready reaction kit as recommended by the manufacturer. The sequences were analyzed using the Auto Assembler software (Applied Biosystems).

In vitro transcription and uracil-DNA glycosylase assays

In vitro transcription/translation was performed with the TNT transcription/translation system with [^{35}S]methionine as recommended by the manufacturer, using 200 ng of the expression constructs per 10 µl reaction volume. The mouse UNG1-pBluescript construct was transcribed from the T3 promoter in the pBluescript vector. The insert of mouse UNG2-pBluescript was amplified by the polymerase chain reaction using Dynazyme PCR kit, ligated into the pGEM-T vector and transcribed from the T7 promoter. The human UNG2-pBluescript was transcribed from the T3 promoter after *Sac*I/*Nhe*I excision of a 79 bp fragment from the polylinker and the 5′-end of cDNA for UNG2. Human UNG1 cDNA was transcribed from the T7 promoter as previously described (2). The samples were run on a 12% denaturing sodium dodecyl sulfate polyacrylamide gel (SDS–PAGE). The gel was dried, autoradiographed overnight and scanned on an LKB Ultroscan XL Enhanced Laser Densitometer. Uracil-DNA glycosylase activity was measured in parallel samples of the *in vitro* transcription/translation assay mixture containing unlabelled amino acids (2).

Preparation of pUNG-EGFP-N1 fusion constructs and localization studies

UNG15 cDNA, which encodes UNG1, in pGEM7Zf+ (pUNG15) (2,7) was digested with *Bcl*I, which cuts at bp 1019 in *UNG15* cDNA, blunted with DNA polymerase I (Klenow fragment), and ligated to an *Age*I linker prepared from the oligonucleotide 5′-ACCGGTGCC-3′ and its complementary copy. The religated pUNG15 containing the *Age*I linker correctly ligated into the *Bcl*I site (verified by sequencing) was digested with *Rsr*II, which cuts at bp 49 in *UNG15* cDNA (7), blunted as above and finally digested with *Age*I. The fragment was then ligated into pEGFP-N1 digested with *Sma*I (blunt) and *Age*I. The construct was sequenced to verify that the construct was in frame with the ATG of the EGFP-N1 fusion protein. The TGA stop codon of pUNG15 was changed to GGA by site-directed mutagenesis performed according to the procedure provided by the manufacturer using ssDNA prepared with R408 phage. Potential $pUNG1_{GGA}$-EGFP-N1 constructs were screened by digestion with *Bcl*I (digests only unmutated plasmids) and verified by sequencing. The correct construct was named pUNG1-EGFP-N1. cDNA for UNG2 (this article) in pBluescript was digested with *Nhe*I, which cuts 54 bp upstream of ATG, and *Eco*NI which cleaves the cDNAs in the sequence that is shared by cDNAs for UNG1 and UNG2 (positions 529 and 520, respectively). The resulting fragment of interest (501 bp) was isolated and ligated to the 5155 bp fragment of *Nhe*I/*Eco*NI-digested pUNG1-EGFP-N1 to obtain pUNG2-EGFP-N1. Transient transfections of HeLa cells were done with the $CaPO_4$-method (Profection, Promega) according to the manufacturer's recommendations. Confocal microscopy (BioRad MRC-600) of HeLa cells and staining of mitochondria with mouse anti human mitochondria antibody (MAB 1273, Chemicon) and Texas Red anti-mouse IgG (Vector) were performed as previously described (10). Examination of HeLa cells transfected with expression plasmids pEGFP-N1, pUNG1-EGFP-N1 or pUNG2-EGFP-N1 was carried out using an excitation wave length of 488 nm and emission wave length >515 nm at 16 h after transfection.

RESULTS

We have screened a human NT2 neuronal precursor cell cDNA library and a mouse embryonic carcinoma cDNA library and discovered a new form of human uracil-DNA glycosylase (human UNG2) encoded by the *UNG* gene, as well as the homologous cDNA from mouse (mouse UNG2). In addition we have isolated the cDNA for the mouse homologue (encoding mouse UNG1) of human UNG1 (7). cDNA for human UNG2 has an ORF encoding 44 N-terminal amino acids not found in human UNG1 whereas cDNA for human UNG1 has an ORF encoding 35 amino acids not found in human UNG2 (Fig. 1). The two forms are identical in the rest of the N-terminal sequence which is not required for enzyme activity, as well as in the catalytic domain, altogether 269 identical consecutive amino acids. cDNAs for human UNG2 and its mouse homologue are apparently as abundant as UNG1 in cDNA libraries from proliferating cells since among 20 cDNA clones that were sequenced 10 were of the UNG2 type and 10 were similar to the previously known UNG1 type. Among four mouse cDNAs sequenced, three were of the UNG2 type and one was of the UNG1 type. However, screening of a human hepatocyte cDNA library with *UNG40* cDNA (7) resulted in the isolation of 80 strongly hybridizing clones and sequencing of 14 of these demonstrated that they were all similar to the previously characterized cDNA for UNG1 or the cDNA *UNG40* (7).

