almost absent, even after 48 hours (Fig. 1A, bands 4 to 6).

Previously, we analyzed the initial damage distribution after high-dose UV irradiation at *p53* codons frequently mutated in skin malignancies (*16*). Mutations are produced when DNA lesions are encountered by DNA polymerase during DNA replication. Thus, the mutation frequency depends not only on initial damage frequency, but also on the repair rate for each individual lesion. To study the relation between repair and mutation hotspots, we examined the repair efficiency at *p53* codons 151, 177, 196, 245, 248, 278, 286, and 294—positions that are frequently mutated in human skin cancers (*3*). The CPD signal at these positions was sufficient to allow repair rates to be analyzed at 12 J/m$^2$. We found that, with the possible exception of position 151, which is repaired at a similar rate as most neighboring positions, all CPDs at mutation hotspots were repaired more slowly than those at surrounding positions on the same strand (Fig. 2 and Table 1). Densitometric quantitation (Table 1) revealed that CPDs on the nontranscribed strand at codons 177, 196, and 278 were only 0 to 33% repaired after 24 hours, whereas other CPDs that are not at mutation hotspots were 70 to 90% repaired at that time. The CPDs at codons 177 and 196 were still less than 40% repaired after 48 hours (Fig. 2A and Table 1). On the transcribed strand, between 23 and 56% of the CPDs remained at codons 245, 248, 286, and 294 after 24 hours, whereas adjacent positions on the same strand in the same exon were almost completely repaired (Fig. 2, B and C, and Table 1).

Mutations induced by UV are generally targeted to dipyrimidine sequences (*1, 5, 17*). CPDs represent 70 to 80% of all photoproducts, but mutations in skin cancer could also be caused by the (6-4) photoproducts. The (6-4) photoproducts are generally repaired faster than CPDs (*18*) and appear less likely to be the principal mutagenic lesion after UV irradiation (*1*). Some of the slow repair sites in the *p53* gene are not mutated in skin cancer. However, not all base changes necessarily result in a mutant protein and there may be a selection pressure that affects the mutation spectrum. Furthermore, CPDs containing T are much less mutagenic than those containing C (*1, 17*).

We have shown that excision repair of CPDs is sequence-specific in human cells. The molecular basis of excision repair slow spots is not known. These slow spots of repair do not share a common primary sequence. Excision repair in human cells is initiated on chromatin but not on naked DNA. Gao and co-workers (*13*) have shown that some slowly repaired sites are within transcription factor binding sites. Transcription factors are not expected to bind to *p53* exons, as these sequences are more likely to be associated with nucleosomes. Chromatin folding may impede repair efficiency at certain sequence positions while allowing higher repair rates at others.

