MOLECULAR AND CELLULAR BIOLOGY, Sept. 1988, p. 3748–3754
0270-7306/88/093748-07$02.00/0

Vol. 8, No. 9

Structures of Spontaneous Deletions in *Caenorhabditis elegans*

ROCK A. PULAK AND PHILIP ANDERSON*

Department of Genetics, University of Wisconsin, 445 Henry Mall, Madison, Wisconsin 53706

Received 16 March 1988/Accepted 26 May 1988

We have investigated the structural features of spontaneous deletions in *Caenorhabditis elegans*. We cloned and sequenced the junctions of 16 spontaneous deletions affecting the *unc-54* myosin heavy-chain gene and compared their sequences with those of the wild type. We analyzed these sequences in an attempt to identify structural features of the gene that are consistently involved in the spontaneous deletion process. Most deletions (15 of 16) removed a single contiguous region of DNA, with no nucleotides inserted or rearranged at the deletion junctions; one deletion was more complex. *unc-54* deletions were small, averaging 600 base pairs in length, and were randomly distributed throughout the gene. Unlike deletions that occur in *Escherichia coli*, spontaneous *unc-54* deletions did not contain statistically significant direct or inverted repeats at or near their termini. Except for their small average size, we have not identified any distinguishing features of their sequence or structure. We discuss these results with regard to the mechanisms for spontaneous deletion in eucaryotic and procaryotic cells.

Our understanding of the mechanisms of spontaneous deletion is based primarily on the properties of deletions in *Escherichia coli*. The termini of most spontaneous deletions in *E. coli* are located within sequences of the wild-type gene that constitute short direct repeats, usually about 5 to 10 base pairs long (1, 14, 54). Material between the direct repeats is deleted, and the resulting junction contains a single copy of what originally was repeated. The association of direct sequence repeats with deletion termini in procaryotes is very striking, and it suggests that during the deletion process intermediate structures form in which one strand of a repeat base pairs with its complement at the second repeat (e.g., "slipped mispairing" [17]). Repair or replication of such structures might then eliminate all sequences between the repeats.

Two lines of evidence indicate that base pairing between direct repeats is important for deletion formation in *E. coli*: (i) mutations that reduce homology between repeats lower the frequency of the corresponding deletions (1); and (ii) many deletions occur at reduced frequencies in *recA* mutants compared with $recA^+$ strains (1). These results suggest that homologous pairing, promoted by $recA^+$ function, is necessary for generating many deletions. Not all deletions are $recA^+$ dependent (15, 16, 27, 41, 57), however, and the termini of *recA*-independent deletions are also located within direct repeats (1). Thus, $recA^+$ function facilitates but is not essential for the deletion process.

Inverted repeats or "quasi-palindromes" (19) are associated with the endpoints of some *E. coli* deletions. Base pairing between inverted repeats or between both direct and inverted repeats simultaneously may align and stabilize structures that are intermediates in the deletion process. Deletions probably occur when such structures are repaired or used as a template for DNA replication. The frequency of a deletion increases as the length of its terminal inverted repeat increases (15), suggesting that stability of a base-paired intermediate influences its fate.

Spontaneous deletions in eucaryotes have not been studied as systematically as those in *E. coli*. The termini of deletions affecting a variety of different eucaryotic genes and organisms have been determined. Although superficially similar to those that occur in *E. coli*, their sequence features are much less striking. Nalbantoglu et al. (43, 44) showed that the termini of seven spontaneous deletions affecting a mammalian *aprt* gene are located within direct repeats of 2 to 5 base pairs. Deletions responsible for certain β-thalassemias, α-thalassemias, and hereditary persistence of fetal hemoglobin occasionally but inconsistently exhibit such repeats (21, 28, 33, 45, 60). Excision of simian virus 40 from the chromosome of its host is formally a deletion process, and the crossover points of several such excisions are located within direct repeats of 2 to 3 base pairs (9); in this case, the repeats may be related to sites of cleavage by DNA topoisomerase I (8). Certain deletion termini in humans and yeasts are associated with transposon-like Alu or delta sequences (21, 23, 31, 32, 42, 45, 51), but the role of these elements in the deletion process is unclear.

