Sequence specificity for removal of uracil from U·A pairs and U·G mismatches by uracil-DNA glycosylase from *Escherichia coli*, and correlation with mutational hotspots

Hilde Nilsen, Siamak Pour Yazdankhah, Ingrid Eftedal, Hans E. Krokan*

UNIGEN Center for Molecular Biology, University of Trondheim, N-7007 Trondheim, Norway

Received 30 January 1995; revised version received 3 March 1995

Abstract The rate of removal of uracil from different positions in double-stranded DNA by uracil-DNA glycosylase from *Escherichia coli* varied more than 15-fold. Consensus sequences for good and poor removal were 5′-(A/T)UA(A/T)-3′ and 5′-(G/C)U(T/G/C)-3′, respectively. In general, the sequence context surrounding U was more important for the rate of removal than whether U was present in U·A pairs or U·G mispairs. Rates of removal of U from sites of amber mutations in the *lacI* gene, where mutation frequencies and deamination rates were known, indicated that the observed variation in removal is biologically significant.

Key words: Uracil-DNA glycosylase; Sequence specificity; Mutational hotspot

1. Introduction

DNA repair processes are known to display heterogeneity at several levels, including preferential repair of active genes [1,2], which can be accounted for at least partly by accelerated repair of the transcribed strand due to a coupling between transcription and repair (reviewed in [3]). Lately the biological significance of sequence-dependent repair within genes after exposure to ultraviolet light was demonstrated in a study where damage and repair rates in known mutational hotspots in exons of the *p53* gene were measured [4]. Seven out of eight hotspots were found to coincide with repair 'slow spots'.

In general, GC-to-AT transition mutations account for a large fraction of mutations both in bacteria and in inherited human disease and cancer cells [5–10]. While many of these mutations may be caused by deamination of 5-methylcytosine to thymine, deamination of cytosine to uracil may also contribute significantly [6–10]. Uracil-DNA glycosylase removes uracil from DNA, thus initiating the base excisison repair pathway for removal of uracil in DNA [11]. Uracil in DNA may result from misincorporation of dUMP during replication [12,13] or from chemical deamination of cytosine residues in DNA [14]. The latter process leads to the generation of U·G mismatches that, unless repaired, will lead to GC-to-AT transition mutations. Deamination of cytosine by cytosine methyltransferase has also been observed [15], but was recently found to be unlikely as a major cause of deamination [16]. U in U·G mismatches in the specific sequences involved has also been shown to be removed by the bacterial very short patch (VSP) system for T·G mismatches [17], and a human T·G mismatch DNA glycosylase efficiently removes U from U·G mismatches in certain sequences [18]. Such systems may contribute to removal of U from some sequences, perhaps those in which uracil-DNA glycosylase works less efficiently. Recently, deficient mismatch repair due to a mutation in the human *mutS* homologue *MSH2* was shown to be associated with colon cancer [19].

We have previously reported considerable heterogeneity in the step of uracil excision from U·A matches in double-stranded DNA in vitro by uracil-DNA glycosylase from calf thymus [20]. In the present study, we have employed uracil-DNA glycosylase (Ung) from *E. coli* to examine the rates of removal of uracil from U·A pairs and U·G mismatches in otherwise identical sequence contexts. In addition, we present data suggesting that slow repair of deaminated C residues may contribute to the occurence of certain mutational hotspots in DNA.

2. Materials and methods

2.1. Enzymes and reagents

E. coli uracil-DNA glycosylase was purchased from Epicentre Technologies (Madison, WI). This preparation is homogeneous as judged from polyacrylamide gel-electrophoresis (PAGE) and did not contain any detectable endonuclease activity on depurinated DNA in the presence of Mg^{2+}. 1 unit is the amount of enzyme that releases 1 nmol of uracil per min at 37°C. All oligodeoxyribonucleotides were custom-made and PAGE purified by Genosys Biotechnologies Inc. (Woodlands, TX). [γ-^{33}P]ATP and [α-^{35}S]dATP were from Amersham (UK). T4 polynucleotide kinase was from New England Biolabs (Beverly, MA). M13 forward primer and USB Sequenase sequencing kit containing modified T7 DNA polymerase were from United States Biochemical (Cleveland, OH).

