DNA Loop Repair by Human Cell Extracts

Asad Umar, Jayne C. Boyer, Thomas A. Kunkel*

An activity in human cell extracts is described that repairs DNA with loops of five or more unpaired bases. Repair is strand-specific and is directed by a nick located 5′ or 3′ to the loop. This repair is observed in a colorectal cancer cell line that is devoid of a wild-type *hMLH1* gene and is deficient in repair of mismatches. However, a cell line with deletions in both *hMSH2* alleles is deficient in repair of both loops and mismatches. Defects in loop repair may be relevant to the repetitive-sequence instability observed in cancers and other hereditary diseases.

Human homologs of the *Escherichia coli* mismatch repair genes *mutS* and *mutL* have been found at loci associated with hereditary nonpolyposis colorectal carcinoma (HNPCC), and mutations in these genes have been found in HNPCC kindreds (*1*, *2*). Mutations in either of these genes are believed to inactivate mismatch repair to yield a mutator phenotype that could alter critical genes and lead to tumor formation. In support of this hypothesis, several endometrial and colorectal carcinoma cell lines have been shown to be defective in mismatch repair (*3*, *4*) and three of these lines contain mutations in either the *mutS* homolog *hMSH2* or the *mutL* homolog *hMLH1* (*1*, *2*, *4*). Furthermore, these cell lines, as well as tumor cells from patients with cancer of the colon and several other tissues, exhibit genome-wide instability in simple repetitive DNA sequence elements (microsatellites) (*5*–*7*). This type of instability is also observed in yeast cells with mutations in homologs of the *E. coli mutS* or *mutL* mismatch repair genes (*8*).

Methyl-directed mismatch repair in *E. coli* does not correct DNA heteroduplexes with more than four consecutive unpaired bases (*9*). However, human DNA contains numerous microsatellites comprised of both long repeating units and many consecutive repeats (*10*). Because the number of unpaired bases resulting from strand slippage during replication (*11*) could exceed four if the misalignment involved more than a single repeat or a single repeat motif of more than four bases, we investigated whether human cell extracts can repair DNA containing several unpaired nucleotides (loops).

We constructed M13mp2 DNA substrates (Fig. 1) containing loops of increasing size within the coding sequence of the *lacZ* α complementation gene. The extra nucleotides were either in the (−) strand, which also contained a nick located several hundred base pairs away, or in the covalently-closed (+) strand. The nick was intended to be a signal to direct repair to one strand, as a nick is known to be a signal for strand-specific repair of single-base mispairs and one- or two-base loops in human cell extracts (*12*, *13*). The two strands encode different M13 plaque colors—blue or colorless. If an unrepaired heteroduplex is introduced into an *E. coli* strain that is unable to repair it (for example, a *mutS* strain), plaques show a mixed phenotype on indicator plates because both strands are expressed (*13*). Repair of the substrate in a cell extract reduces the percentage of mixed plaques and increases the percentage of pure color plaques (*14*). If the nick directs repair to the (−) strand, the (+) strand phenotype increases and the (−) strand phenotype decreases.

We first prepared extracts of two human cell lines, HeLa cells and TK6 lymphoblasts, and assayed their ability to repair the M13mp2 substrates. Both extracts efficiently repaired heteroduplexes containing mismatches or 1 to 16 extra nucleotides (Table 1 and Fig. 2). Repair was preferentially directed to the (−) strand containing the nick, since regardless of the location of the loop in the (+) or (−) strand, the reduction in percentage of mixed plaques was accompanied by a selective decrease in the nicked (−) strand plaque phenotype with a concomitant increase in the covalently closed (+) strand plaque phenotype. That the nick serves as a signal for repair of DNA containing a loop is indicated by the fact that strand-specific repair in an extract of TK6 cells was greatly diminished when a covalently closed circular substrate was used (Table 1). Finally, repair of substrates containing 5, 8, or 16 unpaired bases occurred when the nick was located on either the 3′ or 5′ side of the loop (Fig. 2 and Table 1). This observation suggests that strand-specific repair of heteroduplexes containing loops has bidirectional capability, just as does repair of DNA containing mismatches or one or two extra bases (*2*, *3*, *15*, *16*).