We have investigated the properties of spontaneous deletions in the nematode *Caenorhabditis elegans*. Spontaneous mutations affecting the *unc-54* myosin heavy-chain gene are easily selected (47), and many of these mutations are small deletions (12, 13). We cloned 16 independent deletion alleles of *unc-54* and determined the DNA sequences at the junction of each mutation. By comparing these sequences with those of the wild-type gene (29), we have deduced the breakpoints of each deletion. We report here the DNA sequences of these deletions and an analysis of their structural features. We discuss these results with regard to the mechanisms for formation of spontaneous deletions in *C. elegans*.

MATERIALS AND METHODS

General procedures. The conditions for growth, maintenance, and genetic manipulation of *C. elegans* are described by Brenner (7). Nematode cultures from which we prepared DNA were grown in petri dishes containing NGM medium (7) solidified with 1% agarose rather than agar. Isolation of spontaneous *unc-54* mutants and identification of deletion alleles were as described by Eide and Anderson (12, 13). We have renamed two *unc-54* mutations that were previously described. *unc-54(r102)* and *unc-54(r244)* each contain two separate deletions that affect *unc-54* (12; see below). We renamed these alleles *unc-54(r856r857)* and *unc-54 (r858 r859)*, respectively. The altered nomenclature more accu-

* Corresponding author.

rately indicates that they are double mutants and allows us to discuss independently each of the deletions.

Cloning and sequencing of deletions. Nematode cultures were harvested as described by Sulston and Brenner (59). DNA was extracted essentially as described by Marmur (34) after disruption of the animals with proteinase K. Most deletions were cloned by digesting mutant DNAs with *Sal*I and ligating the products to bacteriophage lambda EMBL3 DNA (18) that had been digested with both *Sal*I and *Bam*HI. The complete *unc-54* gene is contained within a *Sal*I restriction fragment of approximately 18.5 kilobase pairs. Ligated DNAs were packaged into virions (22), and phage carrying the *unc-54* region were identified by plaque hybridization (5) with an *unc-54*-specific hybridization probe. A bacteriophage lambda clone of *unc-54(r102)* was described previously (12).

We analyzed the restriction maps of each lambda clone and identified a small restriction fragment that contained the deletion junction. These fragments were purified after electrophoresis on polyacrylamide gels, and they were subcloned into the DNA sequencing vector M13mp10 or M13mp11 (37). We sequenced each clone by the method of Sanger et al. (53).

RESULTS

Origins of spontaneous *unc-54* deletions. Eide and Anderson (12, 13) described the isolation and molecular analysis of 114 spontaneous *unc-54* mutants, which occur in *C. elegans* variety Bristol at a frequency of approximately 3×10^{-7}. Deletions were identified among these mutations because they exhibited abnormalities on total genomic Southern blots. Such methods identify most or all deletions affecting 50 or more base pairs of DNA. The genetic methods that were used to isolate *unc-54* deletions required that they be viable when homozygous, thus limiting the maximal size of recoverable deletions. We know, however, that at least 17 kilobases of DNA spanning the *unc-54* region are nonessential (see below). Thus, we were able to detect in our screen most or all *unc-54* deletions whose lengths were between 50 base pairs and 17 kilobases. Approximately 15% of spontaneous *unc-54* mutations isolated in *C. elegans* variety Bristol are deletions of this type (12).

DNA sequences at deletion termini. We cloned the *unc-54* gene from 14 independent deletion mutants into bacteriophage lambda vectors. We determined a restriction map for each clone, identified a small restriction fragment that contained each deletion junction, subcloned these junction fragments into bacteriophage M13 vectors, and determined their DNA sequences. We then compared these sequences to those of the wild-type *unc-54* gene (29) and deduced the endpoints of each deletion.

