

# Functional characterization of Ape1 variants identified in the human population

Masood Z. Hadi, Matthew A. Coleman, Krzysztof Fidelis, Harvey W. Mohrenweiser and David M. Wilson III\*

Molecular and Structural Biology Division, L-441, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94551, USA

Received July 7, 2000; Revised August 22, 2000; Accepted August 29, 2000

## ABSTRACT

**Apurinic/apryrimidinic (AP) sites are common mutagenic and cytotoxic DNA lesions. Ape1 is the major human repair enzyme for abasic sites and incises the phosphodiester backbone 5' to the lesion to initiate a cascade of events aimed at removing the AP moiety and maintaining genetic integrity. Through resequencing of genomic DNA from 128 unrelated individuals, and searching published reports and sequence databases, seven amino acid substitution variants were identified in the repair domain of human Ape1. Functional characterization revealed that three of the variants, L104R, E126D and R237A, exhibited ~40–60% reductions in specific incision activity. A fourth variant, D283G, is similar to the previously characterized mutant D283A found to exhibit ~10% repair capacity. The most common substitution (D148E; observed at an allele frequency of 0.38) had no impact on endonuclease and DNA binding activities, nor did a G306A substitution. A G241R variant showed slightly enhanced endonuclease activity relative to wild-type. In total, four of seven substitutions in the repair domain of Ape1 imparted reduced function. These reduced function variants may represent low penetrance human polymorphisms that associate with increased disease susceptibility.**

## INTRODUCTION

Chromosome modifications arise during normal cellular processes (e.g. aberrant replication or by attack of reactive oxygen species generated during cellular metabolism) or from exposure to environmental agents such as ionizing and ultraviolet radiation (1). If left unrepaired, replication mistakes and chromosome damage can promote permanent genetic changes (i.e. mutations) that lead to cellular dysfunction or lethality. The most active process for correcting DNA alterations that arise spontaneously or from attack by endogenous reactive chemicals is the Base Excision Repair (BER) pathway (2–4).

The first step of BER typically involves the removal of a damaged or mismatched base by a DNA glycosylase, generating a baseless site (4–6). The major protein in humans responsible

for repairing abasic sites in DNA is Ape1. The Ape1 protein incises the phosphodiester backbone of DNA immediately 5' to the baseless lesion, leaving a strand break with a normal 3'-hydroxyl group and a non-conventional 5'-abasic terminus. Ape1 is a member of the exonuclease III (ExoIII) family of proteins that includes apurinic/aprimidinic (AP) endonucleases from *Escherichia coli*, *Drosophila*, *Arabidopsis* and all vertebrates examined to date (7,8). Subsequent steps of the BER process involve polymerization, termini processing and DNA ligation, restoring DNA to its original, unmodified state (2,4,9).

Mice engineered to lack a central participant (e.g. Ape1, Pol $\beta$  or XRCC1) in the BER pathway do not survive embryogenesis (reviewed in 2,10). The likely interpretation is that BER is needed to cope with the everyday accumulation of genetic damage, and in its absence, the genome of the developing embryo is damaged beyond compatibility with survival. Thus, BER is an essential process for maintaining genetic integrity and in turn animal viability. Although a direct relationship of BER to human disease has not been demonstrated, mutations in BER genes that result in a reduced repair capacity (but not elimination of function) are proposed to be associated with cancer, premature aging and neurodegeneration (3,11). Notably, yeast defective in AP site repair exhibit an elevated spontaneous mutator phenotype (12), and reduced AP endonuclease activity has been observed in brain tissue from patients with the degenerative disease amyotrophic lateral sclerosis (ALS, or Lou Gehrig's disease) (13).

Genetic factors influence disease susceptibility by affecting the consequences of endogenous and environmental exposures (11,14). Specifically, the efficiency and efficacy of gene products involved in the repair of DNA damage plays a key role in determining the mutagenic and cytotoxic outcome of many (if not all) cancer-causing agents. This connection was first identified in individuals with the rare genetic disorder, xeroderma pigmentosum (XP) (15). Patients suffering from XP display an extreme sensitivity to UV radiation due to a defect in the removal of bulky DNA adducts induced by sunlight (stemming from a severe reduction in nucleotide excision repair activity), and exhibit a significantly elevated risk of developing skin cancer. Other examples of DNA repair-related genes associating with increased cancer incidence include the colorectal cancer-related mismatch repair genes and the breast cancer-related BRCA1 and BRCA2 genes (16,17). However, individuals with a mutation in a cancer-causing gene (e.g. XP) are rare and account for <5% of all cancer cases.

\*To whom correspondence should be addressed. Tel: +1 925 423 0695; Fax +1 925 422 2282; Email: wilson61@llnl.gov

While loss-of-repair function variants are uncommon, studies utilizing lymphocyte based assays have shown that 10–20% of the human population exhibit a 20–35% reduced capacity to repair DNA damage induced by gamma radiation, bleomycin and Benzo[a]pyrene Diol Epoxide (BPDE), relative to the population mean. These damage-specific repair capacity phenotypes are independent traits (18–20), which are heritable (20). Notably, such reductions in repair capacity are also associated with an increased likelihood of developing breast, lung and skin cancer (21–26).

Recent studies have shown that single nucleotide polymorphisms (SNPs) and amino acid substitution variants in DNA repair genes are common in the human population (27,28). Molecular epidemiology studies indicate that several substitution variants directly correlate with the persistence of DNA adducts, the formation of genetic alterations, and cancer risk (11,29–34). To begin building a bridge from genotype to phenotype, we sought to characterize amino acid substitution variants in Ape1 identified in the human population. We report here that four of seven variants in the repair domain of human Ape1 exhibit reduced endonuclease function.

## MATERIALS AND METHODS

### Variant identification

Variants in APE1 were identified by resequencing of genomic DNA from 128 individuals (256 chromosomes) using a previously published protocol (27). Thirty-six samples were from caucasians, 24 from Johns Hopkins University (Baltimore, MD) and 12 from University of Michigan (Ann Arbor, MI). Ninety-two samples were from the 'DNA Polymorphism Discovery Resource' at the Coriell Institute for Medical Research (Camden, NJ). The Coriell samples were selected to represent the ethnic diversity of the United States population, although the ethnic origin of specific individuals is unknown. None of the samples can be associated with a donor, and thus were deemed to be exempt by the Institutional Review Board.

