Minireview

Base excision repair of DNA in mammalian cells

Hans E. Krokan*, Hilde Nilsen, Frank Skorpen, Marit Otterlei, Geir Slupphaug

Institute of Cancer Research and Molecular Biology, Norwegian University of Science and Technology, N-7489 Trondheim, Norway

Received 5 May 2000

Edited by Gunnar von Heijne

Abstract Base excision repair (BER) of DNA corrects a number of spontaneous and environmentally induced genotoxic or miscoding base lesions in a process initiated by DNA glycosylases. An AP endonuclease cleaves at the 5′ side of the abasic site and the repair process is subsequently completed via either short patch repair or long patch repair, which largely require different proteins. As one example, the *UNG* gene encodes both nuclear (UNG2) and mitochondrial (UNG1) uracil DNA glycosylase and prevents accumulation of uracil in the genome. BER is likely to have a major role in preserving the integrity of DNA during evolution and may prevent cancer.

Key words: DNA damage; Base excision repair; DNA glycosylase

1. Introduction

The integrity of DNA depends upon several processes that in part are integrated. These include DNA damage recognition and repair, replication, transcription and cell cycle regulation. In addition, programmed cell death (apoptosis) contributes to the genetic integrity by removing genetically altered cells. Most damaged or inappropriate bases in DNA are removed by excision repair, while a minority are repaired by direct damage reversal [1]. In base excision repair (BER) the damaged nucleotide is removed, the resulting single-stranded area is filled in by DNA synthesis, and the new segment ligated to the preexisting strand downstream of the damaged area.

Small base alterations that do not distort the DNA helix (e.g. uracil and 3-methyladenine (3-meA)) are generally repaired by BER in a multistep process initiated by a damage-specific DNA glycosylase which releases the damaged base. This leaves an abasic site that is subsequently processed further in a multistep pathway that restores the correct DNA sequence [2,3]. Covalent DNA alterations that also distort the DNA helix (e.g. pyrimidine dimers and benzpyrene adducts) are repaired by nucleotide excision repair (NER) which is initiated by an invariant damage-recognising multiprotein complex. In this case, two different endonucleases (XPF-ERCC1 and XPG) associated with the complex cleave 5′ and 3′ of the damage, thus releasing the damaged base as part of an oligonucleotide [1]. The mismatch repair (MMR) system constitutes the third major excision repair pathway. This system recognises single mispairs and mismatches involving several bases. MMR has attracted wide attention because defective MMR genes are responsible for the vast majority of cases of hereditary non-polyposis colon cancer [4].

2. Mammalian DNA glycosylases and BER

BER was discovered by Tomas Lindahl in 1974 as the result of a search for an enzymatic activity that would act on deaminated cytosine [5]. The enzyme he discovered, uracil DNA glycosylase, represented a new type of enzyme that released the damaged base. Subsequently a number of other DNA glycosylases have been reported (Table 1). These generally remove bases that cause minor structural changes in DNA. The damage recognised is most frequently the result of the inherent instability of DNA. This includes deamination of cytosine to uracil, alkylation caused by normal cellular metabolites such as *S*-adenosylmethionine, (e.g. 3-meA), oxidative damage caused by reactive oxygen species from oxidative metabolism (e.g. 8-oxoG and thymine glycol), and errors in DNA replication (misincorporation of dUTP or 8-oxo-dGTP). Many of these lesions are also caused by environmental agents, such as tobacco-specific nitrosamines that alkylate DNA, and ionising radiation that generates reactive oxygen species. All mammalian DNA glycosylases have N-terminal extensions that are not present in the bacterial counterparts. When known, these sequences have important functions in targeting enzymes to nuclei or mitochondria and in interactions with other proteins that may have a role in BER [2,6]. DNA glycosylases are generally small, monomeric proteins that cleave the N–C1′ glycosylic bond between the base and deoxyribose, thus releasing the damaged base and leaving an abasic site that is cytotoxic and mutagenic and must be further processed. Removal of the damaged base is the only catalytic function of monofunctional DNA glycosylases, such as uracil DNA glycosylases (UNG), the mismatch-specific thymine/uracil DNA glycosylase (TDG) and the methylpurine DNA glycosylases (MPG, also called ANPG or AAG). Whereas UNG and TDG have narrow substrate specificities, MPG removes a large array of damaged bases which have a weakened glycosylic bond as their only common feature [2]. Several DNA glycosylases have associated lyase activities that cleave at the 3′ side of the abasic site. Prototype examples of such enzymes are enzymes that remove oxidised purines such as 8-oxoG DNA glycosylase (OGG1) and the human Endo III homologue hNTH1 that removes oxidised pyrimidines.