Our analysis (24) (see below) indicated that two mutants, previously designated as *unc-54(r102)* and *unc-54(r244)*, each contained two separate deletions. We renamed these mutants *unc-54(r856r857)* and *unc-54(r858r859)*, respectively. The altered nomenclature more accurately indicates that they are double mutants and allows us to discuss independently each of the deletions. We sequenced both deletions for each of these double mutants. Inspection of their sequences predicts that any of these four deletions would eliminate *unc-54* gene function. The occurrence of two separate deletions in each of two mutant strains indicates that a minority of spontaneous deletions arise from concerted events, in which multiple breakages and rejoinings occur in unison.

The sizes and locations within *unc-54* of spontaneous deletions are diagrammed in Fig. 1. The DNA sequences at their termini are shown in Fig. 2 and summarized in Table 1. Deletions ranged in size from 38 to 2,631 base pairs. Most deletions (14 of 16) were simple, indicating that (i) a single contiguous region of DNA was deleted, (ii) nucleotides were not inserted or rearranged at the deletion junctions, and (iii) when short repeats were present at deletion termini the breakpoints occurred at equivalent positions within the repeats. Thus, except for the presence of short repeats at the termini of certain deletions (see below), the points of breakage and rejoining for these 14 deletions were unambiguous.

Two of 16 deletions were more complex. (i) The termini of deletion *unc-54(r260)* were both located within the sequence 5′-AAGC-3′, and the deletion junction exhibited the sequence 5′-AAAGC-3′. We interpret this to indicate that the deletion termini of *r260* occurred within a 4-base-pair direct repeat, but that the points of breakage within these repeats were not at equivalent positions. Rather, the breakages occurred at nonequivalent positions within or near the two A residues, such that the resulting junction contained three A residues instead of two. An alternative interpretation for *r260* is that its termini were located within a 1-base-pair direct repeat (wild-type nucleotides 7183 and 7254), and the resulting deletion was simple. (ii) The deletion junction of *unc-54(r857)* contained 18 additional nucleotides that were not found in the wild-type gene near either deletion endpoint. Fifteen of these inserted nucleotides constituted a small, displaced duplication; they were identical to a 15-nucleotide region of the wild-type gene that was about 600 base pairs away from the *r857* deletion junction (nucleotides 5158 through 5172). This distant region was not affected by *r857*. Three additional A residues were inserted adjacent to the displaced duplication; two were inserted on the 5′ side, and one was inserted on the 3′ side. Thus, at least three events occurred during formation of the *r857* junction: the deletion, the displaced duplication, and the insertion of additional A residues. We cannot establish the sequence of these events, and we have excluded *r857* from our analysis of sequence repeats at deletion endpoints. The complex structure of *r857* is similar to those observed infrequently but consistently at other sites of deletion or nonhomologous recombination in eucaryotic cells (3, 20, 28, 33, 45, 56, 61).

Spontaneous deletions are small. The average size of our 16 deletions was 600 ± 183 base pairs (mean ± standard error of the mean). The genetic methods used to isolate deletions required that they be viable when homozygous. This limited the maximal size of detectable deletions. The following observations, however, indicate that at least 17 kilobases of DNA in the *unc-54* region are nonessential. (i) The deletion *unc-54(r259)* removes *unc-54* plus at least 9 kilobases of DNA beyond the *unc-54* 5′ end (12). *unc-54(r259)* is viable when homozygous. (ii) The DNA sequences of approximately one kilobase beyond the *unc-54* 3′ end are known (29). Other than *unc-54*, no protein-coding sequences are present in this region. (iii) The *unc-54* gene itself is 7.4 kilobases long, and it is nonessential. We conclude that at least 17 kilobases of DNA covering the *unc-54* region can be deleted, and the resulting mutant would be viable when homozygous. This number is an underestimate, and the true nonessential region could be much larger. If deletion endpoints were completely random within this 17-kilobase region, we calculate that the average size of deletions that include all or part of *unc-54* would be 5.5 kilobases. This figure is 7 standard deviations larger than the observed

