Novel Insertion Mutation in *Caenorhabditis elegans*

DAVID J. EIDE AND PHILIP ANDERSON*

Department of Genetics, University of Wisconsin, Madison, Wisconsin 53706

Received 6 August 1984/Accepted 13 September 1984

The mutation *e1662* is an allele of the *Caenorhabditis elegans unc-54* gene induced with the difunctional alkylating agent 1,2,7,8-diepoxyoctane. *unc-54* encodes the major myosin heavy chain isozyme of body wall muscle cells. Filter-transfer hybridization and DNA sequence analysis show that *e1662* is an insertion of 288 base pairs of DNA within *unc-54*. The inserted DNA is identical to a 288-base pair region of *unc-54* located ca. 600 base pairs from the insertion site. Thus, *e1662* is a displaced duplication. A 14-base pair sequence located at one end of the duplicated segment is found adjacent to the site of insertion. These homologous sequences are juxtaposed head-to-tail by the insertion event. *e1662* thus contains a tandem direct repeat extending across one of its junctions.

Since the discovery of the mutagenicity of nitrogen mustards (2), the varieties of mutational events caused by alkylating agents have been investigated. The interactions of alkylating agents with DNA and the mechanisms of mutation induction have been extensively reviewed (8). These compounds are classified as mono-, di-, or polyfunctional based on their active group number. Several studies have indicated that some di- and polyfunctional agents generate different spectra of mutations than do monofunctional agents (1, 11, 25, 26).

1,2,7,8-Diepoxyoctane (DEO), a difunctional alkylating epoxide, has been shown to induce a high frequency of deletions in *Neurospora crassa* and *Caenorhabditis elegans* (1, 26). Approximately 42% of DEO-induced *ad3* mutants of *Neurospora* and ca. 25% of DEO-induced *unc-54* mutants of *C. elegans* are large deletions. Monofunctional agents such as ethylmethane sulfonate (EMS) induce deletions at much lower frequencies in these organisms. For example, only 1% of ethylmethane sulfonate-induced *Neurospora ad3* mutants (26) and less than 3% of ethylmethane sulfonate-induced *C. elegans unc-54* mutants are deletions (1). This difference has been attributed to the potential of multifunctional alkylating agents to form intra- and interstrand DNA cross-links (17). DNA cross-linkage may result in chromosomal breakage and lead to DNA rearrangements such as deletions and translocations.

In this report, we present the structure of a particularly interesting mutation induced in *Caenorhabditis elegans* by DEO. *C. elegans* contains several electrophoretically distinguishable myosin heavy chains (10, 20). The *unc-54* gene encodes the most abundant isozyme expressed in body wall muscle cells. *unc-54* mutants are paralyzed due to their defective body wall muscle. As part of our studies on mutagenic specificity in *C. elegans* (1), we have examined the structure of DEO-induced mutations at the *C. elegans unc-54* locus. One mutation, *e1662*, is described in this report. This mutation results from the insertion of a 288-base pair segment into the *unc-54* gene. The inserted sequence is a displaced duplication, present in the inverted orientation, of a nearby region of *unc-54*.

MATERIALS AND METHODS

General procedures. The conditions for growth and maintenance of *C. elegans*, the genetic techniques that select

directly for *unc-54* mutants, and the procedure for DEO mutagenesis have been previously described (1, 6). DEO was purchased from Aldrich Chemical Co., Inc., Milwaukee, Wis.

DNA isolation and Southern hybridizations. Nematode cultures were harvested by the method of Sulston and Brenner (31). DNA was extracted essentially as described by Marmur (22), after disruption of the nematodes with proteinase K (BDH Chemicals). Restriction endonucleases were purchased from New England BioLabs and used in accordance with the supplier's instructions. Filter-transfer hybridizations were performed by the method of Southern (30). DNAs were fractionated on 0.8% agarose gels containing a mixture of high and low gelling temperature agaroses (2:1 Sigma types V:VII) in 1× Tris-borate-EDTA gel buffer (21). Radiolabeled probe DNAs for these experiments were prepared by nick translation (28). Recombinant lambda phages SG23/1 and SG25/3 represent 25 kilobases of *C. elegans* genomic DNA that include the *unc-54* gene and extensive flanking sequences (19).

