Extremophiles (2002) 6:73–86
Digital Object Identifier (DOI) 10.1007/s007920100225

ORIGINAL PAPER

Olav Lanes · Ingar Leiros · Arne O. Smalås
Nils Peder Willassen

Identification, cloning, and expression of uracil-DNA glycosylase from Atlantic cod (*Gadus morhua*): characterization and homology modeling of the cold-active catalytic domain

Received: March 2, 2001 / Accepted: June 28, 2001 / Published online: December 4, 2001

Abstract Two distinct forms of the highly conserved uracil-DNA glycosylase (UNG) have been isolated from Atlantic cod (*Gadus morhua*) liver cDNA by rapid amplification of cDNA ends (RACE). From the cDNA sequences, both forms were deduced to encode an open reading frame of 301 amino acids, with an identical 267-amino-acid C-terminal region and different N-terminal regions of 34 amino acids. By comparison with the human UNG sequences, the two forms were identified as possible mitochondrial (cUNG1) and nuclear (cUNG2) forms. Several constructs of recombinant cUNG (rcUNG) were expressed in *Escherichia coli* in order to optimize the yield. The recombinant enzyme was purified to apparent homogeneity as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Activity and stability experiments showed that rcUNG was similar to cUNG previously purified from Atlantic cod liver, and was more pH- and temperature labile than a recombinant human UNG (rhUNG). Under optimal assay conditions for both rcUNG and rhUNG, the turnover number (k_{cat}) was three times higher for rcUNG compared with rhUNG, with an identical K_M, resulting in a threefold higher catalytic efficiency (k_{cat}/K_M) for rcUNG. These activity and stability experiments reveal cold-adapted features in rcUNG. Homology models of the catalytic domains of Atlantic cod (cUNG) and mouse uracil-DNA glycosylase (mUNG) were built using the human UNG (hUNG) crystal structure as a template. The unique amino acid substitutions observed in cod UNG were mainly located in the N- and C-terminal parts of the sequence. The analysis indicated a more stable N-terminal, a more flexible C-terminal, and a less stabilized core in cUNG as compared with the mammalian UNGs. Substitution of several amino acids in or near the DNA-binding site in cUNG could give rise to a more positively charged surface and a higher electrostatic potential near the active site compared with the mammalian UNGs. The higher potential may increase the electrostatic interactions between the enzyme and DNA, and may explain the increased substrate affinity and, in combination with the higher flexibility, the higher catalytic efficiency observed for rcUNG.

Key words Uracil-DNA glycosylase · UDG · UNG · Subcellular targeting · Cold-adapted · Cold-active · Recombinant gene expression · Homology modeling · *Gadus morhua*

Introduction

The appearance of the RNA base uracil in single- or double-stranded DNA may arise from two different cellular events. One event is the misincorporation of deoxyuridine monophosphate (dUMP) instead of deoxythymidine monophosphate (dTMP) during replication, producing a U:A base pair. The other event arises from spontaneous deamination of cytosine to uracil within DNA to yield a U:G mismatch base pair, which can lead to a C → T transition mutation in the next round of DNA synthesis. If not repaired, these DNA lesions may present a serious threat to the information integrity of the genome. Several different enzymes have been shown to be able to remove uracil from DNA. These include a cyclin-like uracil-DNA glycosylase (UDG2) (Muller and Caradonna 1991), a newly identified family of single-strand-selective monofunctional uracil-DNA glycosylase (SMUG) (Haushalter et al. 1999), a mismatch-specific thymine/uracil-DNA glycosylase (MUG) (Barrett et al. 1998), uracil-DNA glycosylase from *Thermotoga maritima* (TMUDG) (Sandigursky and Franklin 1999), and the major and highly conserved uracil-DNA glycosylase (UDG) first cloned from *E. coli* (Duncan and Chambers 1984). The latter is homologous to the human classic uracil-

Communicated by K. Horikoshi

O. Lanes · N.P. Willassen (✉)
Department of Molecular Biotechnology, Institute of Medical Biology, Faculty of Medicine, University of Tromsø, N-9037 Tromsø, Norway
Tel. +47-77-644651; Fax +47-77-645350
e-mail: nilspw@fagmed.uit.no

I. Leiros · A.O. Smalås
Department of Chemistry, Faculty of Science, University of Tromsø, Tromsø, Norway

DNA glycosylase (UNG; EC 3.2.2.3), encoded by the *UNG* gene (Slupphaug et al. 1995). While the physiological function of the other classes of uracil-DNA glycosylases is somewhat unclear, the function and mechanism of the highly conserved and specific UNG are well documented (Krokan et al. 1997; Parikh et al. 2000). UNG is believed to be the major enzyme for the specific removal of uracil from DNA by hydrolyzing the *N*-glycosylic bond linking the base to the deoxyribose sugar, leaving an abasic site, which is subsequently removed by an apurinic/apyrimidinic (AP)-endonuclease and a phosphodiesterase. The resulting gap is then filled in and the remaining nick sealed by the action of DNA polymerase and DNA ligase, respectively (Kubota et al. 1996; Lindahl et al. 1977; Nicholl et al. 1997; Parikh et al. 1997).

In rat and human, enzymes with uracil-DNA glycosylase activity have been isolated from both the mitochondria and the nucleus (Domena and Mosbaugh 1985; Wittwer and Krokan 1985). Recently, it was demonstrated by Nilsen et al. (1997) that transcription from different positions in the human *UNG* gene and alternative splicing generated two isoforms of UNG (UNG1 and UNG2). The two isoforms have unique N-terminal signal sequences, whereas they are identical in the rest of their sequence. Using transfection experiments, it was shown that the different N-terminal sequences encode signals targeting the enzyme to the mitochondria (UNG1) and nucleus (UNG2) in human cells (Nilsen et al. 1997; Otterlei et al. 1998). So far, this has been observed only for human and mouse UNG (Nilsen et al. 1997, 2000).

The crystal structures of the human, herpes simplex virus 1 (HSV-1), and *E. coli* enzymes have been determined (Mol et al. 1995; Ravishankar et al. 1998; Savva et al. 1995). The UNG structure consists of a central four-stranded parallel β-sheet, surrounded by eight α-helices. The N and C termini are on opposite sides of the central β-sheet, and there are no disulfide bridges in the structure. The active site is located within a groove at the C-terminal end of the parallel β-sheet, encompassed by a positively charged DNA-binding region (Mol et al. 1995). Studies of structures of mutants of hUNG complexed with DNA have established that UNG undergoes conformational changes when forming the productive DNA-enzyme complex, and the damaged base is "flipped out" of the major groove of the DNA helix into the active site of the enzyme (Parikh et al. 1998; Slupphaug et al. 1996). This active "pinch-push-pull" mechanism proposed by Parikh et al. (1998) has recently also been confirmed kinetically (Stivers et al. 1999). The enzyme compresses the DNA phosphate backbone flanking the flipped-out nucleotide through several interactions with three serine- and proline-rich loops (4-Pro loop, 165-PPPPS-169; Gly–Ser loop, 246-GS-247; Leu272 loop, 268-HPSPLS-VYR-276). The side chain of the conserved Leu272 penetrates into the minor groove, pushing the nucleotide out of the major groove of the DNA double helix. Essential enzyme conformational changes occur when forming the productive DNA-enzyme complex. The Leu272 loop moves towards the active site and positions the catalytically important His268 within hydrogen-bonding distance of the incoming uracil base. Specific hydrogen bond-interactions with uracil are formed to stabilize the flipped-out nucleotide, and the tightly formed recognition pocket excludes other bases from entering the catalytic site (Parikh et al. 1998). The residues involved in catalysis, Asp145 and His268, are highly conserved among all UNGs so far examined.

Cold-adapted organisms may achieve enzymatic adaptability in several different ways in order to compensate for reduced catalytic rates at low temperatures. They include increased enzyme production, reduced rate of enzyme degradation, expression of isoenzymes with different temperature optima, or adaptation of preexisting enzymes (Haard 1992). In the latter case, one should expect that selection through evolution would favor enzymes with an increased catalytic potential. This has in fact been shown to be the case for most cold-adapted enzymes analyzed so far (Leiros et al. 2000; Outzen et al. 1996; Smalås et al. 1994). A higher catalytic efficiency can be achieved by a more dynamic and flexible structure with enhanced ability to undergo structural changes during catalysis (Hochachka and Somero 1984). The proposed higher structural flexibility is still the dominant theory for explaining higher catalytic efficiency at low temperatures, although it has been difficult to prove. To achieve a higher catalytic efficiency at low temperatures, enzymes must lower the activation energy barrier during catalysis. This is presumably done by binding the "transition state" more strongly than the substrate itself, and it can be achieved by having a more flexible structure and altering the electrostatic potential at and around the active or substrate-binding sites (Leiros et al. 1999, 2000).

Recently, the sequence of a psychrophilic UDG from a marine bacterium was published (Jaeger et al. 2000). However, this UDG is only 34% identical to the *E. coli* enzyme, and it is therefore difficult to establish common cold-adapted residue determinants. In an attempt to understand how cold-adapted enzymes at low temperatures are able to maintain and even increase their catalytic efficiency at a molecular level, we have initiated a study to compare warm- and cold-adapted UNG (Lanes et al. 2000). In this paper, we describe the cloning of two forms of uracil-DNA glycosylase from the cold-adapted Atlantic cod (*Gadus morhua*) and the expression, purification, and characterization of the common catalytic domain.