The PCR primers used for amplifying the five exons (exon 1 is a non-coding exon) and the immediately adjacent intronic or untranslated regions of APE1 were:

Fex1\_2s.APE 5' GTTTTCCAGTCACGACGAGGCTA-AGCGTCTCCGTCAC 3' (325)

Rex1\_2a.APE 5' AGGAAACAGCTATGACCATCTGCGA-CTTCTTACAAACC 3' (1005)

Fex3s.APE 5' GTTTTCCAGTCACGACGTGTGAA-GAAGTCGCAGGAAC 3' (1028)

Rex3a.APE 5' AGGAAACAGCTATGACCATCCTGAAG-GCTAAACGGAGAA 3' (1071)

Fex4s.APE 5' GTTTTCCAGTCACGACGAATATTGT-GCTGCTTGACTC 3' (1750)

Rex4a.APE 5' AGGAAACAGCTATGACCATGGGAAAG-CAATCAAGAGGTG 3' (2130)

Rex5s.APE 5' AGGAAACAGCTATGACCATTGCTAAT-TCTCTATCTCTG 3' (2126)

Fex5a.APE 5' GTTTTCCAGTCACGACGAGTGTTTAAA-GAAGGAATGG 3' (2759)

Note that exons 1 and 2 are co-amplified using a single primer pair (Fex1\_2s.APE and Rex1\_2a.APE). All primers are specific for intronic regions, located ~50–75 nt from the intron/

exon splice site. M13 forward and reverse primer sequences (underlined) were incorporated into the PCR primers so that the amplified product may be directly sequenced as described (27,35). The reference sequence for numbering the 3'-end nucleotide of each primer (indicated above), and the sites of sequence variation, was GenBank accession no. M92444. All PCR amplification products were sequenced in both directions. The sequencing of rare variants was repeated with an independent amplification.

A direct comparison of the Ape1 protein sequence (GenBank accession no. M92444, considered to be wild-type) with the expressed sequence tag (EST) database ([www.ncbi.nlm.nih.gov/irx/dbST/](http://www.ncbi.nlm.nih.gov/irx/dbST/)) was performed using the tBlastn algorithm (<http://www.ncbi.nlm.nih.gov/BLAST/>) (36). To confirm identified sequence variation, appropriate EST cDNA clones were obtained from the I.M.A.G.E. Consortium (<http://image.llnl.gov/>) and sequenced in both directions using a fluorescent based dye-terminator chemistry (Amersham, Arlington Heights, IL).

### Site-directed mutagenesis and Ape1 purification

To generate specific variants, site-directed mutagenesis was performed using an overlapping PCR method (37), and the following mutant primers (Operon, Alameda, CA; variant codon is underlined):

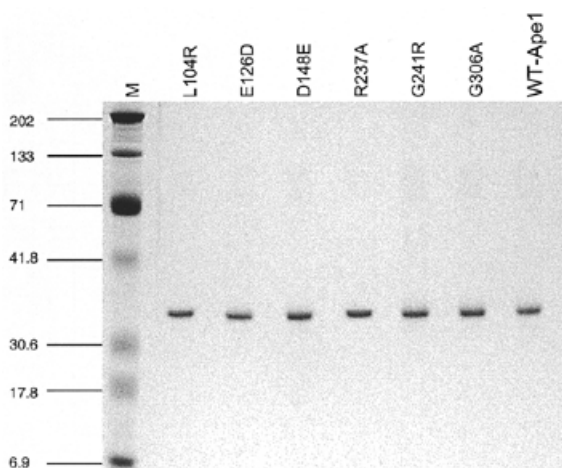
L104R, 5' TCA GAG AAC AAA CGA CCA GCT GAA CTT C 3';  
E126D, 5' CCT TCG GAC AAG GAT GGG TAC AGT GGC 3';  
D148E, 5' AC GGC ATA GGC GAA GAG GAG CAT GAT CAG 3';  
R237A, 5' C ACG CCA CAA GAG GCC CAA GGC TTC GG 3';  
G241R, 5' CAG CAG TAA TTC CCG GAA GCC TTG GCG 3';  
G306A, 5' CA GTG ATC ACT GGC GAG GGC CTT GG 3'.

Recombinant plasmids were purified using the alkaline lysis method (38) and sequenced as above to confirm the presence of the site-directed mutation and the absence of PCR artifacts.

All proteins were expressed in bacteria and purified as described for the wild-type Ape1 protein (39) with minor modifications. Isopropyl β-D-thiogalactopyranoside (IPTG)-inductions were carried out at 37°C, except for R237A, where induction was performed at 28°C for optimal protein solubility/production. Fractions eluting from the cation exchange S10-column containing wild-type or variant Ape1 protein were concentrated using a centricon-10 filtration device (Amicon, Bedford, MA) and further separated on a gel-filtration column (Bio-Silect SEC125–5) in 50 mM HEPES, pH 7.5, 50 mM KCl and 5% glycerol. Protein concentrations were determined simultaneously using molar absorbance coefficients (40) and confirmed on a SDS–polyacrylamide gel with BSA as a standard; comparative visualization of each Ape1 protein sample was performed (Fig. 1).

### Generation of molecular models of Ape1 and analysis of the molecular structure

Before the crystallographic data were publicly available, we modeled Ape1 using the structure of *E.coli* ExoIII as a template (PDB ID 1AKO) (41) to assess potential impact of amino acid variants. A sequence-to-structure alignment was constructed by first generating a sequence alignment of all Ape1/ExoIII homologs exhibiting <80% identity to any other sequence in the set; this prevented domination by a sub-group of sequences (PILEUP program, version 8.0; GCG Inc., Madison, WI, with Blosum50 substitution matrix) (42). The



**Figure 1.** SDS-polyacrylamide gel electrophoresis of wild-type and variant Ape1 proteins. Each lane contains 1  $\mu$ g protein (as indicated). The proteins were separated on a 10% polyacrylamide gel and stained with Coomassie brilliant blue R250. M, molecular weight standard (in kDa) as marked.

alignment was then verified against the spatial and electrostatic constraints of the ExoIII template. The modeled Ape1 structure was built with the Homology module of InsightII (MSI Inc., San Diego, CA), and no attempt was made to generate structure for non-alignable regions (i.e. the unique loop regions) (41). Modeling reliability was assessed with the ProsaII program (43). We have since analyzed the now available crystallographic data of Ape1 (PDB ID IBIX) to reassess the impact of the observed variation on the structure and function of Ape1 (41,44). Comparison of the model with the crystallographic structures of Ape1 revealed no errors in the sequence-to-structure alignment, and no revisions to the functional impact assessments were necessary. Analysis was performed with the molecular visualization program InsightII.