Cloning procedures and DNA sequencing. The *e1662* mutation was cloned as follows. *unc-54(e1152 e1662)* DNA was partially digested with *Mbo*I. The digestion products were size fractionated on NaCl gradients, and 12- to 20-kilobase fragments were collected (29). These fragments were ligated to λ 1059 DNA that had been digested with *Bam*HI, and the products were packaged into virions as described by Hohn (14) and by Karn et al. (16). Hybrid phage carrying *unc-54* sequences were identified by plaque hybridization, using a gel-purified *unc-54* wild-type restriction fragment as probe (4). One phage, designated DE-5, was retained. A 1,381-base pair *Bgl*II-*Eco*RI fragment containing *e1662* was subcloned into pUC9 (32) as follows. DE-5 DNA was double-digested with *Bgl*II and *Eco*RI and size fractionated by agarose gel electrophoresis (0.8% Sigma type VII low gelling temperature agarose in 1× TBE). The 1,381-base pair fragment was purified from the gel by the method of Weislander (34). The fragment was ligated to *Bam*HI-, *Eco*RI-digested pUC9 and transformed into *Escherichia coli* JM83 (13). The recombinant plasmid is designated pDE3. Plasmid DNA was isolated by standard procedures (12).

The DNA sequence of *e1662* was determined by the chemical method of Maxam and Gilbert (23), with minor modifications. The fragment that we sequenced (572 base pairs) was generated by *Hind*III and *Sac*I digestion of pDE3 and isolated after polyacrylamide gel electrophoresis (23).

* Corresponding author.

The *Hind*III site used is present in the pUC9 polylinker. The *Sac*I end (a 3' overhang) was prepared for 5' end labeling by treatment with T4 DNA polymerase (New England BioLabs) (9). The 5' end labeling was performed by using calf intestinal alkaline phosphatase (Boehringer Mannheim Biochemicals), T4 polynucleotide kinase (New England BioLabs), and [α - 32 P]ATP (Amersham Corp.) under conditions recommended by the suppliers.

RESULTS

Isolation of *e1662*. *e1662* was isolated by genetic techniques that select directly for *unc-54* mutants (1). Briefly, this technique is as follows. *unc-54(e1152)* is a strongly dominant allele of *unc-54* (20). Animals heterozygous for *e1152* are paralyzed. The *e1152* allele produces *unc-54* myosin heavy

chains that are defective for assembly. In *e1152/+* heterozygotes, the mutant polypeptide interferes with the process of sarcomere assembly, resulting in muscle disruption and paralysis (24). Mutations that eliminate the synthesis of *e1152* myosin eliminate the dominance of the mutation. Recessive mutations introduced into the *e1152* gene copy are wild type as heterozygotes. These animals are highly motile and can be easily identified among a population of paralyzed siblings. *e1662* was isolated in this way. The double mutant homozygote *unc-54(e1152 e1662)* was subsequently recovered among the self-fertilized progeny of our primary isolate.

Structure of *e1662*. The wild-type *unc-54* gene extends over 7.2 kilobases and contains eight short introns (15). We have analyzed the structure of *e1662* by filter-transfer hybridization techniques. The radiolabeled probe for this experiment included the *unc-54* coding region plus 9 kilobases

FIG. 1. Southern blot analysis of *e1662*. Total genomic DNAs of the wild type and the *unc-54(e1152)* and *unc-54(e1152e1662)* mutants were digested with *Bgl*II and separated on a 0.8% agarose gel. Radiolabeled probe DNA was 1:1 mixture of phages SG23/1 and SG25/3. We have revised slightly the restriction map of MacLeod et al. (19) to reflect more accurately the sizes of certain restriction fragments. The transcribed portion of *unc-54* is indicated above its restriction map. The 2.4-kilobase novel fragment observed in *unc-54(e1152e1662)* (lane 3) is indicated by an arrow.

