MOLECULAR AND CELLULAR BIOLOGY, Oct. 1996, p. 5764–5771 Vol. 16, No. 10
0270-7306/96/$04.00+0

Molecular Cloning of *Drosophila mus308*, a Gene Involved in DNA Cross-Link Repair with Homology to Prokaryotic DNA Polymerase I Genes†

PAUL V. HARRIS, OLGA M. MAZINA, EDITH A. LEONHARDT,‡ RYAN B. CASE,§ JAMES B. BOYD, AND KENNETH C. BURTIS*

Section of Molecular and Cellular Biology, University of California, Davis, California 95616

Received 26 April 1996/Returned for modification 13 June 1996/Accepted 1 July 1996

Mutations in the *Drosophila mus308* gene confer specific hypersensitivity to DNA-cross-linking agents as a consequence of defects in DNA repair. The *mus308* gene is shown here to encode a 229-kDa protein in which the amino-terminal domain contains the seven conserved motifs characteristic of DNA and RNA helicases and the carboxy-terminal domain shares over 55% sequence similarity with the polymerase domains of prokaryotic DNA polymerase I-like enzymes. This is the first reported member of this family of DNA polymerases in a eukaryotic organism, as well as the first example of a single polypeptide with homology to both DNA polymerase and helicase motifs. Identification of a closely related gene in the genome of *Caenorhabditis elegans* suggests that this novel polypeptide may play an evolutionarily conserved role in the repair of DNA damage in eukaryotic organisms.

The *mus308* gene is one of over 30 identified *Drosophila* genes required for resistance to DNA-damaging agents (10, 11, 27, 48). Mutant alleles of *mus308* are, however, unique in conferring strong hypersensitivity to DNA interstrand cross-linking agents such as photoactivated psoralen, diepoxybutane, and nitrogen mustard without conferring hypersensitivity to the monofunctional alkylating agent methyl methanesulfonate (13). It thus seems probable that the wild-type *mus308* gene product functions in a repair pathway that targets DNA cross-links but is not required for either base excision repair of alkylated bases or the repair of DNA strand breaks occurring as a result of methyl methanesulfonate treatment. Very little is known about the mechanism by which interstrand cross-links are repaired in multicellular eukaryotes, and characterization of the *mus308* gene represents an important step toward identifying the essential components involved in this repair pathway.

MATERIALS AND METHODS

Fly stocks. The *mus308* mutant stocks and culture conditions used have been previously described (12, 35). All other stocks are described by Lindsley and Zimm (37).

PCR cloning of the *Men* gene. The degenerate-sense primers coding for the protein sequence IQFEDFA were 5′-AT(ATC)CA(GA)TT(TC)GA(GA)GA(TC)TT(TC)GC-3′. The antisense primers corresponding to the protein sequence FNDDIQG were 5′-CC(CT)TG(TAG)AT(AG)TC(AG)TC(AG)TT(AG)AA-3′. The annealing temperature was 40°C in a 0.1-ml reaction volume containing 50 mM KCl, 10 mM Tris-HCl (pH 9.0), 0.1% Triton X-100, 1.5 mM $MgCl_2$, 50 μM each deoxynucleoside triphosphate (dNTP), 100 nM each primer, 2 U of *Taq* DNA polymerase, and 200 ng of *Drosophila* genomic DNA. The reaction was carried out for 28 cycles. The 92-bp product was reamplified, electrophoresed on a 12% polyacrylamide gel, cut out, minced, and eluted overnight into 0.5 M ammonium acetate. For library screening, 10 ng of PCR product was labeled by 15 cycles of thermal cycling in a 25-μl reaction volume containing the same components as above except that no antisense primer was present, dGTP and dTTP were at 20 μM, and [^{32}P]dCTP and [^{32}P]dATP were each at 2 μM (40 μCi). The product was used to screen the iso-1 *Drosophila* genomic library (kindly provided by J. Tamkun, University of California, Santa Cruz) by standard methods (4, 46).

Isolation and sequencing of genomic and cDNA clones. All genomic clones were isolated from the iso-1 lambda and cosmid libraries. All cDNA clones were isolated from a 2- to 14-h embryonic library of strain Canton-S (Stratagene) or from an ovarian library kindly provided by P. Tolias, Public Health Research Institute, New York, N.Y. After λZAP phage purification, plasmids were autoexcised by the method recommended by Stratagene. Other phage inserts were subcloned into pBluescript vectors prior to sequencing. The complete genomic DNA and cDNA sequences were determined for both strands with ^{35}S-dATP and Sequenase (U.S. Biochemical). Sequencing subclones were generated by exonuclease III-mung bean nuclease digestion (Promega). Gaps in the sequence were filled in with gene-specific primers. The sequences were assembled with PC-Gene (IntelliGenetics). Computerized database similarity searches were done with the BLAST (3) and FASTA (43) algorithms.

Northern blot analysis. Total RNA from ovaries of the control strain *mwh red e* and the various alleles of *mus308* was isolated by hot-phenol extraction (36). Approximately 60 μg per track was electrophoresed, blotted to nitrocellulose (BA85; Schleicher & Schuell), and probed with a restriction fragment labeled with ^{32}P by random primer extension. For normalization, the blot was reprobed with the *Drosophila* ribosomal protein 49 gene (42). Sizes were calculated by comparison of mobility with standards from the GIBCO-BRL 0.24- to 9.5-kb RNA ladder.

In situ hybridization. DNA clones were labeled with biotin-14-dATP (GIBCO-BRL) and hybridized to salivary chromosome squashes (33). Detection was carried out with a streptavidin-alkaline phosphatase conjugate, nitroblue tetrazolium, and 5-bromo-4-chloro-3-indolylphosphate (GIBCO-BRL).

Homology modeling. The C-terminal region of the predicted MUS308 polypeptide was aligned to the C-terminal domains of prokaryotic DNA polymerase I-like enzymes by using a variety of computer algorithms including Bestfit, Pileup (20), Clustal (52), and MACAW (34). With several possible alignments, the Homology module of the Biosym software package was used to model MUS308 onto the structure of DNA polymerase I from *E. coli* (PDB entry 1KLN [6]). Energy minimization was performed in vacuo with a distance-dependent dielectric constant and the consistent-valence force field (CVFF) with cross terms and Morse potentials and a nonbond cutoff of 15 Å (1.5 nm). Models were evaluated with the Profiles module.