#### DNA substrates and biochemical assays

Oligonucleotides (18mer) containing the abasic site analog [Tetrahydrofuran (F), which is considered a model AP site (39,45–47)] were synthesized (GTCACCGTGFTACGACTC) on an Applied Biosystems DNA synthesizer (model 308B). Duplex DNA substrates with only the F-containing DNA strand 5'-[ $^{32}$ P]-labeled were generated as described (39).

To measure DNA binding activity, a modified version of an established electrophoretic mobility shift assay (EMSA) was used (48). Wild-type or variant Ape1 protein was incubated with 200 fmol of 5'-[ $^{32}$ P]-labeled duplex DNA substrate for 5 min on ice in 10  $\mu$ l EMSA buffer (50 mM HEPES-KOH pH 7.5, 50 mM KCl, 100  $\mu$ g/ml BSA, 0.05% Triton X-100, 10% glycerol and 4 mM EDTA). Binding reactions (8  $\mu$ l) were separated on an 8% non-denaturing gel at 8 V/cm (39). Band visualization and quantification of the DNA (bound and unbound) was achieved using a Molecular Dynamics (Sunnyvale, CA) STORM 860 Phosphorimager and Molecular Dynamics ImageQuant v2.10 software.  $K_d$  values were calculated by first determining the equation of the line through the linear portion of the DNA binding curve (i.e. Ape1 protein at 0.1, 0.3, 1, 3 and 10 ng). Using  $y = mx + b$ , where  $m$  = slope of the line and  $b$  =  $y$  intercept,  $x$  (i.e.  $K_d$ ) was calculated at  $y = 40\%$  (i.e. half maximal binding). Values reported are the means and standard deviations of seven different experiments (except for G306A, where three experiments were performed) from three independent purifications of each protein.

AP endonuclease assays were performed with duplex 5'-[ $^{32}$ P]-labeled F-DNA substrate (1 pmol) at 37°C in 10  $\mu$ l reactions containing 50 mM HEPES-KOH pH 7.5, 50 mM KCl, 100  $\mu$ g/ml BSA, 10% glycerol, 0.05% Triton X-100 and 10 mM MgCl<sub>2</sub> (47). Visualization and quantification of labeled DNAs (substrate and product) was achieved as above. Activity units are defined as pmol of F-DNA incised per minute at 37°C.

## RESULTS

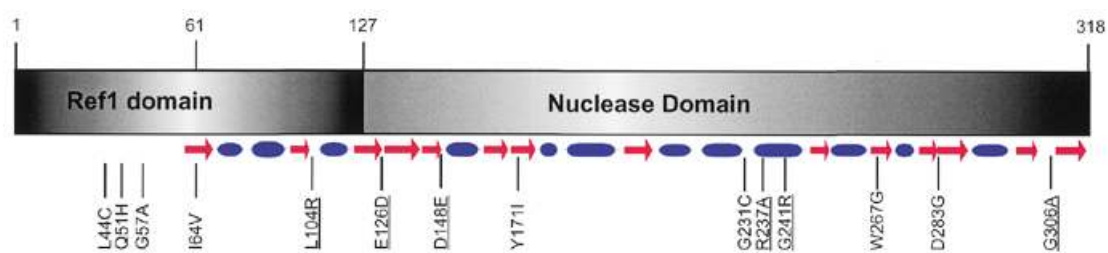
### Identification of amino acid substitutions in Ape1

**Resequencing.** The exonic and immediately adjacent intronic or untranslated regions of APE1 were resequenced in 128 ethnically diverse individuals as described in Materials and Methods. Thirteen SNPs were identified. Four of the SNPs resulted in an amino acid substitution (Table 1). The 148E variant had an allele frequency of 0.38, with the genotypes being distributed as follows: D/D = 54, D/E = 45 and E/E = 25. The other amino acid substitution variants were identified in only 1–8 chromosomes (Table 1), and thus could be rare variants or low frequency polymorphisms with a distribution restricted to a specific subset of the US population. The nine additional SNPs (5'UTR: nucleotide position 340 A/C. Untranslated exon 1: nucleotide positions 448 G/A, 459 C/T and 583 A/G. Intron 1: nucleotide position 774 C/T. Intron 3: nucleotide positions 1368 C/T and 1835 G/T. Exon 3: nucleotide positions 1216 C/G.

**Table 1.** SNPs in APE1 identified by resequencing of genomic DNA

Nucleotide position	Nucleotide substitution (common/variant)	Genomic location	Amino acid position affected	Amino acid substitution (common/variant)	Variant observed/ chromosomes screened (frequency)
1210	G/C	Exon 3	51	Gln/His	8/256 (0.03)
1247	A/G	Exon 3	64	Ile/Val	2/256 (0.008)
2197	T/G	Exon 5	148	Asp/Glu	95/248 (0.38)
2474	G/A	Exon 5	241	Gly/Arg	1/256 (0.004)

The variants Q51H and D148E can be genotyped by allele-specific PCR with *Dde*I and *Bfa*I restriction enzymes, respectively. Note that the 1835 G/T polymorphism (see text) is identical to the 1835 C/A polymorphism reported by Hayward *et al.* (51).



**Figure 2.** Linear schematic diagram of Ape1 protein. Residues 1–127 comprise the Ref1 domain (52,53), while the nuclease domain spans residues 61–318, encoded within exons 3–5 (52). Red arrows depict  $\beta$ -sheet and blue cylinders depict  $\alpha$ -helices. All candidate variants identified (Tables 1 and 2; see text) are labeled, while variants characterized in this study are underlined.

Exon 5: nucleotide position 2560 T/C) were observed in 1–8 chromosomes and again are likely low frequency polymorphisms. None of the sequence polymorphisms were in splice sites or known regulatory regions.

*Literature and database searches.* ALS is a fatal neuromuscular disease that involves elevated intracellular free radical concentrations and the degeneration of nerve cells and pathways in the brain and spinal cord (49), leading to progressive weakness and muscle atrophy. Analysis by Olkowski (50) of cloned APE1 transcripts from nine patients with ALS and two individuals (twins) with familial ALS uncovered five amino acid substitutions: L104R, E126D, D148E (also observed in resequencing and EST database searches, see below), D283G and G306A (also observed in NCBI database comparisons, see below). All substitutions (Table 2) were observed once, excluding E126D, which was observed in both twins with familial ALS. No amino acid changes in Ape1 were detected in the five healthy controls. In an independent study, the D148E variant was found in both ALS and control samples and was shown to associate with sporadic ( $P = 0.027$ ), but not familial ALS (51). No other amino acid substitution variants were identified in the 153 ALS samples screened, although a truncation (i.e. premature stop) variant was identified in a single sporadic ALS patient.