To explore the genetic requirements for repair of heteroduplexes containing ≥5 ex-

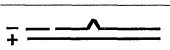
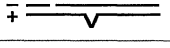
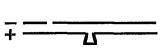
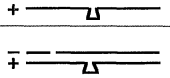
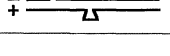
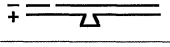
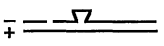
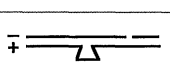
Heteroduplex	Phenotype +/-	Nick relative to unpaired bases	Sequence of heteroduplexes
G•G 88	W/LB		- 3′-G A G T C T -5′ + 5′-C T G A G A -3′
Ω1(-) 88-90	B/W		C - 3′-A C C C T T-5′ + 5′-T G G G A A-3′
Ω2(+) 90-91	B/W		- 3′-A C C T T T-5′ + 5′-T G G A A A-3′ G A
Ω3(+) 155-157	B/LB		- 3′-A T T G C T T -5′ + 5′-T A A C G A A -3′ T A G
Ω4(+) 68-71	B/W		- 3′-G C A A A T -5′ + 5′-C G T T T A -3′ C G T T
Ω5(+) 127-131	B/W		- 3′-G T C G G G-5′ + 5′-C A G C C C-3′ C A C A T
Ω5(-) 127-131	W/B		G T G T A - 3′-G T C G G G-5′ + 5′-C A G C C C-3′
Ω8(+) 60-84	B/W		- 3′-C A C A C A-5′ + 5′-G T G T G T-3′ G T G T G T G T
Ω16(-) 60-76	B/W		C A C A C A C A C A C A C A - 3′-C A C A C A-5′ + 5′-G T G T G T-3′

Fig. 1. Heteroduplex DNA substrates. Substrates were prepared as in (*13*). Heteroduplexes with unpaired bases are depicted by the symbol Ω followed by the number of unpaired bases. The (+) or (−) indicates whether the extra bases are present in the (+) or (−) strand and the numbers indicate the position of the mutation, where position +1 is the first transcribed base of the *lacZ* α-complementation gene in M13mp2. The phosphodiester bond interruption in the (−) strand is shown either to the left (at position −264, 3′-nicked substrate) or to the right (at position +276, 5′-nicked substrate) of the mispair or unpaired base. The phenotypes of the (+) and (−) strands are shown as W for white (colorless), LB for light blue, or B for blue.

Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC 27709, USA.

*To whom correspondence should be addressed.

tra bases, we used extracts of human cell lines that exhibit microsatellite instability and are defective in repair of DNA containing mismatches or an extra base (*4*). Two of these lines contain known genetic defects. In one sporadic colorectal carcinoma cell line, HCT116, the *hMLH1* gene contains a C→A transversion that generates a termination codon, thus yielding a truncated polypeptide and no detectable wild-type gene product (*2*). An extract of HCT116 cells was defective in repair of DNA containing a single mispaired or unpaired base (*3*, *4*) and of DNA containing 2, 3, or 4 unpaired bases (Fig. 2). However, in parallel reactions, the same extract repaired heteroduplexes containing 5, 8, or 16 unpaired bases. This observation suggests that repair of heteroduplexes containing loops of five or more bases is distinct from repair of smaller mismatches, in that the former does not require the wild-type hMLH1 protein. The truncated hMLH1 protein in HCT116 cells may be sufficient for repair of substrates containing loops, or such repair may be completely independent of hMLH1 protein.