## REFERENCES AND NOTES


1. D. E. Brash, *Photochem. Photobiol.* **48**, 59 (1988).
2. F. Urbach, *ibid.* **50**, 507 (1989).
3. /tk;3D. E. Brash *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10124 (1991); W. E. Pierceall, T. Mukhopadhyay, L. H. Goldberg, H. N. Ananthaswamy, *Mol. Carcinog.* **4**, 445 (1991); P. Rady, F. Scinicariello, R. F. Wagner Jr., S. K. Tyring, *Cancer Res.* **52**, 3804 (1992); J.-P. Molès *et al.*, *Oncogene* **8**, 583 (1993); A. Ziegler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4216 (1993).
4. C. Caron de Fromentel and T. Soussi, *Genes Chromosomes Cancer* **4**, 1 (1992).
5. J. H. Miller, *J. Mol. Biol.* **182**, 45 (1985); A. Bredberg, K. H. Kraemer, M. M. Seidman, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8273 (1986); E. A. Drobetsky, A. J. Grosovsky, B. W. Glickman, *ibid.* **84**, 9103 (1987); W. G. McGregor, R. H. Chen, L. Lukash, V. M. Maher, J. J. McCormick, *Mol. Cell. Biol.* **11**, 1927 (1991).
6. V. A. Bohr, C. A. Smith, D. S. Okumoto, P. C. Hanawalt, *Cell* **40**, 359 (1985); D. C. Thomas, A. G. Morton, V. A. Bohr, A. Sancar, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3723 (1988).
7. I. Mellon, G. Spivak, P. C. Hanawalt, *Cell* **51**, 241 (1987); I. Mellon and P. C. Hanawalt, *Nature* **342**, 95 (1989).
8. C. P. Selby and A. Sancar, *Science* **260**, 53 (1993).
9. P. R. Mueller and B. Wold, *ibid.* **246**, 780 (1989); G. P. Pfeifer, S. D. Steigerwald, P. R. Mueller, B. Wold, A. D. Riggs, *ibid.*, p. 810.
10. G. P. Pfeifer, J. Singer-Sam, A. D. Riggs, *Methods Enzymol.* **225**, 567 (1993).
11. G. P. Pfeifer, R. Drouin, A. D. Riggs, G. P. Holmquist, *Mol. Cell. Biol.* **12**, 1798 (1992).
12. S. Tornaletti and G. P. Pfeifer, unpublished observations.
13. S. Gao, R. Drouin, G. P. Holmquist, *Science* **263**, 1438 (1994).
14. After sequential treatment with T4 endonuclease V and photolyase (*11*), the DNA fragments [1.5 μg, quantified by DAPI (4′,6-diamidino-2-phenylindole dihydrochloride) fluorimetry] were subjected to LM-PCR. Synthetic primers specific for the human *p53* gene were as in (*16*). Primer extension, ligation of the linker, PCR (21 cycles), and gel electrophoresis were carried out as in (*10, 16*).
15. M. K. Evans, B. G. Taffe, C. C. Harris, V. A. Bohr, *Cancer Res.* **53**, 5377 (1993).
16. S. Tornaletti, D. Rozek, G. P. Pfeifer, *Oncogene* **8**, 2051 (1993).
17. E. Sage, *Photochem. Photobiol.* **57**, 163 (1993).
18. D. L. Mitchell, *ibid.* **48**, 51 (1988).
19. A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
20. Quantitation was done by video densitometry or phosphoimaging. Most reported mutations (*3*) are at the 3′ pyrimidine of the listed dimer position. Mutations within a run of pyrimidines could, in theory, be caused by two different dimers. However, in most cases mutations were found at the 3′ base of a dipyrimidine sequence [H. C. Hsia, J. S. Lebkowski, P.-M. Leong, M. P. Calos, J. H. Miller, *J. Mol. Biol.* **205**, 103 (1989); J. D. Armstrong and B. A. Kunz, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9005 (1990); (*17*)]. This observation was taken into account when listing individual dimers in Table 1.


21. We thank S. Bates for assistance with cell culture, A. Sancar for photolyase, S. Lloyd for T4 endonuclease V, and G. Holmquist for review. Supported by the National Institute of Environmental Health Sciences (NIH grant ES06070).

16 December 1993; accepted 3 January 1994


# DNA Repair Rates Mapped Along the Human *PGK1* Gene at Nucleotide Resolution


Shuwei Gao, Régen Drouin, Gerald P. Holmquist*



**The repair of cyclobutane pyrimidine dimers (CPDs), DNA lesions induced by ultraviolet light, was studied at nucleotide resolution. Human fibroblasts were irradiated with ultraviolet light and allowed to repair. The DNA was enzymatically cleaved at the CPDs, and the induced breaks along the promoter and exon 1 of the *PGK1* gene were mapped by ligation-mediated polymerase chain reaction. Repair rates within the nontranscribed strand varied as much as 15-fold, depending on nucleotide position. Preferential repair of the transcribed strand began just downstream of the transcription start site but was most pronounced beginning at nucleotide +140 in exon 1. The promoter contained two slowly repaired regions that coincided with two transcription factor binding sites.**


In both prokaryotes and eukaryotes, DNA is repaired at nonuniform rates. Nucleotide excision repair, particularly of ultraviolet (UV)-induced DNA damage, is rapid in domains near active genes (*1*) and even more rapid along the transcribed strand of active genes (*2*). We have been studying the incidence and repair of CPDs, the most frequent class of UV-induced DNA damage. In previous work, we mapped CPD lesions in the 5′ region of the human *PGK1* (phosphoglycerate kinase 1) gene. Specifically, we irradiated fibroblasts or naked DNA with a lethal dose (1000 J/m$^2$) of UV light (254 nm) and determined the CPD frequency at


Beckman Research Institute of the City of Hope, Department of Biology, Duarte, CA 91010, USA.