After removal of the damaged base by a DNA glycosylase and incision by AP endonuclease, BER may proceed by 'short patch repair' or by 'long patch repair' (Fig. 1). In short patch

*Corresponding author. Fax: (47)-73598801.
E-mail: hans.krokan@medisin.ntnu.no

PII: S0014-5793(00)01674-4

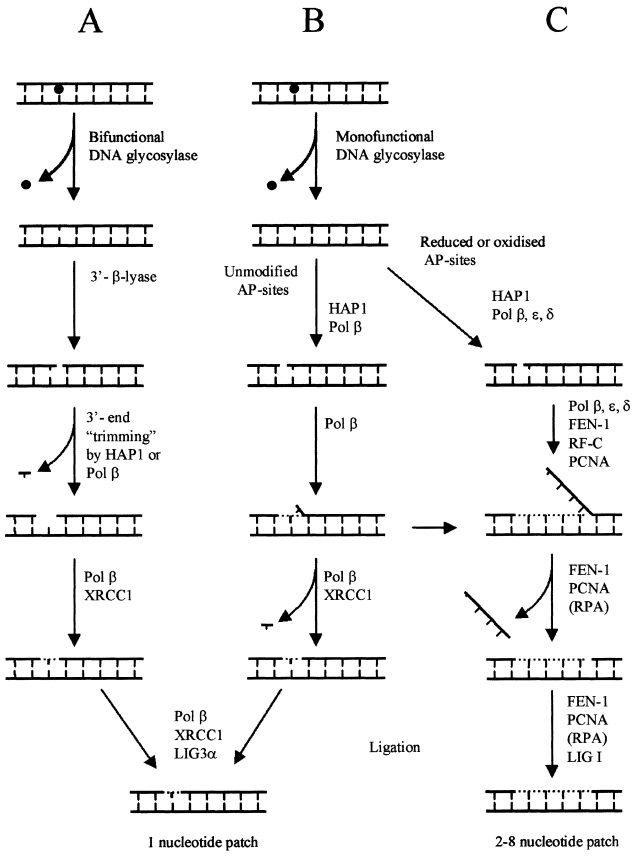


Fig. 1. Alternative BER pathways. A, B: Short-patch BER initiated by bifunctional and monofunctional glycosylases, respectively. Unmodified AP sites result in a one-nucleotide repair patch generated by a common ligation step. C: Processing of reduced or oxidised AP sites proceeds via the PCNA-dependent pathway, and involves cleavage of a 2–8-nucleotide 'flap' by FEN-1.

Fig. 2. Generation of nuclear (UNG2) and mitochondrial (UNG1) forms of UDG. A: The 13.5-kb *UNG* gene consists of seven exons and uses alternative promoters (A and B) and alternative splicing to generate UNG2 and UNG1 respectively. B: Structures of promoters A and B, indicating positive and negative regulatory elements. C: Intracellular transport of UNG1 (mitochondria) and UNG2 (nuclei) is determined by alternative N-terminal presequences. D: Amino acid sequences of unique and common regions in the UNG1 and UNG2 N-terminal domains. Residues 1–29 in UNG1 are cleaved off during mitochondrial import. Binding motifs in UNG2 to PCNA and RPA are indicated. Potential Ser/Thr phosphorylation sites in UNG2 are shown by asterisks. E: Amino acids in the human UNG2 PCNA binding region are conserved in yeast and mouse, and are also found in p21.

