© 1998 Oxford University Press
Nucleic Acids Research, 1998, Vol. 26, No. 20 **4611–4617**

# Nuclear and mitochondrial splice forms of human uracil-DNA glycosylase contain a complex nuclear localisation signal and a strong classical mitochondrial localisation signal, respectively

**Marit Otterlei[1], Terje Haug[1], Toril A. Nagelhus[1,2], Geir Slupphaug[1], Tore Lindmo[2] and Hans E. Krokan[1,*]**

[1]Institute of Cancer Research and Molecular Biology and [2]Department of Physics, Norwegian University of Science and Technology, N-7005 Trondheim, Norway

Received July 2, 1998; Revised and Accepted August 31, 1998

## ABSTRACT

**Nuclear (UNG2) and mitochondrial (UNG1) forms of human uracil-DNA glycosylase are both encoded by the *UNG* gene but have different N-terminal sequences. We have expressed fusion constructs of truncated or site-mutated *UNG* cDNAs and green fluorescent protein cDNA and studied subcellular sorting. The unique 44 N-terminal amino acids in UNG2 are required, but not sufficient, for complete sorting to nuclei. In this part the motif $R_{17}K_{18}R_{19}$ is essential for sorting. The complete nuclear localization signal (NLS) in addition requires residues common to UNG2 and UNG1 within the 151 N-terminal residues. Replacement of certain basic residues within this region changed the pattern of subnuclear distribution of UNG2. The 35 unique N-terminal residues in UNG1 constitute a strong and complete mitochondrial localization signal (MLS) which when placed at the N-terminus of UNG2 overrides the NLS. Residues 11–28 in UNG1 have the potential of forming an amphiphilic helix typical of MLSs and residues 1–28 are essential and sufficient for mitochondrial import. These results demonstrate that UNG1 contains a classical and very strong MLS, whereas UNG2 contains an unusually long and complex NLS, as well as subnuclear targeting signals in the region common to UNG2 and UNG1.**

## INTRODUCTION

Uracil-DNA glycosylase (UDG or UNG) initiates the base excision repair (BER) pathway for removal of DNA uracil resulting from deamination of cytosine or misincorporation of dUMP (1–3). The significance of UNG in mutation avoidance is indicated by a 20-fold increase in spontaneous mutation frequency in yeast cells deficient in UNG activity (4). Mammalian cells lacking UNG activity are not yet available. The 13.5 kb gene (5) for human UNG (*UNG*) consists of seven exons and has two promoters that are required for generation of the mitochondrial preform (UNG1) and nuclear (UNG2) forms of uracil-DNA glycosylase by alternative splicing (6). The two promoters are differentially regulated in human tissues. Thus, UNG2 has the highest expression levels in tissues containing proliferating cells, whereas UNG1 is more widely expressed, with the highest expression in skeletal muscle, heart and testicles (7). UNG1 and UNG2 have 35 and 44 unique N-terminal amino acids, respectively, while the C-terminal 269 amino acids are identical in the two forms. The 220 C-terminal residues form the well-characterized compact catalytic domain which binds DNA and flips the uracil-containing nucleotide into the catalytic pocket by a 'push and pull' mechanism (8,9). The common sequence N-terminal of the catalytic domain contains residues that interact with the 34 kDa subunit of replication protein A (RPA2). This interaction may suggest a role for RPA in the initial step(s) of BER, although we have not been able to demonstrate stimulation of UNG activity by RPA *in vitro* (10). To our knowledge, RPA has not been demonstrated to be present in mitochondria and the possible functional role of RPA in BER must therefore presumably be limited to UNG2.

In the present work we have examined sequence requirements for targeting of UNG1 and UNG2 to mitochondria and nuclei, respectively, as well as sequences involved in subnuclear distribution of UNG2. We demonstrate that the 35 unique N-terminal residues in UNG1 constitute a very strong classical mitochondrial localization signal (MLS) with the potential to form an amphiphilic α-helix and that these residues are required and sufficient for complete mitochondrial sorting. In contrast, the nuclear localization signal (NLS) in UNG2 is rather complex and, in addition to the 44 unique N-terminal residues, which contains basic residues essential for sorting, complete nuclear translocation requires at least another 60 N-terminal residues, many of which are common to UNG1 and UNG2.

## MATERIALS AND METHODS

### Protein fusion constructions

pUNG1$_{1–142}$EGFP and pUNG1$_{29–142}$EGFP were made by inserting an *Age*I linker in the *Eco*NI site of pUNG1 and pUNGΔ28 (lacking the 28 first N-terminal codons of UNG1) (11)

*To whom correspondence should be addressed. Tel: +47 73 598660; Fax: +47 73 598801; Email: hans.kroken@medisin.ntnu.no