Materials and methods

Materials

SuperScript II RNase H^- reverse transcriptase was purchased from Gibco BRL (Carlsbad, CA, USA). Deoxy[5-^{3}H]uridine 5′-triphosphate (17.0 Ci/mmol) was obtained from Amersham (Uppsala, Sweden). The expression vector pTrc99A and fast protein liquid chromatograph (FPLC) system and purification columns were purchased from Amersham Pharmacia Biotech (Uppsala, Sweden), and restriction enzymes were obtained from New England Biolabs (Beverly, MA, USA). The SMART polymerase chain reaction (PCR) cDNA library construction kit, and the Marathon cDNA amplification kit came from Clontech

(Palo Alto, CA, USA). *E. coli* NR8052 [Δ(*pro-lac*), *thi-*, *ara*, *trpE9777*, *ung1*] (Kunkel 1985; Varshney and van de Sande 1989) and purified recombinant human UNG (UNGΔ84) (Slupphaug et al. 1995) was kindly provided by Dr. Hans E. Krokan, Institute for Cancer Research and Molecular Biology, Norwegian University of Science and Technology.

Preparation of cDNA from *Gadus morhua*

mRNA was isolated from 250 mg cod liver using an Oligotex direct mRNA Midi kit (Qiagen, Chatsworth, CA, USA), following the instructions in the manufacturer's protocol. cDNA was made from 250 ng of the isolated poly A^+ RNA using the SMART PCR cDNA library construction kit according to the protocol recommended by the manufacturer. As the template for PCR amplification of the second-strand cDNA, 2 µl of the first-strand reaction was used. The PCR analysis was done in a GeneAmp 9700 thermocycler (Perkin Elmer, Foster City, CA, USA) as follows: 95°C for 1 min followed by 16 cycles of 95°C for 15 s and 68°C for 5 min. To 50 µl of the amplified ds cDNA, 40 µg proteinase K was added and incubated at 45°C for 1 h. Proteinase K was inactivated by incubating the mixture at 90°C for 8 min. To blunt the ds cDNA ends, 15 U of T4 DNA polymerase was added and the mixture was incubated at 16°C for 30 min and at 72°C for 10 min. Finally the ds cDNA was precipitated with ethanol and resuspended in 10 µl H_2O. All tubes were kept on ice if not otherwise stated.

Generation of a 300-bp fragment of the *cUNG* gene

Degenerated oligonucleotide primers were designed from two conserved regions (GQDPYH and VFLLWG) from known UNG amino acid sequences. Codon usage for Atlantic cod was also considered when designing the primers. The cUNG fragment was generated by PCR with 10 ng cod liver cDNA as template, and PCR was carried out in a final volume of 50 µl, containing 10 mM Tris/HCl pH 9.0 (25°C), 50 mM KCl, 2.5 mM $MgCl_2$, 0.1% Triton X-100, 0.2 mM each dATP, dCTP, dGTP and dTTP, 2.0 µM each upstream primer (5′-GGH CAR GAY CCC TAY CA-3′) and downstream primer (5′-DCC CCA SAG SAG RAA VAC-3′), and 2.5 U *Taq* polymerase (Promega, Madison, WI, USA). PCR amplification was done by 94°C for 4 min, 30 cycles of 94°C for 1 min, 60°C for 1 min, and 72°C for 1 min, with a final extension step of 72°C for 5 min.

Rapid amplification of cDNA ends (RACE)

Ligation of RACE adapters to cDNA was done as described in the Marathon cDNA amplification protocol from Clontech, and after ligation the mixture was incubated for 5 min at 70°C to inactivate the ligase. Before RACE, the adapter-ligated cDNA was diluted 50 times in TE buffer and denatured by incubating at 100°C for 2 min; thereafter, it was placed directly on ice. The sequence deduced from the 300-bp fragment of the *UNG* gene was used to design two gene-specific primers for both 3′- and 5′-RACE, creating a small sequence overlap in the two fragments generated. Both 3′- and 5′-RACE reactions were done as recommended by manufacturer, by using 1 µl of the diluted cDNA with RACE adapters as the template and 0.2 µM of the internal gene-specific 3′- (5′-TGT ACC GAC ATT GAT GGC TTC AAG CAT-3′) or 5′- (5′-CCC ATC CGC TTA GAT CTC CAT GTC CAG-3′) RACE primer. RACE was done by 94°C for 30 s followed by 5 cycles of 94°C for 5 s and 72°C for 3 min, 5 cycles of 94°C for 5 s and 70°C for 3 min, and 20 cycles of 94°C for 5 s and 68°C for 3 min.

Isolation of two different UNG cDNAs

Both RACE fragments generated were sequenced using their respective internal RACE primers. Examination of the sequence of the 5′-RACE fragment found a double sequence in a small region near the 5′-end of the fragment. However, at the 5′-end of the fragment, only one sequence appeared, due to a long untranslated region (UTR) in one of the two UNG-sequences. A new primer complementary to this 5′-end was designed (5′-ATG GAA TTC GAT TGA GAT TGG CGC CTT TGG-3′), and a new PCR analysis was carried out using this new primer and the 5′-RACE internal primer, with the 5′-RACE fragment as a template, using the Advantage polymerase mix (Clontech). PCR amplification was done by 94°C for 1 min followed by 30 cycles of 94°C for 30 s, 60°C for 1 min, and 72°C for 1 min. The fragment was sequenced using the internal RACE primer, revealing one of the two cUNG sequences, and the sequence was subtracted from the double-sequence region described above. A primer complementary to the 5′-end of the underlying sequence was designed (consisting of nucleotides from both the SMART sequence and the UNG sequence) (5′-AGA AGG GGA CAT CCG CTT GC-3′), and a new PCR analysis with this primer and the internal 5′-RACE primer was carried out and the product sequenced as described above. From the two UNG sequences in the different 5′-RACE fragments, two final primers (udg1; 5′-ACC ATG GAA TTC ATG TTG TTC AAG TTA GGG TTA TGC C-3′ and udg2; 5′-ACC ATG GAA TTC ATG ATT GGT CAA CAG CAT ATC AAC TC-3′) were made to isolate full-length UNG1 and UNG2, respectively, using 10 ng cDNA as the template, a downstream primer complementary to the C-terminal (udgend2; 5′-GAG CTC GTC GAC TTA GAG TGC TCT CCA GTT TAT AGG-3′), and the same PCR conditions as described above. Both full-length sequences were verified by sequencing.

DNA sequencing

DNA sequencing was done with the Amersham Pharmacia Biotech Thermo Sequence Cy5 dye terminator kit, an ALFexpress DNA sequencer, and ALFwin sequence analyzer version 2.10. Gels were made with a Reprogel Long Read and Reproset UV-polymerizer. All items were purchased from Amersham Pharmacia Biotech (Uppsala, Sweden).

Construction of expression vectors

The catalytic domain of the *cUNG* gene was cloned into the expression vector pTrc99A, which contains a strong trc promoter upstream of a multiple cloning site (Amann et al. 1988). Several different constructs were made by PCR using the Advantage polymerase mix and the buffer supplied by manufacturer, 10 ng cDNA as template, and 0.2 µM each of upstream and downstream primers containing *Eco*RI and *Sal*I restriction sites, respectively. PCR amplification was carried out by 94°C for 4 min followed by 30 cycles of 94°C for 30 s, 60°C for 1 min, and 72°C for 2 min, with a final extension step of 72°C for 5 min. PCR products were purified, digested with *Eco*RI and *Sal*I, and ligated into the pTrc99A expression vector. Competent *E. coli* JM105 was transformed, and plasmid DNA was reisolated from positive clones, and transformed into competent *E. coli* NR8052 used for expression of recombinant cUNG (rcUNG).

Six different expression constructs were made by PCR. rcUNGΔ74 and rcUNGΔ74o and rcUNGΔ81 and rcUNGΔ81o, from which 74 and 81 of the N-terminal amino acids were removed, and had the same length as the human UNGΔ77 and UNGΔ84, respectively (Slupphaug et al. 1995). The rcUNGΔ74o and rcUNGΔ81o constructs had two adjacent arginine codons (AGA 87 and AGA 88) close to the N-terminal of the constructs optimized for expression in *E. coli* by changing them to the more commonly used CGT codon in *E. coli*. In addition, the full-length cUNG1 and cUNG2 were cloned into the vector and expressed. All constructs were made by PCR under the conditions described above. The rcUNGΔ81o was PCR-amplified using 10 ng cDNA as a template and the udgl84 (5′-ACC ATG GAA TTC TTC GGA GAG ACT TGG AGA AGA-3′)and udgend2 primers.

Protein determination

Protein concentrations were determined with Coomassie protein assay reagent G-250 (Pierce, New York, NY) by the method of Bradford (1976), according to the microtiter-plate protocol as described by the manufacturer and using bovine serum albumin (BSA) as a standard.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed as described by Laemmli (1970) with a 12% polyacrylamide separating gel and a 5% stacking gel using a Mini-Protean 3 electrophoresis system (Bio-Rad, Hercules, CA, USA). Gels were stained with Coomassie brilliant blue G250.