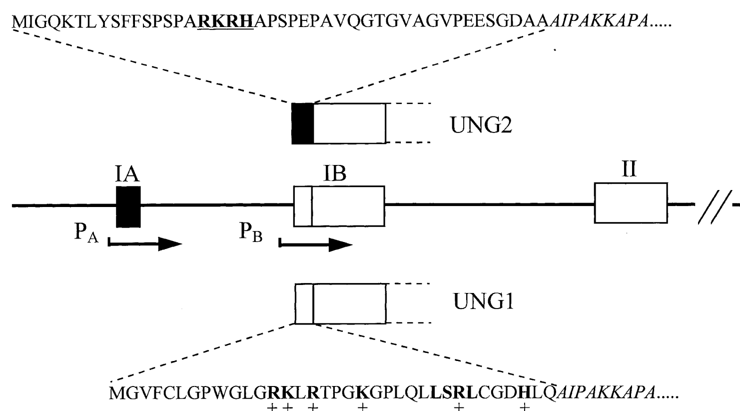

Figure 1. Generation of human UNG1 and UNG2 by transcription from two promoters and alternative splicing. P_B is the previously recognized promoter for transcription of UNG1 (11) and P_A the promoter from which UNG2 is transcribed. Exon 1A encodes 44 amino acids present in UNG2, but absent in UNG1. The 35 N-terminal codons of exon 1B are only present in UNG1. The N-terminal sequence of UNG2 is shown on top with the putative nuclear localization signal underlined. The N-terminal sequence of UNG1 directing mitochondrial import is shown in the bottom line.

Comparison of the human cDNA for UNG2 with the recently published complete human *UNG* sequence (6) revealed the presence of a previously unrecognized exon (exon 1A) located ~650 bp upstream of the previously identified exon 1 (hereafter called exon 1B). Exon 1B forms the leader sequence and codon 1–104 of the mRNA encoding the previously known form UNG1. The mRNA corresponding to the new human cDNA is formed by joining exon 1A (encoding 44 amino acids) into a consensus splice site after codon 35 in exon 1B after which the two human cDNAs are identical. The ORF of human *UNG2* cDNA predicts a protein of 313 amino acids, as compared to 304 amino acids for UNG1 (7). We have also isolated and sequenced genomic clones for the mouse homologue of the *UNG* gene. This has revealed that the splice sites for exons 3, 4, 5 and 6 in the *UNG* genes from mouse and man are in identical positions. Furthermore, PCR analyses have demonstrated that the rest of the mouse gene is structurally similar to the human gene, as expected from the cDNA clones (data not shown).

Figure 1 shows how the alternative forms of mRNA for UNG1 and UNG2 arise as deduced from human cDNAs and the corresponding *UNG* sequences and indicates the presence of a putative nuclear localization signal of four basic residues (RKRH) in the N-terminal end of the new cDNA and putative mitochondrial localization signals in cDNA for UNG1. In addition, and not shown here, both human cDNAs contain a putative nuclear localization signal (RKRHH) in the catalytic domain (residues 258–262 in the ORF of cDNA for UNG1). These residues are located at the surface of the enzyme between α-helix 7 and β-strand 4 (3).

Figure 2 shows the genomic structure of exons 1A and 1B, as well as the structure of the previously characterized promoter (hereafter called P_B), possible elements in the putative promoter upstream of exon 1A (hereafter called P_A) and the alternative splice acceptor site. Promoter P_A probably starts after the 3′-terminal end of two *Alu*-repeats (position 425) and ends immediately upstream of the start of exon 1A. However, it can not be excluded that the promoter is located upstream of the *Alu*-repeats. This would require the presence of an exon encoding a leader that would be joined to exon 1A. We judge this as unlikely since we have not detected promoter motifs upstream of the *Alu*-repeats and since we have not detected transcripts of the required size by Northern analyses (data not shown). Furthermore, the cDNA for UNG2 does not contain sequences from this upstream region.