and subsequently ligating the *Nco*I-blunted/*Age*I fragments into the *Sma*I and *Age*I sites of pEGFP-N1 (Clontech Laboratories Inc., CA). pUNG1EGFP and pUNG2$_{85–313}$EGFP were made by inserting an *Age*I linker into the *Bcl*I site of pUNG1 and pUNG2Δ84 (lacking the residues specifying the 84 N-terminal amino acids in UNG2 and equal to the previously described pUNGΔ75) (11) and inserting the *RsrII*-blunted/*Age*I fragments into the *Sma*I and *Age*I sites of pEGFP-N1. The TGA stop codons of pUNG1EGFP and pUNG2$_{85–313}$EGFP were finally changed to GGA by site-directed mutagenesis. The *Nhe*I/*Eco*NI fragment of pUNG2 from pBluescript replaced the corresponding fragment of pUNG1EGFP to obtain pUNG2EGFP (6) and the *Eco*RI/*Eco*NI fragment of pUNGΔ28 replaced the corresponding fragment of pUNG1EGFP to obtain pUNG1$_{29–304}$EGFP. It was observed that fusion proteins made with small inserts containing their own ATG that were out of frame with the ATG in the pEGFP-N1 vector (as control) still had high green fluorescence, suggesting that translation initiation occurred from the ATG in pEGFP-N1 when the insert was sufficiently short. The first ATG in green flourescent protein (EGFP) was therefore mutated (ATG→CTG) to avoid translation initiation downstream of short inserts, giving pEGFP*NI. All fusion constructs of UNG1 or UNG2 and EGFP were verified to be correct by DNA sequencing. None of the truncated constructs contained alternative in-frame ATG start codons in the UNG1 or UNG2 part. Different oligonucleotides encoding amino acids 1–12, 12–17, 12–20, 12–20+26–29, 12–28, 21–35 and 1–28 of UNG1 were ligated in front of pEGFP*NI, named as stated in Figure 4. An ATG start codon was placed 5′ of the first UNG1 encoding triplet in all oligonucleotides used to prepare constructs expressing N-terminally truncated fusion products. pUNG2$_{1–151}$EGFP was made by ligating the *Nhe*I/*Eco*NI-blunted fragment from pUNG2 into *Nhe*I/*Eco*RI-blunted pEGFP*N1. pUNG1$_{1–39}$EGFP and pUNG2$_{1–48}$EGFP were made from *Eco*RI/*Bgl*I-blunted and *Nhe*I/*Bgl*I-blunted fragments of pUNG1 and pUNG2 ligated to *Eco*RI/*Sma*I- and *Nhe*I/*Sma*I-digested pEGFP*N1, respectively. pUNG1$_{1–39}$UNG2EGFP and pUNG1$_{1–28}$UNG2EGFP were made by ligating the *Eco*RI/*Bgl*I-blunted fragment of pUNG1 and the *Nde*I/*Age*I-blunted fragment of pUNG1$_{1–28}$EGFP in front of pUNG2EGFP (*Eco*RI/*Sma*I and *Nde*I/*Sma*I, respectively). ATG of UNG2 was finally mutated to CTG.

An *Xma*I/*Sma*I site (bp 33) in UNG2 cDNA was removed and the resulting pUNG2EGFP was used to prepare new single (codons 58 or 103) and double (codons 58/85 and 58/103) *Sma*I sites. A *Sal*I site was introduced at codon 92. pUNG2$_{1–58}$EGFP and pUNG2$_{1–103}$EGFP were made from the *Nhe*I/*Sma*I fragments ligated into *Age*I-blunted/*Nhe*I pEGFP*N1. pUNG2$_{1–92}$EGFP was made from the *Sal*I-blunted/*Nhe*I fragment ligated to *Eco*RI-blunted/*Nhe*I pEGFP*N1. pUNG2EGFP with *Sma*I sites at codons 58/85 and 58/103 were digested with *Sma*I and religated to make the deletion mutants pUNG2$_{1–58+85–304}$EGFP and pUNG2 $_{1–58+104–304}$EGFP.

All mutations in UNG1 and UNG2 were made with the Alter Site II *In Vitro* Mutagenesis System (Promega) or the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene). All constructs and mutations of UNG1 and UNG2 are listed in Figures 2 and 4.

## Immunostaining, transient transfection and analysis by confocal microscopy

HeLa cells were immunostained as previously described (12) using the anti-mitochondrial antibody mAb 1273 (Chemicon) as primary antibody. Secondary antibodies were rhodamine (tetramethyl)-conjugated goat-anti-mouse antibodies (Molecular Probes).

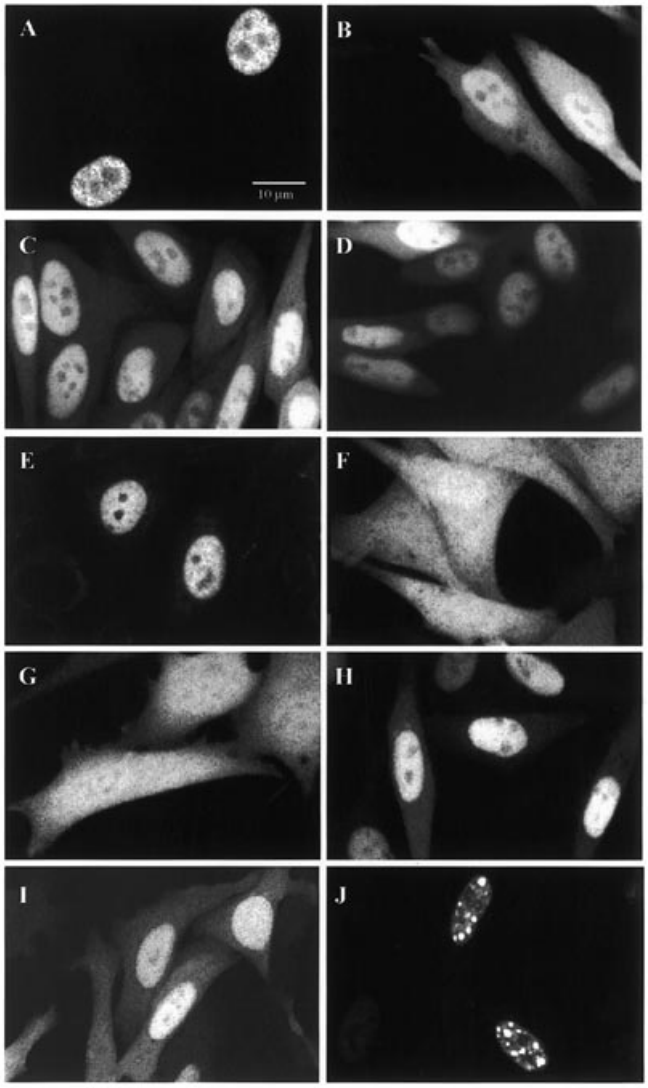


**Figure 1.** Subcellular localisation in HeLa cells of fusion products of EGFP and UNG2 or mutant UNG2 proteins. Cells were transfected with the constructs pUNG2$_{1–313}$EGFP (**A**), pUNG2$_{1–48}$EGFP (**B**), pUNG2$_{1–58}$EGFP (**C**), pUNG2$_{1–103}$EGFP (**D**), pUNG2$_{1–151}$EGFP (**E**), pUNG2$_{85–313}$EGFP (**F**), pUNG2$_{K18N}$EGFP (**G**), pUNG2$_{R17G}$EGFP (**H**), pUNG2$_{K49N/K50N}$EGFP (**I**) or pUNG2$_{K99N/K100N/R122I/K123N}$EGFP (**J**).