Enzyme assay

UNG activity was measured using [^{3}H]-dUMP-labeled DNA made by nick translation (specific activity 1.441×10^5 dpm/µg) or PCR (specific activity 5.9×10^5 dpm/µg) as substrate (Lanes et al. 2000). The UNG assay was carried out in a final volume of 20 µl, containing 70 mM Tris/HCl pH 8.0, 10 mM NaCl, 1 mM ethylenediaminetetraacetate (EDTA), 100 µg/ml BSA, and 230 ng nick substrate or 71 ng PCR substrate. The reaction mixture was incubated for 10 min at 37°C and terminated with the addition of 20 µl of ice-cold single-stranded calf thymus DNA (1 mg/ml; Sigma, St. Louis, MO, USA) and 500 µl 10% (w/v) trichloroacetic acid (TCA). The samples were incubated on ice for 15 min, followed by centrifugation at 16,000 g for 10 min. Supernatants with acid-soluble ^{3}H-uracil were analyzed with a liquid scintillation counter. These are referred to as standard assay conditions. One unit of activity is defined as the amount of enzyme required to release 1 nmol of acid soluble uracil per minute at 37°C.

Small-scale expression and fermentation of uracil-DNA glycosylase

Optimization of expression were done in 1-l baffled Erlenmeyer flasks with 100 ml LB+ medium containing 100 µg/ml ampicillin, inoculated with 5 ml of preculture. Cells were induced with 1 mM of isopropyl thiogalactoside (IPTG) at OD_{600} = 2.0, and induced at various lengths as indicated in the figures. Expression was examined under various temperatures (20°C, 25°C, 30°C, and 37°C).

Fermentation was done in a 15-l Chemap CF 3000 fermenter (Volketswil, Switzerland). A 200-ml preculture of *E. coli* NR8052 transformed with the pTrc99A rcUNGΔ84o construct was inoculated to 7 l of LB medium supplemented with 20 mM glucose and 100 µg/ml ampicillin. Cells were grown to an OD_{600} of 2.0 and induced with 1 mM IPTG for 8 h. Additional glucose [3 × 50 ml of 20% (w/v) glucose] was supplemented during the fermentation to avoid glucose starvation. Cells were harvested and centrifuged at 10,000 g for 10 min, and the cell paste was frozen at –70°C.

Purification of recombinant cod UNG

Cell disruption and all purification steps were performed at 4°C if not otherwise stated. From the fermentation, cell paste from 4 l of *E. coli* NR8052 (68 g wet weight) were resuspended in 400 ml extraction buffer [25 mM Tris/HCl, 10 mM NaCl, 1 mM EDTA, 1% (v/v) glycerol, 1 mM DTT, 1 mM phenylmethylsulfonyl fluoride (PMSF) pH 8.0]. The cells were disrupted five times with a Bio-Neb cell disruptor (Glas-Col, Terre Haute, IN, USA) using 0.689 MPa of nitrogen. The cell extract was centrifuged at 25,000 g for 10 min, and the pellet was resuspended in 100 ml of buffer A (25 mM Tris/HCl, 10 mM NaCl, 1 mM EDTA, 1% glycerol, pH 8.0) and recentrifuged as described above. The two supernatants were combined and filtrated through glass wool.

To the crude extract (460 ml), 60 ml of 2% (w/v) protamine sulfate in buffer A was added and incubated at 4°C for 5 min with stirring. The solution was centrifuged at 25,000 g for 10 min, and the supernatant was collected.

The protamine sulfate fraction (510 ml) was applied on a Q-Sepharose FF column (5.0/10) coupled with an SP-

Sepharose FF column (2.6/10), both equilibrated in buffer A, using a flow rate of 10 ml/min. The columns were washed with 750 ml buffer A, and then the Q-Sepharose column was removed. The SP-Sepharose column was washed with an additional 550 ml buffer A + 60 mM NaCl, and a gradient from 60 to 400 mM NaCl in buffer A was applied to elute the column, using a flow rate of 5 ml/min. Fractions of 10 ml were collected, and fractions containing UNG activity were pooled.

The pooled SP-Sepharose fraction (115 ml) was diluted two times in buffer A, and directly applied to a Blue Sepharose FF column (1.6/5.0) equilibrated in buffer A, using a flow rate of 4 ml/min. The column was then washed with 30 ml of buffer A and 60 ml of buffer A + 110 mM NaCl, and thereafter eluted with 60 ml buffer A + 0.7 M NaCl. UNG-containing fractions were pooled.

Using an Ultrafree 15 unit, (Millipore, Bedford, MA, USA), the Blue Sepharose fraction (24 ml) was concentrated to 3 ml and applied to a Superdex 75 column (2.6/60) equilibrated in buffer A + 0.15 M NaCl. A flow rate of 2 ml/min was used, and fractions of 5 ml were collected. Fractions containing UNG activity were pooled.

The Superdex 75 fraction (30 ml) was diafiltrated in an Ultrafree 15 unit (Millipore), and 20 ml of the fraction was applied to a Source 15S column (2.6/3.0) equilibrated in buffer A at room temperature. The column was washed with 60 ml buffer A + 60 mM NaCl, and then a gradient from 60 to 210 mM NaCl in buffer A was applied to elute the column. Fractions were collected on ice, and UNG-containing fractions were pooled (UNG activity eluted between 105 and 145 mM NaCl).

Characterization experiments

The characterization experiments were carried out as previously described (Lanes et al. 2000). In brief, the pH and NaCl optima were determined under various concentrations of NaCl (0–200 mM) and pH ranging from 9.0–6.5. Temperature optima for rcUNG and recombinant human UNG (rhUNG) were determined by assays in a final volume of 20 µl in 25 mM hydroxyethylpiperazine ethanesulfonic acid (HEPES), 10 mM NaCl, 1 mM EDTA, pH 8.0 at different temperatures. Assays were performed simultaneously in a gradient PCR machine at temperatures from 37°C to 55°C in a manner similar to that described for the standard enzyme assay.

The pH stability was examined by preincubating rcUNG (0.01 U) (in a total volume of 50 µl) for 10 min at 37°C in 10 mM buffer, 10 mM NaCl, 1 mM EDTA, 1% (v/v) glycerol, with pH ranging from 10.0 to 5.5.

Temperature stability was examined by incubating UNG (0.01 U) in 10 mM Tris/HCl, 50 mM NaCl, 1 mM EDTA, 1% (v/v) glycerol, pH 8.0 (in a total volume of 50 µl; pH was adjusted at each temperature). After different time intervals, 5 ml aliquots were transferred to the assay mixtures and residual activity was measured using standard assay conditions and half-lives were determined.

K_M and k_{cat} were measured in the presence of 1.0–3.8 µM [^{3}H]-dUMP containing substrate prepared by nick translation. The UNG amounts used were 9 and 43 pg for rcUNG, and 24 and 121 pg for rhUNG at 37°C and 15°C, respectively. Assays were performed under two different conditions, as previously described, in 25 mM Tris/HCl, 50 mM NaCl, 1 mM EDTA, 100 µg/ml BSA, pH 7.5 and in 25 mM Tris/HCl, 10 mM NaCl, 1 mM EDTA, 100 µg/ml BSA, pH 8.0. The kinetic constants were determined at both 37°C and 15°C for both conditions described.

Model building

The structures of the catalytic domains of cod and mouse uracil-DNA glycosylase were made by homology modeling using the program O (Jones et al. 1991). The crystal structure of the catalytic domain of human uracil-DNA glycosylase (PDB1akz) was used as a template (Mol et al. 1995). Three additional N-terminal amino acids encoded by the expression vector (Met82, Glu83, and Phe84) were included in the 1akz in addition to the catalytic domain (85–304). These three N-terminal amino acids were also included in the two model sequences to make sequences of identical length. Sequence alignment of known vertebrate uracil-DNA glycosylases was performed using Clustal W (Thompson et al. 1994). The structural alignment for homology modeling contained no gaps or inserts. In order to reduce the local strain on the system, a simple energy minimization of the models were performed using the GROMOS96 implementation of Swiss-PdbViewer (Guex and Peitsch 1997). For better comparison of the hydrogen-bonding patterns within the different models, the crystal structure of human uracil-DNA glycosylase was subjected to the same energy minimization protocol as the models of the mouse and cod structures. The energy minimization of the crystal structure of human uracil-DNA glycosylase resulted in only minor changes (rmsd 0.05 Å for all atoms), but was necessary because the hydrogen-bonding pattern was optimized in the same way as for the two homology models. The final models were evaluated using the structure validation software WHAT_CHECK (Hooft et al. 1996).

Analysis of structural features

Hydrogen bonds were identified using HBPLUS (McDonald and Thornton 1994) and ion-pair interactions were determined using programs in the CCP4 program package (Collaborative Computational Project No. 4 1994). Ion-pair interactions between acidic and basic residues were determined using cut-off distances between donor and acceptor atoms of both 4 and 6 Å. Aromatic interactions were determined by manual inspection. Aromatic interactions were defined as when the ring centers of two adjacent aromatic residues were within 7 Å of each other. Classification of internal and external residues was based on water-accessible surfaces found in the cUNG model using programs in the CCP4 program suite (Collaborative Computational Project No. 4 1994). Residues with accessible surfaces of less than 10 Å^2 were defined as internal residues (I); the others were external (E). Estimates of

electrostatic surface potentials were visualized by the computer program Grasp (Nicholls et al. 1991). Helix propensities were calculated from the scale devised by Pace and Scholtz (1998).

Results

Sequences of cUNG1 and cUNG2

Using RACE, we were able identify two different forms of cod uracil-DNA glycosylase. The sequences isolated, cUNG1 and cUNG2, consisted of 1,283 bp and 1,355 bp, respectively. The variation in length was due to a longer 5′-untranslated region (UTR) in cUNG2 than in cUNG1. Both sequences had an open reading frame (ORF) of 906 bp, encoding 301 amino acids. The two sequences had 34 unique N-terminal amino acids, while the 267 C-terminal amino acids were identical.