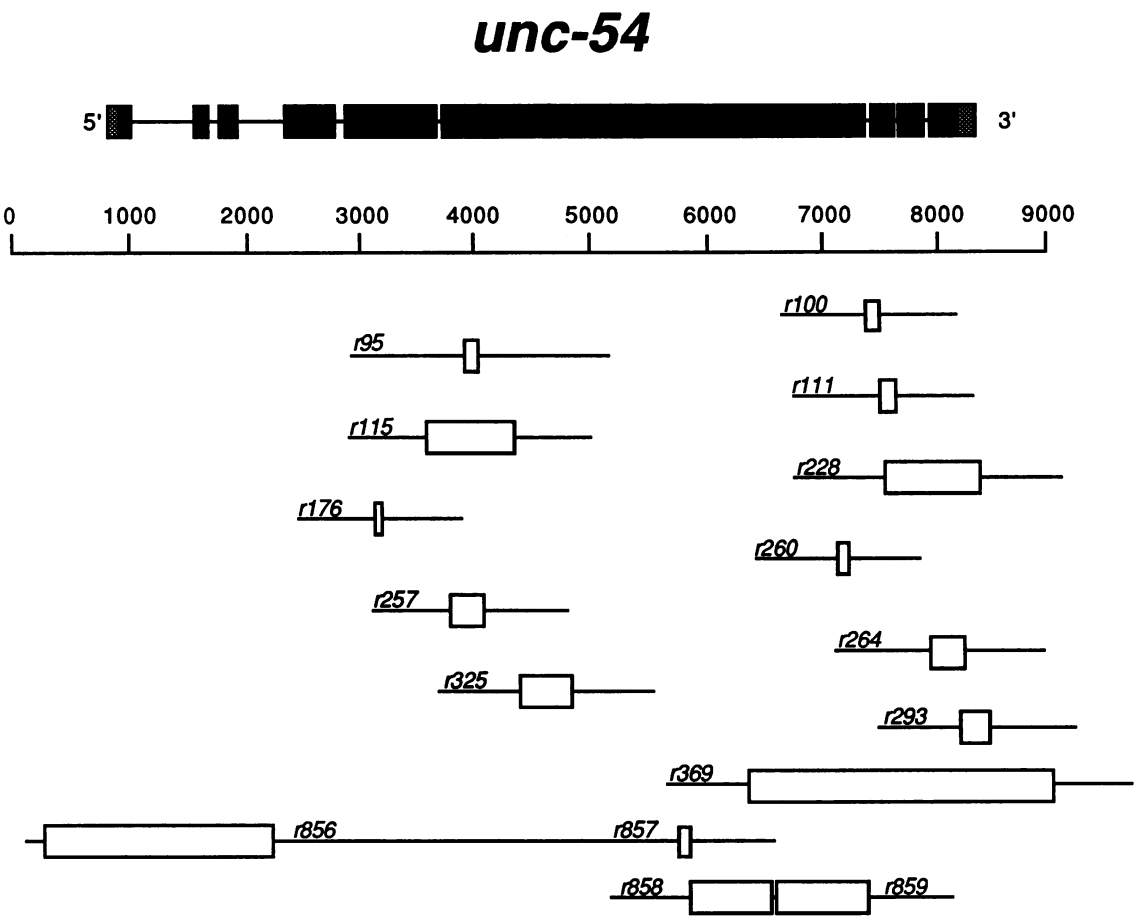

FIG. 1. Spontaneous deletions affecting *unc-54*. The organization of *unc-54*-coding regions is shown at the top. Filled boxes are exons; the 5'- and 3'-untranslated regions are stippled. The DNA sequence coordinates (in base pairs) established by Karn et al. (29) are shown below the gene. The size and position of each deletion are individually diagrammed as open boxes.

average size. We conclude that the small size of spontaneous deletions is statistically significant.

