Nucleic Acids Research, Vol. 20, No. 15 4097–4098

The human gene for apurinic/apyrimidinic endonuclease (HAP1): sequence and localization to chromosome 14 band q12

Bimei Zhao, David K.Grandy[1], Jennifer M.Hagerup, R.Ellen Magenis[2], Leslie Smith[2], Balwant C.Chauhan and W.David Henner*

Division of Hematology and Medical Oncology, [1]Vollum Institute for Advanced Biomedical Research and the Department of Cell Biology and Anatomy and [2]Department of Molecular and Medical Genetics and Child Development and Rehabilitation Center, Oregon Health Sciences University, Portland, OR 97201, USA

Submitted April 29, 1992

GenBank accession no. M92444

Both bacterial and mammalian cells contain multiple forms of apurinic endonucleases, including both class II AP endonucleases and AP lyases. The cDNAs for the major Mg^{2+}-dependent apurinic endonuclease of bovine (1), mouse (2) and human cells (3–5) have all recently been cloned and sequenced. The human cDNA has been variously designated HAP1 (3) or APE (4). The deduced amino acid sequences of the bovine and human cDNAs are nearly identical and have homology to exonuclease III of *E.coli*, exoA of *S.pneumoniae* and the *Drosophila* Rrp1 protein. The HAP1 protein is likely to be important in protection of cells against both spontaneous and induced DNA damages and mutagenesis and the gene coding for this protein could be involved in human diseases of DNA repair deficiency. We have cloned and sequenced the human gene for HAP1 and identified its chromosomal location.

PCR primers TGGTCGAAGACGCGGAAGAG (forward) and TCTTCTTAATCCAGGCTCG (reverse) from the published sequence of BAP1 cDNA (1) were used to amplify a 463 bp segment of bovine genomic DNA which was subcloned into the SmaI site of pTZ18U (6) to yield pBZ2 and sequenced to confirm the presence of BAP1 sequences. Radiolabelled pBZ2 insert was used to screen a human lymohocyte genomic library in EMBL-3 (HL1006d, Clontech) by filter hybridization. 1.8×10^5 plaques were screened and 16 positives were obtained. Adjacent PstI fragments from one of these, lambdaHAP-8, were subcloned into pTZ18U to yield pBZ68 and pBZ60 that contain the 5′ and 3′

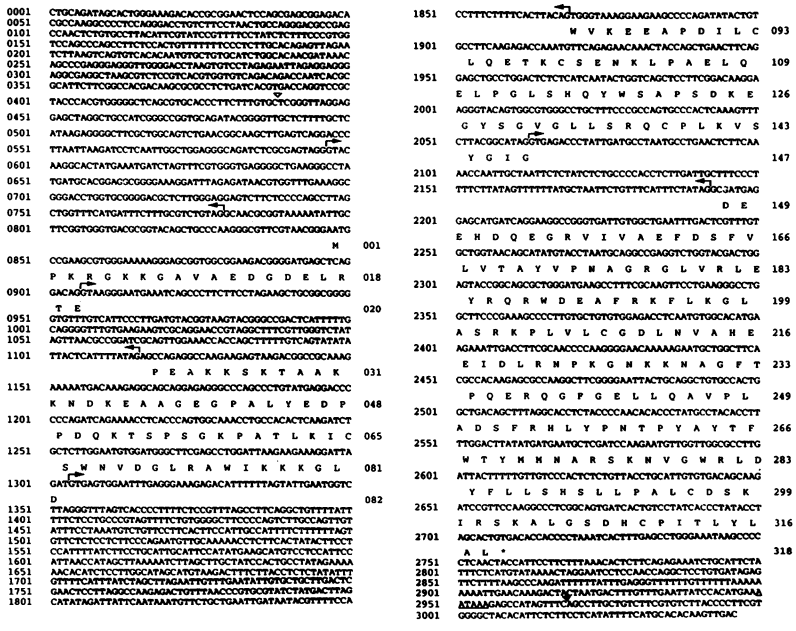

Figure 1. Nucleotide sequence of HAP1 and deduced amino acid sequence. The open triangle indicates the mapped position of the transcription initiation site. The filled triangle indicates the polyadenylation site (3). The polyadenylation signal is underlined. Arrows flank the intron sequences.

* To whom correspondence should be addressed

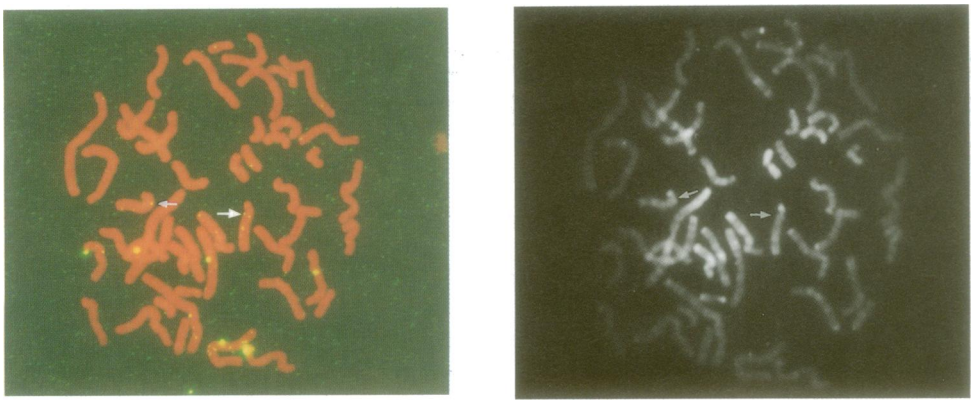

Figure 2. Metaphase chromosomes from a normal male hybridized with HAP1-derived clones pBZ60 and pBZ68. Sequentially stained hybridized metaphase cells, stained first with propidium iodide to show fluorescent signals (left) and then with chromomycinA3/distamycin A to produce R-banding for chromosome identification (right).