of flanking sequences from both the 5' and 3' ends of the gene. Figure 1 shows total genomic Southern blots of *unc-54(e1152e1662)* and control strains. The restriction fragments of *unc-54(e1152)* are indistinguishable from the wild type (Fig. 1, lanes 1 and 2). *e1152* was induced with ethylmethane sulfonate and is probably a point mutation. The restriction fragments of *unc-54(e1152e1662)* (Fig. 1, lane 3) are abnormal. The wild-type, 2.1-kilobase *Bgl*II fragment is replaced by a novel fragment of ca. 2.4 kilobases. Thus, *e1152e1662* contains additional DNA within the 2.1-kilobase *Bgl*II fragment.

A large restriction fragment containing the *unc-54(e1152e1662)* gene was cloned into λ 1059 after partial digestion of *e1152e1662* DNA with *Mbo*I. A 1.4-kilobase fragment containing *e1662* was then subcloned into pUC9. The resultant plasmid, designated pDE3, was digested ap-

propriately, and a 572-base pair fragment that contains *e1662* was purified by gel electrophoresis. To determine the origins of the extra DNA in *e1662*, this fragment was radiolabeled and used as the probe of total genomic Southern blots. Figure 2 shows the results. Since the probe for this experiment contains the *e1662* mutation, homologies to the additional DNA present in *e1662* will be revealed. The results demonstrate that *e1662* is an insertion mutation and that the inserted sequences are derived from *unc-54* itself. The critical data are shown in Fig. 2, lane 3. As expected, the 2.1-kilobase *Bgl*II fragment is ca. 300 base pairs larger in *e1152e1662*. However, the *e1662* probe also hybridizes to the adjacent 3.9-kilobase *Bgl*II fragment. Thus, the DNA insertion of *e1662* is located within the 2.1-kilobase *Bgl*II fragment but is derived from the nearby 3.9-kilobase *Bgl*II fragment.

FIG. 2. Southern blot analysis of *e1662*. Total genomic DNAs of wild-type and *unc-54(e1152e1662)* animals were digested with *Bam*HI or *Bgl*II and separated on a 0.8% agarose gel. The radiolabeled DNA probe was a 572-base pair fragment isolated from pDE3. This fragment contains the *e1662* mutation. Pertinent restriction sites and fragment sizes are indicated. The organization of *unc-54* coding sequences is shown above its restriction map. Boxed regions are exons; the 3' untranslated region is unfilled. The limits of *unc-54* shown here are the AUG translational initiation codon (5') and the AAUAAA polyadenylation signal (3'). The novel *e1662* *Bam*HI and *Bgl*II fragments are indicated by arrows.

```

wild-type: 7424-aagattcacatggaatagactgaaacacagattgcaatattctaaactacatgaatggcttggatttctccctcgagctttctcttggcagaagtcgaagctgctgagcgtt-7305
e1662:      AACCTGACATTCATCATCATCTATTATTAATTCAGGAGGTCATATCTAAACCTTACATGAATGGCTTGGATTCTCCCTCGAGCTTTCTCTTGGCAGAAGTCGAAGCTGCTGAGCAGCTT
wild-type: 7937-AACCTGACATTCATCATCTATTATTAATTCAGGAGG-7972
.....

wild-type: 7304-GGCTTCGTTGGCTTGATCAGAGCATCGGGGCTTCATACTCGGCTTCTCTCGGGCTCTCTCGGGCTGCTTGGCAACCAAAAGTTCTTCTTCTCGGATTGAAGAAGAGTGCGCG-7185
e1662:      GGCTTCGTTGGCTTGATCAGAGCATCGGGGCTTCATACTCGGCTTCTCTCGGGCTCTCTCGGGCTGCTTGGCAACCAAAAGTTCTTCTTCTCGGATTGAAGAAGAGTGCGCGG
wild-type:
.....

wild-type: 7184-CTTCTCGGGCTTGAAGAAATTGTTACGGGTGTCGGCTCATTGCTGCTCTCTCGACTTGCATTTGCAACTTCGGGACTTGTGctcttggtatcttcaagttcttctgggcacgcgc-7065
e1662:      CTTCTCGGGCTTGAAGAAATTGTTACGGGTGTCGGCTCATTGCTGCTCTCTCGACTTGCATTTGCAACTTCGGGACTTGTGCTCGGCAACTTGAACCTCCAGAACTACAAGCAACTC
wild-type:
.....
7973-aactcgccaacttgaacctccagaaagtacaaagcaactc-8010

