DOE/ER/60538--T6

FG03-87ER60538

Annual Progress Report
for the period
June 1, 1992 to June 30, 1993

1. Overview. The most interesting discovery we have made over the past year derives from sequence analysis of cDNAs from the putative *mus308* gene. The theoretical translation product of this gene contains a DNA polymerase domain near the carboxy terminus and DNA/RNA helicase motifs near the amino terminus. There is currently no precedent in the literature for a single polypeptide containing both of these domains. The protein appears to be a novel DNA repair enzyme which should be fruitful ground for future enzymological analysis. We have identified two ORFs by sequence analysis of the transforming fragment containing the *mei-41* gene and of corresponding cDNAs. ORF 1 includes the P element insertion sites and encodes a peptide of 757 amino acids. ORF 2 starts 900 base pairs downstream of ORF 1 and encodes a peptide of 1037 amino acids. This putative peptide shows homology to the yeast DNA repair genes, *rad50* of *S. cerevisiae* and *rad3* of *S. pombe*.

2. Structural characterization and functional identity of the *mus308* gene. During the past year we have sequenced cDNAs and genomic clones which we believe comprise the majority of the *mus308* gene. As shown in fig. 1, the open-reading frame extends for 5.6 kb and spans a region of greater than 7 kb in the genomic DNA. It is comprised of 4 small introns and 1 larger intron containing a nested gene oriented in the opposite direction from *mus308*. This small intronic gene has substantial homology to the drosphila heat-shock gene *Hsp-G2*, the function of which is unknown. Transformation of *mus308* embryos with constructs containing the entirety of this small nested gene failed to produce functional complementation of nitrogen mustard sensitivity, and hence this gene is unlikely to have any relationship to *mus308*. However, we have demonstrated that one of our *mus308* mutants contains an insert in the coding region of the intronic gene which eliminates its transcription. Presumably the effect of this insertion on *mus308* function results from disrupted splicing. This is currently under investigation.

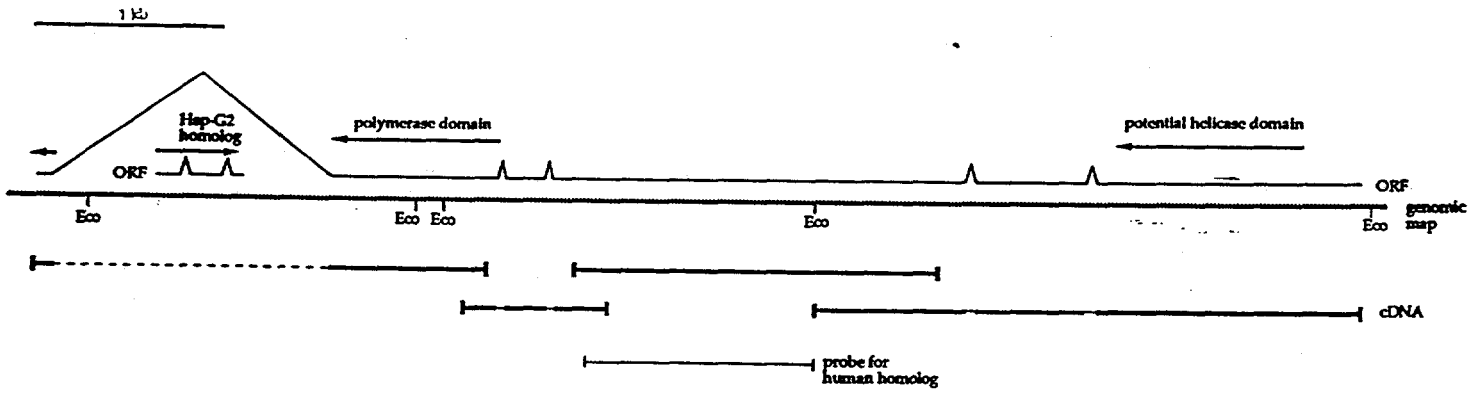

Figure 1. Map of putative *mus308* gene.