HeLa cells were transfected using calcium phosphate (Profection; Promega) according to the manufacturer's recommendation. Transfected cells were examined using a BioRad MRC-600 confocal microscope equipped with 488 (BHS) and 514 nm (GHS) excitation laser lines and a 60× Nikon water immersion objective with NA = 1.2. The 488 nm laser line was used for excitation of EGFP and fluorescence was detected at $\lambda$ >515 nm (BHS filter) ~16 h post-transfection. Two-parameter confocal microscopy analysis was performed in consecutive scans with the 488 and 514 nm laser lines, respectively. FITC/EGFP and rhodamine fluorescence were detected at 525 nm < $\lambda_{FITC/EGFP}$ < 555 nm (A2, BHS and PMT2) and $\lambda_{rhodamine}$ > 600 nm (A2, GHS

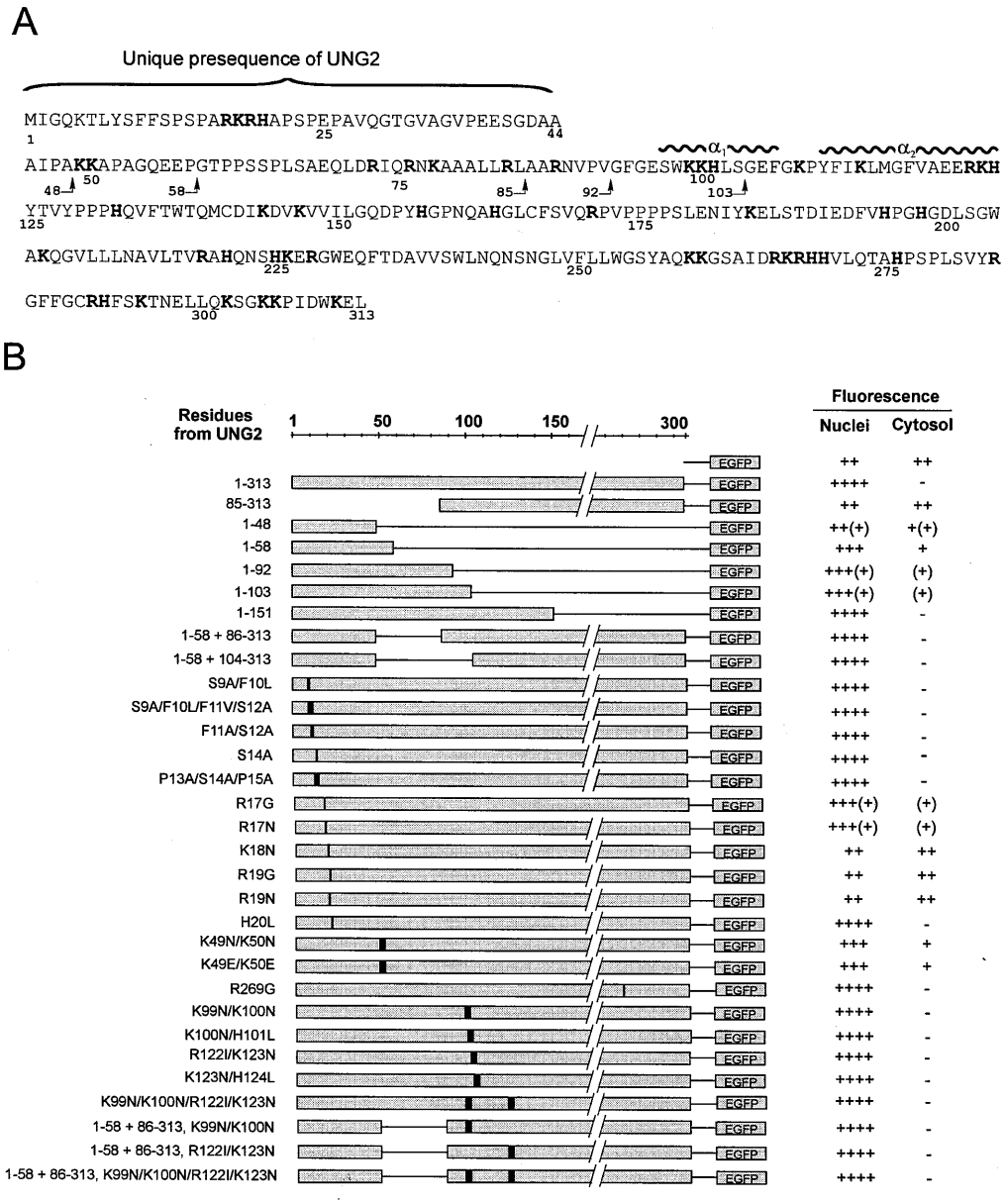


**Figure 2.** Regions in UNG2 important for nuclear localization (**A**). The amino acid sequence of UNG2 with its unique N-terminal sequence of 44 amino acids. Positively charged residues are displayed in bold letters. $\alpha_1$ and $\alpha_2$ indicate α-helices 1 and 2 as determined by X-ray crystallography (8). Positions of deletions are indicated by arrows pointing to the different amino acid positions in UNG2. (**B**) Overview of the various deletion and site-specific mutants and their ability to direct nuclear targeting of EGFP.

and PMT1), respectively. Pinhole sizes 2/15 ($\lambda_{FITC/EGFP}$) and 12/15 ($\lambda_{rhodamine}$) were used for PMT2 and PMT1, respectively, in order to optimize imaging for the specified fluorophores.