DNA sequence repeats at deletion termini are not statistically significant. Direct DNA sequence repeats at the termini of *unc-54* deletions are summarized in Table 1. Direct repeats ranged in size from 0 to 4 base pairs. A 0-base-pair repeat describes a deletion for which there is no ambiguity concerning which phosphodiester bonds were broken and rejoined to form the deletion junction. A repeat of 1 base pair or larger causes ambiguity concerning where the breakages and reunions occurred. For example, with a 1-base-pair repeat the DNA strand breakages could occur either on the 5' side or on the 3' side of the repeated nucleotide. Since nucleotides are not inserted at the junction, both breakages must occur on the same side of the repeat. Similar ambiguities apply to deletions having 2-, 3-, or 4-base-pair repeats at their termini.

We tested whether the DNA sequence repeats at deletion termini were statistically significant. Since the complete DNA sequence of *unc-54* is known (29), we calculated the probability that two randomly chosen breakpoints would occur within a direct repeat. If the observed numbers and sizes of direct repeats are not statistically different from those predicted for random endpoints, we must conclude that DNA sequence repeats are not significant for the deletion process.

A formula that describes predicted repeat sizes for two randomly chosen points in DNA is $P_k = (k + 1)(f_k)(1 - f_1)^2$, where: P_k is the probability that any two randomly chosen points will fall within a direct repeat of length k, k is the length of the repeat, and f_k is the sum of the squares of the frequencies of oligonucleotides that have length k. For example, f_1 equals the sum of the squares of the four mononucleotide frequencies, f_2 equals the sum of the squares of the 16 dinucleotide frequencies, f_3 equals the sum of the squares of the 64 trinucleotide frequencies, etc.

We have used this equation together with the known frequencies of *unc-54* mononucleotides, dinucleotides, trinucleotides, etc., to calculate the sizes of repeats that would be expected among 15 deletions having random endpoints. Calculated and observed distributions of sequence repeats are shown in Table 2. We used the chi-square test to compare the expected and observed distributions (Table 2). The probability that the expected and observed distributions are equivalent equals 0.28. We conclude by this test that the direct sequence repeats present at the endpoints of our deletions are not statistically significant.

Inverted repeats at the termini of spontaneous *unc-54* deletions range from zero to two base pairs long. Similar calculations indicate that they are also not statistically significant.

Are direct and inverted repeats of the size typical for procaryotic deletions present within *unc-54*? We searched the wild-type *unc-54* DNA sequence for all pairs of direct or