The two cod sequences were aligned with the human (hUNG1 and hUNG2) and mouse (mUNG1 and mUNG2) sequences (Fig. 1), and the catalytic domain of cUNG (82–301) (85–304 for hUNG1) was 75% and 77% identical to the catalytic domains of hUNG and mUNG, respectively. It is thus highly conserved among vertebrates. Overall, cUNG was more similar to mouse than human UNG, and the N-terminal domain of UNG2 was more conserved than UNG1 (Table 1).

The N-terminals of the N-terminal domains consisted of a potential subcellular signal sequence (1–34 in cUNG) which differed between UNG1 and UNG2, and an identical C-terminal region (35–81) (Fig. 1).

Expression and purification

The expression of various constructs of the catalytic domain from cUNG (74, Δ74o, Δ81, Δ81o) and of full-length cUNG1 and cUNG2 using the pTrc99A expression vector were compared by measuring the enzyme activity in *E. coli* cell extracts. Small-scale expression was conducted at 30°C, and the highest activity was found using the codon-optimized Δ81o construct (346 U/mg total protein), which gave a 140-fold increase compared with the Δ81 construct. The two full-

Table 1. Identity scores among vertebrate UNG sequences

UNG	Full sequence (%)	N-terminal domain (%)	Catalytic domain (%)
Cod1/human1	64	35	75
Cod1/mouse1	65	34	77
Human1/mouse1	81	56	90
Cod2/human2	64	39	75
Cod2/mouse2	66	41	77
Human2/mouse2	84	70	90

Atlantic cod UNG (cUNG), human (hUNG), and mouse (mUNG)
Full sequence (1–301), N-terminal domain (1–81), and catalytic domain (82–301), following cod numbering

Fig. 1. Multiple alignment of UNG1 and UNG2 from vertebrates. The alignment was performed by using Clustal W (1.08), and corrected manually by inspection. The N-terminal domain is indicated by a *line* above the sequences, and the consensus (100%) is shown below the alignment. The *numbers* on the right of the alignment denote the distance from the N-terminal for UNG1 and UNG2, respectively

length constructs (cUNG1 and cUNG2) yielded 5.3 U/mg and 6.6 U/mg total protein after 6 h induction, respectively.

Expression at various temperatures between 20°C and 37°C using the Δ81o construct showed that a fourfold increase in total UNG activity was obtained when the temperature was reduced from 37°C to 25°C. rcUNGΔ81o was fermented at 30°C, and a typical fermentation yielded approximately 128 U/ml.

The purification results for rcUNGΔ81o are summarized in Table 2. From 41 of the fermentation batch, 4.8 mg of recombinant cUNG was purified to apparent homogeneity as determined by SDS-PAGE (Fig. 2). The enzyme was purified 156-fold with a yield of 34%. The high yield after the Superdex 75 column was verified several times, and may be explained by the removal of a compound with inhibitory effect on UNG. The molecular weight was determined to be 28 kDa by SDS-PAGE, a little higher than the theoretical molecular weight of 25.251 kDa. The specific activity of rcUNG was determined to be 30,092 U/mg using the nick-generated substrate and standard assay conditions. Using near optimal conditions (41°C, 50 mM NaCl, pH 7.5), the specific activity was determined to be 45,750 U/mg. This is approximately twice that of the recombinant human UNG, which had a specific activity of 19,800 U/mg under optimal conditions (45°C, pH 7.9 and 10 mM NaCl) (Slupphaug et al. 1995).When tested in this system, the specific activity of rhUNG was determined to be 25,320 U/mg.

Characterization experiments

The pH optimum for rcUNG was found to be 7.0 with 50 mM NaCl. Both pH and NaCl concentration affect the activity, and the optimal concentration of NaCl changes from 25 mM at pH 9.0 to 50–75 mM at pH 7.0 (Fig. 3). The temperature optima for cUNG and hUNG were determined to be 45°C and 47°C, respectively (data not shown). pH-stability experiments showed that rcUNG was most stable between pH 6.5 and pH 8.5, with maximal stability at pH 7.5. The half-life of rcUNG was examined at different

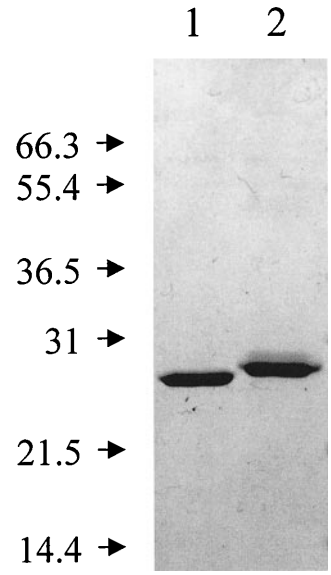

Fig. 2. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis gel of purified recombinant cod UNG. Purification of rcUNGΔ81o was done as described in the experimental procedures. *Lane 1,* 3 μg purified rcUNG; *lane 2*, 3 μg recombinant human UNG (UNGΔ84). The gel was stained with Coomassie brilliant blue G-250

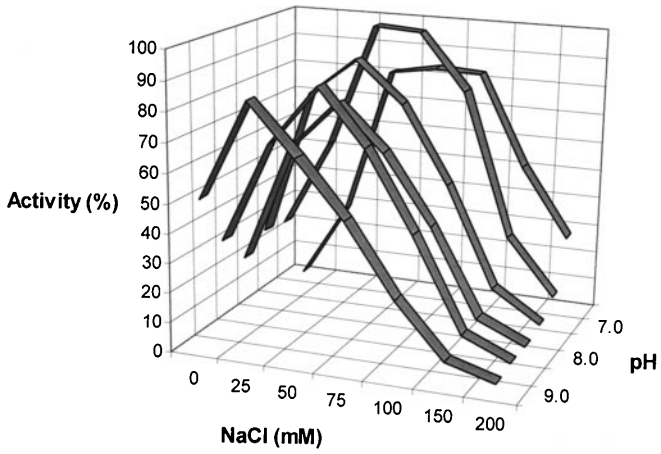

Fig. 3. pH and NaCl optima for rcUNG. Activity of rcUNG was measured in different pH buffers with 0–200 mM NaCl, as described in the experimental procedures. One hundred percent activity is set relative to the highest measured value at pH 7.0 with 50 mM NaCl

Table 2. Purification of recombinant cod uracil-DNA glycosylase (rcUNG)[a]

Step	Volume (ml)	Activity (U/ml)	Protein concentration (mg/ml)	Total activity (U)	Total protein (mg)	Specific activity (U/mg)	Yield (%)	Purification (-fold)
Crude extract	460	938	4.86	431,480	2,236	193	100	1.0
Protamine sulfate	510	822	2.67	419,220	1,362	308	97	1.6
Q/SP-Sepharose FF	115	2,082	0.21	239,430	24.2	9,914	55	51
Blue-Sepharose	24	9,531	0.66	228,744	15.8	14,441	53	75
Superdex 75	30	10,263	0.42	307,890	12.6	24,436	71	127
Source 15S[b]	20	7,222	0.24	144,440	4.8	30,092	34	156

[a]Uracil-DNA glycosylase was purified as described under Materials and methods. Enzyme assays were done using standard conditions, and protein concentration was determined as described in Materials and methods

[b]Approximately 2/3 of the Superdex 75 fraction was applied to the Source 15S column

Table 3. Kinetic constants determined for recombinant human and cod uracil-DNA glycosylase[a]

Condition	UNG	V_{max} (nmol/min/mg)	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}\mu M^{-1}$)	Increase[b] (-fold)
pH 7.5 (37°C)	Human	33,700	860	2.8	304	1
	Cod	77,750	1,963	0.6	3,385	11
pH 8.0 (37°C)	Human	24,500	625	0.6	1,024	1
	Cod	24,000	606	0.6	1,045	1
pH 7.5 (15°C)	Human	3,200	82	2.0	42	1
	Cod	6,800	172	0.5	337	8
pH 8.0 (15°C)	Human	3,700	94	0.6	152	1
	Cod	3,500	88	0.2	384	2.5

[a]Kinetic values were determined as described under experimental procedures. The kinetic constants were calculated from Eadi-Hofstee plots.
[b]Increase in k_{cat}/K_M between rhUNG and rcUNG determined for each of the four different conditions

temperatures and determined to be 0.5 min at 50°C, 20 min at 37°C, 80 min at 25°C and 1.6 h at 4°C (data not shown).

The kinetic constants, K_M and k_{cat}, were determined under two different conditions at both 37°C and 15°C. The two conditions tested, pH 8.0 with 10 mM NaCl and pH 7.5 with 50 mM NaCl, are the optimal conditions for rhUNG (Slupphaug et al. 1995) and rcUNG, respectively, with respect to pH and NaCl concentration (Table 3). Under their respective optimal conditions at 37°C, the catalytic efficiency, k_{cat}/K_M, was approximately threefold higher for rcUNG compared with rhUNG. Under these conditions, the higher catalytic efficiency was due to a threefold higher turnover number, k_{cat}, for rcUNG than for rhUNG, since the K_M values were found to be identical. Generally K_M was lower for rcUNG than for rhUNG and decreased with decreasing temperature. Under the optimum conditions for rhUNG, the catalytic efficiency was found to be in the same range for the two enzymes.