## RESULTS

### Nuclear and subnuclear localization signals in UNG2

To identify amino acids in UNG2 required for translocation to nuclei, we prepared a number of deleted or point mutated constructs of UNG2 cDNA fused to the reading frame of EGFP as a reporter. Transfection of a fusion construct of complete UNG2 and EGFP into HeLa cells resulted in complete nuclear translocation (Fig. 1A). As observed previously with immunostaining (12), nuclear fluorescence in more intensively staining 'spots' is also observed in a fraction of the cells with EGFP–UNG2 fusion constructs (Fig. 1A). Results of transfection with a number of different constructs are summarised in Figure 2. Clusters of basic residues (RKRH) resembling known NLS are present in UNG2 both in the N-terminal unique sequence and C-terminally immediately after α-helix 7 at the surface of the catalytic domain, at amino acid positions 17–20 and 267–270, respectively (8). The N-terminal RKRH motif is surrounded by several prolines, a property observed for NLS sequences in some viruses (13). However, this RKRH motif and surrounding sequences are not sufficient for complete nuclear sorting since the expressed fusion product of $pUNG2_{1–48}EGFP$ was only in part translocated to

nuclei (Fig. 1B). This demonstrates that sequences common to UNG2 and UNG1 are required for complete nuclear sorting. In agreement with this, transfection with a longer construct, pUNG2$_{1–58}$EGFP, resulted in improved nuclear translocation (Fig. 1C). A construct expressing amino acids 1–103 (pUNG2$_{1–103}$EGFP) improved sorting slightly (Fig. 1D) compared with pUNG2$_{1–58}$EGFP, but was similar to pUNG2$_{1–92}$EGFP (summarised in Fig. 2). Complete sorting was observed with a construct expressing UNG2 amino acids 1–151 (pUNG2$_{1–151}$EGFP) (Fig. 1), indicating that nuclear sorting also requires residues between positions 103 and 151. The C-terminal $R_{267}K_{268}R_{269}H_{270}$ motif is neither sufficient nor required for nuclear sorting, since transfection with a fusion construct of UNG2 lacking the 84 N-terminal residues (pUNG2$_{85–313}$EGFP) resulted in homogeneous staining (Fig. 1F) and since the N-terminal 151 amino acids resulted in complete sorting. Furthermore, mutating the C-terminal $R_{267}K_{268}\mathbf{R_{269}}H_{270}$ motif to $R_{267}K_{268}\mathbf{G_{269}}H_{270}$ in a fusion construct containing the complete UNG2 sequence did not affect nuclear sorting (summarised in Fig. 2).

Amino acids K18 and R19 in the $R_{17}K_{18}R_{19}H_{20}$ motif are essential for nuclear translocation because site-directed mutagenesis of either one (K18N or R19N or R19G) resulted in homogeneous staining over the cell (Fig. 1G). In the generally well-conserved mouse UNG2 N-terminal sequence the above motif is $G_{17}K_{18}R_{19}T_{20}$ (6). Mutation of R17 (R17N or R17G) reduced sorting somewhat (Fig. 1H), whereas the mutation H20L had no apparent effect (Fig. 2). These results establish the N-terminal $R_{17}K_{18}R_{19}$-motif as an essential part of the NLS in human UNG2. Although 14 of the 15 N-terminal residues are conserved between mouse and man, single or multiple mutations in residues 9–14 had no effect on nuclear sorting, possibly indicating other functions for this region of UNG2 (Fig. 2). A likely candidate to contribute to the NLS motif in the residues between positions 47 and 58 is the sequence $A_{48}K_{49}K_{50}A_{51}$. Site-directed mutagenesis of K49 and K50 (K49N/K50N or K49E/K50E) clearly reduced nuclear transport of the fusion protein, demonstrating that this motif is part of the complete NLS (Fig. 1I).

RPA binding in the sequence region common to UNG1 and UNG2 (10) is not essential for nuclear import because deletion of involved residues 58–85 (pUNG2$_{1–58+86–313}$EGFP) or 58–103 (pUNG2$_{1–58+104–313}$EGFP) did not reduce nuclear sorting (Fig. 2). In agreement with this, double mutations (K99N+K100N or K100N+H101L) in a potential NLS ($K_{99}K_{100}H_{101}$) in α-helix 1 did not reduce nuclear sorting. Neither was nuclear sorting reduced by mutations in the motif $R_{122}K_{123}H_{124}$ in α-helix 2 (R122I+K123N or K123N+H124L) or simultaneous mutations in both α-helices 1 and 2 (K99N+K100N+R122I+K123N), alone or together with deletion of residues 58–85 (Fig. 2). Interestingly, however, the specified substitutions of the positively charged residues in α-helices 1 and 2 or both, as well as substitutions in both helices together with deletion of residues 58–85, all resulted in a change in subnuclear distribution of UNG2 because the fluorescence intensity and size of 'spots' in a fraction of the cells (~10%) were very significantly enhanced and the fluorescence outside spots was reduced (see Fig. 1J as an example). Large spots were seen very infrequently in cells transfected with intact UNG2 fused to EGFP (<2%). This indicates that certain positively charged amino acids between positions 99 and 123 are required for localisation outside spots. Residues in the catalytic domain of UNG proteins involved in specific interaction with uracil-containing DNA (positions 153–285) (9) are located entirely outside the region in UNG2 found to be sufficient for efficient nuclear sorting. Thus, the DNA-binding region in the catalytic domain of UNG2 is apparently not required for retention of UNG2 in the nucleus.