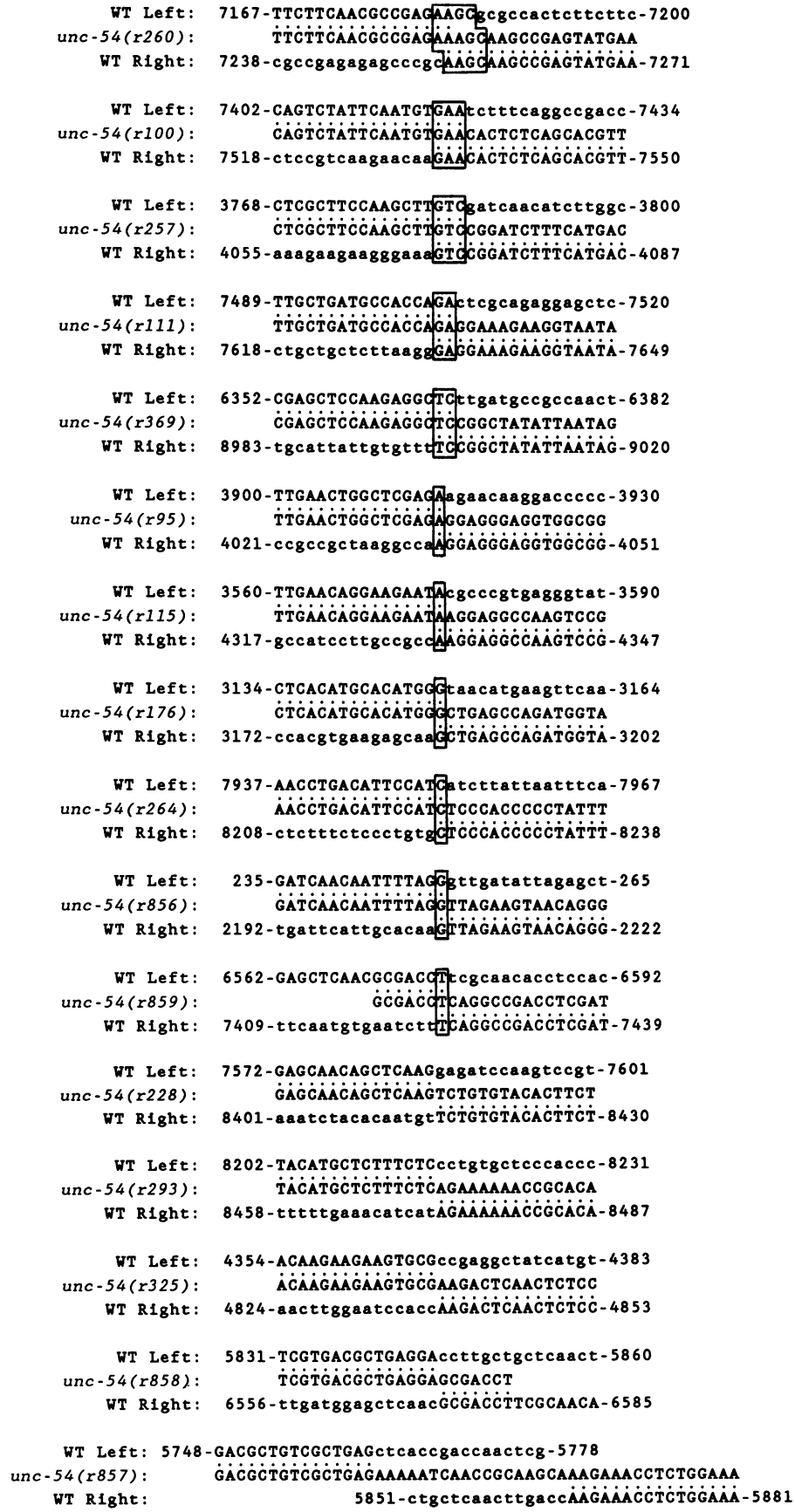

FIG. 2. DNA sequences of 16 spontaneous *unc-54* deletions. The DNA sequences of 16 spontaneous deletions are aligned with the sequences of the wild-type gene at each endpoint. Coordinates are those of Karn et al. (29). Direct repeats, when present at the crossover points, are boxed. Left and right refer to the end of each deletion as drawn in Fig. 1.

TABLE 1. Endpoints of spontaneous deletions affecting *unc-54*[a]

Allele	Deletion endpoints		Size of deletion (bases)	Direct repeat	
	5′ end	3′ end		Length (bases)	Sequence
unc-54(r260)	7183	7253	71	4	AAGC
unc-54(r100)	7417	7532	116	3	GAA
unc-54(r257)	3783	4069	287	3	GTC
unc-54(r111)	7504	7632	129	2	GA
unc-54(r369)	6365	8995	2,631	2	TC
unc-54(r95)	3918	4035	118	1	A
unc-54(r115)	3575	4331	757	1	A
unc-54(r176)	3149	3186	38	1	G
unc-54(r264)	7952	8222	271	1	C
unc-54(r856)	250	2206	1,957	1	G
unc-54(r859)	6577	7423	847	1	T
unc-54(r228)	7587	8415	829	0	
unc-54(r293)	8217	8472	256	0	
unc-54(r325)	4369	4838	470	0	
unc-54(r858)	5846	6570	725	0	
unc-54(r857)[b]	5763	5865	103		

[a] The nucleotide positions are those for the wild-type *unc-54* gene (29). For each deletion, the 5′ and 3′ endpoints indicate the first and last deleted bases, respectively. When direct repeats are located at deletion endpoints, the first and last deleted bases are uncertain. In these cases, the nucleotide positions indicate the nucleotides of each repeat that are closest to the 5′ end of *unc-54*.