*To whom correspondence should be addressed.

each base position along the promoter and first exon of the gene. These studies revealed that transcription factor binding drastically alters CPD frequency (*3*). Here, we irradiated fibroblasts with a sublethal dose of UV light, and by means of a ligation-mediated polymerase chain reaction (LMPCR) assay (*4*), we determined the rate of CPD repair at each nucleotide position in the same region of the *PGK1* gene.

Confluent human foreskin fibroblasts were irradiated (*3*, *5*) with 20 J/m$^2$ of 254-nm UV light at 24°C and allowed to repair for up to 64 hours before analysis. The nuclear DNA was extracted and the CPDs were cleaved with an excess of T4 endonuclease V (*3*, *6*). Photolyase photoreactivation was used to remove a ligation-inhibiting 5′-pyrimidine overhang (*3*). Analysis of the DNA on alkaline 0.8% agarose gels (*3*) revealed that, in the absence of repair, the mass distribution of cellular DNA was maximal at 9 kb; this corresponds to one CPD per 4.5 kb. Almost 50% of the T4 endonuclease V–sensitive sites were repaired in 4 to 8 hours, and 65% were repaired by 16 hours, as reported (*5*).

From LMPCR analysis (Fig. 1), we were able to estimate a repair rate for a sequence position (represented as a band on the membrane transfer) if the intensity of the signal, as quantified with a phosphoimager, increased linearly with percent of DNA from irradiated cells in the standards representing 0, 25, 50, 75, and 100% irradiated cells. The time required to reduce the T4 endonuclease V–dependent signal to that of a 50% mixture of irradiated and unirradiated cells was determined for each measurable dipyrimidine position in both strands (Fig. 2) either visually from autoradiograms or from phosphoimager scans. On average, dipyrimidines in the nontranscribed strand required 5.5 hours to reach a 50% repair level. The repair rates varied 15-fold across the DNA region studied. This variation is likely due to sequence context, as there was no apparent spatial pattern (Fig. 2).

In *Escherichia coli*, transcribed and nontranscribed DNA strands are repaired at different rates. This "strand-specific repair" (*2*) occurs because a transcription-repair coupling factor (TRCF) (*7*) recognizes an RNA polymerase complex stalled at a DNA lesion, displaces the polymerase, and forms a high-affinity site for binding of repair enzymes (*8*). The likely human homolog of TRCF, ERCC-6, is deficient in patients with Cockayne's syndrome B, who are deficient in strand-specific repair (*9*). We found evidence of strand-specific repair just downstream from the transcription start site of the *PGK1* gene. However, the effect became most pronounced (50% repair of the transcribed strand achieved in ≤2 hours) beginning at approximately position +140 in exon 1 (Fig. 2).