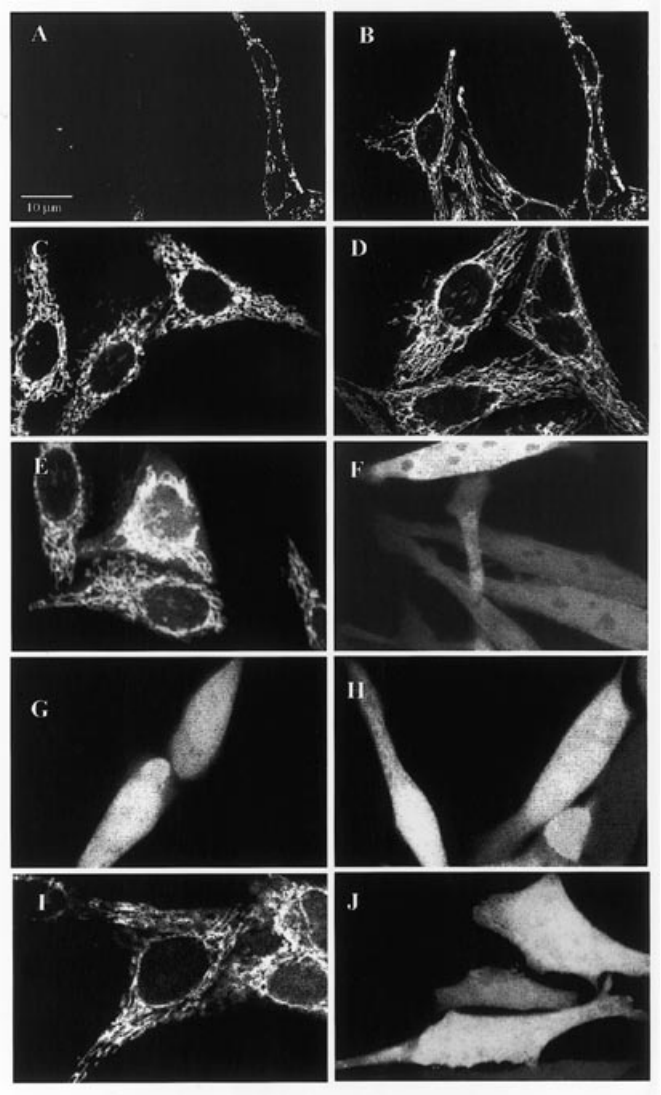


**Figure 3.** Subcellular localization in HeLa cells of UNG1 and different UNG1 mutant proteins fused to EGFP protein. The cells were transfected with constructs pUNG1$_{1–304}$EGFP (**A**), pUNG1$_{1–142}$EGFP (**C**), pUNG1$_{1–39}$EGFP (**D**), pUNG1$_{1–28}$EGFP (**E**), pUNG1$_{29–304}$EGFP (**F**), pUNG1$_{1–12}$EGFP (**G**), pUNG1$_{12–28}$EGFP (**H**), pUNG1$_{1–28}$UNG2EGFP (**I**) or the control pEGFP-NI vector (**J**). (**B**) Staining of the mitochondria in (A) with rhodamine-labelled anti-mitochondrial antibodies.

## The MLS in UNG1

In summary, the 35 N-terminal residues unique to UNG1 were found to be necessary and sufficient for mitochondrial import and certain amino acid residues were critical for import. Double staining with rhodamine-labelled anti-mitochondrial antibodies clearly identified the spots in the transfected cells as mitochondria (Fig. 3B). Note that not all cells are transfected, consequently, more

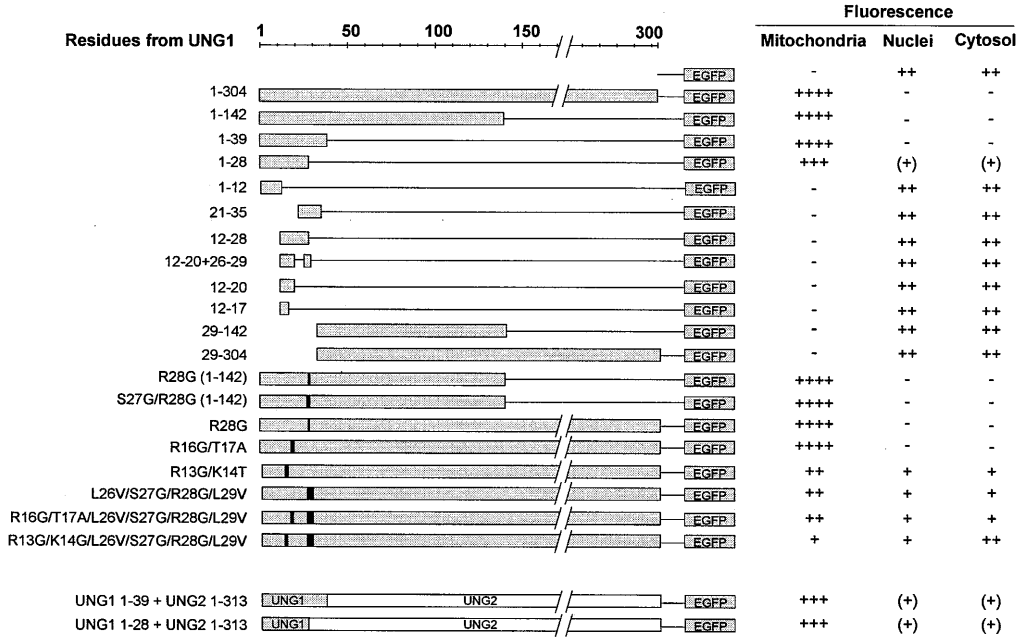


**Figure 4.** Mitochondrial sorting of UNG1 mutants. Overview of various deletion and site-specific (black areas) mutants of UNG1 and summary of their ability to target the EGFP fusion partner to mitochondria in HeLa cells. The numbers of + indicate the relative fluorescence within the cellular compartments, with ++++ indicating complete sorting, + very poor sorting and – no apparent sorting to the indicated compartment.