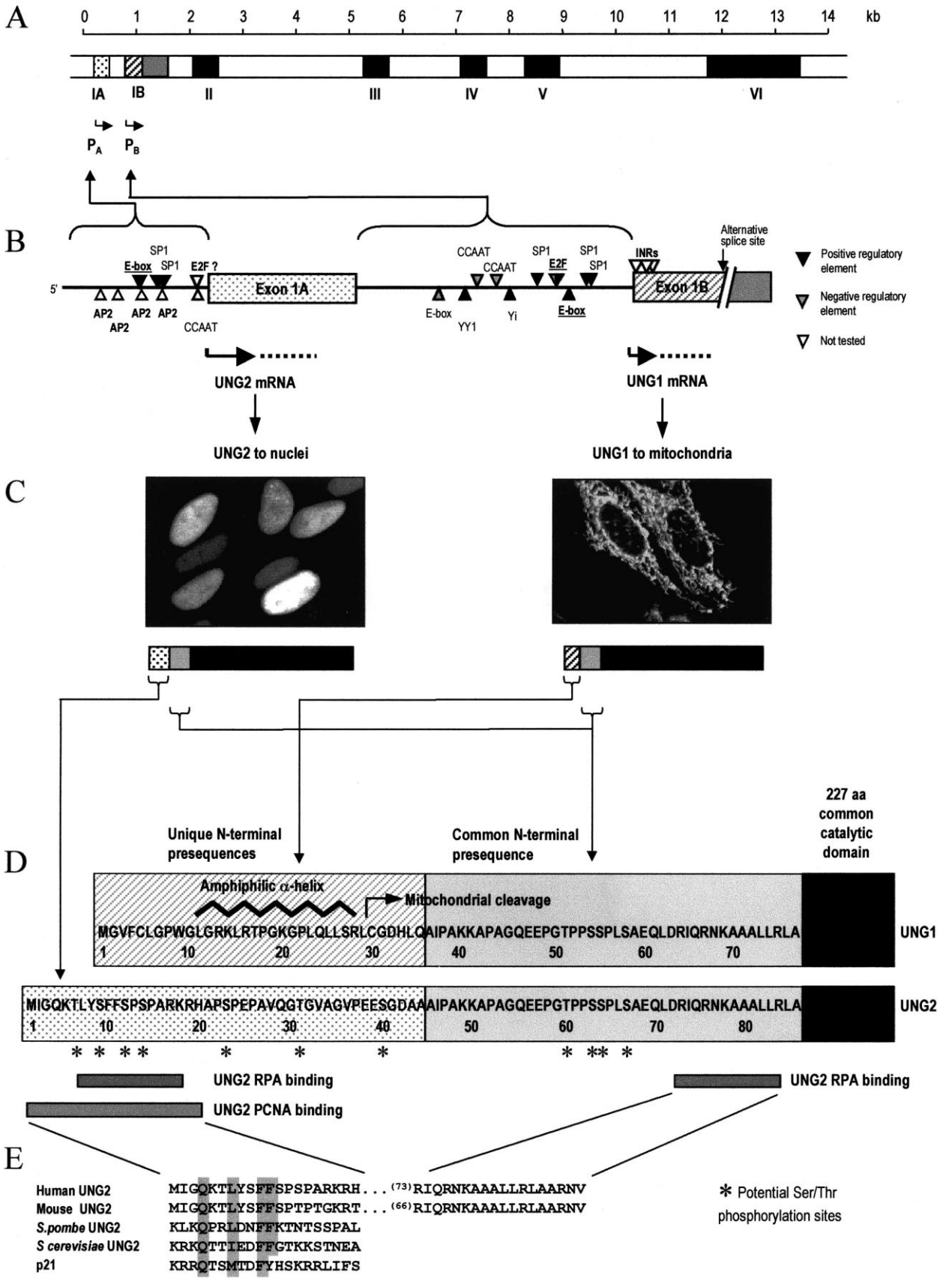
→

repair the repair gap is only one nucleotide, while in long patch repair the gap is 2–8 nucleotides in size [7,8]. The type of DNA glycosylase is one determinant for selection of the BER pathway. Thus, in vitro repair in HeLa cell extracts revealed that removal of two substrates by the monofunctional MPG/ANPG takes place by via both short and long patch repair, while repair of 8-oxoG initiated by the bifunctional OGG1 glycosylase/lyase takes place mainly via short patch repair [9]. It is also possible that the cell cycle stage is a determinant. Thus, removal of misincorporated uracil takes place in replication foci which contain all factors required for long patch repair, but not DNA polymerase β, the preferred polymerase in short patch repair [6]. However, DNA polymerase β has been shown to stimulate long patch repair for uracil removal in cell extracts where it displaces the damaged strand and co-operates with 'flap endonuclease' FEN1 in excision of the incised damaged strand [10]. If this mechanism operates in vivo it must presumably take place outside replication foci.