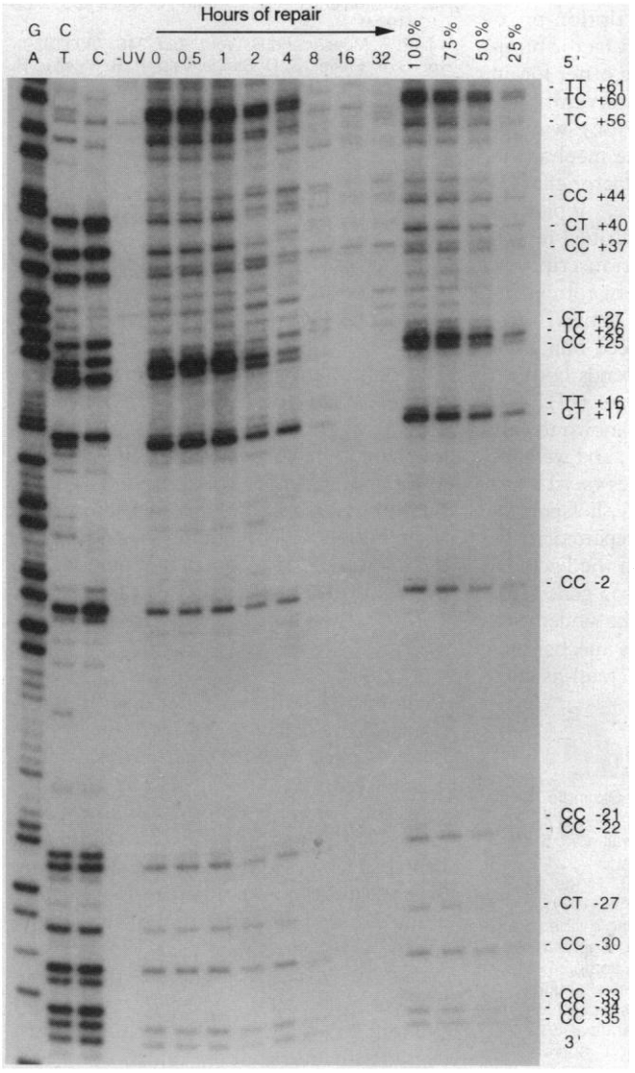


**Fig. 1.** Repair of CPDs at individual nucleotide positions in the transcribed strand of the human *PGK1* gene. Primer set E (*10*) was used to display sequences around the transcription initiation site. Sequence standards (left three lanes) contain 0.8 μg of HeLa cell DNA chemically cleaved (*17*) to 400 bp. Experimental lanes contain 1.0 μg DNA from UV-irradiated cells allowed to repair for the indicated times. Repair standards (rightmost four lanes) were 1-μg mixtures of DNA from irradiated and unirradiated (100 to 0%) cells. The autoradiogram displays the frequency of ligatable breaks generated by UV irradiation with subsequent enzymatic cleavage of CPDs.

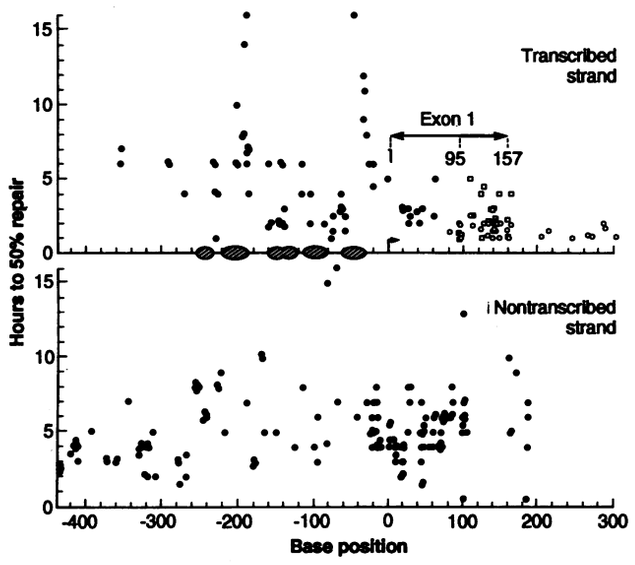


**Fig. 2.** Rates of CPD repair along the promoter and exon 1 of the human *PGK1* gene. Twelve different primer sets (*3*, *10*, *15*) were used, one set for each LMPCR reaction. Many positions were sampled several times because data ranges from different primer sets overlap. Open symbols (□ and ○) are data from primer sets K and N, respectively. Initiation sites for transcription (+1) and translation (+95) were taken from sequence maps in (*18*) as modified in (*10*). Transcription factor binding sites (*10*) are indicated by striped ellipses.