2.2. Preparation of long DNA with uracil in U·A base pairs

The template for generation of long substrates was a single-stranded M13 derivative containing the *E. coli lacI* gene, mRS81, generously provided by Dr. Roel M. Schaaper [21]. Double-stranded DNA containing dUMP in one strand was synthesized using Sequenase polymerase with dUTP partly substituting for dTTP in addition to the three other normal dNTPs. The ratio of dUTP to dTTP was 1:10 with the sum of dTTP and dUTP always being 12.5 μM, equal to the concentration of each of the other dNTPs. The primer was 5′ end-labelled with [γ-^{33}P]ATP using T4 polynucleotide kinase prior to annealing to the template. The reaction conditions were otherwise as recommended by the manufacturer of the Sequenase system. Synthesis was carried out for 15–20 min at 37 °C and terminated by heating to 65°C for 15 min.

2.3. Oligodeoxyribonucleotide substrates with uracil in U·A pairs and U·G mismatches

Oligodeoxyribonucleotides of 19–22 nucleotides, each containing one central dUMP residue, were 5′ end-labeled using [γ-^{33}P]ATP and T4 polynucleotide kinase. The following oligos were used, (sequences are given 5′→3′): Amber oligos: 5, TGTCTCTTATUAGACCGTTTC; 9, GGCGGGCAAAUAGTCGTTGCT; 16, TGTGTCTGACUAGACACCCAT; 19, ATTGGGTCACUAGCAAATCGC; 23, CAATCA-

*Corresponding author. Fax: (47) (73) 598 705.

SSDI 0014-5793(95)00244-8

AATTUAGCCGATAGC; 24, CATGGCACTCUAGTCGCCTTC; 26, TGCCAACGATUAGATGGCGCT. M13 oligos: 90, CTGTGTGAAAUTGTTATCCGTC; 91, CTGTGTGAAATUGTTATCCGTC; 93, TGAAATTGUTATCCGCTCA; 94, TGAAATTGTUATCCGCTCA; 141, CATAAAGTGUAAAGCCTGG; and 291, GCGCCAGGGUGGTTTTTCT. For each of these oligos, complementary oligos containing either A (match) or G (mismatch) opposite to U were annealed to their uracil-containing counterparts at a molar ratio of 1:1 by heating for 2 min at 65°C and then slowly cooling to room temperature. Labeled double-stranded substrates were then mixed with equal amounts of 5 other unlabeled double-stranded uracil-containing oligos to a total of 5 pmol. The unlabelled oligos were the same in all experiments.

2.4. Excision of uracil by uracil-DNA glycosylase and cleavage of apyrimidinic sites

0.5 μg of long dUMP DNA substrate (mRS81) or 1 pmol of mixed double-stranded oligodeoxyribonucleotides, out of which one (0.17 pmol) was 5′ end-labeled with ^{33}P, was mixed with uracil-DNA glycosylase in buffer containing 60 mM Tris-HCl, pH 8.0, 1 mM DTT, 0.1 mg/ml BSA and 10 mM EDTA to a total volume of 20 μl. The reaction was allowed to proceed at 37°C, and subsequently stopped by addition of 100 μl 1.2 M piperidine, and transferred to 90°C to cleave the apyrimidinic sites [22]. Prior to electrophoresis, the samples were resuspended in deionized water and Stop solution from the Sequenase kit.

2.5. Electrophoresis and determination of relative frequencies of uracil removal

Standard dideoxy sequencing reactions were performed in parallel in the case of excision of U from M13 substrates. Samples of long DNA were electrophoresed on 7% polyacrylamide gels containing 7 M urea at 2000 V for 2–5 h. Short oligonucleotides were electrophoresed on 10% polyacrylamide gels containing 7 M urea at 200 V. The gels were subsequently fixed in 10% acetic acid, dried and either developed on β-max films (Amersham, UK) which were scanned using a LKB Ultroscan XL laser densitometer (Pharmacia), or the gels were analysed directly on a PhosphorImager (Molecular Dynamics). Control experiments demonstrated that the two systems gave essentially identical results under the present conditions.