DISCLAIMER

The most surprising and exciting result of the sequence analysis is a clue to the function of the *mus308* gene. The predicted protein product from a 720 bp region near the 3' end of the gene posseses striking homology to prokaryotic DNA polymerase I (33% identity, 54% similiarity to pol I from *E. coli*). All of the amino acids identified as critical to polymerase function are extremely well conserved, leaving little doubt that this is a DNA polymerase domain. There is much lower homology to other cloned DNA polymerases, either prokaryotic or eukaryotic, except for DNA polymerase gamma from yeast (16% identity, 41% similarity). The 5' portion of the gene is proving equally interesting. The predicted protein product from a 1030 bp region contains five motifs which are conserved in the family of DNA/RNA helicases. No functional domain encoded by the central region of the gene has yet been identified.

3. **Recovery of human sequences homologous to *mus308*.** Given the phenotypic similarity between *mus308* and the human disorder Fanconi anemia, a primary emphasis of our research in the past year has been to recover the human homolog of *mus308*. Initially we used the pREP4 human cDNA library which Manuel Buchwald's lab used successfully to obtain the gene for the C complementation group of Fanconi anemia. A screen of this library with Drosophila DNA from the region indicated in the figure yielded more than 20 positively hybridizing clones. All of these clones proved to be identical. Unfortunately, they also proved to contain large deletions and rearrangements in the vector, rendering them useless for complementation analysis in Fanconi anemia cell lines. Nonetheless, we now possess a portion of a human gene with substantial homology to the *mus308* gene.

4. **Structural Analysis of *mei-41*.** During the course of this analysis we re-evaluated the restriction map of the lE1 clone and identified an error in the region of fragments "c" and "d". The fragment "c" in the earlier map is replaced by fragment "d" and *vice versa*. Also the orientation of the fragment "c" is reversed such that the HindIII site is now toward the 5' end of the fragment (Fig. 2).

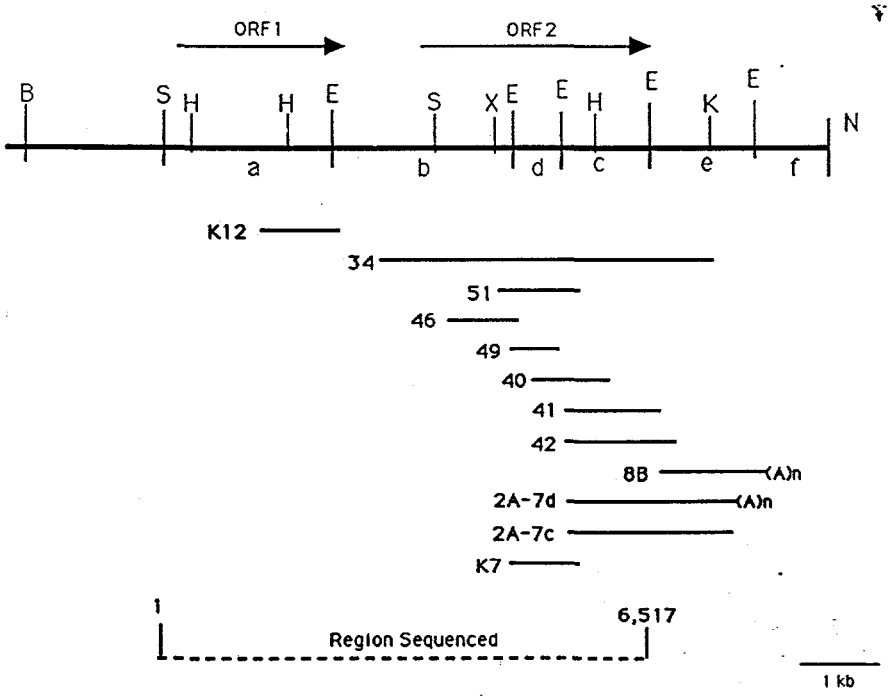

Figure 2. Map of *mei-41* gene.

Analysis of cDNAs selected by screening of two cDNA libraries suggests the presence of two different species of cDNAs. To identify the coding regions corresponding to these cDNAs, we have so far sequenced 6.5 kb of the transforming genomic DNA fragment and also cDNAs. Two open reading frames are identified. ORF 1 starts at nucleotide 235 and spans most of the fragment "a" and the first 375 bp of fragment "b" ending at nucleotide 2506. It includes P element insertion sites and codes for a putative peptide of 757 amino acids. The encoded peptide shows weak homology to the large subunit of ribonucleotide reductase from Vaccinia virus (18% identity and 46% similarity). In the 3' untranslated region of this ORF a stretch of AU rich sequences is present. Such sequences have been implicated in the destabilization of mRNA in several eukaryotic genes. We have isolated a 1.2 kb cDNA (K12) which corresponds to this region.