Structural alignment and amino acid substitutions

The sequence comparison, including the vertebrate mUNG, hUNG, and cUNG sequences, was based on the sequence derived from the 3-D structure of the catalytic domain of hUNG. For the hUNG sequence, three N-terminal residues (MEF) encoded from the expression vector were also used in the comparison and calculations. These residues were also added to the mUNG and cUNG sequences for proper comparison. Each of the three sequences consists of 223 amino acids (Fig. 4). The enzymes from mouse and human sources are referred to as "mammalian" and the cod enzyme as "cold-active" in the following discussion. Of the 56 amino acid substitutions between cod UNG and the mammalian UNGs, 31 are unique, 16 amino acid substitutions share the same properties, and the remaining 9 amino acids are found in one or the other of the mammalian sequences. Furthermore, the majority of the substitutions, 31 amino acids, are found in secondary structural elements, while 25 amino acids are found in loop regions. The unique substitutions in the cod enzyme are mainly localized in the N-terminal part of the structure, including the first three α-helices (α1, α2, and α3), and in the C-terminal part, including the last three α-helices. The sequence comparison of the UNGs from the different vertebrates reveals several potentially interesting substitutions in the cold-adapted enzyme. The substitutions may be divided into four classes: hydrophobicity and packing substitutions; flexibility substitutions; charged substitutions; and helical propensities and N- or C-capping substitutions.

Based on the accessible surface, 149 residues were defined as external, two were classified as external/internal, and 72 as internal residues. More than 90% (50 amino acids) of the substitutions are for external residues, while only six are found internally. The internal substitutions are Thr127 to Ser, Val230 to Ile, Leu240 to Val, and Val274 to Ala (Fig. 4). In addition, Tyr174 present in hUNG and cUNG is replaced by a Phe in mUNG, and Thr287 in hUNG is substituted by Ala in cUNG and mUNG. In total, cUNG has three fewer methyl groups in the hydrophobic core compared with the mammalian UNGs, which may lead to an increased thermolability of cUNG.

All 14 proline residues in cUNG are fully conserved in both mammalian enzymes, which have one additional proline that is replaced by an alanine in cUNG (Ala122). There are 17, 18, and 19 glycine residues in cUNG, hUNG, and mUNG, respectively. Of these, 15 are fully conserved in all sequences, while 18 residues are conserved between the mammalian enzymes. The C-terminal Gly289 is unique to cUNG, while three N-terminal glycine residues, Gly95, Gly98, and Gly107, are unique to the mammalian UNGs.

Interestingly, the distribution of glycine residues was found to be different in cUNG as compared with the mammalian UNGs. Of the four glycines present in the N-terminal part of the mammalian UNGs (82–130), including the first three α-helices (α1–α3), only one is present in cUNG. In the C-terminal catalytic part, the opposite is observed, and cUNG has more glycine residues than the mammalian UNGs.

The total number of charged residues ranges from 45 to 48 residues, and most of these are well conserved (Table 4). Among the 20–22 acidic amino acid residues in the three sequences, 16 are fully conserved in all of the sequences. In the mammalian UNGs, three acid residues are unique, Glu171, Glu289, and Glu303, while five are unique to cUNG. Of the 25 to 26 basic residues in the analyzed sequences, 19 are conserved. Four residues are unique for mammalian UNGs and six for cUNG. It is also worth men-

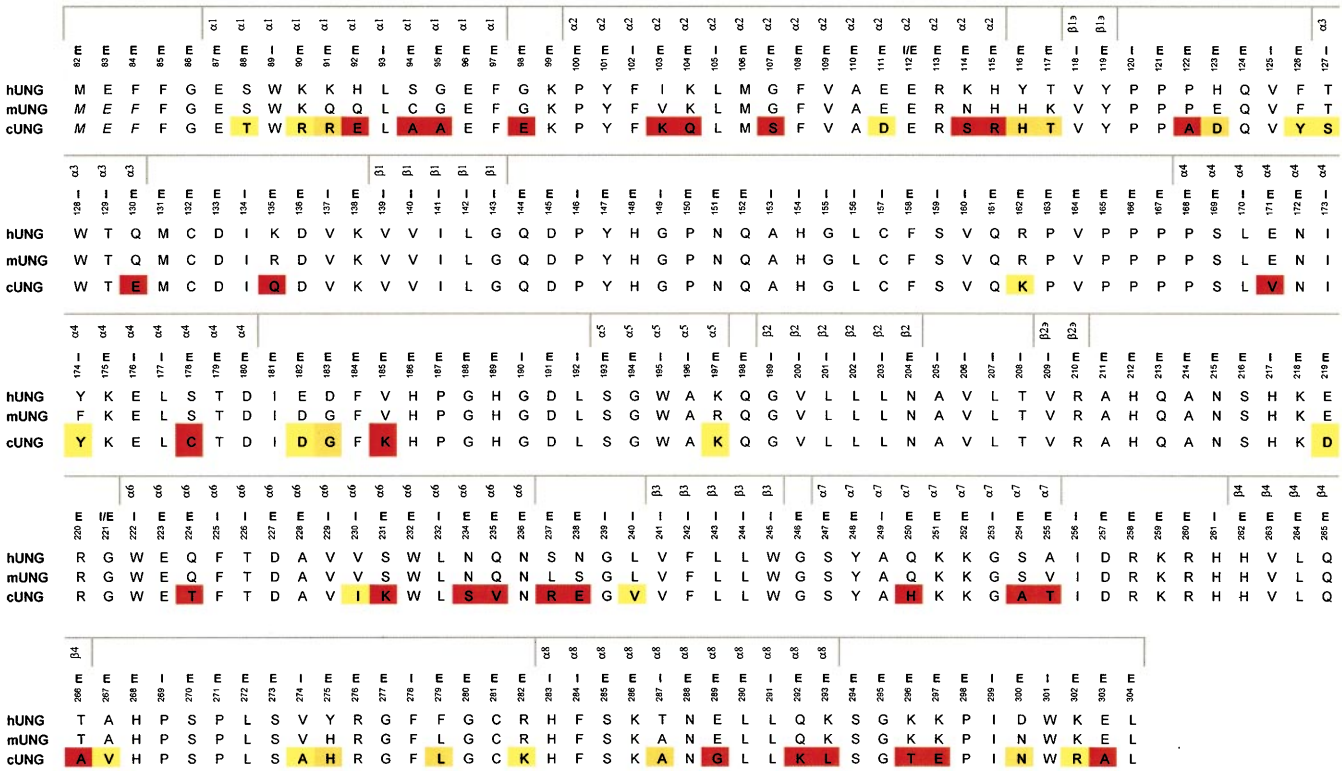

Fig. 4. Structural sequence alignment of the three vertebrate UNGs. The residues M, E, and F (in *italics*) were added to the *mUNG* and *cUNG* sequences to facilitate comparison. *E* and *I* denote external and internal residues, respectively. Residues boxed in *red* are unique to *cUNG*, residues boxed in *yellow* have the same properties as those of *mUNG* and *hUNG*, whereas residues boxed in *orange* can be found in either one or the other of the mammalian sequences

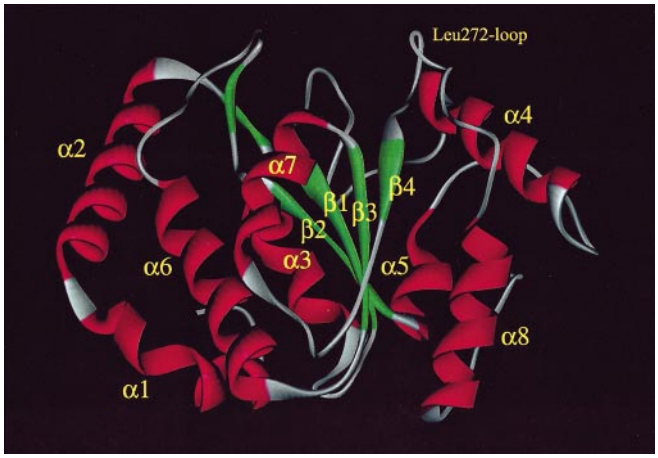

Fig. 5. Homology model of uracil-DNA glycosylase from Atlantic cod

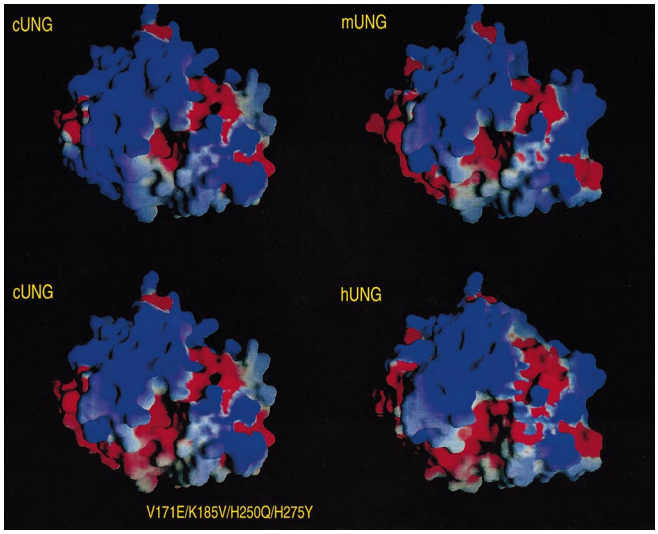

Fig. 6. Electrostatic surface potential of the three vertebrate UNGs and the mutant of cUNG (V171E, K185V, H250Q, H275Y). *Red* indicates a negative potential, whereas *blue* indicates positive potential

tioning that the mammalian UNGs have a highly charged C-terminal compared with the cod enzyme. Mammalian UNGs have six or seven charged residues among the last 16 residues, while cUNG have only three charged residues. The arginine content and the ratio of Arg/Arg+Lys are within the same range for all UNG sequences analyzed.