Repair of promoter sequences in the transcribed strand showed a striking pattern (Fig. 2). Two slowly repaired regions (50% repair attained in 10 to 16 hours), at nucleotides −34 to −47 and at −187 to −193, occurred on the transcription-proximal side of two transcription factor binding sites, one for SP1 and the other for an unknown factor (*10*). An intervening region (nucleotides −59 to −86) was repaired at a very fast rate. The mechanism underlying the transcription factor effect is unknown. In this instance, repair is strand-specific for a nontranscribed region and is probably coupled to transcription; as such, this effect confuses historical repair nomenclature for active genes.

The frequency of UV-induced mutations at any nucleotide position depends both on damage frequency and repair rate (*11*). The frequency of CPDs was already known to vary considerably along genes (*3, 6*), and we have shown here that CPD repair rates vary 15-fold along the *PGK1* gene. Recently, hotspots for UV-induced mutations were reported to be positions of slow CPD repair in the bacterial *lacI* gene (*12*) and the human *p53* gene (*13*). Mutation frequency can now be understood by its component mutagenesis mechanisms because the final link (repair rate) is now amenable to mapping.

## REFERENCES AND NOTES


1. V. A. Bohr, C. A. Smith, D. S. Okumoto, P. C. Hanawalt, *Cell* **40**, 359 (1985).
2. I. Mellon, G. Spivak, P. C. Hanawalt, *ibid.* **51**, 241 (1987).
3. G. P. Pfeifer, R. Drouin, A. D. Riggs, G. P. Holmquist, *Mol. Cell. Biol.* **12**, 1798 (1991).
4. Ligation-mediated PCR is a single-sided PCR method (*14, 15*). A set of three oligonucleotide primers are synthesized that are complementary to three overlapping sites within a sequenced region of DNA. Genomic DNA is nicked to generate broken molecules with terminal 5′-phosphate groups. After denaturation, primer 1 is annealed and extended with Sequenase to the previously induced nicks, thereby generating blunt ligatable ends. Ligation of a double-stranded oligonucleotide linker onto the 5′ end of each DNA molecule provides a common sequence at all previously broken ends. Primer 2, whose sequence partially overlaps that of primer 1, is then used in conjunction with the common linker sequence in a PCR reaction. After 19 to 21 PCR cycles, the amplified DNA fragments are separated on a sequencing gel, electroblotted onto a nylon membrane, and hybridized to a single-stranded, gene-specific $^{32}$P-labeled probe [synthesized by PCR with primer 3 (*16*)]. On the final autoradiogram, each band represents a nucleotide position where a break was induced, and the intensity of the band reflects the number of DNA molecules with ligatable ends terminating at that position.
5. G. J. Kantor and R. B. Setlow, *Cancer Res.* **41**, 819 (1981).
6. L. K. Gordon and W. A. Haseltine, *J. Biol. Chem.* **255**, 12047 (1980).
7. C. P. Selby, E. M. Witkin, A. Sancar, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11574 (1992).
8. C. P. Selby and A. Sancar, *Science* **260**, 53 (1993).
9. C. Troelstra *et al.*, *Cell* **71**, 939 (1992).
10. G. P. Pfeifer and A. D. Riggs, *Genes Dev.* **5**, 1102 (1991).
11. A. A. van Zeeland *et al.*, in *Mutation and the Environment. Part A*, M. L. Mendelsohn and R. J. Albertini, Eds. (Wiley-Liss, New York, 1990), pp. 249–254.
12. S. Kunala and D. E. Brash, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 11031 (1992).
13. S. Tornaletti and G. P. Pfeifer, *Science* **263**, 1436 (1994).
14. P. R. Mueller and B. Wold, *ibid.* **246**, 780 (1989).
15. G. P. Pfeifer, S. D. Steigerwald, P. R. Mueller, B. Wold, A. D. Riggs, *ibid.*, p. 810.
16. G. P. Pfeifer, J. Singer-Sam, A. D. Riggs, *Methods Enzymol.* **225**, 567 (1993).
17. A. M. Maxam and W. Gilbert, *ibid.* **65**, 499 (1980).
18. J. Singer-Sam *et al.*, *Gene* **32**, 409 (1984).