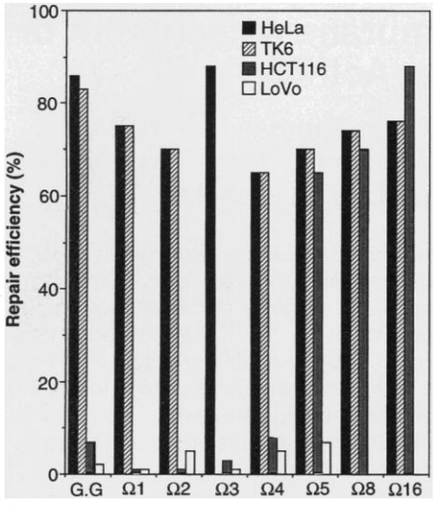


Fig. 2. Repair efficiencies in extracts of four human cell lines. Reactions were incubated for 30 min. Results are expressed as total repair efficiency and are based on counts of several hundred to several thousand plaques per variable (see Table 1). Data reproducibility is indicated by the fact that measurements with multiple substrates gave consistent repair or lack thereof, and by the fact that multiple determinations with the same substrate gave similar values. For example, the mean and standard deviation for repair in a HeLa cell extract was 72 ± 11% (four determinations) for a G·G mispair at position 88, and 68 ± 8% (five determinations) for repair of a substrate containing a five-base loop. Repair values between 0 and 10% represent experimental fluctuation in this assay. Thus, we cannot conclude that extracts of LoVo cells are completely devoid of repair. When no bar is shown (for example, Ω3 with TK6 cell extracts), this variable was not examined. For Ω5 repair, the loop was in the (+) strand.

Extracts of HCT116 cells efficiently repaired a substrate containing a five-base loop in one sequence context but did not repair a substrate with the four-base loop in a different surrounding sequence (Fig. 2). To distinguish whether this difference resulted from the one additional base in the loop or from a difference in flanking nucleotide sequences, we examined repair of two substrates that had the same sequence context but differed by one extra base in the loop. Both substrates were efficiently repaired (Fig. 3). Thus, although a HCT116 cell extract did not repair heteroduplexes containing mismatches or loops of 1, 2, or 3 bases (Fig. 2) (*3*, *4*), it did repair heteroduplexes containing loops of 5, 8, or 16 bases and one of two heteroduplexes containing a 4-base loop. The efficiency of repair of looped heteroduplexes in HCT116 extracts may therefore depend not only on the number of nucleotides in the loop but also on the surrounding sequence.

In a second sporadic colorectal carcinoma cell line, LoVo, both alleles of the *hMSH2* gene contain deletions (*4*). Like HCT116 extracts, an extract of LoVo cells was defective in repair of DNA containing a single mispaired or unpaired base (Fig. 2) (*4*, *17*). However, in contrast to HCT116 extracts, the LoVo extract was also defective in repair of a heteroduplex containing two to five extra bases (Fig. 2). This result suggests that repair of looped heteroduplexes may require the hMSH2 protein, although it is possible that LoVo cells con-

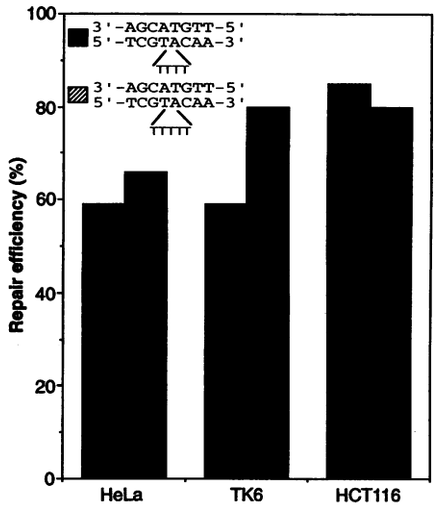


Fig. 3. Repair efficiencies of heteroduplex substrates containing four or five unpaired bases in the identical sequence context. Reactions were performed as in Fig. 2. Substrates contained either four or five unpaired bases between nucleotide positions 70 and 74 of the *lacZα* gene, within the same surrounding sequence. The phenotype of the (+) strand is colorless and the phenotype of the (−) strand is blue. The nick is present in the (−) strand at position −264.