```

FIG. 3. Sequence analysis of *e1662*. The nucleotide sequence of *e1662* is shown on the middle line of each panel. It is aligned with portions of the wild-type *unc-54* sequence (15) to show both the site of insertion and the source of inserted DNA. The duplicated segment (indicated by vertical lines) extends from nucleotide 7101 through 7388 of the wild-type sequence. This sequence is shown in the upper line of each panel as capital letters; sequences surrounding the duplicated segment are lower case. In *e1662*, this 288-base pair segment is inserted after wild-type nucleotide 7972. Sequences that surround the site of insertion in the wild-type gene are shown in the lower line of each panel. The AA dinucleotide that is deleted in *e1662* is underlined and shown as lower case. The sequence homologies contributing to the tandem direct repeat are indicated by arrows.

We have determined the nucleotide sequence of *e1662* and compared it with the sequence of the wild-type *unc-54* gene (15). The results confirm the conclusions described above and indicate precisely the nature of the *e1662* insertion. The DNA sequence of *e1662* is shown in Fig. 3 and summarized in Fig. 4. Several features of *e1662* are noteworthy. (i) *e1662* results from an insertion of 288 base pairs of DNA. (ii) The inserted DNA is identical to 288 base pairs of *unc-54* located ca. 600 base pairs away from the site of insertion. *e1662*, therefore, is a displaced duplication. (iii) The displaced duplication is present in an inverted orientation relative to its source. (iv) Two base pairs of target DNA adjacent to the site of insertion are deleted in *e1662*.

We have searched for DNA sequence homologies among the breakpoints involved in generating *e1662* by using a computer-aided DNA sequence analysis program (5; F. Blattner and J. Schroeder, personal communication). A 14-base pair sequence located at one end of the duplicated segment is closely related to 14 base pairs adjacent to the site of insertion (12/14 base pair homology; Fig. 3). These sequences are juxtaposed head-to-tail by the insertion event. *e1662* thus contains a tandem direct repeat extending across one of its two junctions. The tandemly repeated sequence is actually 12 base pairs in length since the two base pairs deleted in *e1662* are part of the 14-base pair sequence.

DISCUSSION

e1662 was isolated after mutagenesis with DEO, a difunctional alkylating agent. DEO is a potent mutagen in *C. elegans*, increasing the frequency of *unc-54* mutations greater than 200 times the background spontaneous frequency (1; Eide and Anderson, Genetics, in press). Although *e1662* was isolated only once, it is reasonable to assume that it was generated by the mutagenic properties of DEO.