**Table 2.** Candidate variant Ape1 proteins identified in ALS patients or in the NCBI GenBank database

Amino acid variation	Identified in ALS patients <sup>a</sup>	Identified in NCBI database
Q51H		X
G57A		X
L104R	X	
E126D	X	
D148E	X	
R237A		X
D283G	X	
G306A	X	X

Method for identifying amino acid substitution is marked with an "X". Frequency of the variants listed here cannot be determined.

<sup>a</sup>Olkowski *et al.* (50) and Hayward *et al.* (51).

Four variants were found upon comparison of the various GenBank APE1 sequencing reports (either full-length cDNA or genomic clones): Q51H (accession number NP\_001632; also observed by resequencing), G57A (A41631), R237A (S34422; also observed in EST database searches, but found to be a read error, see below) and G306A (A41631; also observed in an ALS patient) (Table 2). For R237A, while we believe it is unlikely that the GC→CG sequence alteration is necessary to create such an amino acid change, we included this variant in the set for characterization.

During the interrogation of 139 sequences from the EST database, we identified 18 potential nucleotide substitutions that cause an amino acid substitution in Ape1. Six of these variants were observed more than once: L44C (CTG→TGT) appeared twice, D148E (GAT→GAG) five times, Y171I (TAT→ATT) twice, G231C (GGC→TGC) four times, R237A (CGC→GCC) three times and W267A (TGG→GGG) twice. Confirmatory studies (not performed on D148E, which was observed in the resequencing effort, and L44C, which falls outside of the DNA repair endonuclease domain) found that the sequence reports of Y171I, G231C, R237A and W267G were inaccurate (i.e. sequence read errors), clearly indicating the importance of direct EST verification. Variants Y171I, G231C and W267G were therefore eliminated from further consideration. Although we were unable to confirm the R237A variation by sequencing of the EST clones, its presence in the NCBI and a potential relationship of R237C to endometrial cancer (Maura Pieretti, University of Kentucky, personal communication) prompted us to characterize this variant.

### Biochemical characterization of selected Ape1 variants

L44C, Q51H, G57A and I64V variants fall outside of the Ape1 endonuclease domain (Fig. 2), and thus were not characterized biochemically (52,53). Whether these amino acid changes affect the ability of Ape1 to regulate the DNA binding activity of proteins such as Fos, Jun and p53 via its Ref1 function (54,55) remains to be determined. Given the similar nature of variant D283G to a previously characterized Ape1 reduced-function mutant (D283A), we elected not to essentially reproduce the prior experiments. Mutating D283 to alanine reduces the repair capacity of Ape1 to ~10% of wild-type (56), and we have found that a D283N mutant exhibits a similar 10-fold reduction in incision capacity (data not shown). The D283G Ape1 variant protein is likely to exhibit a similarly reduced activity, as the negatively-charged aspartate is necessary to

**Table 3.** Conservation of variant residue and predicted impact of amino acid substitution on Ape1 repair endonuclease function based on molecular modeling evaluation

Amino acid variant	Conservation of residue among the ExoIII family	Predicted impact of amino acid substitution
L104R	Always a hydrophobic residue; identical in mouse	Positioned between $\beta$ -sheet no. 2 and $\alpha$ -helix no. 3; changing Leu to Arg may give rise to a localized structural change in the protein by disrupting the hydrophobic interaction with L72, L108 and W119 and introducing new charge–charge interactions with E106 and E107
E126D	Not conserved; identical in mouse	A surface residue involved in a repulsive electrostatic interaction with the DNA backbone; changing Glu to a similar Asp should not affect binding (is difficult to predict in detail)
D148E	Not conserved; E in mouse	Positioned prior to the start of helix no. 4, on the protein surface; conservative substitution unlikely to affect enzymatic activity
R237A	Identical among ExoIII family	Positioned within helix no. 9; R237 stabilizes two negatively charged E216 and E217 residues; an alanine substitution would likely promote local structural instability
G241R	Not conserved; identical in mouse	Positioned in the middle of helix no. 9; a non-conservative substitution that may slightly stabilize the helix due to the stronger helix forming propensity of Arg
D283G	Identical among ExoIII family	Active site functional residue (orients H309 and acts to stabilize its positive charge); D283A or D283N imparts a 10-fold reduced incision activity
G306A	Not conserved; identical in mouse	Positioned prior to $\beta$ -sheet no. 14; alanine substitution could reduce catalytic efficiency by altering the phi, psi angles (126,–7 in the ExoIII crystal structure) of G306 (which are selectively accessible to glycine), thus promoting a conformational change in this region that moves the catalytic pocket (and active site residue H309) out of ideal alignment with the incoming DNA

Conservation is based upon comparison of *E. coli* exonuclease III, *Drosophila melanogaster* Rrp1, *Arabidopsis thaliana* Arp, mouse Apex (accession no. D38077) and human Ape1 (defined here as the ‘ExoIII family’). Impact was determined using molecular models and crystal structures of Ape1 as described in Materials and Methods.

hydrogen bond with H309, facilitating catalysis (57) or protein–DNA complex stability (44).

We assessed the potential impact of the remaining variants (L104R, E126D, D148E, R237A and G306A) prior to functional characterization by examining amino acid conservation among the ExoIII family and by employing molecular modeling techniques (Table 3). These studies suggested a probable negative impact for the L104R and R237A variants, and a less certain negative impact for G306A. For G241R, a potential stabilizing effect was predicted. The D148E and E126D variants were predicted not to impact structure or function, yet given the high frequency of the 148E allele and its potential association with sporadic ALS (51), the functional impact of this variant was of particular interest.