2.6. Determination of ung⁻/ung⁺ ratio of mutation for various amber sites

We have used data from [5,6] to obtain the *ung⁻/ung⁺* ratio of mutation for 8 amber sites. Data from [5] was used directly. Similar data from [6] stated only the number of mutants but not frequencies. Data from [6] was normalized to be comparable with those in [5] as follows: the mutation frequencies from the amber sites having the highest number of mutants (amber site 6 for *ung⁺* cells and amber site 9 for *ung⁻* cells) were assumed to be similar for [5] and [6] and the relative frequencies at other sites determined from the number of mutants at each site. Data used in the results are the weighted average of those in [5] and [6].

3. Results

Removal of U in U·A matches in M13 DNA varied more than 15-fold when limiting amounts of Ung were used. Table 1 summarizes these results. The sequences are listed according to the rates determined such that the sequences from which U is removed fastest are at the top of the table. In Fig. 1 an example demonstrating the results from one experiment is

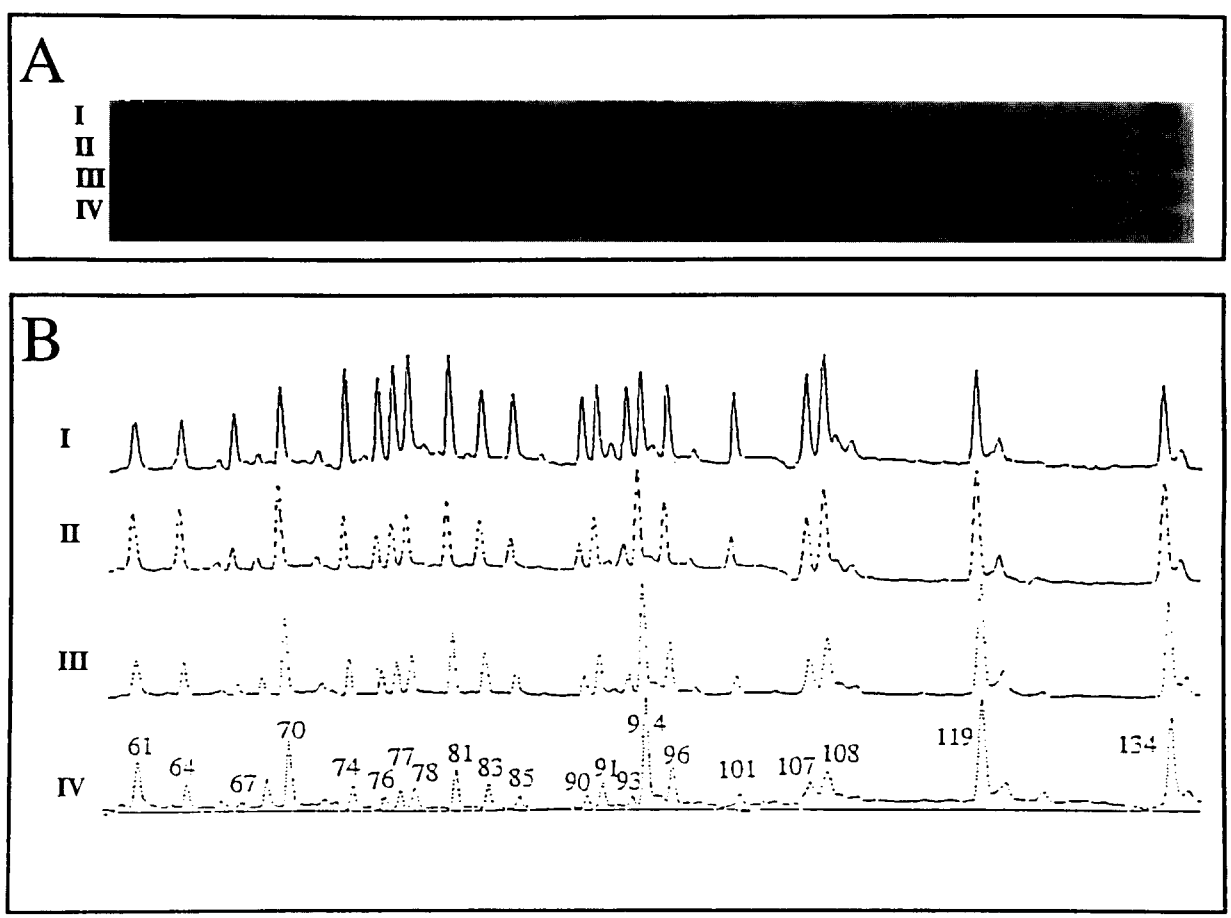

Fig. 1. Removal of uracil from different positions in double-stranded M13 DNA. (A) Autoradiogram of a polyacrylamide gel. (B) Laser scans of lanes in A. Incubation times were 60 min (I), 30 min (II), 15 min (III) and 5 min (IV). 0.0006 units of Ung was used for each reaction, except for I for which 0.006 units were used to ensure complete removal of U. Numbers in B refer to the distance of uracil relative to the M13 universal primer.