The most striking feature of *e1662* is the tandem direct repeat generated at one junction of the insertion (Fig. 3 and 4). Unfortunately, this homology does not suggest an obvious mechanism for the formation of *e1662* by homologous recombination. Homologous recombination between these repeated sequences in the wild-type gene does not generate the structure found at that breakpoint of *e1662*. Nevertheless, the homology seems too striking to be insignificant. How was *e1662* formed? Errors of replication or recombination seem plausible. Multifunctional alkylating agents cause intra- and interstrand cross-links in DNA (7, 17), chromosomal rearrangements (1, 11, 26), and chromosomal breaks (11, 25). Free ends of DNA molecules are highly recombinogenic in yeast (27). A free end generated within *unc-54* near the duplicated segment might undergo strand invasion into a homologous region of another *unc-54* gene copy. This may be responsible for DNA associations that ultimately resulted in the direct repeat observed at one junction of *e1662*. Unfortunately, it is not possible to be more precise mechanistically.

e1662 is similar in structure to the mutations *Cbx*¹ and *pbx*¹ of the *Drosophila bithorax* complex (3, 18). These mutations arose together after X-ray mutagenesis. *pbx*¹ is associated with a 17-kilobase deletion. *Cbx*¹ has an insertion of the same 17-kilobase segment at a nearby site in the *bithorax* complex. Since these mutations were recovered on the same chromosome, they were generated by intra-chromosomal events. We have not recovered a deletion that corresponds to the material duplicated in *e1662*. It is likely, therefore, that *e1662* arose after interactions between sister

FIG. 4. Summary of *el662*. *el662* is the result of an insertion of 288 base pairs into the *unc-54* gene. The inserted segment is a displaced duplication, present in the inverted orientation, of a nearby region of *unc-54*. The positions of the inserted and duplicated sequences relative to the protein coding regions of *unc-54* are diagrammed. Open boxes indicate exons. The cross-hatched box represents the 3' untranslated region. The relative orientations of the duplicated segment are indicated by long arrows. The locations and orientations of the sequence homologies contributing to the tandem direct repeat are indicated by small arrows.

chromatids or chromosome homologs. It is interesting to note that the *bithorax* complex mutations were isolated after X-ray mutagenesis. Ionizing radiation causes double-strand breaks and DNA cross-links (33). Therefore, the *Cbx¹-pbx¹* double mutant and *el662* may have arisen by similar mechanisms.

ACKNOWLEDGMENTS

We thank Amy Bejsovic, Claudia Cummins, and Timothy Schedl for critical reading of the manuscript.

This work was supported by Public Health Service grant GM30132 from the National Institutes of Health and by a University of Wisconsin Training grant in Cell and Molecular Biology.