Table 1
Mammalian DNA glycosylases [1–3]

Name[a]	Size (aa)	Lyase activity	Cellular localisation	Chromosome localisation	Known substrates[b]
hUNG1	304	No	Mitochondria	12q24.1	ssU > U:G > U:A, 5-FU (poor: 5-hydroxyU, isodialuric acid, alloxan)
hUNG2	313	No	Nuclei	12q24.1	ssU > U:G > U:A, 5-FU (poor: 5-hydroxyU, isodialuric acid, alloxan)
mUNG1	295	No	Mitochondria	–	U, 5-FU
mUNG2	306	No	Nuclei	–	U, 5-FU
hSMUG1	270	No		12q13.1–q14	ssU > U:A, U:G
hTDG	410	No	Nuclei	12q24.1	U:G > εC:G > T:G
hMBD4	580	?	–	3q21	U or T in U/TpG:5-meCpG
hUDG2	327	No	–	5	U:A
hMPG	293	No	–	16p (telomere)	3-mA, 7-mA, 3-mG,7-mG, 8-oxoG, hypoxanthine, εA, εG
mMPG	–	No	–	–	3-mA, 7-mA, 3-mG
hOGG1[c]	345	Yes	Nuclei (1a)	3p25	Me-fapyG:C ≫ fapyG:C > 8-oxoG:C ≫ 8-oxoG:T
	424	Yes	Mitochondria (2a)		
mOGG1	345	Yes	–		8-oxoG:C ≫ 8-oxoG:T > 8-oxoG:G
hMYH[d]	521	Yes?	Mitochondria	1p32.1–p34.3	A:G, A:8-oxoG ≫ C:A, 2-OH-A
	535	Yes?	Nuclei	1p32.1–p34.3	A:G, A:8-oxoG ≫ C:A, 2-OH-A
hNTH1	312(304)	Yes	Nuclei (+mitoch?)	16p13.2–13.13	T/C-glycol, dihydrouracil, fapy,
mNTH1	300	Yes	–	17A3	T-glycol, urea

[a]h = human, m = mouse.
[b]In mismatches, the target base is at the left.
[c]Thirteen alternative OGG1 transcripts have been demonstrated, of which only 1a encodes a nuclear localisation signal.
[d]Ten alternative hMYH transcripts have been demonstrated.


A
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 kb
IA IB II III IV V VI
P_A P_B
B
5'
SP1
E-box SP1
E2F ?
Exon 1A
AP2 AP2 AP2
AP2
CCAAT
CCAAT
SP1
E2F
SP1
INRs
Alternative splice site
Exon 1B
E-box
YY1
Yi
E-box
Positive regulatory element
Negative regulatory element
Not tested
UNG2 mRNA
UNG1 mRNA
UNG2 to nuclei
UNG1 to mitochondria
C
D
Unique N-terminal presequences
Common N-terminal presequence
227 aa common catalytic domain
Amphiphilic α-helix
Mitochondrial cleavage
MGVFCLGPWGLGRKLRTPGKGPLQLLSRLCGDHLQAIPAKKAPAGQEEPGTPPSSPLSAEQLDRIQRNKAAALLRLA
1 10 20 30 40 50 60 70
UNG1
MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAKKAPAGQEEPGTPPSSPLSAEQLDRIQRNKAAALLRLA
1 10 20 30 40 50 60 70 80
UNG2
UNG2 RPA binding
UNG2 PCNA binding
UNG2 RPA binding
E
Human UNG2 MIGQKTLYSFFSPSPARKRH...(73)RIQRNKAAALLRLAARNV
Mouse UNG2 MIGQKTLYSFFSPTPTGKRT...(66)RIQRNKAAALLRLAARNV
S.pombe UNG2 KLKQPRLDNFFKTNTSSPAL
S cerevisiae UNG2 KRKQTTIEDFFGTKKSTNEA
p21 KRRQTSMTDFYHSKRRLIFS
* Potential Ser/Thr phosphorylation sites

3. Mammalian 3-meA DNA glycosylase (MPG, AAG, ANPG)

MPG has a wide substrate specificity that includes several alkylated purines, 1-*N*6-ethenoadenine and hypoxanthine. It shares this property with the bacterial enzyme AlkA although they are not related in their amino acid sequences [2]. The 3D structure of MPG revealed intercalation into the minor groove of DNA, flipping of the damaged nucleotide into the electron-rich active site, where a bound water molecule is poised for nucleophilic attack [11]. Knockout mice deficient in MPG have demonstrated that the enzyme is not essential for development or survival, and somewhat surprisingly, spontaneous mutation frequencies were not increased [12,13]. This indicates that alkylation by endogenous metabolites is probably not a major mutagenic load. MPG-deficient cells are, however, hypersensitive to killing by alkylating agents [12] and demonstrate a 3–4-fold increased mutation frequency in the *hprt* gene after exposure to alkylating agents [13].