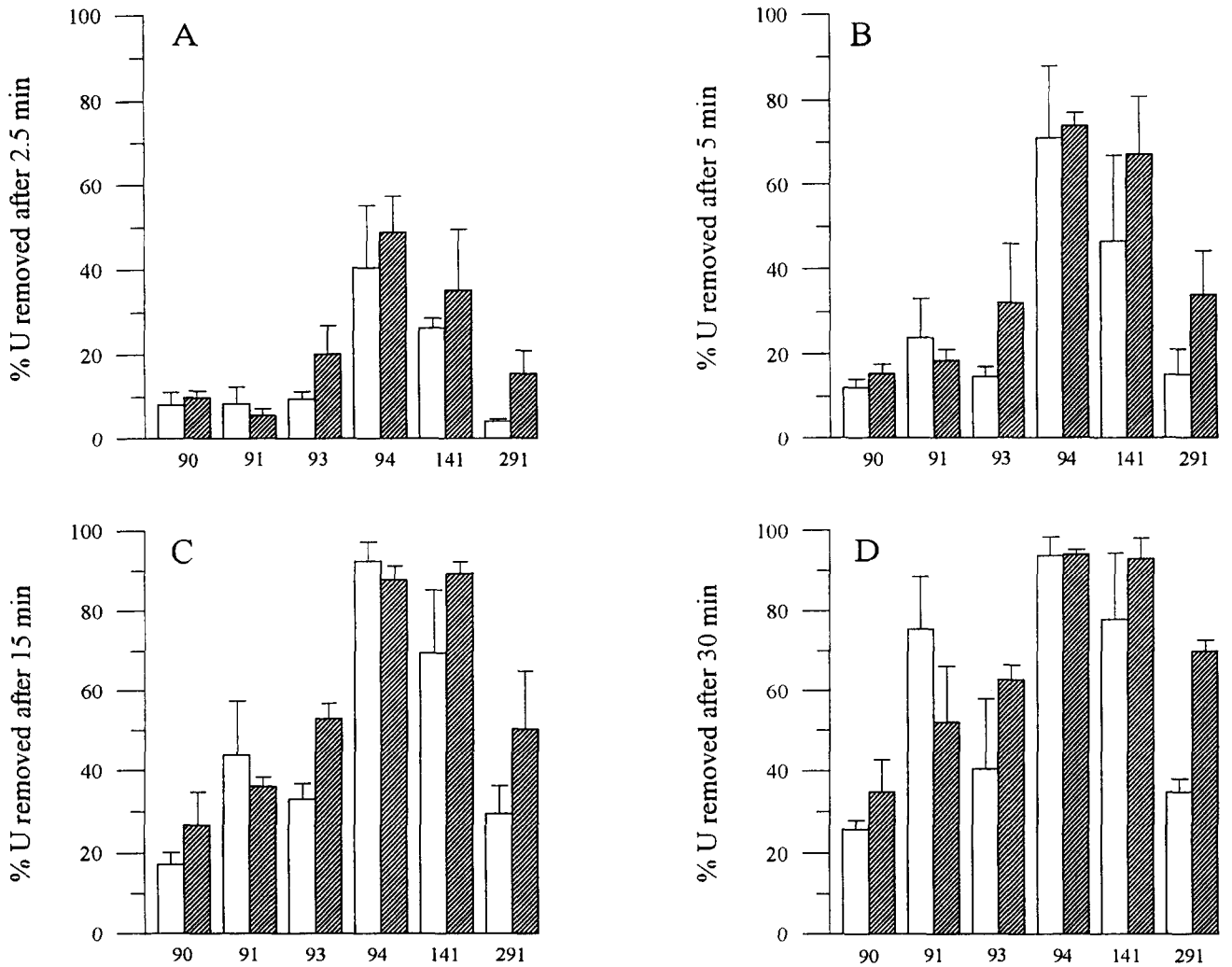

Fig. 2. Removal of uracil from double-stranded oligonucleotides. The numbers used refer to oligos corresponding to sequence positions of U relative to the primer, as in Fig. 1. The amount of Ung used was 0.0006 units. Open columns are data for U·A pairs, and hatched columns are data for U·G mispairs. The results are means of 5 independent experiments.

shown. Consensus sequences for good removal are 5′-(A/T)UA(A/T)-3′, whereas that for poor removal are 5′-(G/C)U-(T/G/C)-3′. In addition, sequences in which a T was located 3′ of U were mostly poorly repaired even when A was located 5′ of U. Therefore, GC richness is not the only determinant for poor removal of U. The rate of removal of U is, in general, intermediate in sequences corresponding to hybrids of sequences for good and poor removal. Although the majority of sequences behave according to the consensus, we have noticed a few exceptions from these rules. Thus, from uracil positions 81, 83 and 74 (Table 1), which all have G or C surrounding U, the rates of removal were as high as 57%, 40% and 33% of the maximal rates, respectively. It is possible that A- or T-rich sequences at some distance from U may increase the removal rates. Alternatively, other structural elements may have an effect. Thus, in these sequences alternating pyrimidines are found in stretches of 5–6 nucleotides and these might take on a Z-DNA configuration.