19. We thank S. Bates for culturing and irradiating fibroblasts, A. Sancar for photolyase, S. Lloyd for T4 endonuclease V, and G. Pfeifer for review. Supported by a Centennial Fellowship of the Canadian Medical Research Council (to R.D.) and NIH Grant CA54773 (to G.P.H.).

16 December 1993; accepted 3 January 1994


# Coding of Hemolysins Within the Ribosomal RNA Repeat on a Plasmid in *Entamoeba histolytica*


Åslög Jansson, Frances Gillin, Ulrika Kagardt, Per Hagblom*



**The pathogenesis of amoebic dysentery is a result of cytolysis of the colonic mucosa by the parasitic protozoan *Entamoeba histolytica*. The cytolysis results in extensive local ulceration and allows the amoeba to penetrate and metastasize to distant sites. Factors involved in this process were defined with three clones that express hemolytic activities in *Escherichia coli*. These potential amoebic virulence determinants were also toxic to human colonic epithelial cells, the primary cellular targets in amoebal invasion of the large intestine. The coding sequences for the hemolysins were close to each other on a 2.6-kilobase segment of a 25-kilobase extrachromosomal DNA element. The structural genes for the hemolysins were within inverted repeats that encode ribosomal RNAs.**


Amoebic dysentery results from destruction of the colonic epithelium by the human pathogen *Entamoeba histolytica*. The parasite resides primarily in the large intestine, where it colonizes mucosal surfaces (*1*). Penetration of the colon wall may lead to spread of the amoebae to other organs, most commonly causing hepatic abscesses, and dissemination of the disease (*2*). To invade and cause disease, the amoebae must first penetrate the mucus blanket and adhere to the intestinal epithelium (*3*). Second, the pathogen must degrade the extracellular matrix to release anchored cells. Finally, target cells are lysed by a contact-dependent process and the parasites phagocytose cellular debris. Virulent amoebae are also erythrophagocytic and hemolytic (*4*). Amoebic virulence determinants that are involved in lysis of host cells have also been described and these include a pore-forming protein, amoebaporin (*5*), and hemolytic proteins (*6*). Although the gene for amoebaporin has been identified (*7*), the molecular basis for cytolysis by *E. histolytica* has not been defined. To isolate genes that encode lytic proteins and thereby identify factors required for tissue destruction, we assayed for the functional expression of amoebal hemolysins in *Escherichia coli*.

In the isolation of the hemolysin genes, we took advantage of the predicted biological activity of the virulence factor to detect hemolysins produced from a phage expression gene bank in *E. coli*. The gene bank was constructed by partial cleavage of genomic DNA with Eco RI* to produce randomly digested products for ligation into the λ ZAP/pBluescript vector (*8*). Phage lysis of the bacterial host releases the amoebic proteins and makes them available for detection. The hemolysis assays were done directly on bacterial agar plates containing red blood cells (RBCs). With this approach, the genes can be isolated and the virulence factors characterized in *E. coli* without purification of the proteins from the parasite. Small plaques with relatively large clearing zones were selected at two different pH values that correspond to optima of known *E. histolytica* virulence determinants (*9*). We isolated 12 hemolytic clones that were grouped in five classes by restriction enzyme analysis and size of the inserts. The clones isolated at pH 7.5, represented by pHLY5, were all identical, whereas four different classes were found at pH 6.7. A more detailed restriction enzyme analysis of the inserts showed that representative clones from three of the classes overlapped (Fig. 1). The largest insert (pHLY5) was 2.6 kb in size and contained three Eco RI fragments. Two other clones, pHLY1 and


Å. Jansson, U. Kagardt, P. Hagblom, Department of Microbiology, Biomedical Center, Box 581, Uppsala University, S-751 23 Uppsala, Sweden.
F. Gillin, Department of Pathology, University of California, Medical Center 8416, San Diego, CA 92103, USA.

*To whom correspondence should be addressed.