4. Mammalian DNA glycosylases for oxidative lesions

Oxidative stress from normal metabolism or ionising radiation causes oxidative damage to both purines and pyrimidines. 8-oxoG opposite C in DNA results from oxidation of template Gs. This lesion is removed by OGG1. This relatively abundant lesion [14] is mutagenic because it may direct incorporation of either dCMP or dAMP. This may result in a GC to TA transversion. However, if template 8-oxoG escapes repair and dAMP is incorporated, hMYH glycosylase, which specifically removes A opposite 8-oxoG, may serve as a secondary defence mechanism [2]. This gives the DNA polymerase a second chance to incorporate dCMP, and subsequently 8-oxoG opposite C may be removed by OGG1. Like MPG-deficient mice, knockout mice deficient in OGG1 develop normally and do not have an obvious clinical phenotype, although they accumulate 8-oxoG in the genome and show slightly increased spontaneous mutation frequencies [14].

Oxidised pyrimidines, e.g. thymine glycol and cytosine glycol, are removed by hNTH1, a human homologue of *Escherichia coli* Endo III encoded by the *nth* gene. It removes at least five different oxidised pyrimidines [15]. This enzyme has little activity alone, but is stimulated several-fold by XPG, an endonuclease that has an important function in NER. In BER XPG has a different role and apparently aids hNTH1 binding to target DNA [16,17]. Thus, XPG has a verified role in both BER and NER.

5. DNA glycosylases that remove uracil from DNA

Enzymes that remove uracil from DNA are collectively called uracil DNA glycosylase (UDG). Deamination of cytosine creates mutagenic U:G mispairs at a frequency of 100–500 events per cell per day. In addition dUTP, a normal intermediate present in low concentrations in the cell, may be incorporated opposite template adenine during replication [2,18]. While UDG encoded by the *UNG* gene encodes the majority of the total UDG activity [19], several other UDG activities may be important as backup enzymes, or they may serve specialised functions that are not yet fully understood. These enzymes include the T(U):G mismatch DNA glycosylase, TDG [20], SMUG1 [21], MBD4 [22] and a cyclin-like enzyme [23]. Among these uracil-releasing activities only the UNG proteins and TDG have been characterised in more detail so far. TDG requires a T(U):G mismatch, and in contrast to UNG it is essentially inactive with U:A pairs and single-stranded DNA. It also removes 3-*N*(4)-ethenocytosine opposite G [24]. The structure of MUG, the bacterial homologue of TDG, in complex with target DNA has been solved. MUG/TDG enzymes are structurally related to the common catalytic core domain of UNG proteins, but have a very different mechanism for substrate recognition [25]. TDG is also larger than UNG because it requires double-stranded DNA and interacts with both the damage-containing strand and the complementary strand. Human TDG and the homologous enzyme MUG have very low catalytic rates compared with the UNG proteins.

The *UNG* gene contains seven exons, is approximately 13.5 kb and is located in chromosome 12q24.1 [26]. It contains two differentially expressed TATA-less promoters [27]. Using the two promoters and alternative splicing, the *UNG* gene encodes both the mitochondrial (UNG1) and nuclear (UNG2) forms of the major human UDG (Fig. 2). Unique N-terminal sequences in UNG1 and UNG2 are essential for correct subcellular targeting [28]. The mitochondrial preprotein UNG1 of 304 residues is processed to mature mitochondrial forms that are 29 and 75/77 amino acids shorter [29]. mRNA for UNG2 encodes a protein of 313 amino acids. The 3D structure of the common catalytic domain has identified a positively charged DNA binding groove and a tight fitting catalytic pocket tailored for binding of flipped-out uracil and exclusion of normal pyrimidines in DNA. The enzyme-assisted flipping and release of uracil have revealed a very complex mechanism for positioning of uracil and catalysis. This involves the formation of a destabilised uracil oxyanion intermediate, attack on the glycosyl bond by activated water nucleophile and strain on the glycosyl bond by conformational changes at the enzyme–DNA interface [30–33]. The N-terminal amino acids in UNG2 that are not required for catalytic activity are essential for complete nuclear translocation and for interaction with replication protein A (RPA) and proliferating cell nuclear antigen (PCNA). Furthermore, a large fraction of UNG2 co-locates with RPA and PCNA in replication foci and UNG2 is important for rapid removal of dUMP residues incorporated during DNA replication. This suggests a distinct function for this protein in immediate postreplicative removal of misincorporated uracil in DNA. Knockout mice deficient in the murine *Ung* gene develop normally, have no obvious phenotype after 18 months and have a less than twofold increase in spontaneous mutation frequencies. However, nuclei from *Ung*−/− cells show a strong deficiency in removal of misincorporated uracils and have a steady-state level of approximately 2000 uracil residues per cell [34].