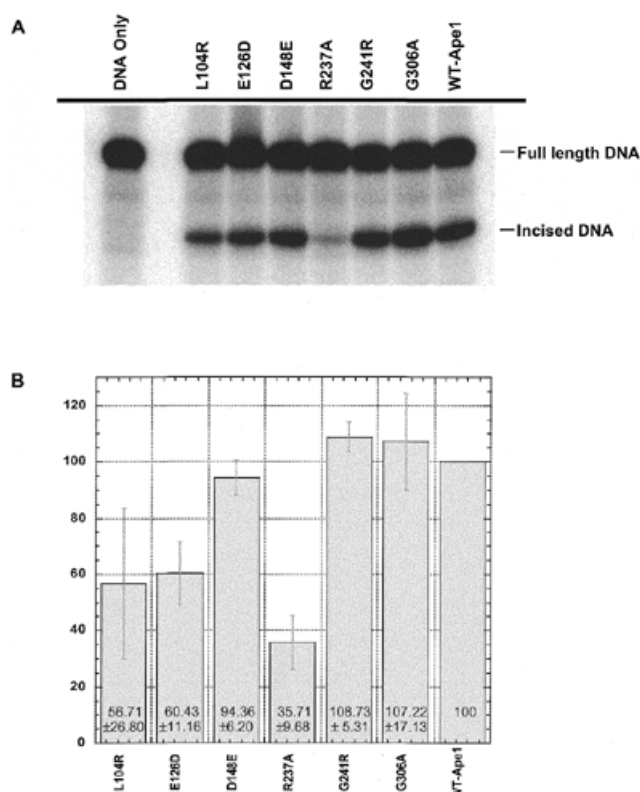
Wild-type and the Ape1 variants L104R, E126D, D148E, R237A, G241R and G306A were purified to >95% homogeneity and protein concentrations accurately quantified (Fig. 1). To measure DNA binding activity, <sup>32</sup>P-labeled duplex DNA substrates harboring a ‘centrally’ located F residue were incubated with varying amounts of Ape1 protein and the binding reactions resolved on a non-denaturing gel. Binding affinity was determined by quantifying the percentage of protein–DNA complex formed as a function of protein concentration (see Materials and Methods). These studies revealed that the Ape1 variants L104R, E126D, D148E, R237A, G241R, G306A and wild-type Ape1 protein bound duplex DNA substrates with apparent  $K_d$  values of  $54.3 \pm 19.7$ ,  $44.0 \pm 7.2$ ,  $20.3 \pm 3.4$ ,  $26.5 \pm 14.1$ ,  $20.1 \pm 0.1$ ,  $35.3 \pm 13.9$  and  $25.8 \pm 12.2$  nM, respectively. Incision capacity was measured by monitoring the ability of wild-type and variant Ape1 proteins to convert <sup>32</sup>P-end-labeled 18mer F-containing oligonucleotide substrates to 9mer DNA products (Fig. 3A). These studies found L104R, E126D, D148E, R237A, G241R and G306A to exhibit an incision

activity that is 0.56, 0.60, 0.94, 0.35, 1.08 and 1.07 relative to wild-type (Fig. 3B).

### Molecular modeling predictions of variant impact

Given their reduced activity, insights into the consequences of L104R and R237A substitutions were gleaned by analyzing the known molecular structures of Ape1 (41,44). Residue L104 is positioned in the loop between  $\beta$ -sheet number 2 and  $\alpha$ -helix number 3, adjacent to the general DNA binding/recognition region (Fig. 4A). An L104R substitution likely perturbs the local structure of this loop, leading to the suboptimal AP–DNA binding and reduced repair capacity (see above). Specifically, hydrophobic interactions between L104 and its neighbors (L72, L108 and W119) are replaced by less-than-optimal contacts with the hydrophobic part of R104 (Fig. 4B). One possible orientation of R104, where the charged end of the side chain is partially buried (Fig. 4B, right), indicates that a cavity could form, diminishing the normally tight hydrophobic packing (Fig. 4B, left). The R104 substitution may also compromise the native local structure by (i) requiring solvation of its charged side chain and (ii) promoting interactions with the negatively charged residues in its immediate vicinity, namely E101 and E107.

R237, which is conserved among the ExoIII family, lies at the beginning of  $\alpha$ -helix number 9 (a region adjacent to the DNA binding groove; Fig. 4A), and appears to stabilize two negatively charged residues (E216 and E217) positioned sequentially along the main chain. These two glutamates form an H-bonding network with R237 and, in the Ape1 crystal structure, co-ordinate a non-active site samarium ion (41). An R237A substitution likely reduces structural stability by disrupting this H-bonding circuitry (Fig. 4C). Our finding that the R237A variant was <10% soluble at 37°C is consistent with



**Figure 3.** (A) Representative incision reaction of the variant and wild-type Ape1 proteins. 2.5  $\mu$ g of protein was incubated with 1 pmol of DNA substrate as described in Materials and Methods. An aliquot of the incision reaction was separated on a 16% polyacrylamide 8 M urea-denaturing gel and visualized with the phosphorimager. (B) Quantification of incision reactions. The substrate and product bands were quantified by phosphorimager analysis. Amount of product formed was divided by the total substrate in the reaction mixture and percentage conversion calculated. The percentage values shown are relative numbers compared to wild-type, set at 100%. Wild-type specific activity was determined to be  $4.3 \times 10^7$  pmol/min, similar to previous reports (47). Data shown are the values of seven different experiments from three independent purifications of each protein.

a protein folding problem at this temperature. Furthermore, this variant exhibited a different chromatographic profile compared to the other Ape1 proteins when eluted from a cation exchange column (i.e. it eluted at lower salt concentration), likely due to an uncompensated local surface charge and/or a global structural change. R237A also exhibits increased sensitivity to multiple freeze-thaw cycles and displays a reduced incision activity half-life at 37°C *in vitro* relative to the wild-type protein (data not shown). We conclude that this substitution results in a localized structure perturbation that makes the protein unstable. It is noteworthy that a similar effect would occur with an R237C substitution, since it would also lead to a loss of the H-bonding network described above.

Residue E126 is positioned in a loop between  $\beta$ -sheet number 3 and 4 (Fig. 4A). An E126D substitution preserves the negative charge and shortens the side chain by one  $-\text{CH}_2-$  group, leading to a relatively minor structural modification. Taken together with the high mobility of this region, as seen in crystallography studies (41,44) and molecular dynamics

simulations (58), it is unclear how this substitution would cause a reduced endonuclease activity. However, it is noteworthy that in the crystal structure of the Ape1/AP-DNA complex, E126 is near the DNA phosphate backbone (44), perhaps to provide an important repulsive interaction. A role in DNA binding is consistent with the reduced AP-DNA binding affinity measured biochemically (see above).

D148E and G241R as predicted had essentially no effect on the repair function of Ape1 (Table 3). Since the G306A variant displayed wild-type activity, the possibility that this amino acid substitution might alter the alignment of the functional residue H309 with respect to AP-DNA, and thus affect incision, was not borne out, emphasizing the importance of biochemical studies. Notably, the conservation and/or modeling pre-screening strategies agreed with the measured biochemical activities in six of seven cases, with E126D being the only significant deviation.