Structural comparison

The models of the catalytic domain of cod UNG and mouse UNG were compared with the crystal structure of human UNG with respect to hydrogen-bond interactions, ion-pair interactions, hydrophobicity, and electrostatic surface potentials. An overview of the secondary structural elements in UNG is shown in Fig. 5. Comparing the cod sequence with the human sequence, 53 of the total of 223 amino acids are substituted. All residues involved in the

Table 4. Amino acid composition for warm and cold UNG

	mUNG		hUNG		cUNG	
pI	9.05		9.04		8.85	
Net charge	5		5		4	
Amino acid	No	%	No	%	No	%
Charged	45	20.2	47	21.1	48	21.5
Acidic	20	8.96	21	9.42	22	9.87
Basic	25	11.2	26	11.7	26	11.7
Polar	109	48.9	113	50.7	109	48.9
Hydrophobic	95	42.6	92	41.3	97	43.5
Aromatic	24	10.8	27	12.1	24	10.8
Ala (A)	10	4.48	10	4.48	16	7.17
Cys (C)	4	1.79	3	1.34	4	1.79
Asp (D)	8	3.58	9	4.03	11	4.93
Glu (E)	12	5.38	12	5.38	11	4.93
Phe (F)	13	5.82	13	5.82	11	4.93
Gly (G)	19	8.52	18	8.07	17	7.62
His (H)	13	5.82	13	5.82	13	5.82
Ile (I)	6	2.69	7	3.13	7	3.13
Lys (K)	15	6.72	18	8.07	15	6.72
Leu (L)	22	9.86	20	8.96	21	9.41
Met (M)	3	1.34	3	1.34	3	1.34
Asn (N)	9	4.03	8	3.58	7	3.13
Pro (P)	15	6.72	15	6.72	14	6.27
Gln (Q)	14	6.27	12	5.38	9	4.03
Arg (R)	10	4.48	8	3.58	11	4.93
Ser (S)	14	6.27	15	6.72	13	5.82
Thr (T)	6	2.69	8	3.58	9	4.03
Val (V)	19	8.52	17	7.62	18	8.07
Trp (W)	7	3.13	7	3.13	7	3.13
Tyr (Y)	4	1.79	7	3.13	6	2.69
Total	223	100	223	100	223	100
Arg/Arg+Lys	0.596		0.614		0.609	
Leu + Ile/Leu + Ile + Val	0.400		0.308		0.423	

Theoretic isoelectric point (pI) values were calculated from the sequences. Net charge is the difference between basic and acidic residues (KR–ED)

Residues included in the different classes are charged (KRED), basic (KR), acidic (ED), polar (QNCSTYRKHDE), hydrophobic (AILFPMWV), and aromatic (FWY)

Table 5. Total number of hydrogen bonds and ion-pair interactions

Number of	hUNG	mUNG	cUNG
Backbone–backbone hydrogen bonds	125	127	127
Backbone–side-chain hydrogen bonds	63	62	55
Side-chain–side-chain hydrogen bonds	19	18	17
Total number of hydrogen bonds	207	207	199
Ion-pair interactions <4 Å	11	8	7
Ion-pair interactions <6 Å	23	22	18

formation of the substrate-binding site and the recognition pocket and the catalytic residues (Mol et al. 1995; Parikh et al. 1998; Slupphaug et al. 1996) are well conserved.

A total of 199 hydrogen bonds were identified in cUNG compared with 207 in the mammalian UNGs. The backbone–backbone interactions were found to be virtually identical for all three structures, and variations in the hydrogen-bond pattern were mainly due to a different number of side-chain-backbone interactions (Table 5). The number of potential ion-pair interactions was also found to be higher in the mammalian UNGs than in the cold-active cUNG. Interestingly, the hydrogen-bond distribution was found to be significantly different for cUNG as compared with the mammalian UNGs. In the N-terminal part of the enzyme, including the three first α-helices (residues 82–130), cUNG seems to be stabilized by an increased number of hydrogen bonds and potential ion-pair interactions as compared with the mammalian UNGs. This is in contrast to the C-terminal part (residues 131–304), where more unique hydrogen bonds and ion pairs were identified in the mammalian UNGs. All four ion-pair interactions unique for cUNG are located within the first two α-helices, and three of these interactions are internal α-helix ion pairs, possibly stabilizing the two helices. In mammalian UNGs, all five unique ion pairs are located after the first three helices, mainly in connection with the secondary structures α4 and α8.

Aromatic-aromatic interactions have also been shown to contribute to protein stability (Burley and Petsko 1985; Serrano et al. 1991). Only three aromatic to nonaromatic substitutions were found: Tyr116, Tyr275, and Phe279 in human UNG are substituted by two histidines and Leu, respectively, in the mouse and cod sequences. In hUNG, Phe279 is stacked perpendicularly to Tyr275, possibly forming an aromatic-aromatic interaction between the two residues in the Leu272 loop. Overall, the aromatic interactions do not seem to be a major factor responsible for the observed differences in stability among the UNGs.

Approximately 38% of the amino acid residues in the UNGs are located in the eight α-helices. Since amino acids have different propensities to form helical structures, the composition in helical regions may affect not only the helix stability but also the overall stability of the protein. Calculations of the α-helical sequences based on the scale of Pace and Scholtz (1998) for the three homologous enzymes show a higher degree of stabilization for cUNG compared with hUNG and mUNG. Only two of the eight α-helices (α3 and α6) indicate a lower degree of stabilization for cod UNG.

All N-capping amino acids in the eight helices in both human and cod UNGs are well conserved. Only one mutation in the C-cap of cUNG can account for the lower degree of stabilization. Lys293 in mammalian UNGs is replaced with an Ile in the cod enzyme. Large hydrophobic residues are thought to destabilize helices by shielding the terminal carbonyl groups, thereby destabilizing the helix structure.

The higher efficiency observed for cold-adapted enzymes might arise from a lower activation energy as a result of an improved accommodation of the substrate and from a reduction in the cost of conformational changes in the structure during catalysis at low temperatures (Hochachka and Somero 1984).

In UNG, the movement of the Leu272 loop is essential for bringing the catalytically important residue, His268, within hydrogen-bonding distance of uracil O2 and for accomplishing the formation of the uracil-recognition pocket (Parikh et al. 1998). In cUNG, four residues in con-

nection with this loop are substituted when compared with the mammalian homologues. Two residues anchoring the Leu272 loop to β4 in the mammalian UNGs are substituted in cUNG: Thr266 and Ala267 by Ala and Val, respectively. In the middle of the loop, Val274 in the mammalian UNGs is substituted by an Ala in cUNG, and hUNG Tyr275 is substituted by His in both cod and mouse. In the mammalian UNGs, the Oγ1 atom of the Thr266 side chain forms a hydrogen bond with the backbone N atom of Ala267, which is not present in cUNG, probably anchoring the loop to β4. The backbone oxygen of Thr266 also forms a hydrogen bond with the side chain N ∈ 2 of Gln250, which is not present in cUNG.

Positively charged residues dominate the UNG active site where DNA binds to the enzyme. In the DNA-binding site, cUNG has only a few substitutions when compared with the mammalian UNGs. In cUNG, Arg162 is substituted by Lys, Gln250 by His, Ala267 by Val, and Val274 by Ala. As mentioned above, Tyr275 is substituted by His in both cUNG and mUNG. Also the residues Val185 and Glu171, which are substituted by Lys and Val, respectively, in cUNG, are localized close to this area and may affect the electrostatic potential around the DNA-binding sites. When the electrostatic surface potential of the three enzymes was contoured, the surface potential appeared to be more positive around the active and DNA-binding sites in cUNG compared with both hUNG and mUNG. A modeled mutant of cUNG, where four cod residues were substituted compared with hUNG (Val171 to Glu, Lys185 to Val, His250 to Gln, and His275 to Tyr), revealed that these four residues could account for the higher electrostatic potential around the DNA-binding site (Fig. 6). A higher potential could promote stronger electrostatic interactions between the enzyme and the DNA.

Discussion

The UNG enzymes are highly conserved among vertebrates

The human UNG gene was the first mammalian gene shown to encode both a nuclear (UNG2) and a mitochondrial (UNG1) isoform by a mechanism involving transcription from two different promoters and alternative splicing (Nilsen et al. 1997). The alternative splicing of the same gene to give two different UNGs has been proposed previously for the fish *Xiphophorus*, and the exon-intron boundaries in UNGs from human, mouse, and *Xiphophorus* were found to be conserved (Krokan et al. 1997; Nilsen et al. 1997; Walter and Morizot 1996). However, no sequence has yet been published, and *Gadus morhua* represents the first fish species and lower vertebrate from which two different UNG sequences have been described. From the identification of the two different forms of UNG and from sequence alignment, we suggest that the lower vertebrate represented by the fish *G. morhua* may also share the alternative splicing mechanism, and that this mechanism may have been evolved before the two taxa were separated more than 450 million years ago. In prokaryotes, the N-terminal domain is absent, and so far two different UNGs have been determined only for human, mouse, and cod.