LITERATURE CITED

- Anderson, P., and S. Brenner. 1984. A selection for myosin heavy chain mutants in the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U.S.A. **81**:4470-4474.
- Auerbach, C., and J. M. Robson. 1947. The production of mutations by chemical substances. Proc. R. Soc. Edinburgh Sect. B **62**:271-283.
- Bender, W., M. Akam, F. Karch, P. A. Beachy, M. Peifer, P. Spierer, E. B. Lewis, and D. S. Hogness. 1983. Molecular genetics of the *bithorax* complex in *Drosophila melanogaster*. Science **221**:23-29.
- Benton, W. D., and R. W. Davis. 1977. Screening λ gt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180-182.
- Blattner, F. R., and J. L. Schroeder. 1984. A computer package for DNA sequence analysis. Nucleic Acids Res. **12**:615-617.
- Brenner, S. 1974. The genetics of *Caenorhabditis elegans*. Genetics **77**:71-94.
- Brookes, P., and P. D. Lawley. 1964. Alkylating agents. Br. Med. Bull. **20**:91-95.
- Calabresi, P., and R. E. Parks, Jr. 1980. Antiproliferative agents and drugs used for immunosuppression, p. 1256-1313. In A. G. Gilman, L. S. Goodman, and A. Gilman (ed.), Pharmacological basis of therapeutics. Macmillan Publishing Co. Inc., New York.
- Challberg, M. D., and P. T. Englund. 1980. Specific labeling of 3' termini with T4 DNA polymerase. Methods Enzymol. **65**:39-43.
- Epstein, H. F., R. H. Waterson, and S. Brenner. 1974. A mutant affecting the heavy chain myosin of *Caenorhabditis elegans*. J. Mol. Biol. **90**:291-300.
- Fahmy, O. G., and M. J. Fahmy. 1960. Cytogenetic analysis of the action of carcinogens and tumor inhibitors in *Drosophila melanogaster*. X. The nature of the mutations induced by the mesyloxy esters in relation to molecular cross-linkage. Genetics **46**:447-458.
- Godson, G. N., and D. Vapnek. 1973. A simple method of preparing large amounts of ϕ X174 RFI supercoiled DNA. Biochim. Biophys. Acta **299**:516-520.
- Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557-580.
- Hohn, B. 1979. *In-vitro* packaging of lambda and cosmid DNA. Methods Enzymol. **68**:299-309.
- Karn, J., S. Brenner, and L. Barnett. 1983. Protein structural domains in the *Caenorhabditis elegans unc-54* myosin heavy chain gene are not separated by introns. Proc. Natl. Acad. Sci. U.S.A. **80**:4253-4257.
- Karn, J., S. Brenner, L. Barnett, and G. Cesareni. 1980. Novel bacteriophage λ cloning vector. Proc. Natl. Acad. Sci. U.S.A. **77**:5172-5176.
- Lawley, P. D., and P. Brookes. 1963. The action of alkylating agents on deoxyribonucleic acid in relation to biological effects of the alkylating agents. Exp. Cell Res. **9**(Suppl.):512-520.
- Lewis, E. B. 1954. The theory and application of a new method for detecting chromosomal rearrangements in *Drosophila melanogaster*. Am. Nat. **88**:225-239.
- MacLeod, A. R., J. Karn, and S. Brenner. 1981. Molecular analysis of the *unc-54* myosin heavy chain gene of *Caenorhabditis elegans*. Nature (London) **291**:386-390.
- MacLeod, A. R., R. H. Waterston, R. M. Fishpool, and S. Brenner. 1977. Identification of the structural gene for a myosin heavy chain in *Caenorhabditis elegans*. J. Mol. Biol. **114**:133-140.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Marmur, J. 1961. A procedure for the isolation of deoxyribonucleic acid from micro-organisms. J. Mol. Biol. **3**:208-218.
- Maxam, A., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499-559.
- Moerman, D. G., S. Plurad, R. H. Waterston, and D. L. Baillie. 1982. Mutations in the *unc-54* myosin heavy chain of *Caenorhabditis elegans* that alter contractility but not muscle structure. Cell **29**:773-781.
- Nawar, M. M., C. F. Konzak, and R. A. Nilan. 1971. Comparative studies of the biological effectiveness of nitrogen mustards, ethyleneimines and γ -rays. Mutat. Res. **11**:339-346.
- Ong, T., and F. J. DeSerres. 1975. Mutation induction by difunctional alkylating agents in *Neurospora crassa*. Genetics **80**:475-482.
- Orr-Weaver, T. L., J. W. Szostak, and R. W. Rothstein. 1981. Yeast transformation: a model system for the study of recombination. Proc. Natl. Acad. Sci. U.S.A. **78**:6354-6358.
- Rigby, P. W. J., M. Diekmann, C. Rhodes, and P. Berg. 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro*

- by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
29. **Slightom, J. L., A. E. Blechl, and O. Smithies.** 1980. Human fetal γ - and α -globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21**:627–638.
30. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments by gel electrophoresis. J. Mol. Biol. **98**:503–517.
31. **Sulston, J. E., and S. Brenner.** 1974. The DNA of *Caenorhabditis elegans*. Genetics **77**:95–104.
32. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
33. **Ward, J. F.** 1975. Molecular mechanisms of radiation-induced damage to nucleic acids. Adv. Radiat. Biol. **5**:181–239.
34. **Weislander, L.** 1979. A simple method to recover intact high molecular weight RNA and DNA after electrophoresis in low gelling temperature agarose gels. Anal. Biochem. **98**:305–309.