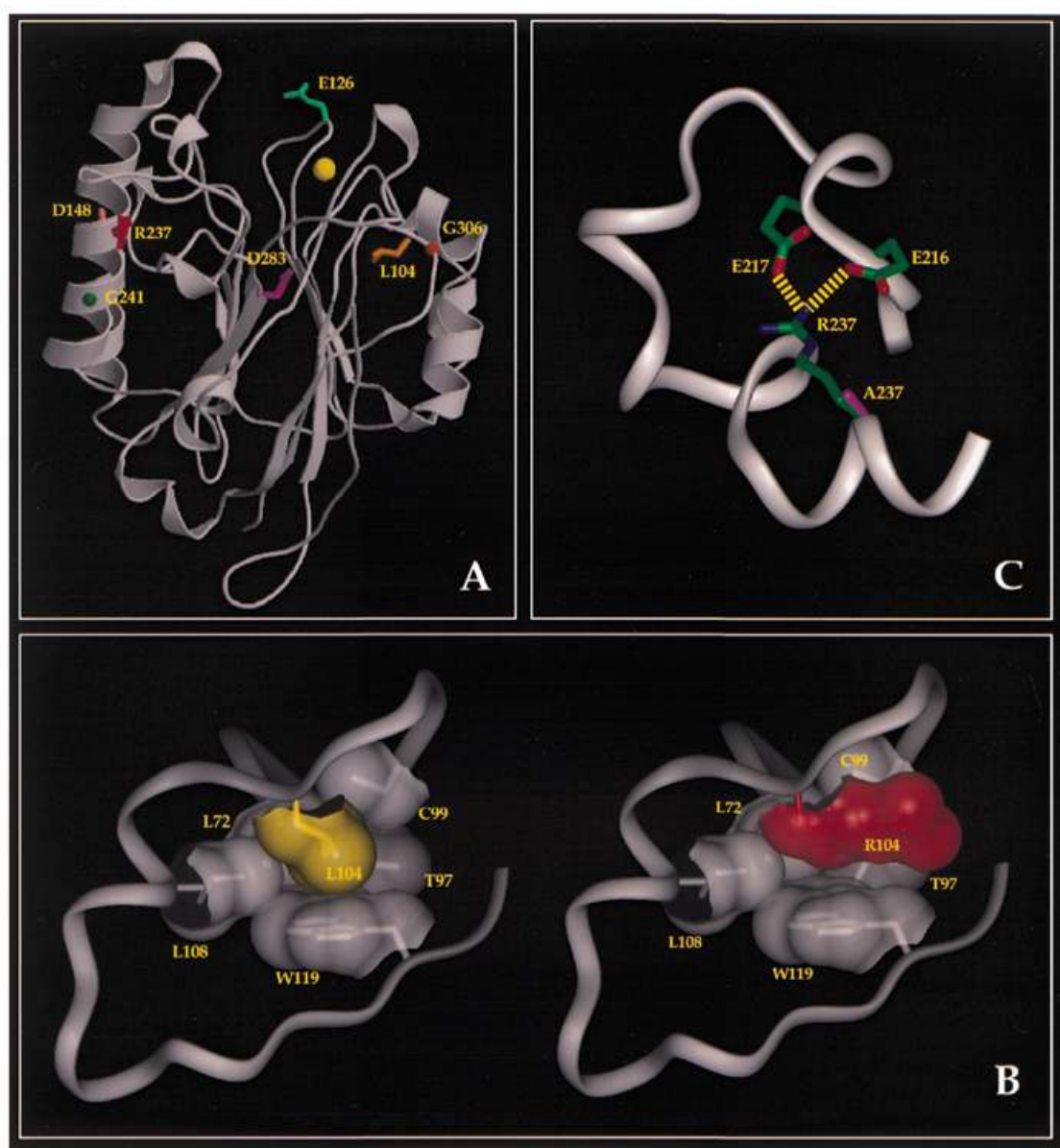
## DISCUSSION

Cancer risk is the product of individual genetic susceptibility and environmental (lifestyle) exposure. The mutagenic or cytotoxic outcome of a particular agent depends upon mutagen activation, detoxification and its effectiveness at generating cellular damage, as well as the ability of the cell to respond and repair potentially deleterious damage prior to replication and cell division. Since most environmental compounds elicit their harmful effects through chemical modification of DNA, repair systems that correct DNA lesions and thus prevent mutagenic or cytotoxic events are directly related to disease susceptibility (9,59).

Loss-of-function variants, generally observed as highly penetrant alleles that segregate in cancer families and associate with very high individual cancer risk, contribute to only a small fraction of cancer cases in the human population. Moreover, a role for genetics is observed in sporadic cancer, as familial relationship is a risk factor independent of environmental exposure. Thus, an effort has been spearheaded in recent years to identify candidate susceptibility markers, i.e. polymorphic genetic traits with low penetrance that are linked to disease proneness (60). While several high frequency DNA repair gene polymorphisms have been shown to associate with cancer risk (11,29-34), little is known about whether or how these genetic differences affect DNA repair function. We report here the identification of genetic variability in the major human AP endonuclease gene, APE1, and demonstrate that variants L104R, E126D and R237A exhibit reduced nuclease function. These, as well as the D283G variant found in one ALS patient (50), represent potential disease susceptibility alleles, and their relationship to cancer and neurodegenerative diseases warrants further investigation. Although we are skeptical that the R237A variant reported is real (for the reason mentioned earlier), it is noteworthy that an independent study has found association of an R237C Ape1 substitution with endometrial cancer (M.Pieretti, personal communication). While reduced AP endonuclease activity has been observed in samples from some ALS patients (13,50), the link between variant D148E and sporadic ALS (51) does not appear to stem from a compromised endonuclease function.