ORF 2 starts 900 bp downstream of ORF 1, starting at nucleotide 3400, and spans the entire region sequenced, up to nucleotide 6517. Several overlapping cDNAs corresponding to this ORF have been selected. Two of these cDNAs (8B and 2A-7d) contain poly A tails which map 800 bp apart on the genomic map. We have also identified a 4.4 kb cDNA corresponding to ORF 2, which we are currently sequencing. By compairing the restriction pattern of the genomic and overlapping cDNAs, no introns could be detected in this region. This ORF encodes a putative peptide of 1037 amino acids, which shows an overall homology to *rad50* of *S. cerevisiae* (20% identity and 44% similarity) and to *rad3* of *S. pombe* (23% identity and 47 % similarity).

5. Functional Analysis of *mei-41*. To identify the role(s) of the products of these ORFs in DNA repair, we have constructed two truncated molecules. In one, an approximately 1 kb KpnI-NotI fragment is deleted from the 3' end of the original transformation vector (Fig. 2). In the second construct, most of the coding sequences for the ORF 2, flanked by XhoI and KpnI, are removed, leaving the ORF 1 and its 3' sequences intact. Currently we are using these constructs to transform flies. The transformants will be tested for complementation of the *mei-41* defects.

Continuation Proposal
Period: July 1, 1993 to Jun 30, 1994

During the coming year we will fill in all remaining gaps in the *mus308* gene sequence, both genomic and cDNA. To definitively prove that the *mus308* gene encodes the cloned sequence, we will demonstrate one or more of the following: complementation of *mus308* mutants following transformation with the complete genomic sequence; alteration of transcripts in *mus308* mutants; or alterations in the sequence in *mus308* mutants.

The possibility that *mus308* is related to DNA polymerase gamma has led to a collaboration with Dr. Laurie Kaguni at Michigan State University, whose laboratory is engaged in enzymological analysis of this polymerase in Drosophila and is also attempting to clone the gene. She has succeeded in overexpressing the *mus308* polymerase domain in E. coli using one of our cDNAs. Western analysis of the overexpressed protein using antibody to Drosophila gamma polymerase has thus far failed to yield a positive signal. As an additional test, she will use antibodies generated against the fusion protein to determine its subcelluar localization in drosophila. A mitochondrial localization would point to-a role in mitochondrial DNA metabolism , whereas a nuclear localization would suggest that this is a new eukaryotic DNA repair polymerase which may function analogously to pol I in E. coli. The availability of antibody will also provide a means of affinity purification of the native protein from Drosophila.

One of our primary goals during the coming year is to recover a complete cDNA of the human homolog of the *mus308* gene. This effort is currently underway both in this laboratory and in the laboratory of one of our collaborators, Jean-Michel H. Vos at the University of North Carolina at Chapel Hill. If successful, the cDNA will be tested for capacity to complement the hypersensitivity of Fanconi anemia cells to DNA cross-linking agents following transfection. We will also attempt to recover a complete cDNA of the *mus308* gene either by probing additional cDNA libraries or by splicing together our exisiting partial cDNA's.

With respect to the *mei-41* gene, in the comming year we will complete the sequence of the transformation fragment and of the relevent cDNAs. We will also identify the 5' and the 3' ends of these ORFs by nuclease protection assay and by primer extenssion.

Publications

Published within the past year:

Harris, P. V. and J. B. Boyd. 1993. Re-evaluation of excision repair in the *mus304, mus306, and mus308* mutants of Drosophila. Mutation Res. 301: 51-55.

Harosh, I., M. Mezzina, P.V. Harris and J.B. Boyd. 1992. Purification and characterization of a mitochondrial endonuclease from *Drosophila melanogaster* embryos. Eur. J. Biochem. 210: 455-460.

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.

Leonhardt, E.A. and J.B. Boyd. 1993. Identification of a new locus, *mus115*, in *Drosophila melanogaster*. Mutation Res. 301: 121-124.

Submitted

Henderson, D.S., S.S. Banga, T.A. Grigliatti and J.B. Boyd. Mutagen sensitivity and suppression of position-effect variegation result from mutations in mus209, the drosophila gene encoding proliferating cell nuclear antigen (PCNA).