regions of the human HAP1 gene, respectively. The sequence of HAP1 was obtained by dideoxy sequencing both strands of the DNA from an upstream PstI site to a downstream HincII site and is shown in Figure 1 (GenBank accession no. M92444). By reference to the published cDNA sequences (3–5) the HAP1 cDNA can be divided into 5 exons and 4 introns. The exons of this sequence match exactly the cDNA sequence reported by Cheng *et al.* (5).

The site for initiation of transcription in the HAP1 gene was determined by the technique of primer extension. A primer complementary to nucleotides 526 to 550 of the genomic sequence was 5′ end-labelled with ^{32}P and extended by reverse transcriptase on cytoplasmic RNA extracted from the human lymphoblastoid cell line WIL2-NS. The results of this analysis indicated that the major site for initiation of transcription is located at nucleotide position 439. The sequence upstream from the transcription initiation site has several of the characteristics of 'housekeeping' genes, with no canonical TATA box or CAAT box and a highly GC rich region. The 5′ region contains a 'CpG island' (7). The analysis also identified a cluster of five uteroglobin transcription factor binding sites (8) near the transcription initiation site at positions 516, 507, 456, 447 and 437.

The human gene HAP1 was assigned to chromosome 14 by two independent methods. A primer pair (GGAATGCCGAA-GCGTGG forward and TCTTCTTAATCCAGGCTCG reverse) was selected that amplified a 437 bp DNA product from human DNA but which did not amplify a similarly-sized product when mouse or hamster DNAs were used as the target. This PCR product was confirmed to contain the expected portions of the HAP1 gene exon 2, intron 2 and exon 3 sequence by multiple restriction digestions. Polymerase chain reactions were then performed with 10 ng of DNA from each of the 18 DNAs from the NIGMS human/rodent somatic cell hybrid mapping panel no. 1. None of the hybrids that lacked chromosome 14 were positive in the PCR analysis and all but one (GM/NA09934) of the hybrids that contained chromosome 14 were positive. Afthough hybrid culture GM/NA09934 contained human chromosome 14, it was present in only 2% of the metaphases. All other chromosomes showed at least three PCR reactions that were discordant. The HAP1 gene was therefore tentatively assigned to chromosome 14.

Two contiguous HAP1 genomic clones were used as probes for *in situ* hybridization, pBZ68 and pBZ60 with inserts of 2.5 kb and 2.0 kb, respectively. Metaphases from normal human lymphocytes were probed with both biotin-labelled plasmids simultaneously and detected by fluorescence microscopy as previously described (9). A total of 42 metaphase cells were examined: all of these had 'double' fluorescent signals, one on each chromatid, towards the centromere of at least one acrocentric D-sized chromosome. Of the 42 metaphases, 5 cells had double fluorescent signals on two acrocentrics, 25 cells had one double and one single signal on acrocentrics and 12 had a double signal on only one of the acrocentrics. No cells had double signals on chromosomes other than a D-sized acrocentric and the low background of other fluorescent signals was never observed more than once for any other location. The same cells were then R-banded and examination revealed that, in all cases, the double signals were from hybridization to chromosome 14 and that the band hybridized was consistently q12, very near to the junction of bands q11.2 and q12 (Figure 2). Together the PCR and *in situ* hybridization results indicate that the HAP1 gene is located on chromosome 14 band q12.

ACKNOWLEDGEMENTS

We thank Dr O.Civelli for helpful discussions, Y.Zhang for excellent technical assistance, Dr M.Litt for the gift of the rodent/human hybrid DNA panel and Dr I.Hickson for communicating results prior to publication. This work was supported by a grant to W.D.H. from the National Institutes of Health (CA35767).

REFERENCES

1. Robson,C.N., Milne,A.M., Pappin,D.J.C. and Hickson,I.D. (1991) *Nucleic Acids Res.* **19**, 1087–1092.
2. Seki,S., Akiyama,K., Watanabe,S., Hatsushika,M., Ikeda,S. and Tsutsui,K. (1991) *J. Biol. Chem.* **266**, 20797–20802.
3. Robson,C.N. and Hickson,I.D. (1991) *Nucleic Acids Res.* **19**, 5519–5523.
4. Demple,B., Herman,T. and Chen,D.S. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 11450–11454.
5. Cheng,X., Bunville,J. and Patterson,T.A. (1992) *Nucleic Acids Res.* **20**, 370.
6. Mead,D.A., Szczesna-Skorupa,E. and Kemper,B. (1986) *Prot. Engin.* **1**, 67–74.
7. Gardiner-Garden,M. and Frommer,M. (1987) *J. Mol. Biol.* **196**, 261–282.
8. Jantzen,K., Fritton,H.P., Igo-Kemenes,T., Espel,E., Janich,S., Cato,A.C.B., Mugele,K. and Beato,M. (1987) *Nucleic Acids Res.* **15**, 4536–4552.
9. Magenis,R.E., Maslen,C.L., Smith,L., Allen,L. and Sakai,L.Y. (1991) *Genomics* **11**, 346–351.