Accession numbers. The GenBank accession number of *mus308* is L76559. The *mus-1* homolog from *Caenorhabditis elegans* spans cosmids R12B2 (U00066) and W03A3 (U50184). The C-terminal 3.3 kb of the *mus-1* gene is contained within cDNAs yk20d11 (D35423) and yk47d11 (D67135 and D64204). The gene *ceh-10* resides within a putative 2.66-kb intron of *mus-1*. Database sequences with helicase motifs similar to MUS308 have accession numbers Z49325, P35207, U35242, P32639, and U22156 (*Saccharomyces cerevisiae*); D36993, D68935, D37217, and U20861 (*C. elegans*); Z42372, D29641, U09877, and D31241 (*Homo sapiens*); and U42580 (*Chlorella* virus). Of these, the most closely related to MUS308 are D36993, D68935, and Z42372.

* Corresponding author. Phone: (916) 752-4188. Fax: (916) 752-1185. Electronic mail address: kcburtis@ucdavis.edu.

† Dedicated to the memory of James B. Boyd.

‡ Present address: Radiation Oncology, University of California, San Francisco, CA 94103.

§ Present address: Department of Biochemistry, University of California, San Francisco, CA 94143.

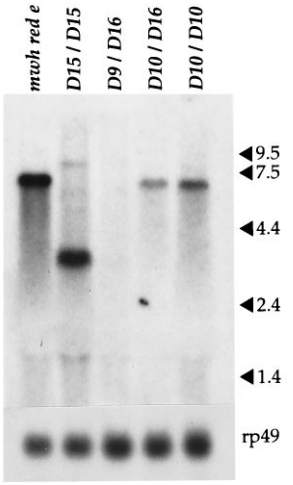

FIG. 1. Northern blot analysis of *mus308* mutants. Total RNA from ovaries of the control strain *mwh red e* and the indicated alleles of *mus308* was electrophoresed, blotted, and probed with a 2.2-kb *Eco*RI-*Sac*I cDNA fragment (Fig. 2). The migration and sizes (in kilobases) of standard RNA markers are indicated to the right of the blot. For normalization of loading, the blot was reprobed with the *Drosophila* ribosomal protein 49 gene (42). The $mus308^{D16}$ allele is a deficiency that removes the entire *mus308* gene and surrounding sequences (35).

RESULTS

The *mus*308 gene was cytogenetically localized to the 87D1,2 region of the third chromosome and shown to be closely linked to the *Men* locus (35), which codes for malic enzyme. Although the latter gene had not been cloned in *Drosophila melanogaster*, sequence information for the homologous gene in other organisms was available and revealed regions of perfectly conserved amino acid sequence among diverse species. Degenerate oligonucleotide primers were designed to amplify a small region of the *Drosophila Men* gene from genomic DNA by PCR. The resulting fragment was used to probe a lambda library of *D. melanogaster* genomic DNA, and one of the selected clones mapped to the 87D1,2 region by in situ hybridization to polytene chromosomes. We carried out a bidirectional chromosomal walk from this position that covered approximately 70 kb and crossed two deficiency breakpoints [Df(3R)kar^{SZ-29} and Df(3R)kar^{SZ-13} (35)] that established the orientation of the walk and delimited the proximal extent of the *mus308* gene. By in situ hybridization, the distal breakpoint of Df(3R)kar^{SZ-13} was mapped to a region 13.5 to 16 kb from the subsequently established 3′ end of the *mus308* gene. By genomic Southern blot analysis, the PCR-generated fragment from the putative *Men* gene mapped to a region 1.8 to 2.0 kb from the 5′ end of the *mus308* open reading frame.

Southern blot analyses of restriction fragment length polymorphisms (data not shown) and Northern (RNA) blot analyses of DNA and RNA from flies homozygous or hemizygous for several mutant alleles established the position of *mus308* within the walk (Fig. 1). The $mus308^{D15}$ allele is associated with a genomic insertion of approximately 4 kb that results in the synthesis of a truncated transcript of approximately 3.4 kb and low levels of a 9.0-kb mRNA that is likely to represent a large chimeric "readthrough" transcript (Fig. 1, lane 2). The $mus308^{D9}$ allele has a different insertion of similar size in the same genomic region (Fig. 2) that results in undetectable transcript levels (Fig. 1, lane 3). The $mus308^{D10}$ allele contains an insertion at the 3′ end of the gene (Fig. 2) that leads to reduced levels of a slightly truncated 6.7-kb transcript (Fig. 1, lanes 4 and 5). The $mus308^{D14}$ mutant expresses a normal-size transcript at 10% of the usual level, and the $mus308^{D2}$ mutant expresses transcripts of approximately 8.5 and 5.5 kb along with a moderately reduced level of normal-size transcript (data not shown). The nature of the mutations in these two alleles is unknown.

Sequence analysis of genomic clones and 13 partial but overlapping cDNA clones revealed a gene structure consisting of seven exons and a predicted open reading frame of 6,177 bases (Fig. 2). The precise transcriptional start site has not been determined, although Northern blot analysis indicates that it occurs within a region of four tandem repeats, each containing a consensus TATA sequence. The first ATG of the large open reading frame is situated 241 bp downstream of the most proximal TATA sequence. The sequence downstream from this ATG predicts a 2059-amino-acid, 229-kDa gene product.