The best-characterized nuclear localization signals (NLS) are the simple basic or "classic" monopartite and the basic bipartite signals (Mattaj and Englmeier 1998). These signals are generally characterized by one or more clusters of basic amino acids, but they do not fit a tight consensus. The vertebrate UNG2s seems to share the common "classic" signal region, with three to four basic amino acids, starting with proline on the N-terminal side and followed by an acidic residue or proline at the C-terminal end (Fig. 1). In human UNG2, the motif RKR (residues 17–19) among the unique 44 N-terminal residues was found to be essential for sorting (Otterlei et al. 1998). The cod UNG2 also has a consensus PCNA-binding site (QxIxxFF), which is also conserved in human, mouse, and yeast (Otterlei et al. 1999).

Mitochondrial matrix import signals are usually between 20 and 35 residues long and rich in hydroxylated and basic amino acids, and they can fold into a amphipathic α-helix or sheet (Schatz and Dobberstein 1996). The signal usually lacks acidic amino acids, and the sequence is generally removed on the *trans* side of the membrane by a mitochondrial processing peptidase (MPPase). The mitochondrial signal sequences observed in the different UNG1s are more heterogeneous than the nucleus signal sequence, and only three positions are conserved (Fig. 1). The length of the mitochondrial signal sequence ranges from 35 to 45 amino acids, compared with the nucleus signal where the length differs by only five amino acids.

Although the C-terminal part (35–81) of the N-terminal domain is not necessary for activity, it contains highly conserved regions (Fig. 1), which are biased to positively charged amino acids. Since no structural information is available for this region, several secondary prediction methods were used to identify secondary structural elements in this part of the sequence. The first half of this region (35–50) gave no clear indications, but in the last half (50–86), a 21-amino-acid-long amphipathic α-helix (55–78) was predicted by all programs. It is likely that this region is involved in an interaction with other proteins participating in the DNA base excision repair (BER) pathway, as has been shown for the human UNG (Nagelhus et al. 1997; Otterlei et al. 1998, 1999).

rcUNG is more catalytically efficient than rhUNG

The interactions between DNA-interacting enzymes and DNA are basically electrostatic in nature, and many of the enzymes operate in a processive manner (Dodson et al. 1994; Lohman 1986; von Hippel and Berg 1989). UNG has been shown to operate in both a processive and distributive manner (Bennett et al. 1995; Higley and Lloyd 1993; Purmal et al. 1994), and processivity is probably dependent on a certain pH and NaCl concentration for proper binding to the DNA. The different optima, with respect to pH and NaCl concentration, between the cod and human enzymes could reflect a variation in their electrostatic surface potentials, giving them different DNA-binding characteristics.

The kinetic constants were determined at optimum conditions for both rcUNG and rhUNG. The catalytic efficiency, k_{cat}/K_M, of rcUNG was found to be higher than that of rhUNG. Overall, the higher catalytic efficiency depended on both a higher turnover, k_{cat}, and a lower K_M. The K_M seems to decrease with decreasing temperature for both rcUNG and rhUNG, and under optimal conditions K_M seems to be around 0.6 μM for both enzymes. Generally, electrostatic interactions increase with decreasing temperature, and an increase in the electrostatic interactions between the enzyme and DNA can be one of the reasons for the decreasing K_M with decreasing temperature. However, at low temperatures, rcUNG seems to be able to lower the K_M more than rhUNG in order to increase the catalytic efficiency.

rcUNG is less pH- and temperature-stable than rhUNG, but is similar to native cUNG

Cold-adapted enzymes are often characterized by lower pH and temperature stability (Smalås et al. 2000), and this has also been shown to be the case for native cUNG (Lanes et al. 2000). The recombinant cUNG was found to be identical to the native cUNG previously purified from cod liver, and both enzymes are more pH and temperature labile than rhUNG. While rcUNG was most stable from pH 6.5 to 8.5, rhUNG was stable from pH 6.5 to at least pH 10. The half-lives of rcUNG and rhUNG were determined to be 0.5 min and 8.0 min, respectively, at 50°C.

rcUNG might be more flexible than the mammalian UNGs

The strategy of lowering the number of proline residues and increasing the number of glycine residues to obtain a higher structural flexibility does not seem to apply to cUNG. However, the localization of these residues seems to be important. The higher content of Met residues found in many cold-adapted enzymes is believed to increase their flexibility (Aittaleb et al. 1997; Leth-Larsen et al. 1996). The mechanism by which Met destabilizes the proteins is not known, but mutation studies reveal a less stable structure when methionines are introduced (Gassner et al. 1996). A high number of methionines is not only unique to cold-adapted enzymes, but common among enzymes of marine origin (Leiros et al. 1999). However, no such differences were found in the three UNG structures compared. A higher arginine content and hence a higher Arg/Arg+Lys ratio have often been used an indicator of higher structural stability (Leiros et al. 1999). Arginine is thought to stabilize protein structure through its ability to form ion pairs and hydrogen bonds (Mrabet et al. 1992). The arginine content and Arg/Arg+Lys ratio are in the same range for all the UNGs, which may indicate that these are not important for the lower stability observed for cUNG compared with hUNG. However, the stabilizing effect has been questioned, since arginine is typically an external residue and is often involved in interactions with water (Leiros et al. 1999).

A replacement of larger amino acids by smaller in the hydrophobic core can lead to the creation of small cavities and hence to a reduced packing density. It has been shown that the removal of methyl groups in the core can destabilize the protein if it leads to formation of cavities (Eriksson et al. 1992; Jackson et al. 1993). The reduced number of methyl groups in cUNG may also affect the stability and flexibility.

The three N-terminal α-helices in cUNG seem to be stabilized by an increased number of ion pairs and hydrogen bonds in comparison with the mammalian UNGs. These N-terminal α-helices are not involved in DNA binding or catalysis and probably have a mainly structural function. In addition, larger and more polar residues in this part of the enzyme are replaced by glycine residues in the mammalian UNGs. The C-terminal part of the enzyme is more important for substrate binding and catalysis, and the mammalian UNGs have several ion pairs and hydrogen bonds not present in this part of the structure in cUNG. These are both intrahelical interactions, stabilizing the secondary structural elements, and interactions that connect helices.

The active "pinch-push-pull" mechanism for excision of uracil from DNA involves conformational changes of UNG during catalysis (Parikh et al. 1998). Analysis of crystal structures of hUNG complexed with DNA have shown that the Leu272 loop moves toward the active site of UNG when forming the productive enzyme-DNA complex (Parikh et al. 1998; Slupphaug et al. 1996). The loop movement is also accompanied by movement of several secondary structural elements, including β4 and α8, which are connected to the N- and C-terminals of the loop, respectively, and β3, β4, α4, and α7. The α8-helix preceding the Leu272 loop and the C-terminal is probably more flexible in cUNG because of several substitutions. Glu289 is substituted by the more flexible glycine in cUNG, which probably also decreases the stability of the helix by the disruption of a possible ion-pair interaction (Glu289-Lys293) that is present in the mammalian UNGs. The α7-helix, in which the N-terminal interacts with DNA, in mammalian UNGs is linked to three other secondary structure elements, α6, β3, and β4, by hydrogen bonds from the side chain of Asn234 Oδ1 to the backbone N of Asp257, and from the side chain of Gln250 N ∈ 2 to the backbones of Leu244 O and Thr266 O. None of these hydrogen bonds are found in the cold-adapted cUNG.

Many of the determinants believed to be important for obtaining higher flexibility and hence lower stability, such as higher content of Gly and Met residues, lower content of Pro and Arg residues, and destabilization of helices, seem not to be involved in the observed decreased stability of cUNG. On the other hand, the reduced number of ion-pair interactions and hydrogen bonds and the reduced packing density of the core are likely to reduce the stability of cUNG. Cod UNG may have a more flexible catalytic C-terminal part compared to the mammalian UNGs, and this flexibility may have to be compensated by stabilizing the N-terminal part of the structure. The flexibility given by a reduced number of ion pairs and hydrogen bonds in the catalytic part of the enzyme may partly explain the higher catalytic efficiency observed for cUNG as compared with hUNG. The higher positive electrostatic potential in the DNA-binding site of cUNG as

compared with the mammalian UNGs could be one of the reasons for the lower K_M observed. However, there is a delicate balance between large contributions of stabilizing and destabilizing interactions involved in the formation of the native three-dimensional structure of a protein. It is likely that this balance can be adjusted to environmental changes by just one or a few changes in the amino acid sequence, which are presumably responsible for optimal function at low temperatures.

The nucleotide sequence data published here have been deposited in the EMBL Nucleotide Sequence Database and are available under accession numbers AJ275971 and AJ275972.

Acknowledgments We thank Dr. Steinar Johansen, Dr. Dag Rune Gjellesvik, and Chris Fenton for helpful comments and critical reading of the manuscript. This work was supported by the Research Council of Norway and Biotec ASA, Tromsø, Norway.

References

Aittaleb M, Hubner R, Lamotte-Brasseur J, Gerday C (1997) Cold adaptation parameters derived from cDNA sequencing and molecular modelling of elastase from Antarctic fish *Notothenia neglecta*. Protein Eng 10:475–477

Amann E, Ochs B, Abel KJ (1988) Tightly regulated *tac* promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene 69:301–315

Barrett TE, Savva R, Panayotou G, Barlow T, Brown T, Jiricny J, Pearl LH (1998) Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. Cell 92:117–129

Bennett SE, Sanderson RJ, Mosbaugh DW (1995) Processivity of *Escherichia coli* and rat liver mitochondrial uracil-DNA glycosylase is affected by NaCl concentration. Biochemistry 34:6109–6119

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254

Burley SK, Petsko GA (1985) Aromatic-aromatic interaction: a mechanism of protein structure stabilization. Science 229:23–28

Collaborative Computational Project No. 4 (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D 50:760–763

Dodson ML, Michaels ML, Lloyd RS (1994) Unified catalytic mechanism for DNA glycosylases. J Biol Chem 269:32709–32712

Domena JD, Mosbaugh DW (1985) Purification of nuclear and mitochondrial uracil-DNA glycosylase from rat liver. Identification of two distinct subcellular forms. Biochemistry 24:7320–7328

Duncan BK, Chambers JA (1984) The cloning and overproduction of *Escherichia coli* uracil-DNA glycosylase. Gene 28:211–219

Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. Science 255:178–183

Gassner NC, Baase WA, Matthews BW (1996) A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. Proc Natl Acad Sci U S A 93:12155–12158

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723.