[b] *r857* is a complex deletion, and we have not considered its repeat length (see text).

inverted repeats that were eight nucleotides or longer, a size typical for deletions in *E. coli*. We are confident that most or all deletions greater than 50 base pairs in length and having at least one endpoint within a protein-coding region of the gene would be detected as an *unc-54* mutant. There were 1,388 pairs of direct repeats and 685 pairs of inverted repeats within *unc-54* that satisfied these criteria. Repeats as long as 15 nucleotides were present. Thus, a very large number of direct and inverted repeats was present within *unc-54*, yet none of the repeats was associated with deletion endpoints.

Positions of deletions within *unc-54* are random. If "hotspots" or regions especially prone to deletion were present in *unc-54*, then the distribution of deletion breakpoints would be nonrandom. We analyzed the distribution of deletion breakpoints within *unc-54* and found them to be random. We divided the *unc-54* gene into eight conceptual intervals of equal size, and we used the chi-square test to compare the number of endpoints within each interval with that expected for a uniform distribution. We considered the 5′ and 3′ endpoints separately, because the small size of deletions would necessarily cause total endpoints to cluster as pairs. Considering 5′ termini, χ^2 equals 6.7 ($P = 0.46$); considering 3′ termini, χ^2 equals 4.5 ($P = 0.72$). We conclude that the observed distribution of endpoints within the gene is random.

DISCUSSION

Spontaneous deletions in *C. elegans* are unlike those that occur in *E. coli*. Direct or inverted sequence repeats are a striking feature of deletion termini in *E. coli*, but the direct sequence repeats at the termini of *C. elegans* deletions are not statistically significant (Table 2). The mechanism(s) of spontaneous deletion, therefore, must be different. What roles do sequence repeats play in the deletion process of *E. coli*? The immense variety of sequence repeats at *E. coli* deletion termini argues that the primary sequence of a repeat is itself not important. Rather, one repeat must base pair with another at some step in the process. Direct base pairing between the termini of prospective deletions probably stabilizes structures that are intermediates to deletion formation. Such slipped mispairing (17) or pairing of quasi-palindromes (19) might generate transient single-stranded loops or stem-loops in DNA, which when replicated or repaired lead to deletion of sequences between the repeats (1, 17, 19).

TABLE 2. Chi-square test of the significance of direct sequence repeats at the termini of spontaneous *unc-54* deletions[a]

Size of direct repeat (bases)	No. of deletions	
	Expected	Observed
0	8.9	4
1	4.5	6
2	1.9	2
3	0.8	2
4	0.3	1

[a] The expected numbers of direct repeats were calculated by assuming random endpoints within *unc-54*. The observed numbers of direct repeats are taken from Table 1. The expected and observed distributions are not significantly different ($\chi^2 = 2.53$; $P = 0.28$). Fifteen deletions were tested.

The wild-type *unc-54* gene contains over 2,000 pairs of direct or inverted repeats that are 8 nucleotides or longer and which, if used for deletion, would generate a detectable mutation. None of these repeats is located at the endpoints of our deletions. Similarly, we are unable to find any compelling examples of quasi-palindromes associated with *unc-54* deletion endpoints. We conclude that base pairing of regions near deletion termini is not essential for deletion formation in *C. elegans*.