Comparison of *mus308* sequences with the GenBank database revealed the existence of genomic cosmid sequences from *C. elegans* encoding a polypeptide that is clearly homologous to the *mus308* product. Sequencing of a partial cDNA clone (kindly provided by Y. Kohara, National Institute of Genetics, Mishima, Japan) permitted us to determine the location of exons that make up a portion of the mRNA encoded by this gene, which expresses a transcript approximately 5.5 kb in length (data not shown). We have tentatively named this *C. elegans* gene *mus-1.*

The C-terminal regions of the MUS308 and MUS-1 polypeptides share substantial identity (up to 38%) and similarity (up to 59%) with the C-terminal subdomain of prokaryotic DNA polymerase I (Pol I) enzymes (Fig. 3). Sequence conservation is particularly striking in regions that are highly conserved within the bacterial and related bacteriophage sequences. This degree of sequence identity is sufficient to permit

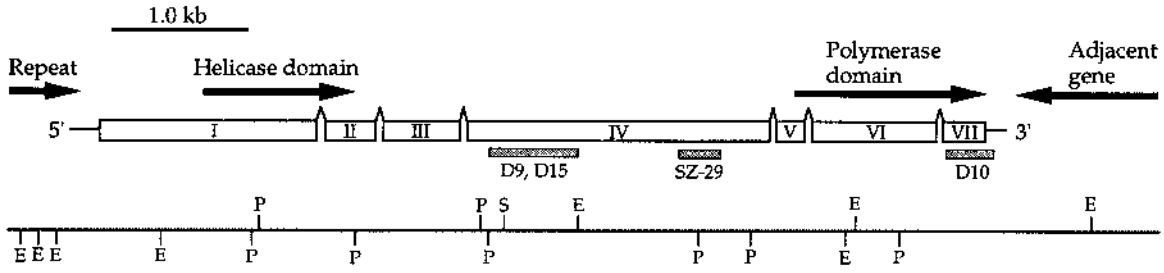

FIG. 2. Schematic representation of the *mus308* gene structure. The seven exons and six introns are shown to scale. Arrows show the extent of the 5′ tandem repeat, the regions of helicase and DNA polymerase homology (detailed in Fig. 4 and 5), and the location of an immediately adjacent gene. The bottom line indicates sites for restriction endonucleases *Eco*RI (E), *Pst*I (P), and *Sac*I (S). Gray bars delimit restriction fragments containing insertion sites in the indicated *mus308* alleles or the distal breakpoint of deficiency Df(3R)kar^{SZ-29}.

FIG. 3. Similarity of the C-terminal domains of MUS308 and MUS-1 to bacterial Pol I. The predicted amino acid sequence was aligned with the polymerase domains from *E. coli* (E.c.), *Streptococcus pneumoniae* (S.p.), and *Bacillus caldotenax* (B.c.). Capital letters indicate that at least four of the five residues are identical or similar, and boldface type indicates 100% identity. Asterisks denote residues that are highly conserved among bacterial and bacteriophage Pol I sequences (15, 18) and which have been implicated in polymerase function (2, 30).

homology modeling based on the known crystal structure of the Pol I Klenow fragment from *Escherichia coli* (6). Modeling suggests a substantial conservation of structure between the polymerizing subdomain of Pol I and the last 400 amino acids of the MUS308 protein (Fig. 4). The largest apparent changes in MUS308 relative to Pol I are a deletion of 8 residues in the loop region following helix H1, an insertion of 7 residues in a loop between helices J and K, and an insertion of 7 residues near the C-terminal end of helix Q. Insertions and deletions in these regions are also evident in one or more of the other bacterial or bacteriophage Pol I sequences. As judged by the method of Lüthy et al. (38), the modeled structure has no seriously misfolded regions within the major secondary-structure elements that form the DNA-binding cleft (Fig. 5), and we propose that the C-terminal domain of MUS308 is homologous to prokaryotic Pol I. Further support for this contention derives from the numerous in vitro mutagenesis, cross-linking, and modeling analyses of Pol I that have identified amino acid residues critical for polymerase function. Of 23 amino acid residues that are both highly conserved throughout the Pol I family and implicated in function (Fig. 3), 21 are conserved in both MUS308 and MUS-1 (Table 1).

No convincing sequence alignment can be detected between MUS308 or MUS-1 and prokaryotic Pol I proteins outside of the polymerase domain. Pol I from *E. coli* is organized into three domains with three distinct enzymatic activities, an N-terminal 5′-nuclease, a central 3′ → 5′ exonuclease, and the C-terminal polymerase domain. Not all Pol I-like DNA polymerases possess the proofreading 3′ → 5′ exonuclease activity. Those which do possess the activity have three moderately conserved regions containing four side chain carboxylates that ligate the divalent metal ions which participate in catalysis of phosphoryl transfer (7–9), and we find little evidence of such sequences in the region immediately amino-terminal of the MUS308 or MUS-1 polymerase domains. However, closer to the N terminus is a region of 212 residues (737 to 948 in MUS308) that may represent a 5′-nuclease domain. This region is 41% identical and 61% similar between MUS308 and MUS-1, suggesting a significant conservation of function. All sequenced bacterial DNA polymerases of the Pol I family and related phage exonucleases contain six conserved motifs characteristic of the N-terminal 5′-nuclease domain (24). These motifs include 10 conserved acidic residues, most of which are closely juxtaposed in the recently solved crystal structure of *Taq* Pol I from *Thermus aquaticus* (31). The carboxylates of at least six of these side chains form three divalent metal ion-binding sites that may participate in the catalysis of phosphoryl transfer as proposed for the 3′ → 5′ exonuclease domain of *E. coli* Klenow fragment. Four of these acidic residues are highly conserved not only in prokaryotic 5′-nucleases but also in several eukaryotic proteins (the RAD2/XPG family and the FEN-1 family) that can catalyze endonucleolytic cleavage at