DNA glycosylases create mutagenic and cytotoxic abasic sites. Interestingly, dissociation of UNG from the abasic site is a rate-limiting step in catalysis, which is stimulated by HAP1/APE1, the major AP endonuclease [29,32]. This is also the case for TDG [35]. This indicates that DNA glycosylases may remain attached to the abasic site until the next player, HAP1, is recruited. Furthermore, HAP1 interacts with DNA polymerase β that comes next. In turn, DNA polymerase β interacts with XRCC1, a scaffold protein, which interacts with the last player in short patch repair, DNA ligase III [1–3]. Thus, BER is a concerted action where the successive

players interact functionally and/or by protein–protein interactions.

6. Concluding remarks

BER mechanisms are highly conserved in evolution and apparently an ancient mechanism of defence that counteracts spontaneous decay of DNA [36]. In addition, BER protects the genome from chemical and physical threats from the environment. Apparently, absence of each of the DNA glycosylases separately results in small or moderate changes in spontaneous mutation frequencies, and no obvious immediate effects at the individual level. It is likely that DNA glycosylases have a major role in protecting the long-term integrity of the genome.

Acknowledgements: This work was supported by The Research Council of Norway, The Norwegian Cancer Society, The Cancer Research Fund at the Regional Hospital in Trondheim and the Svanhild and Arne Must Fund for Medical Research.