While an ~50% reduction in repair capacity, as observed for L104R, E126D and R237A, may seem unimpressive at first





**Figure 4.** Amino acid variants of Ape1. (A) Amino acids within the Ape1 DNA repair domain found variant in the human population (see text). Residue positions are shown in the context of the Ape1 molecular structure determined (41). L104 (orange), E126 (green), D148 (peach), R237 (red), G241 (green), D283 (purple) and G306 (maroon). DNA binding groove is at the top, with the active site metal cofactor indicated by the yellow ball. (B) Packing of hydrophobic residue L104 (yellow). Wild-type hydrophobic pocket is shown to left, and possible variant (R104, red) positioning is depicted to the right. The size and charge of R104 would diminish the tight residue packing and may introduce new charge–charge interactions with neighboring residues (not shown, see text). (C) Hydrogen bonding network of R237. The relative positioning of variant A237 (purple) is shown. The shorter side-chain of A237 would eliminate the hydrogen bond interactions. All relevant amino acids are indicated, and hydrogen bonds are shown by dashed lines. The images were generated using InsightII, MOLSCRIPT (64,65) and Raster3D (66).

glance, we emphasize that 20–35% reductions in repair capacity (equivalent to a heterozygote individual possessing one 50% active allele and a fully functional allele) have been shown to associate with a 4- to 6-fold higher risk of developing cancer (19,21–25,61). One must also keep in mind that proteins often communicate as members of larger protein networks. Thus, an amino acid substitution may not only affect a specific catalytic function, but also impact interactions with other pathway partners, leading to a more global reduction in

repair efficiency. For instance, while Ape1 binding or repair activity may be reduced 2-fold, its ability to stimulate downstream repair events (62) may be hindered as well, resulting in a more pronounced reduction in BER effectiveness. For many of the Ape1 variants described within, the reduced repair capacity likely stems from a protein structural change, which could have an even greater effect on protein–protein interactions. For Ape1 proteins where we did not observe a defect in endonuclease activity (D148E, G241R and G306A), there

may be a reduced ability to communicate with other BER proteins giving rise to a reduced BER efficiency and thus a potential link to disease susceptibility. Lastly, it is important to consider that reduced function variants may bind substrate DNA *in vivo* and impede normal AP site repair, and in this way, act as dominant-negative factors.

Significantly, >25% of the individuals examined by resequencing (population size of 36) have been found to exhibit amino acid substitutions in more than one protein of the short patch BER pathway (APE1, XRCC1 and POL $\beta$ ; H.W.Mohrenweiser, unpublished results). We are presently working to reconstitute this pathway *in vitro* to examine the efficiency of different combinations of protein variants, since BER in many cases appears to depend kinetically on pathway communication (63). The studies described within demonstrate the power and quandaries of the various approaches for identifying DNA sequence variants, and represent a working model towards defining a link between genetic variation, DNA repair capacity and cancer susceptibility.

## ACKNOWLEDGEMENTS

The authors would like to thank all members of the Wilson laboratory (B. Lee, L. Nguyen, P. Beernink, B. Hamaoka, L. Schild and M. Fernandes) for helpful comments and suggestions during the course of this work. In addition we would like to especially thank members of the Mohrenweiser laboratory (T. Xi, J. Schafer, K. Burkhart-Schultz and S. Duarte) for their sequencing support. This work was carried out under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under contract no. W-7405-ENG-48 and was supported by an NIH (CA79056) grant to D.M.W. III.