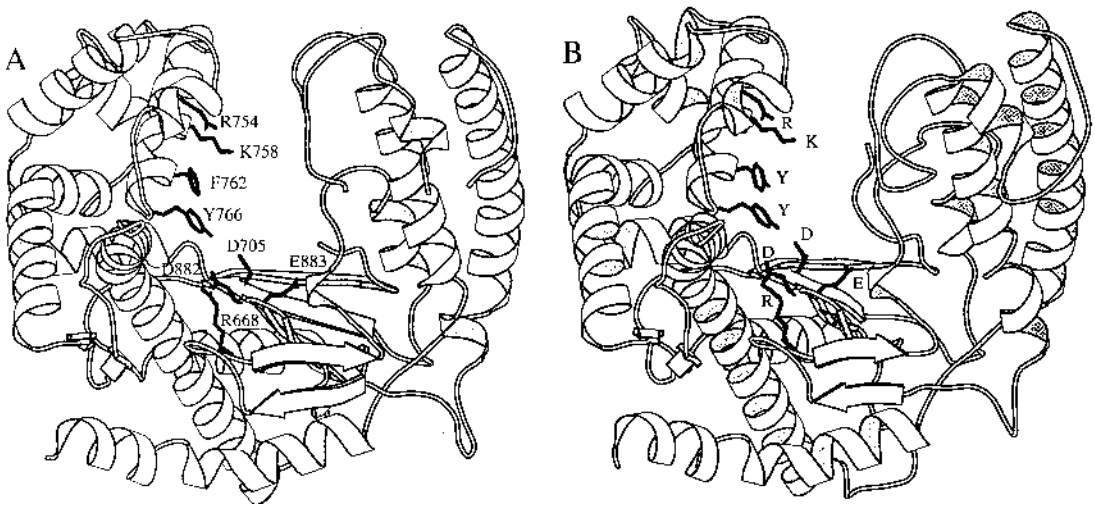

FIG. 4. Homology modeling of the C-terminal region of MUS308. (A) Schematic (32) of the experimentally derived X-ray diffraction structure of the polymerizing domain (amino acids 521 to 928) of *E. coli* Pol I (6). This was used as a template for modeling. Certain functionally critical side chains are shown (Table 1). (B) Schematic of the modeled MUS308 protein from residues 1644 to 2059.

DNA single-strand–duplex junctions, as can the 5′-nuclease of Pol I (16, 25, 26, 39, 44). A similar cluster of four acidic residues is found in the 212-residue conserved region of MUS308 and MUS-1 (Fig. 6). However, confirmation of the functional role of this domain of MUS308 will require biochemical data.

Remarkably, the N-terminal domains of MUS308 and MUS-1 contain the seven motifs characteristic of most members of "superfamily 2" DNA and RNA helicases (Fig. 7) (21, 22). A BLAST search of the protein and translated nucleic acid databases by using these domains detects significant alignment (smallest sum probabilities, <0.05) with a large number of known or suspected helicases. Additional analysis by multiple sequence alignment indicates that three of the helicase motifs in MUS308 and MUS-1 occur in variant form, which serves to define a subfamily of putative helicases that also includes four other sequences from *C. elegans*, five sequences from *S. cerevisiae*, and the apparent human homologs of these sequences (accession numbers are given in Materials and Methods). Within most members of this subfamily, motif I (the Walker A box) is unusual in having a serine in place of the glycine found in almost all other known or putative helicases. In SwissProt (release 31), 3,077 protein sequences contain an annotated putative Walker A-type NTP-binding motif. Of these, relatively few have a serine in the same position as in members of the MUS308 family. Several are involved in replication, recombination, and/or repair, e.g., bacterial RecA protein and DnaB replicative helicase, yeast Rad57, and mammalian parvovirus NS-1 replication accessory protein. Another distinguishing characteristic of the MUS308 subfamily is a conserved threonine in motif V. This is almost invariably a hydrophobic residue in other known superfamily 2 members. An additional unique feature of the MUS308 subfamily is a methionine in motif VI that replaces what is commonly an arginine in other helicases. Finally, the subfamily also contains a conserved motif preceding motif V (called IVa in Fig. 7), which, although present in some other helicases, is uniquely characterized by a histidine residue that is almost invariably a hydrophobic residue in other helicases (unpublished observations).

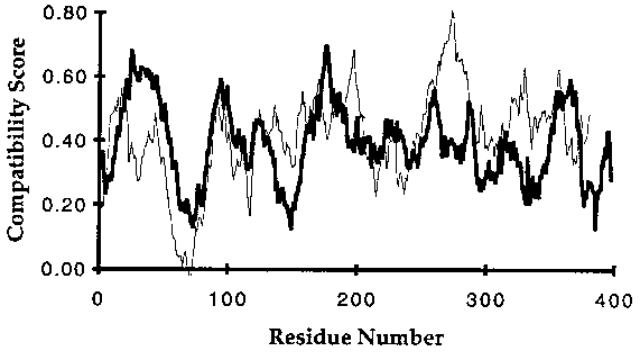

FIG. 5. Profile analysis of the modeled three-dimensional structure of the C-terminal 416 residues of MUS308 (heavy line) and the corresponding region of the experimentally determined structure of the Pol I Klenow fragment from *E. coli* (lighter line) (6). The method of Lüthy et al. (38) was used to evaluate the local environment of every amino acid side chain in the structure and score the suitability of that side chain for its environment. The window size shown is 20 amino acids. Potentially misfolded regions are indicated by scores approaching zero. The apparently misfolded region in both structures from 60 to 85 corresponds to a region of missing and poorly resolved residues in the Pol I template structure. Other low-scoring regions in the MUS308 model correspond to insertions in the sequence relative to Pol I. These areas were modeled de novo as loops and not extensively evaluated.

DISCUSSION

The C-terminal regions of MUS308 and MUS-1 are strikingly homologous to the polymerizing domains of the prokaryotic Pol I family. They are unlike any other sequenced eukaryotic DNA polymerase, with the possible exception of mitochondrial DNA polymerase γ (Pol γ), which is believed to be an evolutionary homolog of bacterial DNA polymerase I. The C-terminal region of MUS308 shares considerably less sequence similarity with yeast Pol γ than it does with Pol I. Indeed, MUS308 is no more similar to Pol γ than Pol I is to Pol γ, indicating that Pol γ is unlikely to be the evolutionary progenitor of MUS308. Aside from sequence homology, two other lines of evidence strongly support close homology between Pol I and MUS308. First, the three-dimensional structure of the C-terminal domain of MUS308 can be successfully modeled with Pol I as a template. We verified the legitimacy of this model by using a program which evaluates the local three-

TABLE 1. Conservation of key residues in the polymerase domain of Pol I and MUS308