References

[1] Lindahl, T. and Wood, R.D. (1999) Science 286, 1897–1905.
[2] Krokan, H.E., Standal, R. and Slupphaug, G. (1997) Biochem. J. 325, 1–16.
[3] Mol, C.D., Parikh, S.S., Putnam, C.D., Lo, T.P. and Tainer, J.A. (1999) Annu. Rev. Biophys. Biomol. Struct. 28, 101–128.
[4] Lynch, H.T. and de la Chapelle, A. (1999) J. Med. Genet. 36, 801–818.
[5] Lindahl, T. (1974) Proc. Natl. Acad. Sci. USA 71, 3649–3653.
[6] Otterlei, M., Warbrick, E., Slupphaug, G., Nagelhus, T., Haug, T., Akbari, M., Aas, P.A., Steinsbekk, K., Bakke, O. and Krokan, H.E. (1999) EMBO J. 18, 3834–3844.
[7] Pascucci, B., Stucki, M., Jonsson, Z.O., Dogliotti, E. and Hübscher, U. (1999) J. Biol. Chem. 274, 33696–33702.
[8] Matsumoto, Y., Kim, K., Hurwitz, J., Gary, R., Levin, D.S., Tomkinson, A.E. and Park, M.S. (1999) J. Biol. Chem. 274, 33703–33708.
[9] Fortini, P., Parlanti, E., Sidorkina, O.M., Laval, J. and Dogliotti, E. (1999) J. Biol. Chem. 274, 15230–15236.
[10] Prasad, R., Dianov, G.L., Bohr, V.A. and Wilson, S.H. (2000) J. Biol. Chem. 275, 4460–4466.
[11] Lau, A.Y., Scharer, O.D., Samson, L., Verdine, G.L. and Ellenberger, T. (1999) Cell 95, 249–258.
[12] Engelward, B.P., Weeda, G., Wyatt, M.D., Broekhof, J.L., de Wit, J., Donker, I., Allan, J.M., Gold, B., Hoeijmakers, J.H. and Samson, L.D. (1997) Proc. Natl. Acad. Sci. USA 94, 13087–13092.
[13] Elder, R.H., Jansen, J.G., Weeks, R.J., Willington, M.A., Deans, B., Watson, A.J., Mynett, K.J., Bailey, J.A., Cooper, D.P., Rafferty, J.A., Heeran, M.C., Wijnhoven, S.W., van Zeeland, A.A. and Margison, G.P. (1998) Mol. Cell. Biol. 18, 5828–5837.
[14] Klungland, A., Rosewell, I., Hollenbach, S., Larsen, E., Daly, G., Epe, B., Seeberg, E., Lindahl, T. and Barnes, D.E. (1999) Proc. Natl. Acad. Sci. USA 96, 13300–13305.
[15] Dizdaroglu, M., Karahalil, B., Senturker, S., Buckley, T.J. and Roldan-Arjona, T. (1999) Biochemistry 38, 243–246.
[16] Cooper, P.K., Nouspikel, T., Clarkson, S.G. and Leadon, S.A. (1997) Science 275, 990–993.
[17] Klungland, A., Hoss, M., Gunz, D., Constantinou, A., Clarkson, S.G., Doetsch, P.W., Bolton, P.H., Wood, R.D. and Lindahl, T. (1999) Mol. Cell 3, 33–42.
[18] Lindahl, T. (1993) Nature 362, 709–715.
[19] Slupphaug, G., Eftedal, I., Kavli, B., Bharati, S., Helle, N.M., Haug, T., Levine, D.W. and Krokan, H.E. (1995) Biochemistry 34, 128–138.
[20] Neddermann, P. and Jiricny, J. (1994) Proc. Natl. Acad. Sci. USA 91, 1642–1646.
[21] Haushalter, K.A., Todd Stukenberg, M.W., Kirschner, M.W. and Verdine, G.L. (1999) Curr. Biol. 9, 174–185.
[22] Hendrich, B., Hardeland, U., Ng, H.H., Jiricny, J. and Bird, A. (1999) Nature 401, 301–304.
[23] Muller, S.J. and Caradonna, S. (1991) Biochim. Biophys. Acta 1088, 197–207.
[24] Hang, B., Medina, M., Fraenkel-Conrat, H. and Singer, B. (1998) Proc. Natl. Acad. Sci. USA 95, 13561–13566.
[25] Barrett, T.E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J. and Pearl, L.H. (1998) Cell 92, 117–129.
[26] Haug, T., Skorpen, F., Kvaløy, K., Eftedal, I., Lund, H. and Krokan, H.E. (1996) Genomics 36, 408–416.
[27] Haug, T., Skorpen, F., Aas, P.A., Malm, V., Skjelbred, C. and Krokan, H.E. (1998) Nucleic Acids Res. 26, 1449–1457.
[28] Nilsen, H., Otterlei, M., Haug, T., Solum, K., Nagelhus, T.A., Skorpen, F. and Krokan, H.E. (1997) Nucleic Acids Res. 25, 750–755.
[29] Bharati, S., Krokan, H.E., Kristiansen, L., Otterlei, M. and Slupphaug, G. (1998) Nucleic Acids Res. 26, 4953–4959.
[30] Mol, C.D., Arvai, A.S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H.E. and Tainer, J.A. (1995) Cell 80, 869–878.
[31] Slupphaug, G., Mol, C.D., Kavli, B., Arvai, A.S., Krokan, H.E. and Tainer, J.A. (1996) Nature 384, 87–92.
[32] Parikh, S.D., Mol, C.D., Slupphaug, G., Bharati, S., Krokan, H.E. and Tainer, H.E. (1998) EMBO J. 17, 5214–5226.
[33] Parikh, S., Walcher, G., Jones, G.D., Slupphaug, G., Krokan, H.E., Blackburn, G.M. and Tainer, J.A. (2000) Proc. Natl. Acad. Sci. USA 97, 5083–5088.
[34] Nilsen, H., Rosewell, I., Robins, P., Skjelbred, C.E., Andersen, S., Slupphaug, G., Daly, G., Krokan, H.E., Lindahl, T. and Barnes, D.E. (2000) Mol. Cell, in press.
[35] Waters, T.R., Gallinari, P., Jiricny, J. and Swann, P.F. Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1, (1999) J. Biol. Chem. 274, 67–74.
[36] Eisen, J.A. and Hanawalt, P.C. (1999) Mutat. Res. 435, 171–213.