Figure 3 shows an alignment of predicted amino acid sequences of the human and mouse enzymes. Note that UNG1 proteins and UNG2 proteins have been aligned separately in the parts of the proteins that are derived from different exons (up to codon 45 in human UNG2). The catalytic domains of human and mouse UNG proteins (residues 64–313 in human UNG2) are highly conserved (91% identity), while the N-terminal sequences of the homologous forms are less well conserved, but clearly related. Amino acids that have been found to be critical for catalytic activity or formation of the uracil-binding pocket (3,12) or DNA binding (5) are completely conserved in mouse (residues Q144, D145, P146, Y147, F158, S169, N204, S247, H268, S270, L272, S273, Y275 and R276 in UNG1).

Table 1. Relative specific activities of different forms of UNG after translation in rabbit reticulocyte lysates[a]

Protein	d.p.m.[a]	Area (mm^2)	Activity (d.p.m./area)
human UNG1	1291	0.054	23907
human UNG2	6360	0.268	23731
mouse UNG1	921	0.061	15098
mouse UNG2	856	0.051	16784

[a]Relative specific activities were calculated from measured d.p.m. values ([^{3}H]uracil released in uracil-DNA glycosylase assays) and areas under the curve of scanned bands on SDS–PAGE gels after subtraction of background values of 123 d.p.m.

To compare the promoter activity of promoter P_B alone, promoter P_A alone and promoters P_B and P_A in combination, we prepared promoter–luciferase gene constructs and performed transient transfection experiments with HeLa cells. These studies verified the promoter activity of promoter P_B alone (11), demonstrated that promoter P_A alone is at least as active as promoter P_B and showed that when both promoters are present in the construct, the luciferase activity increased by ~50% when compared with promoter P_B alone. This was as expected from the abundance of UNG2 cDNA in proliferating cells (data not

Figure 2. Structure of the 5′-terminal part of the human *UNG* gene. Bold letters indicate exons (1A and 1B).

shown). Coupled transcription/translation of the two forms of human and mouse cDNA resulted in easily measurable uracil-DNA glycosylase activity for both forms from mouse and man. For calculations of the relative specific activities, the radioactivity released in uracil-DNA glycosylase assays was compared to band intensities on an SDS–PAGE gel from a transcription/translation reactions using [^{35}S]methionine (Table 1).

To examine whether human UNG1 and UNG2 were translocated to different subcellular compartments, we prepared constructs expressing fusion proteins of the UNG proteins and a red shifted variant of green fluorescent protein (EGFP-N1). These were used for transient transfection experiments with HeLa cells. The major advantage of the green fluorescent protein (over the use of antibodies) is that this method relies on the autofluorescence of this protein alone, and thus possible cross reaction of an antibody with epitopes in irrelevant proteins is not a problem. The control (pEGFP-N1) shows that the green fluorescent protein displays a homogeneous staining over the cells (Fig. 4A). In contrast, the UNG2–EGFP-N1 fusion protein is exclusively located in the nuclei (Fig. 4C) and the UNG1–EGFP-N1 fusion protein (Fig. 4D) is mainly, if not exclusively, located in extranuclear spots that have the same appearance as mitochondria stained with Texas red (Fig. 4B). The few spots of green fluorescence visible over the nucleus in Figure 4D are probably due to the presence of mitochondria over and under the nuclei in this summation image of optical sections. These results provide convincing experimental evidence that UNG2 is a nuclear protein and UNG1 a mitochondrial protein.

DISCUSSION

In the present paper we describe a distinct form of human nuclear uracil-DNA glycosylase (UNG2) encoded by the *UNG* gene as well as the mouse homologs of UNG1 and UNG2. UNG1 (the mitochondrial form) and UNG2 have identical catalytic domains, but very different N-terminal sequences. The two forms of uracil-DNA glycosylase from the human *UNG* gene are generated using two promoters and splicing of an additional exon (exon 1A), transcribed from the putative upstream promoter P_A, into the first exon transcribed from the lower promoter P_B. This observation requires an update of the recently published organization of the human *UNG* gene (6). In a previous study from our laboratory (9) we considered alternative splicing as a less likely mechanism for generation of nuclear and mitochondrial forms because only one transcript was observed (7,13). However, the cDNAs for human UNG1 (2061 bp) and UNG2 (2058 bp) are of very similar sizes and the corresponding mRNAs would probably not be separated on standard gels for Northern analysis.