Table 1. Repair by human cell extracts of heteroduplexes containing unpaired bases. Reactions (25 μl) contained 30 mM Hepes (pH 7.8); 7 mM $MgCl_2$; 4 mM adenosine triphosphate (ATP); 200 μM each of cytidine triphosphate (CTP), guanosine triphosphate (GTP), and uridine triphosphate (UTP); 100 μM each deoxy-ATP, deoxy-GTP, deoxythymidine triphosphate (dTTP), and deoxy-CTP; 40 mM creatine phosphate; 100 μg/ml creatine phosphokinase; 15 mM sodium phosphate (pH 7.5); 1 fmol of heteroduplex DNA; and 50 μg of extract protein. After incubation at 37°C for 15 min, the reactions were processed for repair (*4*). Total repair efficiency was calculated (*13*) as 100 × (1 minus the ratio of percentages of mixed bursts from extract-treated and untreated samples). For calculation of (−) strand-specific repair, the fraction of (+) strand phenotype plaques observed without extract treatment was multiplied by (1 minus the total repair efficiency). This value was then subtracted from the observed percentage of plaques with the pure (+) strand phenotype after extract treatment. Repair of the (+) strand was then obtained by subtracting (−) strand repair efficiency from total repair efficiency.

Cell extract	Loop		Location of nick	Total plaques	Plaque phenotype (%)			Repair efficiency (%)		
	Size	Strand			Mixed	(−)	(+)	Total	(+)	(−)
None	5	(+)	3′	524	45	43	12	0		
HeLa	5	(+)	3′	725	22	27	51	51	6	45
TK6	5	(+)	3′	350	15	17	68	67	3	64
None	5	(−)	3′	692	33	31	36	0		
HeLa	5	(−)	3′	496	18	19	63	46	2	44
TK6	5	(−)	3′	470	10	16	74	70	7	63
None	8	(+)	5′	772	21	71	8	0		
HeLa	8	(+)	5′	442	10	52	38	52	18	34
TK6	8	(+)	5′	866	6	24	70	71	3	68
None	16	(−)	5′	942	19	61	20	0		
HeLa	16	(−)	5′	432	8	45	47	56	18	38
TK6	16	(−)	5′	577	5	29	66	74	5	61
None	5	(+)	3′	550	37	50	13	0		
TK6	5	(+)	3′	771	6	11	83	84	3	81
None	5	(+)	None*	736	32	48	20	0		
TK6	5	(+)	None*	768	27	42	31	16	2	14

*This substrate was covalently closed circular DNA, prepared and gel-purified as in (*13*).

tain other mutations that are responsible for the loop repair defect. We examined loop repair in 13 additional cell extracts (*18*), including several made from tumor cell lines that exhibit microsatellite instability. All extracts efficiently repaired a heteroduplex containing a five-base loop. One extract was from an endometrial tumor cell line (HEC59) known to be defective in mismatch repair and to contain a mutation in one *hMSH2* allele (*4*). Thus, whether hMSH2 participates in loop repair remains to be resolved.

Candidate gene products that may be required for repair of DNA with loops include putative mismatch repair proteins already identified, such as other MSH or MLH homologs, or proteins yet to be discovered. The latter could include a human homolog of a yeast protein that specifically binds to DNA substrates containing loops of three to nine bases, a protein found even in yeast *msh2* and *msh3* mutants (*19*). The possible existence of mutant cell lines defective in some but not all forms of heteroduplex repair is suggested by reports indicating both qualitative and quantitative differences in the stability of various microsatellite alleles in tumor cells and tumor cell lines (*5*, *6*). Identification of extracts defective in repair of loops but not mismatches would reinforce the suggestion that mismatch and loop repair activities have one or more distinct requirements.

Note added in proof: The recent demonstration that purified human MSH2 protein binds to DNA containing loops of up to 14 nucleotides (*20*) is consistent with a possible role for MSH2 protein in loop repair activity.