cells are seen with immunostaining than with EGFP-transfected cells. Transfection of HeLa cells with constructs encoding full-length UNG1 (pUNG1EGFP) (Fig. 3A), the N-terminal 142 amino acids of UNG1 (pUNG1$_{1–142}$EGFP) (Fig. 3C) and the 39 N-terminal amino acids (pUNG1$_{1–39}$EGFP) (Fig. 3D) all resulted in identical and complete mitochondrial sorting, whereas the 28 N-terminal residues (pUNG1$_{1–28}$EGFP) resulted in almost complete sorting (Fig. 3E). Encoded proteins that lacked the 28 N-terminal amino acids of UNG1 (Fig. 3F) failed to translocate to mitochondria. Not only did the unique N-terminal sequence promote mitochondrial translocation, but also precluded nuclear import, compared with the homogeneous staining with the control EGFP (Fig. 3J) and constructs lacking the 28 N-terminal residues in UNG1. This could be due to either a very efficient MLS and/or a specific nuclear exclusion signals (NES), although such signals, often lysine-rich, are usually found in proteins that shuttle between nucleus and cytoplasm. Residues 1–12 contain a consensus NES motif (14), whereas residues 21–35 are lysine-rich. However, when expressed in fusion with EGFP these resulted in homogeneous staining similar to EGFP alone (Figs 3G and 4). Thus, there is apparently not a specific signal in the UNG1 presequence that excludes UNG1 from nuclei; rather the unique N-terminal region in UNG1 constitutes a strong MLS sufficient for exclusion of UNG1 from the nucleus. Constructs expressing residues 12–28 (Fig. 3H), 12–17, 12–20 or 12–20+26–29 fused N-terminal to EGFP also resulted in homogeneous staining, demonstrating that the minimum sequence required for mitochondrial targeting are residues 1–28. A strong MLS was also indicated by transfection of constructs containing the 28 or 39 N-terminal amino acids in UNG1 N-terminal of UNG2 containing the complete NLS (pUNG1$_{1–28}$UNG2EGFP or pUNG1$_{1–39}$UNG2EGFP). These fusion products translocated almost exclusively to mitochondria (Figs 3I and 4). Thus, the MLS is dominant in constructs that contain both the MLS and the NLS.

When the human UNG1 residues 11–29 are plotted as an α-helical wheel, a striking amphiphilicity emerges (Fig. 5B). Site-directed mutagenesis of R28 (R28G) alone or R28 (R28G) together with S27 (S27G/R28G) did not affect the sorting efficiency of the fusion proteins (summarised in Fig. 4). However, mutations of amino acids L26–L29 (L26V/S27G/R28G/L29V) reduced mitochondrial sorting, demonstrating that this LSRL motif is part of the MTS in UNG1 (Fig. 6A). In addition, mutation of R13 and K14 (R13G/K14T) also strongly reduced mitochondrial sorting (Fig. 6B), whereas mutation of R16 and T17 (R16G/T17A) did not cause a significant reduction in mitochondrial sorting (Fig. 6C). Mutations of R16, T17 and all amino acids in the LSRL motif in one construct had essentially the same effect as mutations in the LSRL region alone (Fig. 4), whereas simultaneous mutation of R13, K14 and all residues in the LSRL motif strongly reduced mitochondrial sorting (Fig. 6D). In conclusion, mitochondrial targeting requires the unique N-terminal residues in UNG1 and the MLS in addition prevents nuclear import.

Detectable fluorescent material was observed from ~7 h after transfection for both UNG1 and UNG2 fusion constructs and increased steadily at least until 36 h after transfection. Fusion constructs were rapidly sorted to their respective organelles and no temporary accumulation in the cytosol was observed. Targeting patterns for constructs tested were not dependent on time after transfection. The constructs used for kinetics experiments included both fusions of EGFP with complete UNG1 or UNG2 and several mutated constructs (data not shown).

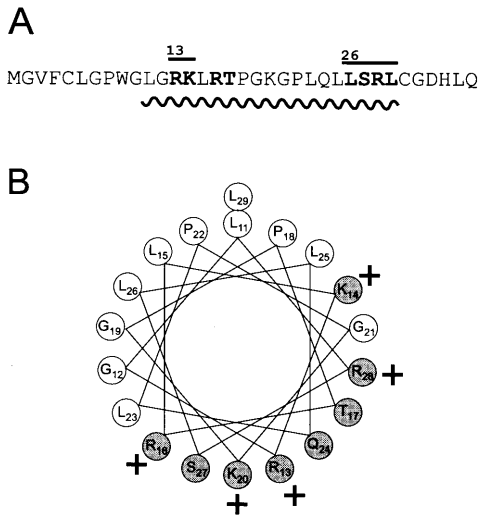


**Figure 5.** Unique N-terminal sequence of human UNG1. (**A**) Regions important for mitochondrial sorting are indicated by straight lines, while the wavy line indicates the region potentially forming an amphiphilic helix. (**B**) Amino acids 11–29 comprising the regions potentially forming an amphiphilic helix drawn as α-helical wheels with 3.6 amino acids/turn. Charged and polar residues are shaded, and positively charged residues are indicated by +.