Haard NF (1992) A review of proteolytic enzymes from marine organisms and their applications in the food industry. J Aquat Food Prod Technol 1:117–125

Haushalter KA, Todd Stukenberg MW, Kirschner MW, Verdine GL (1999) Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. Curr Biol 9:174–185

Higley M, Lloyd RS (1993) Processivity of uracil DNA glycosylase. Mutat Res 294:109–116

Hochachka PW, Somero GN (1984). Biochemical adaptation. Princeton University Press, New Jersey

Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381:272

Jackson SE, Moracci M, elMasry N, Johnson CM, Fersht AR (1993) Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. Biochemistry 32:11259–11269

Jaeger S, Schmuck R, Sobek H (2000) Molecular cloning, sequencing, and expression of the heat-labile uracil-DNA glycosylase from a marine psychrophilic bacterium, strain BMTU3346. Extremophiles 4:115–122.

Jones TA, Zou JY, Cowan SW, Kjeldgaard (1991) Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 47:110–119.

Krokan HE, Standal R, Slupphaug G (1997) DNA glycosylases in the base excision repair of DNA. Biochem J 325:1–16

Kubota Y, Nash RA, Klungland A, Schar P, Barnes DE, Lindahl T (1996) Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein. EMBO J 15:6662–6670

Kunkel TA (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci U S A 82:488–492

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Lanes O, Guddal PH, Gjellesvik DR, Willassen NP (2000) Purification and characterization of a cold-adapted uracil-DNA glycosylase from Atlantic cod (*Gadus morhua*). Comp Biochem Physiol B 127:399–410.

Leiros HK, Willassen NP, Smalås AO (1999) Residue determinants and sequence analysis of cold-adapted trypsins. Extremophiles 3:205–219

Leiros HKS, Willassen NP, Smalås AO (2000) Structural comparison of psychrophilic and mesophilic trypsins. Elucidating the molecular basis of cold-adaptation. Eur J Biochem 267:1–12

Leth-Larsen R, Asgeirsson B, Thorolfsson M, Norregaard-Madsen M, Hojrup P (1996) Structure of chymotrypsin variant B from Atlantic cod, *Gadus morhua*. Biochim Biophys Acta 1297:49–56

Lindahl T, Ljungquist S, Siegert W, Nyberg B, Sperens B (1977) DNA N-glycosidases: properties of uracil-DNA glycosidase from *Escherichia coli*. J Biol Chem 252:3286–3294

Lohman TM (1986) Kinetics of protein-nucleic acid interactions: use of salt effects to probe mechanisms of interaction. Crit Rev Biochem 19:191–245

Mattaj IW, Englmeier L (1998) Nucleocytoplasmatic transport: the soluble phase. Annu Rev Biochem 67:265–306

McDonald IK, Thornton JM (1994) Satisfying hydrogen bonding potential in proteins. J Mol Biol 238:777–793.

Mol CD, Arvai AS, Slupphaug G, Kavli B, Alseth I, Krokan HE, Tainer JA (1995) Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. Cell 80:869–878

Mrabet NT, Van den Broeck A, Van den brande I, Stanssens P, Laroche Y, Lambeir AM, Matthijssens G, Jenkins J, Chiadmi M, van Tilbeurgh H, et al. (1992) Arginine residues as stabilizing elements in proteins. Biochemistry 31:2239–2253

Muller SJ, Caradonna S (1991) Isolation and characterization of a human cDNA encoding uracil-DNA glycosylase. Biochim Biophys Acta 1088:197–207

Nagelhus TA, Haug T, Singh KK, Keshav KF, Skorpen F, Otterlei M, Bharati S, Lindmo T, Benichou S, Benarous R, Krokan HE (1997) 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 272:6561–6566

Nicholl ID, Nealon K, Kenny MK (1997) Reconstitution of human base excision repair with purified proteins. Biochemistry 36:7557–7566

Nicholls A, Sharp KA, Honig B (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 11:281–296

Nilsen H, Otterlei M, Haug T, Solum K, Nagelhus TA, Skorpen F, Krokan HE (1997) Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from

different positions in the UNG gene. Nucleic Acids Res 25:750–755

Nilsen H, Steinsbekk KS, Otterlei M, Slupphaug G, Aas PA, Krokan HE (2000) Analysis of uracil-DNA glycosylases from the murine *Ung* gene reveals differential expression in tissues and in embryonic development and a subcellular sorting pattern that differs from the human homologues. Nucleic Acids Res 28:2277–2285.

Otterlei M, Haug T, Nagelhus TA, Slupphaug G, Lindmo T, Krokan HE (1998) Nuclear and mitochondrial splice forms of human uracil-DNA glycosylase contain a complex nuclear localisation signal and a strong classical mitochondrial localisation signal, respectively. Nucleic Acids Res 26:4611–4617

Otterlei M, Warbrick E, Nagelhus TA, Haug T, Slupphaug G, Akbari M, Aas PA, Steinsbekk K, Bakke O, Krokan HE (1999) Post-replicative base excision repair in replication foci. EMBO J 18:3834–3844

Outzen H, Berglund GI, Smalås AO, Willassen NP (1996) Temperature and pH sensitivity of trypsins from Atlantic salmon (*Salmo salar*) in comparison with bovine and porcine trypsin. Comp Biochem Physiol B 115:33–45

Pace CN, Scholtz JM (1998) A helix propensity scale based on experimental studies of peptides and proteins. Biophys J 75:422–427

Parikh SS, Mol CD, Tainer JA (1997) Base excision repair enzyme family portrait: integrating the structure and chemistry of an entire DNA repair pathway. Structure 5:1543–1550

Parikh SS, Mol CD, Slupphaug G, Bharati S, Krokan HE, Tainer JA (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. EMBO J 17:5214–5226

Parikh SS, Putnam CD, Tainer JA (2000) Lessons learned from structural results on uracil-DNA glycosylase. Mutat Res 460:183–199.

Purmal AA, Lampman GW, Pourmal EI, Melamede RJ, Wallace SS, Kow YW (1994) Uracil DNA *N*-glycosylase distributively interacts with duplex polynucleotides containing repeating units of either TGGCCAAGCU or TGGCCAAGCTTGGCCAAGCU. J Biol Chem 269:22046–22053

Ravishankar R, Bidya Sagar M, Roy S, Purnapatre K, Handa P, Varshney U, Vijayan M (1998) X-ray analysis of a complex of *Escherichia coli* uracil DNA glycosylase (EcUDG) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG. Nucleic Acids Res 26:4880–4887

Sandigursky M, Franklin WA (1999) Thermostable uracil-DNA glycosylase from *Thermotoga maritima* a member of a novel class of DNA repair enzymes. Curr Biol 9:531–534

Savva R, McAuley Hecht K, Brown T, Pearl L (1995) The structural basis of specific base-excision repair by uracil-DNA glycosylase. Nature 373:487–493

Schatz G, Dobberstein B (1996) Common principles of protein translocation across membranes. Science 271:1519–1526

Serrano L, Bycroft M, Fersht AR (1991) Aromatic–aromatic interactions and protein stability. Investigation by double-mutant cycles. J Mol Biol 218:465–475

Slupphaug G, Eftedal I, Kavli B, Bharati S, Helle NM, Haug T, Levine DW, Krokan HE (1995) Properties of a recombinant human uracil-DNA glycosylase from the *UNG* gene and evidence that *UNG* encodes the major uracil-DNA glycosylase. Biochemistry 34:128–138

Slupphaug G, Mol CD, Kavli B, Arvai AS, Krokan HE, Tainer JA (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. Nature 384:87–92

Smalås AO, Heimstad ES, Hordvik A, Willassen NP, Male R (1994) Cold adaption of enzymes: structural comparison between salmon and bovine trypsins. Proteins 20:149–166

Smalås AO, Leiros HK, Os V, Willassen NP (2000) Cold adapted enzymes. Biotechnol Annu Rev 6:1–57

Stivers JT, Pankiewicz KW, Watanabe KA (1999) Kinetic mechanism of damage site recognition and uracil flipping by *Escherichia coli* uracil DNA glycosylase. Biochemistry 38:952–963

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Varshney U, van de Sande JH (1989) Characterization of the ung1 mutation of *Escherichia coli*. Nucleic Acids Res 17:813

von Hippel PH, Berg OG (1989) Facilitated target location in biological systems. J Biol Chem 264:675–678

Walter RB, Morizot DC (1996) Conservation of genome and gene structure from fishes to mammals. Adv Struct Biol 4:1–24

Wittwer CU, Krokan H (1985) Uracil-DNA glycosylase in HeLa S3 cells: interconvertibility of 50 and 20 kDa forms and similarity of the nuclear and mitochondrial form of the enzyme. Biochim Biophys Acta 832:308–318