A cDNA, HAPlh, encoding a human apurinic/apyrimidinic (AP) endonuclease from HeLa cells has been cloned and sequenced. The predicted HAPlh protein consists of 318 amino acids and shows 97% similarity and 93% identity to a bovine AP endonuclease (BAP1) (1). It also shows homology to other DNA repair enzymes such as Drosophila Rrpl protein (2), E.coli exonuclease III (ExoIII) (3), and S.pneumoniae exonuclease A (ExoA) (4). The strategy used to obtain the HAPlh cDNA included: (a) PCR amplification of the coding region of HAPlh cDNA using primers designed according to the sequence of BAP1; and (b) amplification of the 5'- and 3'-ends of HAPlh cDNA from a HeLa cell lambda gt11 cDNA library (Clontech) using primers derived from the coding sequences of the HAPlh cDNA and lambda gt11 vector sequences. Several independent PCR amplification products were cloned and sequenced for each of the three regions (coding region, 5'-, and 3'-untranslated) of HAPlh cDNA.

While this work was in progress, the HAPl cDNA sequence, which also encodes a protein homologous to the BAP1 protein, was reported (5). The HAPl cDNA was isolated from a human melanoma cDNA library. A comparison of the two cDNA sequences (HAPlh and HAPl) revealed two differences in the coding region (CR1 and CR2) and one in the 5'-untranslated region (5'UTR) (Figure 1). We have verified our sequence in these regions by restriction analyses of the corresponding PCR products (see Figure 1). Thus, it is unlikely that the differences are due to errors in sequencing of the HAPlh cDNA or to errors arising from PCR amplification. One of the differences (CR 2) predicts an arginine residue in HAPlh but an alanine in HAPl. We have compared the amino acid sequences of HAPlh and HAPl with other DNA repair enzymes exhibiting AP endonuclease activity from different species. As shown in Figure 2, the arginine residue (in bold) predicted to be in the HAPlh protein by the CR2 difference is conserved within all of the endonucleases examined except HAPl. We believe that the differences between HAPlh and HAPl cDNA are due to allelic variation because of the overall identity we have seen between the two cDNAs.

**REFERENCES**


**Figure 1.** Sequence differences between the HAPl and HAPlh cDNAs. The HAPl sequence is shown on top. Numbers in the brackets represent base positions in the respective cDNAs. Differences are in bold and underlined. N.D.: not determined. –: None . Complete: complete digestion.

**Figure 2.** Amino acid sequence comparison within one region of the BAP1, HAPl, HAPlh, Rrpl, ExoA, and ExoIII proteins. Numbers in the brackets indicate the amino acid positions in the respective proteins.