It has previously been documented by several methods that the nuclear and the mitochondrial forms of human uracil-DNA glycosylase are strongly related (5,10). Transfection studies demonstrated that the N-terminal sequence in UNG1 directed transport to mitochondria, whereas the absence of this sequence resulted in nuclear transport (9). Therefore, the putative nuclear localization signal within the catalytic domain is presumably sufficient to direct nuclear transport in the absence of the mitochondrial localization signal in the N-terminal sequence. Alternatively this small basic protein may not require a signal for entering the nucleus. The discovery of UNG2 containing an N-terminal sequence that probably contains the nuclear localization signal indicates, however, that this new form may represent the true nuclear form. However, formally it can not be excluded that the absence of a mitochondrial targeting signal in UNG2 results in nuclear localization. Furthermore, it is possible that the N-terminal sequence and the putative nuclear localization signal in the catalytic domain are both required for nuclear import. The predicted size of UNG2 (~36 kDa) corresponds to the

```
hUNG1           MGVFCLGPWGLGRKLRTPGKGPLQLLSRLC----GDHLQA   36
mUNG1           MGV-------LGRRSLR--LARRAGLRSLTPNPDSDSRQA   31
hUNG2     MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAA   45
mUNG2     MIGQKTLYSFFSPTPTGKRTTRSPEP-VPGSGVAA--EIGGDAVA   42

hUNG2/1   IPAKKAPAGQEEPGTPPSSPLSAEQLDRIQRNKAAALLRLAARNV   90/81
mUNG2/1   SPAKKARVEQNEQG----SPLSAEQLVRIQRNKAAALLRLAARNV   83/72

hUNG2/1   PVGFGESWKKHLSGEFGKPYFIKLMGFVAEERKHYTVYPPPHQVF   135/126
mUNG2/1   PAGFGESWKQQLCGEFGKPYFVKLMGFVAEERNHHKVYPPPEQVF   128/117

hUNG2/1   TWTQMCDIKDVKVVILGQDPYHGPNQAHGLCFSVQRPVPPPPSLE   180/171
mUNG2/1   TWTQMCDIRDVKVVILGQDPYHGPNQAHGLCFSVQRPVPPPPSLE   173/162

hUNG2/1   NIYKELSTDIEDFVHPGHGDLSGWAKQGVLLLNAVLTVRAHQANS   225/216
mUNG2/1   NIFKELSTDIDGFVHPGHGDLSGWARQGVLLLNAVLTVRAHQANS   218/207

hUNG2/1   HKERGWEQFTDAVVSWLNQNSNGLVFLLWGSYAQKKGSAIDRKRH   270/261
mUNG2/1   HKERGWEQFTDAVVSWLNQNLSGLVFLLWGSYAQKKGSVIDRKRH   263/252

hUNG2/1   HVLQTAHPSPLSVYRGFFGCRHFSKTNELLQKSGKKPIDWKEL     313/304
mUNG2/1   HVLQTAHPSPLSVYRGFLGCRHFSKANELLQKSGKKPINWKEL     306/295
```

Figure 3. Alignment of UNG proteins from man and mouse. Note that UNG1 and UNG2 proteins have been aligned separately down to the common splice site corresponding to codon 44 in human UNG2. The N-terminal sequence not present in the catalytically active form of human placental uracil-DNA glycosylase originally isolated, residues 1–77 in human UNG1 (8) is shown in bold letters. Downstream of the alternative splice site (⬇) used for generating UNG2 forms (from 45 in human UNG2), the sequences of the two forms are identical in each species. Residues that make up walls of the uracil-binding pocket or which are directly involved in catalysis are marked with a star. Residues that are involved in DNA-binding (except those involved in uracil-binding) are marked with a triangle.