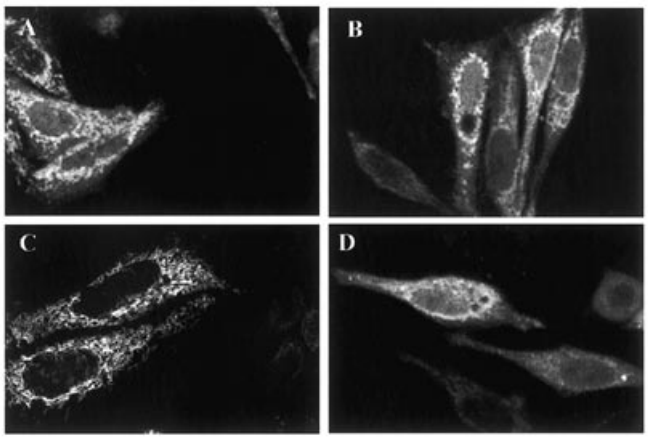


**Figure 6.** Subcellular localization in HeLa cells of UNG1 mutant proteins fused to EGFP protein. The cells were transfected with the constructs pUNG1$_{L26V/S27G/R28G/L29V}$EGFP (**A**), pUNG1$_{R13G/K14T}$EGFP (**B**), pUNG1$_{R16T/T17A}$EGFP (**C**) or pUNG1$_{R13G/K14T/L26V/S27G/R28G/L29V}$EGFP (**D**).

## DISCUSSION

### Nuclear and mitochondrial isoforms generated from one gene may be more common than previously assumed

The recently reported mechanism for generation of UNG1 and UNG2 by the use of two promoters and alternative splicing was the first example of this mechanism for generation of mitochondrial and nuclear isoforms of a protein (6). A very similar mechanism was shortly thereafter also shown for mitochondrial and nuclear isoforms of human dUTPase (15). In general, proteins involved in DNA metabolism may be prime candidates for use of such a mechanism. Interestingly, different splice forms have been reported for two other DNA glycosylases; the human OGG1 enzyme that removes mutagenic 8-oxoguanine residues resulting from oxidative stress (16) and the MPG enzyme that removes 3-methyladenine and some other alkylation products (17). One of the OGG1 splice forms, type 1a, contains a potential NLS and was recently found to have a nuclear localization (18). An enzymatically closely related mitochondrial rat OGG has been identified and partially purified (19). Recently, a predominant nuclear localization of OGG1 type 1a was confirmed, but in addition some mitochondrial localization was observed. Other splice forms of OGG1 were localized to mitochondria, as were the human DNA glycosylase hMYH, while the human glycosylase hNTH1 was localized to both nucleus and mitochondria (20). It may also be likely that one of the MPG splice forms may represent a mitochondrial enzyme, since mitochondria are apparently competent in repair of alkylation damage (21) and probably in other types of BER as well, but they are not competent in nucleotide excision repair (22). Recently, complete *in vitro* repair of abasic sites with mitochondrial enzymes was reported, further strongly suggesting that mitochondria are competent in BER (23). There are a few other examples of generation of nuclear and mitochondrial protein isoforms from one gene. Thus, a single mouse gene encodes the mitochondrial transcription factor A and a testis-specific nuclear HMG box protein, but the mechanism generating the two isoforms is not clear (24). In addition, the yeast *TRM1* gene encodes a tRNA methyltransferase that is sorted to both mitochondria and nuclei, apparently because the protein contains both an NLS and an MLS (25). Thus, generation of nuclear and mitochondrial isoforms of proteins from one gene may be more common than thought previously and it may take place by at least two different mechanisms; alternative splicing and dual import signals.

### Nuclear, subnuclear and mitochondrial targeting of UNG proteins

NLSs have been identified for a number of proteins and usually contain one or more clusters of positively charged amino acids and may be localized both in the N- and C-terminal regions (reviewed in 26–28). In addition, phosphorylation of amino acids flanking the NLS (reviewed in 29), as well as neutral and even acidic residues (30), may be important for nuclear import. Signals for subnuclear localization, as well as for nuclear export, have also been identified (28,29). The NLS in UNG2 was found to be surprisingly complex and extends some 100 amino acids downstream from the unique N-terminal sequence, which itself is absolutely required, but not sufficient, for nuclear targeting. In addition to a bipartite motif of basic residues (positions 17–19 and 49–50) separated by an unusually long spacer of 30 amino acids, less well-characterized sequence regions extending into the catalytic domain were required for complete nuclear sorting. Interestingly, several double substitutions of basic residues in UNG2 interspersed between sequence regions required for nuclear targeting do not affect nuclear translocation, but strongly enhance nuclear spot intensity, pointing to a role of these residues in subnuclear distribution. Thus, the N-terminal part of UNG2, extending into the second α-helix of the catalytic domain, has complex functions in nuclear and subnuclear targeting, as well as in RPA binding.

Import of proteins to mitochondria requires N-terminal MLSs which do not display significant primary sequence homologies, but carry a net positive charge and may form amphiphilic helical secondary structures (31,32; reviewed in 33,34), possibly aided by a chaperonin (35) or by interaction with lipid surfaces (36). Our present study demonstrates that the unique N-terminal end of UNG1 may form an amphiphilic helix as an essential part of the complete MLS. Although a part of the sequence that is common to UNG1 and UNG2 overlaps with the complex NLS, UNG1 is efficiently excluded from nuclei, apparently not due to the presence of nuclear export signals, but rather due to the strength of the MLS which, even when fused to complete UNG2, prevents nuclear import. No single mutation of residues dramatically changed mitochondrial import, in contrast to the situation for UNG2, where certain single mutations in basic residues completely abolished sorting.

## ACKNOWLEDGEMENTS

We would like to thank H. Nilsen for supplying the UNG2 clone, B. Kavli for supplying UNG1 deletion clones, P. A. Aas, I. B. Folstad, C. Brekken and M. K. Holsve for technical assistance and Dr R. Standal for help with computer analyses. This work was supported by The Research Council of Norway, The Norwegian Cancer Society and the Cancer Fund at the Regional Hospital, Trondheim.