Residue in *E. coli* Pol I	Location	Possible function	Conserved in MUS308	Conserved in MUS-1
Asn-579	Loop	DNA P_i binding	Yes	Yes
Ser-582	Loop	DNA P_i binding	Yes	Yes
Thr-609	Helix H2	DNA binding	Yes	Yes
Arg-631	Helix I	DNA P_i binding	Yes	Yes
Lys-635	Helix I	DNA P_i binding	No	No[a]
Arg-668	β-Hairpin loop	DNA binding, catalysis	Yes	Yes
Asn-675	Loop	DNA P_i binding	Yes	Yes
Asn-678	Helix J	DNA P_i binding	Yes	Yes
Arg-682	Loop	DNA binding	No[b]	Yes
Arg-690	Helix K	DNA binding	Yes	Yes
Asp-705	β-Strand 9	Metal binding/catalysis	Yes	Yes
Gln-708	Loop	dNTP binding?	Yes	Yes
Glu-710	Helix L	dNTP binding	Yes	Yes
His-734	Helix N	DNA binding	No	No
Arg-754	Helix O	dNTP P_i binding	Yes	Yes
Lys-758	Helix O	dNTP binding	Yes	Yes
Phe-762	Helix O	dNTP ribose binding	Yes[c]	Yes[c]
Tyr-766	Helix O	DNA binding	Yes	Yes
Arg-841	Helix Q	DNA/dNTP binding	Yes	Yes
Asn-845	Helix Q	dNTP binding	Yes	Yes
Gln-849	Helix Q	DNA binding, catalysis	Yes	Yes
His-881	β-Hairpin loop	dNTP binding	Yes	Yes
Asp-882	β-Hairpin loop	Metal binding/catalysis	Yes	Yes
Glu-883	β-Strand 13	Metal binding	Yes	Yes

[a] This is a histidine in *C. elegans* and bacteriophage SP1 and an arginine in *Mycobacterium tuberculosis* and *M. leprae.*
[b] This residue is not 100% conserved in the Pol I family. An adjacent lysine residue in *Drosophila* MUS308 may play the same functional role.
[c] This is a tyrosine in MUS308 and also in Pol I from some other bacterial species.

dimensional environment of every amino acid residue in the structure and scores the suitability of that side chain for its environment. If the modeled structure contains incorrect folds, this will be reflected in a low score for that particular region (38). On average, the MUS308 model scores almost as highly as does the empirically determined structure of Pol I Klenow fragment. This provides a high degree of confidence that the structure of the C-terminal region of MUS308 is very similar to that of Pol I. Second, most of the amino acid residues known to be critical to polymerase function are conserved in both MUS308 and MUS-1. The two exceptions are Lys-635 (*E. coli* numbering) and His-734. Chemical modification of Lys-635 in *E. coli* Pol I results in a functional polymerase with reduced processivity (5). Mutation of His-734 to alanine significantly reduces both DNA binding and $k_{cat(dTTP)}$ for the polymerase reaction, although the extent of reduction in the latter is dependent on the dTTP concentration (2). The corresponding residue in both MUS308 and MUS-1 is phenylalanine. The effect of this substitution cannot currently be predicted. Phenylalanine 762 of *E. coli* Pol I is replaced by tyrosine in both MUS308 and MUS-1, as well as in some other bacterial Pol I sequences. Mutation of this residue to tyrosine in Pol I from *E. coli* and *Thermus aquaticus* greatly decreases the ability of the enzyme to discriminate deoxynucleotides from dideoxynucleotides (51).

```
   Taq   gyEaDdvlasLakkaekegyevriltaDkDlyqll  149
   XPG   pmEaDaqcaiLdltdqtsg----titdDsDiwlfg  817
 FEN-1   psEaEascaaLv----kagkvyaaateDmDcltfg  186
MUS308   vlEsElhavyLv-----tpysvcyqlqDiDw-lly  797
 MUS-1   alDtElhmlyLv-----tpinvsv-wqEcDwhhlf
```

FIG. 6. Alignment of metal-ion-binding residues of *Taq* DNA polymerase with a conserved region of MUS308. The acidic residues shown in boldface type correspond to residues 117, 119, 142, and 144 in *Taq* DNA polymerase. Also shown is a region of the human XPG repair nuclease and the FEN-1 protein, both of which possess 5′-nuclease activity with similarities to Pol I.

The N-terminal domains of MUS308 and MUS-1 are clearly similar to the large superfamily of helicases and helicase-like proteins. We have identified 15 other sequences in GenBank which have uniquely characteristic sequence similarity to this region of MUS308 and MUS-1; however, only 1 of these, the SKI2 antiviral protein of *S. cerevisiae*, has any known biological function (55). The MUS308 subfamily, as a group, is not obviously more closely related to any one helicase family than to another but, rather, appears to combine motifs from various families. For example, the sequence of motif II is characteristic of the so called DEXH family of helicases, which is highly diverse in function and includes "repair/recombination" helicases such as Rad3, Rad25, and Rad54 of *S. cerevisiae* and RecG and RecQ of *E. coli*. The sequence of motif IV, however, is more characteristic of the DEAD family of RNA helicases, and the asparagine in motif V is found in the SNF2/RAD54 family of helicase-like ATPases potentially involved in transcriptional regulation and DNA repair.

Biochemical and genetic analyses in both prokaryotes and eukaryotes indicate that DNA interstrand cross-links are repaired by an excision-recombination mechanism (17, 23, 29, 40, 41, 47, 53, 54). In prokaryotes, removal of a DNA interstrand cross-link is initiated by the UvrABC endonuclease complex, which incises one strand on each side of the cross-link (53). In eukaryotes, this initial step is presumably carried out by an analogous complex of proteins that is required for nucleotide excision repair (reviewed in reference 19). At least one allele of *mus308* is proficient in the initial incision step of nuclear DNA cross-link repair (13), and thus MUS308 does not appear to be involved in incision.