## REFERENCES

- Lindahl, T. (1993) *Nature*, **362**, 709–715.
- Wilson, D.M. III and Thompson, L.H. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 12754–12757.
- Lindahl, T. (2000) *Mutat. Res.*, **462**, 129–135.
- Wilson, S.H. (1998) *Mutat. Res.*, **407**, 203–215.
- Krokan, H.E., Standal, R. and Slupphaug, G. (1997) *Biochem. J.*, **325**, 1–16.
- McCullough, A.K., Dodson, M.L. and Lloyd, R.S. (1999) *Annu. Rev. Biochem.*, **68**, 255–285.
- David, S.S. and Williams, S.D. (1998) *Chem. Rev.*, **98**, 1221–1261.
- Demple, B. and Harrison, L. (1994) *Annu. Rev. Biochem.*, **63**, 915–948.
- Lindahl, T. and Wood, R.D. (1999) *Science*, **286**, 1897–1905.
- Friedberg, E.C. and Meira, L.B. (2000) *Mutat. Res.*, **459**, 243–274.
- Lunn, R.M., Langlois, R.G., Hsieh, L.L., Thompson, C.L. and Bell, D.A. (1999) *Cancer Res.*, **59**, 2557–2561.
- Ramotar, D., Popoff, S.C., Gralla, E.B. and Demple, B. (1991) *Mol. Cell. Biol.*, **11**, 4537–4544.
- Kisby, G.E., Milne, J. and Sweatt, C. (1997) *Neuroreport*, **8**, 1337–1340.
- Collins, F.S., Guyer, M.S. and Charkravarti, A. (1997) *Science*, **278**, 1580–1581.
- Cleaver, J.E. (1968) *Nature*, **218**, 652–656.
- Edelmann, W., Umar, A., Yang, K., Heyer, J., Kucherlapati, M., Lia, M., Kneitz, B., Avdievich, E., Fan, K., Wong, E., Crouse, G., Kunkel, T., Lipkin, M., Kolodner, R.D. and Kucherlapati, R. (2000) *Cancer Res.*, **60**, 803–807.
- Osorio, A., Barroso, A., Martinez, B., Cebrian, A., San Roman, J.M., Lobo, F., Robledo, M. and Benitez, J. (2000) *Br. J. Cancer*, **82**, 1266–1270.
- Wu, X., Gu, J., Amos, C.I., Jiang, H., Hong, W.K. and Spitz, M.R. (1998) *Cancer*, **83**, 1118–1127.
- Wu, X., Gu, J., Hong, W.K., Lee, J.J., Amos, C.I., Jiang, H., Winn, R.J., Fu, K.K., Cooper, J. and Spitz, M.R. (1998) *J. Natl Cancer Inst.*, **90**, 1393–1399.
- Cloos, J., Nieuwenhuis, E.J., Boomsma, D.I., Kuik, D.J., van der Sterre, M.L., Arwert, F., Snow, G.B. and Braakhuis, B.J. (1999) *J. Natl Cancer Inst.*, **91**, 1125–1130.
- Cheng, L., Eicher, S.A., Guo, Z., Hong, W.K., Spitz, M.R. and Wei, Q. (1998) *Cancer Epidemiol. Biomarkers Prev.*, **7**, 465–468.
- Helzlsouer, K.J., Harris, E.L., Parshad, R., Perry, H.R., Price, F.M. and Sanford, K.K. (1996) *J. Natl Cancer Inst.*, **88**, 754–755.
- Wu, X., Gu, J., Patt, Y., Hassan, M., Spitz, M.R., Beasley, R.P. and Hwang, L.Y. (1998) *Cancer Epidemiol. Biomarkers Prev.*, **7**, 567–570.
- Parshad, R., Price, F.M., Bohr, V.A., Cowans, K.H., Zujewski, J.A. and Sanford, K.K. (1996) *Br. J. Cancer*, **74**, 1–5.
- Kovacs, E. and Almendral, A. (1987) *Eur. J. Cancer Clin. Oncol.*, **23**, 1051–1057.
- Wu, X., Zhao, Y., Kemp, B.L., Amos, C.I., Siciliano, M.J. and Spitz, M.R. (1998) *Int. J. Cancer*, **79**, 490–493.
- Shen, M.R., Jones, I.M. and Mohrenweiser, H. (1998) *Cancer Res.*, **58**, 604–608.
- Mohrenweiser, H.W. and Jones, I.M. (1998) *Mutat. Res.*, **400**, 15–24.
- Sturgis, E.M., Castillo, E.J., Li, L., Zheng, R., Eicher, S.A., Clayman, G.L., Strom, S.S., Spitz, M.R. and Wei, Q. (1999) *Carcinogenesis*, **20**, 2125–2129.
- Duell, E.J., Wiencke, J.K., Cheng, T.J., Varkonyi, A., Zuo, Z.F., Ashok, T.D., Mark, E.J., Wain, J.C., Christiani, D.C. and Kelsey, K.T. (2000) *Carcinogenesis*, **21**, 965–971.
- Dybdahl, M., Frenzt, G., Vogel, U., Wallin, H. and Nexø, B.A. (1999) *Mutat. Res.*, **433**, 15–22.
- Dybdahl, M., Vogel, U., Frenzt, G., Wallin, H. and Nexø, B.A. (1999) *Cancer Epidemiol. Biomarkers Prev.*, **8**, 77–81.
- Møller, P., Wallin, H., Dybdahl, M., Frenzt, G. and Nexø, B.A. (2000) *Cancer Lett.*, **151**, 187–192.
- Lunn, R.M., Helzlsouer, K.J., Parshad, R., Umbach, D.M., Harris, E.L., Sanford, K.K. and Bell, D.A. (2000) *Carcinogenesis*, **21**, 551–555.
- Nickerson, D.A., Tobe, V.O. and Taylor, S.L. (1997) *Nucleic Acids Res.*, **25**, 2745–2751.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) *Nucleic Acids Res.*, **25**, 3389–3402.
- Erzberger, J.P. and Wilson, D.M. III. (1999) *J. Mol. Biol.*, **290**, 447–457.
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1999) In Ausubel, F.M. (ed.), *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, 4th ed. Wiley, New York.
- Erzberger, J.P., Barsky, D., Schärer, O.D., Colvin, M.E. and Wilson, D.M. III. (1998) *Nucleic Acids Res.*, **26**, 2771–2778.
- Pace, C.N., Vajdos, F., Fee, L., Grimsley, G. and Gray, T. (1995) *Protein Sci.*, **4**, 2411–2423.
- Gorman, M.A., Morera, S., Rothwell, D.G., de La Fortelle, E., Mol, C.D., Tainer, J.A., Hickson, I.D. and Freemont, P.S. (1997) *EMBO J.*, **16**, 6548–6558.
- Henikoff, S. and Henikoff, J.G. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 10915–10919.
- Sippl, M.J. (1993) *Proteins*, **17**, 355–362.
- Mol, C.D., Izumi, T., Mitra, S. and Tainer, J.A. (2000) *Nature*, **403**, 451–456.
- Takeshita, M., Chang, C.N., Johnson, F., Will, S. and Grollman, A.P. (1987) *J. Biol. Chem.*, **262**, 10171–10179.
- Matray, T.J. and Kool, E.T. (1999) *Nature*, **399**, 704–708.
- Wilson, D.M. III, Takeshita, M., Grollman, A.P. and Demple, B. (1995) *J. Biol. Chem.*, **270**, 16002–16007.
- Wilson, D.M. III, Takeshita, M. and Demple, B. (1997) *Nucleic Acids Res.*, **25**, 933–939.
- Roehstein, J.D. (1995) In Serratrice, G.T. and Munsat, T.L. (eds), *Excitatory Mechanism in Pathogenesis of ALS. Pathogenesis and Therapy of Amyotrophic Lateral Sclerosis*. Lippincott-Raven, Philadelphia.
- Olkowski, Z.L. (1998) *Neuroreport*, **9**, 239–242.
- Hayward, C., Colville, S., Swingler, R.J. and Brock, D.J. (1999) *Neurology*, **52**, 1899–1901.
- Walker, L.J., Robson, C.N., Black, E., Gillespie, D. and Hickson, I.D. (1993) *Mol. Cell. Biol.*, **13**, 5370–5376.
- Izumi, T. and Mitra, S. (1998) *Carcinogenesis*, **19**, 525–527.
- Xanthoudakis, S., Miao, G., Wang, F., Pan, Y.C. and Curran, T. (1992) *EMBO J.*, **11**, 3323–3335.
- Jayaraman, L., Murthy, K.G., Zhu, C., Curran, T., Xanthoudakis, S. and Prives, C. (1997) *Genes Dev.*, **11**, 558–570.



56. Masuda, Y., Bennett, R.A. and Demple, B. (1998) *J. Biol. Chem.*, **273**, 30360–30365.
57. Barzilay, G., Mol, C.D., Robson, C.N., Walker, L.J., Cunningham, R.P., Tainer, J.A. and Hickson, I.D. (1995) *Nat. Struct. Biol.*, **2**, 561–568.
58. Nguyen, L.H., Barsky, D., Erzberger, J.P. and Wilson, D.M. III. (2000) *J. Mol. Biol.*, **298**, 447–459.
59. Yu, Z., Chen, J., Ford, B.N., Brackley, M.E. and Glickman, B.W. (1999) *Environ. Mol. Mutagen.*, **33**, 3–20.
60. Roberts, S.A., Spreadborough, A.R., Bulman, B., Barber, J.B., Evans, D.G. and Scott, D. (1999) *Am. J. Hum. Genet.*, **65**, 784–794.
61. Wei, Q., Cheng, L., Hong, W.K. and Spitz, M.R. (1996) *Cancer Res.*, **56**, 4103–4107.
62. Bennett, R.A., Wilson, D.M. III, Wong, D. and Demple, B. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 7166–7169.
63. Wilson, S.H. and Kunkel, T.A. (2000) *Nat. Struct. Biol.*, **7**, 176–178.
64. Esnouf, R.M. (1997) *J. Mol. Graph. Model.*, **15**, 112–113.
65. Esnouf, R.M. (1997) *J. Mol. Graph. Model.*, **15**, 132–134.
66. Merritt, E. and Murphy, M. (1994) *Acta Crystallogr.*, **D50**, 869–873.