## REFERENCES

1 Kubota,Y., Nash,R.A., Klungland,A., Schar,P., Barnes,D.E. and Lindahl,T. (1996) *EMBO J.*, **15**, 6662–6670.
2 Nicholl,I.D., Nealon,K. and Kenny,M.K. (1997) *Biochemistry*, **36**, 7557–7566.
3 Krokan,H.E., Standal,R. and Slupphaug,G. (1997) *Biochem. J.*, **325**, 1–16.
4 Impellizzeri,K.J., Anderson,B. and Burgers,P.M.J. (1991) *J. Bacteriol.*, **173**, 6807–6810.
5 Haug,T., Skorpen,F., Kvaløy,K., Eftedal,I., Lund,H. and Krokan,H.E. (1996) *Genomics*, **36**, 408–416.
6 Nilsen,H., Otterlei,M., Haug,T., Solum,K., Nagelhus,T.A., Skorpen,F. and Krokan,H.E. (1997) *Nucleic Acids Res.*, **25**, 750–755.
7 Haug,T., Skorpen,F., Aas,P.A., Malm,V., Skjelbred,C. and Krokan,H.E. (1998) *Nucleic Acids Res.*, **26**, 1449–1457.
8 Mol,C.D., Arvai,A.S., Slupphaug,G., Kavli,B., Alseth,I., Krokan,H.E. and Tainer,J.A. (1995) *Cell*, **80**, 869–878.
9 Slupphaug,G., Mol,C.D., Kavli,B., Arvai,A.S., Krokan,H.E. and Tainer,J.A. (1996) *Nature*, **384**, 87–92.
10 Nagelhus,T.A, Haug,T., Singh,K.K., Keshav,K.F., Skorpen,F., Otterlei,M., Bharati,S., Lindmo,T., Benichou,S., Benarous,R. and Krokan,H.E. (1997) *J. Biol. Chem.*, **272**, 6561–6566.
11 Slupphaug,G., Eftedal,I., Kavlie,B., Bharati,S., Helle,N.M., Haug,T., Levine,D.W. and Krokan,H.E. (1995) *Biochemistry*, **34**, 128–138.
12 Nagelhus,T.A., Slupphaug,G., Lindmo,T. and Krokan,H.E. (1995) *Exp. Cell Res.*, **220**, 292–297.
13 Dang,C.V. and Lee,W.M. (1989) *J. Biol. Chem.*, **264**, 18019–18023.
14 Nix,D.A. and Beckerle,M.C. (1997) *J. Cell Biol.*, **138**, 1139–1147.
15 Ladner,R.D. and Caradonna,S.J. (1997) *J. Biol. Chem.*, **272**, 19072–19080.
16 Aburatani,H., Hippo,Y., Ishida,T., Takashima,R., Matsuba,C., Kodama,T., Takao,M., Yasui,A., Yamamoto,K. and Asano,M. (1997) *Cancer Res.*, **57**, 2151–2156.
17 Pendlebury,A., Frayling,I.M., Santibanez Koref,M.F., Margison,G.P. and Rafferty,J.A. (1994) *Carcinogenesis*, **15**, 2957–2960.
18 Bjørås,M., Luna,L., Johnsen,B., Hoff,E., Haug,T., Rognes,T. and Seeberg,E. (1997) *EMBO J.*, **16**, 6314–6322.
19 Croteau,D.L., ap Rhys,C.M., Hudson,E.K., Dianov,G.L., Hansford,R.G. and Bohr,V.A. (1997) *J. Biol. Chem.*, **272**, 27338–27344.
20 Takao,M., Aburatani,H., Kobayashi,K. and Yasui,A. (1998) *Nucleic Acids Res.*, **26**, 2917–2922.
21 LeDoux,S.P., Wilson,G.L., Beecham,E.J., Stevnsner,T., Wassermann,K. and Bohr,V.A. (1992) *Carcinogenesis*, **13**, 1967–1973.
22 Bohr,V.A. (1995) *Carcinogenesis*, **16**, 2885–2892.
23 Pinz,K.G. and Bogenhagen,D.F. (1998) *Mol. Cell. Biol.*, **18**, 1257–1265.
24 Larsson,N.G., Garman,J.D., Oldfors,A., Barsh,G.S. and Clayton,D.A. (1996) *Nature Genet.*, **13**, 296–302.
25 Li,J.M., Hopper,A.K. and Martin,N.C, (1989) *J. Cell. Biol.*, **109**, 1411–1419.
26 Dingwall,C. and Laskey,R.A. (1991) *Trends Biochem. Sci.*, **16**, 478–481.
27 Hicks,G.R. and Raikhel,N.V. (1995) *Annu. Rev. Cell Dev. Biol.*, **11**, 155–188.
28 Görlich,D. and Mattaj,I.W. (1996) *Science*, **271**, 1513–1518.
29 Jans,D.A. and Hübner,S. (1996) *Physiol. Rev.*, **76**, 651–685.
30 Makkerh,J.P.S., Dingwall,C. and Laskey,R.A. (1996) *Curr. Biol.*, **6**, 1025–1027.
31 von Heijne,G. (1986) *EMBO J.*, **5**, 1335–1342.
32 Roise,D., Theiler,F., Horvath,S.J., Tomich,J.M., Richards,J.H., Allison,D.S. and Schatz,G. (1988) *EMBO J.*, **7**, 649–653.
33 Schatz,G. (1996) *J. Biol. Chem.*, **271**, 31763–31766.
34 Neupert,W. (1997) *Annu. Rev. Biochem.*, **66**, 863–917.
35 Landry,S.J. and Gierasch,L.M. (1991) *Biochemistry*, **30**, 7359–7362.
36 Hammen,P.K., Gorenstein,D.G. and Weiner,H. (1996) *Biochemistry*, **35**, 3772–3781.