The first postincision step in cross-link repair in *E. coli* is thought to be the generation of a gap at the site of incision

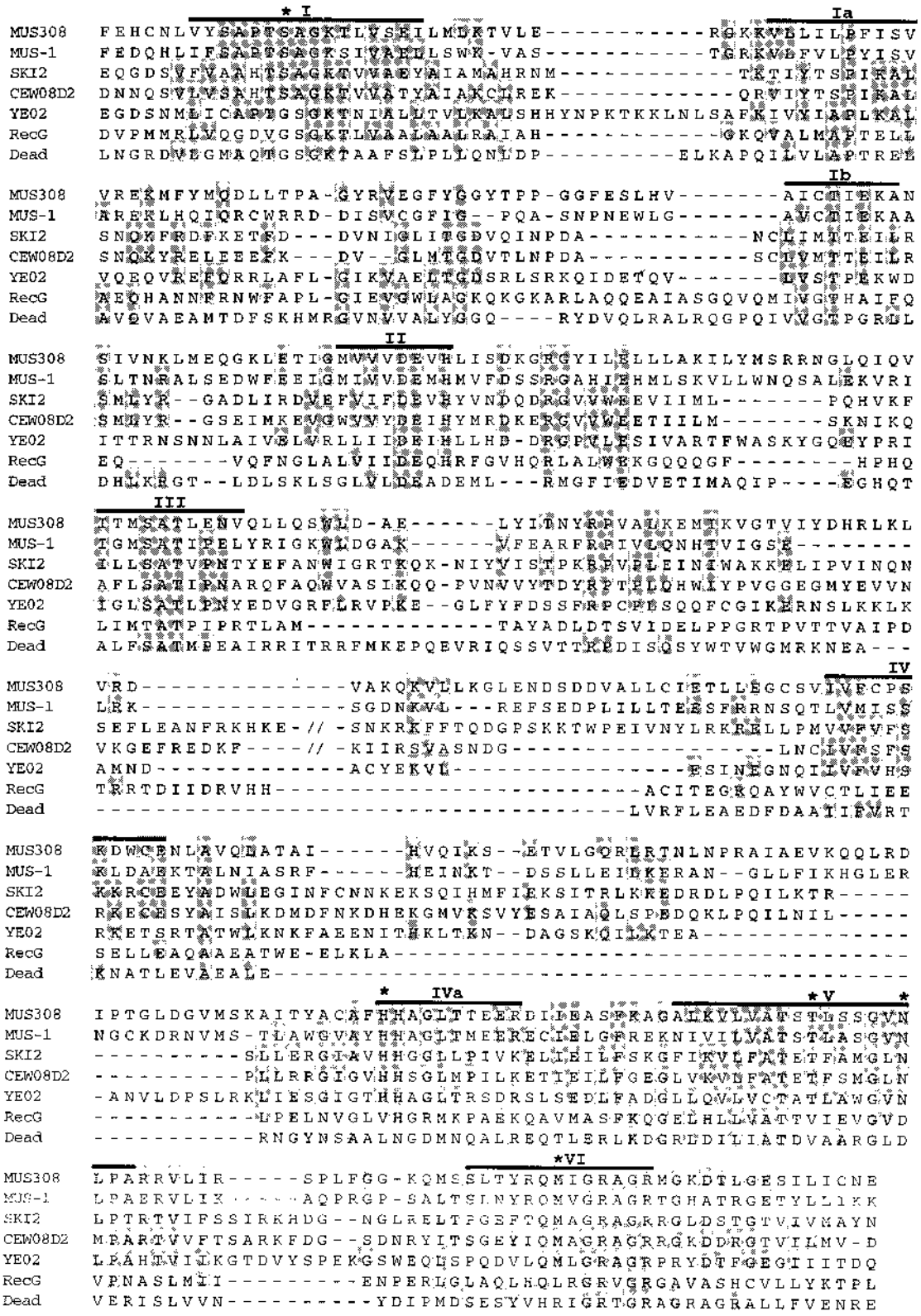

FIG. 7. Presence of helicase-like motifs in the N-terminal domain of MUS308. The predicted sequence was aligned with MUS-1 (predicted from the cDNA and genomic sequence), with related sequences from *S. cerevisiae* (SKI2 and YE02) and *C. elegans* (CEW08D2), and with a known DNA helicase (RecG) and RNA helicase (Dead) from *E. coli*. The eukaryotic sequences all belong to the MUS308 subfamily. Roman numerals indicate regions conserved in many DNA and RNA helicases, and the bars mark their approximate extent. Regions Ib and IVa are not commonly recognized motifs but are characteristic of the MUS308 subfamily. The asterisks denote residues that are highly unusual or unique among known or suspected helicases and thus might serve as markers for this subfamily. Shading marks regions of identity in at least three of the seven sequences. A portion of SKI2 and CEW08D2 is not shown and is marked by //.

through the action of the 5′-nuclease of Pol I, perhaps in concert with the UvrD helicase. The resulting single-stranded region provides a substrate for binding of RecA protein and the initiation of homologous pairing and strand exchange (47). The polymerizing activity of Pol I can then carry out repair synthesis with the undamaged homolog as a template. Pol I can also fill in the gap left by subsequent excision of the cross-link from the other strand.

The polymerase domain of the MUS308 protein may play a role analogous to bacterial Pol I in gap fill-in during recombinational repair of DNA cross-links. A 5′-nuclease, if present, might function to create the gap necessary for binding of proteins that promote homologous pairing and strand exchange. The helicase domain could function similarly to one of several helicases of *E. coli* (RecBCD, UvrD, RecG, and RuvAB) known to participate in various aspects of recombinational repair. The MUS308 helicase domain has some sequence similarity to the RecG helicase, particularly in motif V and in the lack of a nearly invariant arginine in motif VI (Fig. 7). Few DNA polymerases can catalyze strand displacement from a nick or small gap in the absence of a helicase or other accessory factor, and covalent coupling between a helicase and a polymerase would be one mechanism to coordinate duplex unwinding and polymerization.

Two lines of evidence suggest that MUS308 is involved in the repair not only of interstrand cross-links but also of other types of DNA damage in which the coding potential of both strands is simultaneously compromised. Such damage can occur as a result of DNA synthesis on a damaged template, with the formation of daughter strand gaps opposite lesions (reviewed in reference 19). It has been observed that *mus308*^{D2} mutant primary cell cultures display a reduced ability to synthesize high-molecular-weight DNA in vivo on a UV-damaged template (14), indicative of a defect in repairing such damage. A similar role for MUS308 was recently proposed on the basis of the hypermutability of a *mus308* mutant in response to mutagens that produce persistent lesions in DNA (1).