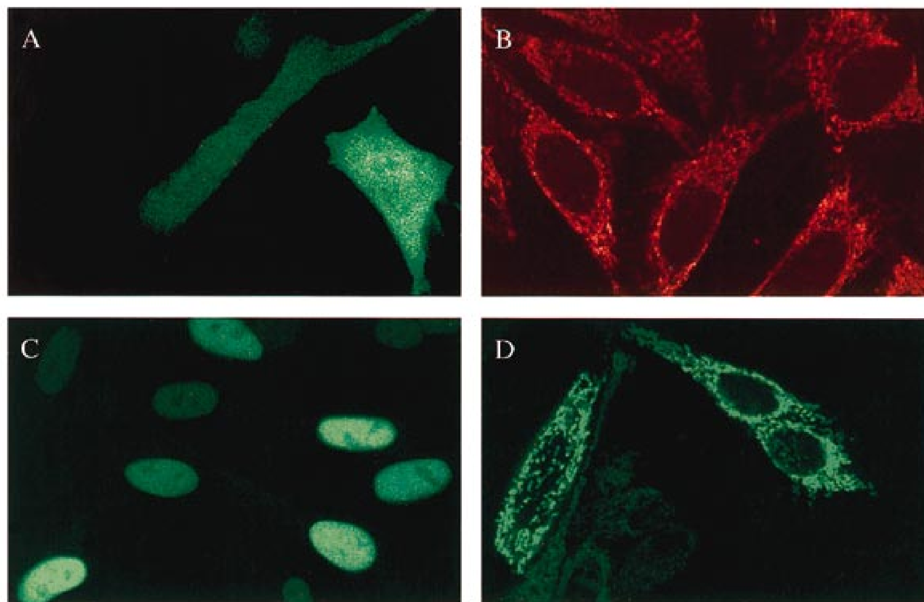

Figure 4. Subcellular localization in HeLa cells of UNG2–EGFP-N1 and UNG1–EGFP-N1 fusion products. HeLa cells were transfected with constructs expressing pUNG2-EGFB-N1 (**C**), pUNG1-EGFP-N1 (**D**) or the control pEGFP-N1 (**A**), all expressed from the CMV promoter, and processed for confocal microscopy. (**B**) Staining of mitochondria with Texas red.

size reported for a highly purified nuclear form of rat uracil-DNA glycosylase (14), while others have reported a size corresponding to the catalytic domain lacking the N-terminal sequence of uracil-DNA glycosylase from human tumor cells (15,16) and human placenta (8). Proteolytic cleavage of the N-terminal sequence for generation of the nuclear form of uracil-DNA glycosylase may have a physiological role, but it is also possible that this is an artificial situation only seen in tumor cells that in general are rich in proteases, and in other protease-rich tissues, such as placenta. Alternatively, the proteolytical removal of the N-terminal sequence may be an artefact of protein purification. We have recently found that treatment of a purified recombinant form of human UNG1 lacking the 28 N-terminal amino acid residues (Δ28UNG1) with proteinase K (or cell extracts) removes some additional 50 amino acid residues, leaving the catalytic domain intact and fully active (unpublished data). Thus, the different sizes previously reported for the nuclear form of UNG may be explained by the action of protease activity released after preparation of cell extracts for purification.

Two other cDNAs claimed to encode human nuclear uracil-DNA glycosylase have been reported (17,18). These are unre-

lated to uracil-DNA glycosylases of the UNG family, as well as to each other. Since polyclonal antibodies raised against a recombinant form of human UNG (UNGΔ84) essentially completely inhibit total uracil-DNA glycosylase activity in crude cell extracts (2) the unrelated uracil-DNA glycosylases are unlikely to contribute very much, if at all, to uracil-DNA glycosylase activities in human cells. In addition, deletion of carboxy-terminal amino acids of a human T(U)/G-mismatch DNA glycosylase results in an enzyme that has lost the thymine-DNA glycosylase activity, but retained uracil-DNA glycosylase activity for U/G mispairs. In addition, bacterial proteins with homology to the core of the human glycosylase were identified (19). This indicates that this mismatch uracil-DNA glycosylase is an ancient enzyme that might represent a backup for the uracil-DNA glycosylase from the *UNG* gene. Consistent with this idea, we have found that when crude cell extracts are incubated with neutralizing antibodies to the UNG-proteins, the removal of uracil from oligonucleotides containing U/A is completely abolished, but a small residual activity (<5%) is found for removal of U from U/G mismatches (unpublished data).

The *UNG* gene in the fish *Xiphophorus* is strongly related to mammalian *UNG* genes and is also similarly organized (20, and R. B. Walter, personal communication). The N-terminal sequence of a putative *Xiphophorus* UNG2 protein predicted from the gene structure is homologous to mammalian UNG2, but much shorter. However, the N-terminal sequence in the putative UNG1 protein in *Xiphophorus* is only distantly related (data not shown). The *Xiphophorus UNG* gene has consensus splice sites located in exactly the same positions as those verified to be splice sites in the human *UNG* gene (6) as well as in mouse (data not shown). It was therefore originally suggested that the *UNG* gene of *Xiphophorus* has six exons (20), but the present work strongly indicates that it encodes seven exons. Other repair gene structures are also highly conserved from fish to man. Thus, 22 of 23 exons of *ERCC2/XPD* in Chinese hamster ovary and *Xiphophorus* are of identical sizes. Since fish and mammals separated more than 450 million years ago, this indicates a very high degree of conservation of coding sequences and splicing patterns for DNA repair genes in vertebrates (6,20).