Experiments carried out with a set of double-stranded oligonucleotides corresponding to selected sequences from Table 1 gave qualitatively similar results (Fig. 2). In addition, and most importantly, the sequence specificities were similar, but not identical, for sequences containing U·A matches and U·G mismatches. The rates of removal were frequently slightly faster from mismatches than from matches, and in one case (oligo 291) the initial rate was several fold faster. However, faster removal of U from A·U matches was also seen. In these experiments the reaction mixtures also contained, in addition to the labelled double-stranded oligonucleotide a mixture of 5 different unlabelled double-stranded oligonucleotides containing a central dUMP residue. This was done to simulate a situation where several possible substrates compete for UDG, as is the case when the M13 substrates are used. Qualitatively, the results were similar when the reaction mixture contained no unlabelled substrate, but the differences in the rate of removal of U were somewhat smaller. This indicates that UDG tends to remove U from the 'easiest' sequences first and then from the more 'difficult' sequences (data not shown).

Previously, the mutational spectra in the *lacI* gene has been examined in *ung*$^-$ and *ung*$^+$ *E. coli* cells [5,6]. As expected, GC-to-AT transitions resulting in amber mutations were much more frequent in *ung*$^-$ than in *ung*$^+$ cells, except in 5-MeC hotspots where the mutations were found at equal frequencies. However, these mutations did not occur randomly. In *ung*$^-$ cells