The phenotype of *mus308* mutants is comparable in certain respects to that seen in Fanconi anemia, a typically fatal human genetic disorder that is characterized by cellular hypersensitivity to DNA-cross-linking agents and a failure of bone marrow stem cells to proliferate normally, leading to pancytopenia. The gene for one of the four known complementation groups of Fanconi anemia has been cloned. However, the predicted amino acid sequence has thus far provided no clue to the function of this gene (49, 50). We are currently investigating the possibility that *mus308* is a homolog of one of the Fanconi anemia genes. Other than *mus308* and possibly one or more of the genes associated with Fanconi anemia, the only reported example of a gene specifically associated with unique sensitivity to cross-linking agents is the *PSO2* gene (also known as *SNM1*) (28, 45). The predicted amino acid sequence of this gene is also uninformative with respect to function.

Should biochemical analysis confirm that the MUS308 protein is in fact a helicase-polymerase or a helicase-nuclease-polymerase, it will represent the first example of these activities residing on the same polypeptide. Biochemical characterization of these activities and the identification of homologs in other eukaryotes should prove useful in elucidating the mechanism of DNA cross-link repair in eukaryotes.

ACKNOWLEDGMENTS

We thank O. Chesnokova, A. Lehman, G. Donovan, M. DeRisi, R. Jain, and A. Wapner for technical support. This work was supported by Public Health Service grant HL-50068 from the National Heart, Lung, and Blood Institute, and by the Fanconi Anemia Research Fund.

REFERENCES

1. **Aguirrezabalaga, I., L. M. Sierra, and M. A. Comendador.** 1995. The hypermutability conferred by the *mus308* mutation of Drosophila is not specific for cross-linking agents. Mutat. Res. **336:**243–250.
2. **Astatke, M., N. D. Grindley, and C. M. Joyce.** 1995. Deoxynucleoside triphosphate and pyrophosphate binding sites in the catalytically competent ternary complex for the polymerase reaction catalyzed by DNA polymerase I (Klenow fragment). J. Biol. Chem. **270:**1945–1954.
3. **Atschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
4. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Siedman, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. Greene Publishing Associates and Wiley-Interscience, New York.
5. **Basu, S., A. Basu, and M. J. Modak.** 1988. Pyridoxal 5′-phosphate mediated inactivation of *Escherichia coli* DNA polymerase I: identification of lysine-635 as an essential residue for the processive mode of DNA synthesis. Biochemistry **27:**6710–6716.
6. **Beese, L. S., V. Derbyshire, and T. A. Steitz.** 1993. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. Science **260:**352–355.
7. **Beese, L. S., and T. A. Steitz.** 1991. Structural basis for the 3′-5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. EMBO J. **10:**25–33.
8. **Bernad, A., L. Blanco, J. M. Lazaro, G. Martin, and M. Salas.** 1989. A conserved 3′-5′ exonuclease active site in prokaryotic and eukaryotic DNA polymerases. Cell **59:**219–228.
9. **Blanco, L., A. Bernad, and M. Salas.** 1991. MIP1 DNA polymerase of *S. cerevisiae*: structural similarity with the *E. coli* DNA polymerase I-type enzymes. Nucleic Acids Res. **19:**955.
10. **Boyd, J. B., M. D. Golino, T. D. Nguyen, and M. M. Green.** 1976. Isolation and characterization of X-linked mutants of Drosophila melanogaster which are sensitive to mutagens. Genetics **84:**485–506.
11. **Boyd, J. B., M. D. Golino, K. E. Shaw, C. J. Osgood, and M. M. Green.** 1981. Third-chromosome mutagen-sensitive mutants of Drosophila melanogaster. Genetics **97:**607–623.
12. **Boyd, J. B., J. M. Mason, A. H. Yamamoto, R. K. Brodberg, S. S. Banga, and K. Sakaguchi.** 1987. A genetic and molecular analysis of DNA repair in Drosophila. J. Cell Sci. Suppl. **6:**39–60.
13. **Boyd, J. B., K. Sakaguchi, and P. V. Harris.** 1990. *mus308* mutants of Drosophila exhibit hypersensitivity to DNA cross-linking agents and are defective in a deoxyribonuclease. Genetics **125:**813–819.
14. **Boyd, J. B., and K. E. S. Shaw.** 1982. Postreplication repair defects in mutants of *Drosophila melanogaster*. Mol. Gen. Genet. **186:**289–294.
15. **Braithwaite, D. K., and J. Ito.** 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res. **21:**787–802.
16. **Cloud, K. G., B. Shen, G. F. Strniste, and M. S. Park.** 1995. XPG protein has a structure-specific endonuclease activity. Mutat. Res. **347:**55–60.
17. **Cole, R. S., and R. R. Sinden.** 1975. Repair of cross-linked DNA in *Escherichia coli*. Basic Life Sci. **5B:**487–495.
18. **Delarue, M., M. Poch, N. Tordo, D. Moras, and P. Argos.** 1990. An attempt to unify the structure of polymerases. Protein Eng. **3:**461–467.
19. **Friedberg, E. C., G. C. Walker, and W. Siede.** 1995. DNA repair and mutagenesis. ASM Press, Washington, D.C.
20. **Genetics Computer Group.** 1994. Program manual for the Wisconsin package, version 8 ed. Genetics Computer Group, Madison, Wis.
21. **Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov.** 1988. A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. FEBS Lett. **235:**16–24.
22. **Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov.** 1989. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. **17:**4713–4730.
23. **Gruenert, D. C., and J. E. Cleaver.** 1985. Repair of psoralen-induced cross-links and monoadducts in normal and repair-deficient human fibroblasts. Cancer Res. **45:**5399–5404.
24. **Gutman, P. D., and K. W. Minton.** 1993. Conserved sites in the 5′-3′ exonuclease domain of *Escherichia coli* DNA polymerase. Nucleic Acids Res. **21:**4406–4407.
25. **Habraken, Y., P. Sung, L. Prakash, and S. Prakash.** 1995. Structure-specific nuclease activity in yeast nucleotide excision repair protein Rad2. J. Biol. Chem. **270:**30194–30198.
26. **Harrington, J. J., and M. R. Lieber.** 1994. Functional domains within FEN-1 and RAD2 define a family of structure-specific endonucleases: implications for nucleotide excision repair. Genes Dev. **8:**1344–1355.
27. **Henderson, D. S., D. A. Bailey, D. A. R. Sinclair, and T. A. Grigliatti.** 1987. Isolation and characterization of second chromosome mutagen-sensitive mutations in *Drosophila melanogaster*. Mutat. Res. **177:**83–93.
28. **Henriques, J. A., and E. Moustacchi.** 1980. Isolation and characterization of *pso* mutants sensitive to photo-addition of psoralen derivatives in *Saccharomyces cerevisiae*. Genetics **95:**273–288.
29. **Jachymczyk, W. J., R. C. von Borstel, M. R. Mowat, and P. J. Hastings.** 1981. Repair of interstrand cross-links in DNA of *Saccharomyces cerevisiae* re-