REFERENCES AND NOTES

1. R. Fishel *et al.*, *Cell* **75**, 1075 (1993); F. S. Leach *et al.*, *ibid.*, p. 1215; C. E. Bronner *et al.*, *Nature* **368**, 258 (1994).
2. N. Papadopoulos *et al.*, *Science* **263**, 1625 (1994).
3. R. Parsons *et al.*, *Cell* **75**, 1227 (1993).
4. A. Umar *et al.*, *J. Biol. Chem.* **269**, 14367 (1994).
5. D. Shibata, M. A. Peinado, Y. Ionov, S. Malkhosya, M. Perucho, *Nature Genet.* **6**, 278 (1994).
6. R. Wooster *et al.*, *ibid.*, p. 152.
7. Y. Ionov, M. A. Peinado, S. Malkhosyan, D. Shibata, M. Perucho, *Nature* **363**, 558 (1993); L. A. Aaltonen *et al.*, *Science* **260**, 812 (1993); S. N. Thibodeau, G. Bren, D. Schaid, *ibid.*, p. 816; A. Lindblom, P. Tannergard, B. Werelius, M. Nordenskjöld, *Nature Genet.* **5**, 279 (1993); R. A. Lothe *et al.*, *Cancer Res.* **53**, 5849 (1993); H.-J. Han, A. Yanagisawa, Y. Kato, J.-G. Park, Y. Nakamura, *ibid.*, p. 5087; J. Risinger *et al.*, *ibid.*, p. 5100; P. Peltomäki *et al.*, *ibid.*, p. 5853; N. M. Mironov *et al.*, *Cancer Res.* **54**, 41 (1994); R. Honchel, K. C. Halling, D. J. Schaid, M. Pittelkow, S. N. Thibodeau, *ibid.*, p. 1159; C. J. Yee, N. Roodi, C. S. Verrier, F. F. Parl, *ibid.*, p. 1641; L. A. Aaltonen *et al.*, *ibid.*, p. 1645; V. Shridhar, J. Siegfreid, J. Hunt, M. M. Alonso, D. I. Smith, *ibid.*, p. 2084; A. Merlo *et al.*, *ibid.*, p. 2098; M.-G. Rhyu, W.-S. Park, S. J. Meltzer, *Oncogene* **9**, 29 (1994); C. Wu *et al.*, *ibid.*, p. 991.
8. M. Strand, T. A. Prolla, R. M. Liskay, T. D. Petes, *Nature* **365**, 273 (1993).
9. B. O. Parker and M. G. Marinus, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1730 (1992); M. Carraway and M. G. Marinus, *J. Bacteriol.* **175**, 3972 (1993).
10. J. S. Beckman and J. L. Weber, *Genomics* **12**, 627 (1992).
11. T. A. Kunkel, *Biochemistry* **29**, 8003 (1990).
12. J. Holmes, S. Clark, P. Modrich, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5837 (1990).
13. D. C. Thomas, J. D. Roberts, T. A. Kunkel, *J. Biol. Chem.* **266**, 3744 (1991).
14. Even with unrepaired DNA heteroduplexes, plaques that are exclusively blue or colorless are obtained upon electroporation of an *E. coli* strain that is defective in mismatch repair. This is believed to reflect selective strand loss in vivo.
15. W.-h. Fang and P. Modrich, *J. Biol. Chem.* **268**, 11838 (1993).
16. Two explanations for repair of looped heteroduplexes can be considered other than repair of DNA containing several consecutive unpaired bases. One is that the sequences of the extra bases allow formation of multiple, adjacent shorter loops that are recognized by the same system that recognizes mismatches. However, except for the substrate containing four extra bases (Fig. 1), the sequences of the looped heteroduplexes used here make this possibility unlikely. A second explanation is that these substrates inadvertently contain an unknown mismatch elsewhere that signals concomitant repair of the loop. This possibility is also unlikely because these same heteroduplexes are efficiently repaired in an extract that is defective in mismatch repair.
17. That these extracts are not generally inactive for all DNA transactions is indicated by the fact that they are competent for simian virus 40 (SV40) origin-dependent DNA replication. The relative replication efficiency of the extracts was: HeLa, 100%; TK6, 120%; HCT116, 90%; LoVo, 41% (*4*).
18. Additional cell lines examined include the transformed lymphoblast line MT1 [V. S. Goldmacher, R. A. Cuzick, W. G. Thilly, *J. Biol. Chem.* **261**, 12462 (1986)]; the colon tumor cell lines SW48 and SW480 [A. Leibovitz *et al.*, *Cancer Res.* **36**, 4562 (1976)], DLD1 and HCT15 [D. L. Dexter, J. A. Barbosa, P. Calabresi, *Cancer Res.* **39**, 1020 (1979)], LS 180 and LS174T [B. H. Tom *et al.*, *In Vitro* **12**, 180 (1976)]; the endometrial tumor cell lines HEC59 [T. Morisawa, *J. Jpn. Soc. Clin. Cytol.* **26**, 433 (1987)] and AN_3CA [C. J. Dawe, W. G. Banfield, W. D. Morgan, M. S. Slat- ick, H. O. Curth, *J. Natl. Cancer Inst.* **33**, 441 (1964)]; the pancreatic cancer cell line HPAF-II [R. S. Metzgar *et al.*, *Cancer Res.* **42**, 601 (1982)]; the breast cancer cell lines T47D [H. C. Freake, C. Marcocci, J. Iwasaki, I. MacIntyre, *Biochem. Biophys. Res. Commun.* **101**, 1131 (1981)] and ZR-75-1 [L. W. Engel *et al.*, *Cancer Res.* **38**, 3352 (1978)]; and the ovarian teratocarcinoma cell line PA-1 [J. Zeuthen *et al.*, *Int. J. Cancer* **25**, 19 (1980)].
19. J. J. Miret, B. O. Parker, R. S. Lahue, *Environ. Mol. Mutagen.* **23** (suppl. 23), 45 (1994).
20. R. Fishel, A. Ewel, S. Lee, M. K. Lescoe, J. Griffith, *Science*, in press.