Table 1
Sequence specificity of uracil excision

U position	Sequence 5' - 3'	% removal of U[b] ± SD
94	ATTGTUATCCG	100 ± 0
238	TGCATUAATGA	100 ± 0
70	GGTCAUAGCTG	65 ± 26
119	CAACAUACGAG	63 ± 8
179	TACATUAATTG	63 ± 18
279	CGTATUGGGCG	61 ± 8
206	CGCTTUCCAGT	59 ± 16
276	TTGCGUATTGG	58 ± 17
81	TTTCCUGTGTG	57 ± 21
188	TGCGTUGCGCT	57 ± 20
96	TGTTAUCCGCT	53 ± 5
174	TAACTUACATT	50 ± 13
134	AAGCAUAAAGT	50 ± 3
301	TTTCTUTTCAC	47 ± 23
233	CCAGCUGCATT	46 ± 9
272	CGGTTUGCGTA	45 ± 5
221	AAACCUGTCGT	45 ± 11
205	CCGCTUTCCAG	44 ± 12
78	CTGTTUCCTGT	43 ± 23
336	GCCCTUCACCG	42 ± 19
183	TTAATUGCGTT	42 ± 16
302	TTCTTUTCACC	41 ± 21
278	GCGTAUTGGGC	41 ± 13
83	TCCTGUGTGAA	40 ± 16
330	CTGATUGCCCT	40 ± 19
241	ATTAAUGAATC	37 ± 5
326	ACAGCUGATTG	35 ± 20
237	CTGCAUTAATG	35 ± 7
178	TTACAUTAATT	35 ± 14
108	ACAATUCCACA	34 ± 11
169	TGAGCUAACTT	34 ± 10
77	GCTGTUTCCTG	34 ± 18
74	ATAGCUGTTTC	33 ± 16
91	GAAATUGTTAT	32 ± 14
350	GGCCCUGAGAG	32 ± 16
64	AATCAUGGTCA	32 ± 15
303	TCTTTUCACCA	31 ± 16
344	CCGCCUGGCCC	29 ± 15
226	TGTCGUGCCAG	28 ± 4
245	ATGAAUCGGCC	26 ± 2
223	ACCTGUCGTGC	26 ± 6
298	GTTTTUCTTTT	26 ± 15
141	AAGTGUAAAGC	26 ± 3
211	TCCAGUCGGGA	25 ± 6
197	GTCACUGCCCG	25 ± 9
204	CCCGCUTTCCA	24 ± 5
296	TGGTTUTTCTT	24 ± 14
85	CTGTGUGAAAT	23 ± 8
271	GCGGTUTGCGT	23 ± 4
297	GGTTTUTCTTT	23 ± 14
182	ATTAAUTGCGT	22 ± 10
76	AGCTGUTTCCT	22 ± 13
310	ACCAGUGAGAC	21 ± 9
193	TGCGCUCACTG	21 ± 15
300	TTTTCUTTTCA	20 ± 13
61	CGTAAUCATGG	20 ± 10
101	TCCGCUCACAA	19 ± 9
295	GTGGTUTTTCT	19 ± 12
107	CACAAUTCCAC	19 ± 4
187	TTGCGUTGCGC	19 ± 11
93	AATTGUTATCC	18 ± 9
335	TGCCCUTCACC	18 ± 10
329	GCTGAUTGCCC	15 ± 8
270	GGCGGUTTGCG	13 ± 3
139	TAAAGUGTAAA	13 ± 10
90	TGAAAUTGTTA	12 ± 5
67	CATGGUCATAG	12 ± 6
173	CTAACUTACAT	11 ± 4
291	CAGGGUGGTTT	7 ± 5
293	GGTGGUTTTTC	6 ± 4

[a] uracil position is the distance from the M13 universal primer to the uracil residue. [b] % removal of uracil is estimated from 3-5 independent experiments. Incubation times was 5 min.

the spontaneous mutation frequencies at different amber sites varied up to 45-fold, which most likely reflects varying deamination rates [5,6]. In *ung*$^+$ cells the non-random mutation spectra may be due to differences in the rate of deamination, as well as to differences in the rate of repair in various sequences. To test the hypothesis that slow repair may contribute to the mutational spectra, we examined the rate of uracil removal in oligonucleotides corresponding to known amber sites, in which U replaced the corresponding C. A high ratio of mutations in *ung*$^-$ compared with *ung*$^+$ would be expected if repair at a particular site is good, whereas a low ratio would be expected when repair is slow. We observed a linear correlation between *ung*$^-$/*ung*$^+$ ratio and repair rate in all but one case (amber site 19) (Fig. 3). The correlation coefficient (r) was 0.83 when data for amber site 19 was excluded, and 0.41 when this site was included in the analysis. These results are consistent with the view that sequence-specific variation in DNA repair rates is an important factor to consider in mutagenesis, although clearly other not yet identified mechanisms may also contribute. No direct inverse correlation was observed between mutation frequencies in *ung*$^+$ cells and repair rates, probably due to the high variation in damage induction (deamination rates) at different amber sites.