quires two systems for DNA repair: the RAD3 system and the RAD51 system. Mol. Gen. Genet. **182:**196–205.
30. **Joyce, C. M., and T. A. Steitz.** 1994. Function and structure relationships in DNA polymerases. Annu. Rev. Biochem. **63:**777–822.
31. **Kim, Y., S. H. Eom, J. Wang, D. Lee, S. W. Suh, and T. A. Steitz.** 1995. Crystal structure of *Thermus aquaticus* DNA polymerase. Nature (London) **376:**612–616.
32. **Kraulis, P. J.** 1991. MolScript—a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. **24:**946–950.
33. **Langer, S. P., M. Levine, and D. C. Ward.** 1982. Immunological method for mapping genes on Drosophila polytene chromosomes. Proc. Natl. Acad. Sci. USA **79:**4381–4385.
34. **Lawrence, C. E., S. F. Altschul, M. S. Boguski, J. S. Liu, A. F. Neuwald, and J. C. Wootton.** 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. Science **262:**208–214.
35. **Leonhardt, E. A., D. S. Henderson, J. E. Rinehart, and J. B. Boyd.** 1993. Characterization of the *mus308* gene in *Drosophila melanogaster*. Genetics **133:**87–96.
36. **Lepesant, J. A., L. J. Kejzlarova, and A. Garen.** 1978. Ecdysone-inducible functions of larval fat bodies in Drosophila. Proc. Natl. Acad. Sci. USA **75:**5570–5574.
37. **Lindsley, D. L., and G. G. Zimm.** 1992. The genome of *Drosophila melanogaster*. Academic Press, Inc., San Diego, Calif.
38. **Lüthy, R., J. U. Bowie, and E. Eisenberg.** 1992. Assessment of protein models with three-dimensional profiles. Nature (London) **356:**83–85.
39. **Lyamichev, V., M. A. Brow, and J. E. Dahlberg.** 1993. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. Science **260:**778–783.
40. **Magana-Schwencke, N., J. A. Henriques, R. Chanet, and E. Moustacchi.** 1982. The fate of 8-methoxypsoralen photoinduced crosslinks in nuclear and mitochondrial yeast DNA: comparison of wild-type and repair-deficient strains. Proc. Natl. Acad. Sci. USA **79:**1722–1726.
41. **Miller, R. D., L. Prakash, and S. Prakash.** 1982. Genetic control of excision of *Saccharomyces cerevisiae* interstrand DNA cross-links induced by psoralen plus near-UV light. Mol. Cell. Biol. **2:**939–948.
42. **O'Connell, P., and M. Rosbash.** 1984. Sequence, structure, and codon preference of the Drosophila ribosomal protein 49 gene. Nucleic Acids Res. **12:**5495–5513.
43. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
44. **Robins, P., D. J. C. Pappin, R. D. Wood, and T. Lindahl.** 1994. Structural and functional homology between mammalian DNase IV and the 5′-nuclease domain of *Escherichia coli* DNA polymerase I. J. Biol. Chem. **269:**28535–28538.
45. **Ruhland, A., M. Kircher, F. Wilborn, and M. Brendel.** 1981. A yeast mutant specifically sensitive to bifunctional alkylation. Mutat. Res. **91:**457–462.
46. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
47. **Sladek, F. M., M. M. Munn, W. D. Rupp, and P. Howard-Flanders.** 1989. In vitro repair of psoralen-DNA cross-links by RecA, UvrABC, and the 5′-exonuclease of DNA polymerase I. J. Biol. Chem. **264:**6755–6765.
48. **Smith, P. D.** 1976. Mutagen sensitivity of *Drosophila melanogaster*. III. X-linked loci governing sensitivity to methyl methanesulfonate. Mol. Gen. Genet. **149:**73–85.
49. **Strathdee, C. A., A. M. Duncan, and M. Buchwald.** 1992. Evidence for at least four Fanconi anaemia genes including FACC on chromosome 9. Nat. Genet. **1:**196–198.
50. **Strathdee, C. A., H. Gavish, W. R. Shannon, and M. Buchwald.** 1992. Cloning of cDNAs for Fanconi's anaemia by functional complementation. Nature (London) **356:**763–767. (Erratum, **358:**434.)
51. **Tabor, S., and C. C. Richardson.** 1995. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. Proc. Natl. Acad. Sci. USA **92:**6339–6343.
52. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
53. **Van Houten, B., H. Gamper, S. R. Holbrook, J. E. Hearst, and A. Sancar.** 1986. Action mechanism of ABC excision nuclease on a DNA substrate containing a psoralen crosslink at a defined position. Proc. Natl. Acad. Sci. USA **83:**8077–8081.
54. **Vuksanovic, L., and J. E. Cleaver.** 1987. Unique cross-link and monoadduct repair characteristics of a xeroderma pigmentosum revertant cell line. Mutat. Res. **184:**255–263.
55. **Widner, W. R., and R. B. Wickner.** 1993. Evidence that the SKI antiviral system of *Saccharomyces cerevisiae* acts by blocking expression of viral mRNA. Mol. Cell. Biol. **13:**4331–4341.