21. We thank R. M. Schaaper and J. W. Drake for comments on the manuscript.

29 July 1994; accepted 11 October 1994

Ligands for EPH-Related Receptor Tyrosine Kinases That Require Membrane Attachment or Clustering for Activity

Samuel Davis,* Nicholas W. Gale, Thomas H. Aldrich, Peter C. Maisonpierre, Vladimir Lhotak, Tony Pawson, Mitchell Goldfarb, George D. Yancopoulos*

The EPH-related transmembrane tyrosine kinases constitute the largest known family of receptor-like tyrosine kinases, with many members displaying specific patterns of expression in the developing and adult nervous system. A family of cell surface–bound ligands exhibiting distinct, but overlapping, specificities for these EPH-related kinases was identified. These ligands were unable to act as conventional soluble factors. However, they did function when presented in membrane-bound form, suggesting that they require direct cell-to-cell contact to activate their receptors. Membrane attachment may serve to facilitate ligand dimerization or aggregation, because antibody-mediated clustering activated previously inactive soluble forms of these ligands.

Intercellular communication is often mediated by protein factors produced in one cell and recognized by receptors on the surface of other cells. Many of these factors, such as insulin and nerve growth factor, bind to and activate cell surface receptors with intrinsic protein tyrosine kinase activity (*1*). Ligand-mediated activation of these receptor tyrosine kinases regulates cell growth, survival, and differentiation in various cell types (*1*). There remain numerous receptor-like tyrosine kinases whose ligands have yet to be identified, and many of these orphan receptors are specifically expressed in the nervous system (*2*). The EPH-related kinases constitute the largest known family of orphan receptor-like tyrosine kinases, with several members of this family displaying specific expression in the developing and adult nervous system (*2*–*17*). In the adult a number of EPH-related kinases, such as EHK1 (*16*), EHK2 (*16*), and ELK (*7*), are restricted in their expression to discrete neuronal populations, including locus coeruleus neurons and the dopaminergic neurons in the substantia nigra. To identify ligands that might func-

S. Davis, N. W. Gale, T. H. Aldrich, P. C. Maisonpierre, M. Goldfarb, G. D. Yancopoulos, Regeneron Pharmaceuticals, 777 Old Saw Mill River Road, Tarrytown, NY 10591, USA.

V. Lhotak and T. Pawson, Division of Molecular and Developmental Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada.

*To whom correspondence should be addressed.