4. Discussion

In the present paper we demonstrate that in double-stranded DNA, uracil-DNA glycosylase from *E. coli* removes uracil from different sequence contexts with widely varying efficiency. This is true both for U·A matches and U·G mismatches. Previously, the rate of removal of U by *E. coli* uracil-DNA glycosylase from single-stranded DNA was found to show a limited sequence dependency [23]. Uracil-DNA glycosylase from different sources removes U from single-stranded DNA with greater efficiency than from double-stranded. However, the preference for single-stranded substrate is only 3-fold or less [24,25], and this is not sufficient to explain the wide variation in removal rates between different sequences, even if U·G mismatches in AT-rich sequences may significantly destabilize the double-stranded structure. Verri et al. [26] reported that U was removed much faster from U·G mismatches than from U·A matches by HSV1, human and *E. coli* uracil-DNA glycosylases. However, in this study only one sequence context was studied and this may not allow a general conclusion concerning substrate preference. Our results are not contradicting those of Verri et al. [26], but they demonstrate that the substrate preference is more subtle than previously assumed. It can not be excluded, however, that the differences could be due to minor variations in assay conditions between the laboratories. Uracil residues in loops were recently shown to be very slowly removed [27]. We have analyzed possible loop formation for all oligos used and, under the present salt and temperature conditions, loop formation was found to be unlikely.

Although methylated C-residues constitute the most pronounced mutational hotspots in the *lacI* gene in *E. coli*, the mutational frequency varies at least 20-fold also between other cytosine residues in the *lacI* gene and other sequences in *E. coli* [28], as well as in yeast [29]. The rates of deamination and the efficiency of repair are the two most obvious parameters that could contribute to these differences. Our observed correlation between DNA repair rates and the relative mutation frequen-

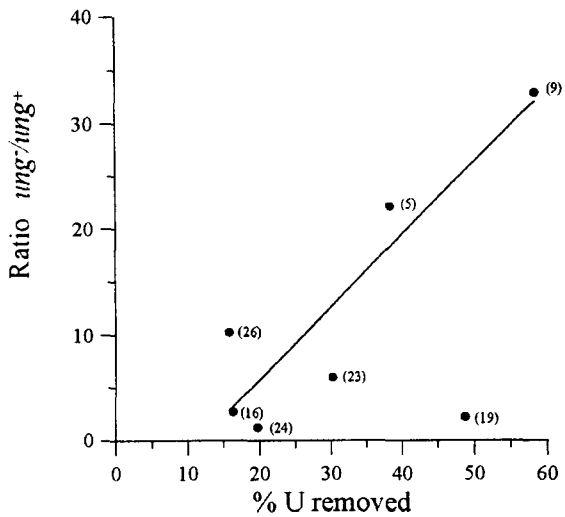

Fig. 3. Correlation between *ung⁻/ung⁺* ratio of mutations at amber sites in the *lacI* gene and rates of removal of U from U·G mispairs in oligonucleotides. All incubations were for 2.5 min and 0.0006 units of Ung were used. Data for amber site 19 is shown, but not included in the correlation curve.

cies in *ung⁻* and *ung⁺* cells (*ung⁻/ung⁺* ratio) indicates that sequence-dependent variation in repair may be a significant parameter in mutagenesis related to cytosine deamination. The sequence-dependent variation in repair may be most important for those sequences that are hotspots for deamination. Deamination frequencies are highest for C residues in AT-rich contexts, and, in general, uracil-DNA glycosylase works best in such sequences. However, our finding that sequences where a T is located immediately 3′ of U are poorly repaired indicate that some C-residues in AT-rich sequences may be more prone to mutations than others.

Slow DNA repair of pyrimidine dimers has been observed at *p53* mutation hotspots in skin cancer [4], and similar observations were reported for the human phosphoglycerate kinase gene [30]. Damage induction after UV exposure is apparently less sequence specific than deamination rates, so a more stringent inverse correlation between repair rates and mutation induction would be expected, and is in fact observed, after UV exposure [4,30]. A non-random pattern of repair has also been observed for AP-endonucleases [31] and redoxyendonucleases from several species [32]. These studies, as well as the present study, strongly indicate that sequence specificity of DNA repair is an important parameter to consider in mutagenesis.

Acknowledgements: This work was supported by the Norwegian Research Council and the Norwegian Cancer Society. SPC support from the Norwegian Technical College and SINTEF is also highly appreciated.