In conclusion, the human *UNG* gene encodes two forms of uracil-DNA glycosylase generated by alternative transcription starts, making use of an exon specific for the N-terminal end of the nuclear form, and alternative splicing. The different N-terminal amino acid sequences generated by this mechanism result in one form that enters the nucleus (UNG2) while the other form enters the mitochondria (UNG1).

ACKNOWLEDGEMENTS

This work was supported by the Norwegian Cancer Society, the Research Council of Norway and the Cancer Fund at the Regional Hospital in Trondheim. We would like to thank Dr Geir Slupphaug for helpful discussions and Dr Ronald B. Walter of the South West State University, San Marcos, Texas, for permission to discuss sequence information on the *UNG* gene from *Xiphophorus*.

REFERENCES

1 Lindahl,T. (1993) *Nature* **362**, 709–715.
2 Slupphaug,G., Eftedal,I., Kavli,B., Bharati,S., Helle,N.M., Haug,T., Levine,D.W. and Krokan,H.E. (1995) *Biochemistry* **34**, 128–138.
3 Mol,C.D., Arvai,A.S., Slupphaug,G., Kavli,B., Alseth,I., Krokan,H.E. and Tainer,J.A. (1995) *Cell* **80**, 869–878.
4 Mol,C.D., Arvai,A.S., Sanderson,R.J., Slupphaug,G., Kavli,B., Krokan,H.E., Mosbaugh,D.W. and Tainer,J.A. (1995) *Cell* **82,** 701–708.
5 Slupphaug,G., Mol,C.D., Kavli,B., Arvai,A.S., Krokan,H.E. and Tainer,J.A. (1996) *Nature* **384**, 87–92.
6 Haug,T., Skorpen,F., Kvaløy,K., Eftedal,I., Lund,H. and Krokan,H.E. (1996) *Genomics* **36**, 408–416.
7 Olsen,L.C., Aasland,R., Wittwer,C.U., Krokan,H.E. and Helland,D.E. (1989) *EMBO J.* **8**, 3121–3125.
8 Wittwer,C.U., Bauw,G. and Krokan,H.E. (1989) *Biochemistry* **28**, 780–784.
9 Slupphaug,G., Markussen,F.H., Olsen,L.C., Aasland,R., Aarsæther,N., Bakke,O., Krokan,H.E. and Helland,D.E. (1993) *Nucleic Acids Res.* **21**, 2579–2584.
10 Nagelhus,T.A., Slupphaug,G., Lindmo,T. and Krokan,H.E. (1995) *Exp. Cell Res.* **220**, 292–297.
11 Haug,T., Skorpen,F., Lund,H. and Krokan,H.E. (1994) *FEBS Lett.* **353**, 180–184.
12 Kavli,B., Slupphaug,G., Mol,C.D., Arvai,A.S., Petersen,S.B., Tainer,J.A. and Krokan,H.E. (1996) *EMBO J.* **15**, 3442–3447.
13 Slupphaug,G., Olsen,L.C., Helland,D., Aasland,R. and Krokan,H.E. (1991) *Nucleic Acids Res.* **19**, 5131–5137.
14 Domena,J.D. and Mosbaugh,D.W. (1985) *Biochemistry* **24**, 7320–7328.
15 Caradonna,S. and Cheng,Y.-C. (1980) *J. Biol. Chem.* **255**, 2293–2300.
16 Myrnes,B. and Wittwer,C.U. (1988) *Eur. J. Biochem.* **173**, 383–387.
17 Muller,S.J. and Caradonna,S. (1991) *Biochim. Biophys. Acta* **1088**, 197–207.
18 Meyer-Siegler,K., Mauro,D.J., Seal,G., Wurzer,J., DeRiel,J.K. and Sirover,M.A.(1991) *Proc. Natl. Acad. Sci. USA* **88,** 8460–8464.
19 Gallinari,P. and Jiricny,J. (1996) *Nature* **383**, 735–738.
20 Walter,R.B. and Morizot,D.C. (1996) *Adv. Struct. Biol.* **4**, 1–24.