References

[1] Bohr, V.A., Smith, C.A., Okumoto, D.S. and Hanawalt, P.C. (1985) Cell 40, 359–369.
[2] Mellon, I., Bohr, V.A., Smith, C.A. and Hanawalt, P.C. (1986) Proc. Natl. Acad. Sci. USA 83, 8878–8882.
[3] Hanawalt, P.C. and Mellon, I. (1993) Curr. Biol. 3, 67–69.
[4] Tornaletti, S. and Pfeifer, G.P. (1994) Science 263, 1436–1438.
[5] Duncan, B.K. and Miller, J.H. (1980) Nature 287, 560–561.
[6] Fix, D.F. and Glickman, B.W. (1986) Mutat. Res. 175, 41–45.
[7] Fix, D.F. and Glickman, B.W. (1987) Mol. Gen. Genet. 209, 78–82.
[8] Coulondre, C. and Miller, J.H. (1977) J. Mol. Biol. 117, 577–606.
[9] Coulondre, C., Miller, J.H., Farabaugh, P.J. and Gilbert, W. (1978) Nature 274, 775–780.
[10] Atkinson, J. and Robin, M. (1994) Nucleic Acids Res. 22, 1327–1334.
[11] Lindahl, T. (1974) Proc. Natl. Acad. Sci. USA 71, 3649–3653.
[12] Tye, B.K., Nyman, P.O., Lehman, I.R., Hochhauser, S. and Weiss, B. (1977) Proc. Natl. Acad. Sci. USA 74, 154–157.
[13] Wist, E., Unhjem, O. and Krokan, H.E. (1978) Biochim. Biophys. Acta 520, 253–270.
[14] Lindahl, T. and Nyberg, B. (1974) Biochemistry 13, 3405–3410.
[15] Shen, J.-C., Rideout III, W.M. and Jones, P.A. (1992) Cell 71, 1073–1080.
[16] Wyszynski, M., Gabbara, S. and Bhagwat, A.S. (1994) Proc. Natl. Acad. Sci. USA 91, 1574–1578.
[17] Gabbara, S., Wyszynski, M. and Bhagwat, A.S. (1994) Mol. Gen. Genet. 243, 244–248.
[18] Nedderman, P. and Jiricny, J. (1994) Proc. Natl. Acad. Sci. USA 91, 1642–1646.
[19] Fishel, R., Lescoe, M.K., Rao, M.R.S., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M. and Kolodner, R. (1993) Cell 75, 1027–1038.
[20] Eftedal, I., Guddal, P.H., Slupphaug, G., Volden, G. and Krokan, H.E. (1993) Nucleic Acids Res. 21, 2095–2101.
[21] Schaaper, R.M., Danford, B.N. and Glickman, B.W. (1985) Gene 39, 181–189.
[22] Brash, D. E. (1988) in: DNA Repair, A Laboratory Manual of Research Procedures, vol 3 (Friedberg, E.C. and Hanawalt, P.C. eds.) pp. 327–347, Marcel Dekker, New York.
[23] Delort, A.-M., Duplaa, A.-M., Molko, D., Teoule, R., Leblanc, J.-P. and Laval, J. (1985) Nucleic Acids Res. 13, 319–335.
[24] Wittwer, C.U. and Krokan, H.E. (1985) Biochim. Biophys. Acta 832, 308–318.
[25] Domena, J.D. and Mosbaugh, D. (1985) Biochemistry 24, 7320–7328.
[26] Verri, A., Mazzarello, P., Spadari, S. and Focher, F. (1992) Biochem. J. 287, 1007–1010.
[27] Kumar, N.V. and Varshney, U. (1994) Nucleic Acids Res. 22, 3737–3741
[28] Duncan, B.K. and Weiss, B. (1982) J. Bacteriol. 151, 750–755.
[29] Impellizzeri, K.J., Anderson, B. and Burgers, P.M.J. (1991) J. Bacteriol. 173, 6807–6810.
[30] Gao, S., Drouin, R. and Holmquist, G.P. (1994) Science 263, 1438–1440.
[31] Haukanes, B.-I., Helland, D.E. and Kleppe, K. (1989) Nucleic Acids Res. 17, 1493–1509.
[32] Doetsch, P.W., Henner, W.D., Cunningham, R.P., Toney, J.H. and Helland, D.E. (1987) Mol. Cell Biol. 7, 26–32.