Perhaps structures or sites other than terminal repeats are necessary for *C. elegans* deletion mutagenesis. Because of their occasional association with deletion termini, a variety of sequences or structures other than direct repeats have been suggested to play a role in the deletion process. We examined the *unc-54* gene and our deletions for such sequences; we found none of the following features. (i) Sequences that are complementary to the junction of a deletion might serve to align its termini (19), but we did not find significant examples of such alignment sequences for any of the deletions. (ii) DNA topoisomerases may be involved in the formation of both procaryotic and eucaryotic deletions (8, 24–26, 35, 38), but the consensus recognition sequences for *Drosophila* topoisomerase II (52), and *Tetrahymena*, rat, and wheat germ topoisomerase I (4, 6) are not enriched near the termini of *unc-54* deletions. (iii) Regions of dyad symmetry are located near the termini of spontaneous deletions in cultured mammalian cells (43), but such structures are not enriched near the termini of *unc-54* deletions. (iv) Certain tri- and tetranucleotides recur at deletion termini in cultured mammalian cells (43), but the representation of di-, tri-, and tetranucleotides at *unc-54* deletion termini is random. (v) Alternating purine-pyrimidine simple sequence repeats have been associated with the endpoints of certain deletions and gene conversions (33, 40, 55, 58), but such repeats are not consistently located at or near the termini of *unc-54* deletions. (vi) The termini of two mutagen-induced deletions of *C. elegans* are located within identical nonamer-pentamer sequences (11), but there was no significant association of these sequences with our deletions. We conclude that, except for their small average size, spontaneous deletions in *unc-54* have random endpoints. The endpoints are random both with respect to their position within the gene and with respect to the DNA sequences at or near their termini.

Spontaneous *unc-54* deletions probably result from intra-

molecular rearrangements. If deletion formation had involved two DNA molecules (either homologous chromosomes or sister chromatids), and if the breakages and rejoinings were reciprocal, then deletions and tandem direct duplications would be equally frequent products. We are confident that most tandem duplications similar in size and position to our spontaneous deletions (600-base-pair average size, both endpoints within *unc-54* coding sequences) would have been detected by our genetic methods. We identified two tandem duplications among 114 spontaneous mutants (12), but their sizes and positions were very unlike those of the deletions. Our failure to identify short, tandem duplications indicates that intermolecular rearrangements are not involved in deletion formation or that such crossovers are nonreciprocal.

How, then, are *C. elegans* deletions formed? The deletion junctions that we describe are very similar to ligation junctions described by Roth and Wilson (50). Mammalian cells are very proficient at ligating free ends of DNA (10, 30, 39, 48, 49, 62). Such activities may function to repair potentially lethal double-strand breaks (46) and may be responsible for the inherent "stickiness" or "healing" of broken chromosomes (36). Roth and Wilson (50) described three distinct pathways for ligation of free DNA ends in mammalian cells. Two of these pathways, template-directed and post-repair ligation, are mediated by very short but significant homologies between protruding single-strand termini. Homologies that align such ligations are often only 1 or 2 base pairs in length. The third pathway, single-strand ligation, is homology independent. The overall efficiencies of these three pathways are about equal, and the nature of the termini to be ligated determines which pathway is most often utilized. Perhaps most ligations of *C. elegans* termini occur via the single-strand ligation (homology-independent) pathway, or perhaps the sequences of additional *C. elegans* deletions will demonstrate that the very short repeats are significant. Either way, we speculate that spontaneous deletions in *C. elegans* result from random double-strand breaks, followed by ligation of those breaks via the pathways described by Roth and Wilson (50). The small average size of *C. elegans* deletions would indicate that sequences are more likely to be joined if they occur in close proximity. Our failure to isolate tandem direct duplications at frequencies similar to those of deletions would indicate that sequences to be joined usually reside on the same DNA molecule.

Could the mechanism outlined above be related to that in *E. coli*? Events that initiate deletion formation in *E. coli* and the intermediates whose structures are stabilized by base pairing are unknown. Most models for *E. coli* deletion formation involve mispaired structures that serve as primers, templates, or substrates for DNA replication and/or repair (1, 14, 16, 19, 54). We believe that repair of double-strand breaks according to the model outlined above could equally well account for the association of direct repeats with *E. coli* deletion termini. Double-strand breaks are probably natural occurrences in all cells, and only those breaks that get successfully repaired yield viable chromosomes. If an *E. coli* system for repair of double-strand breaks were dominated by the homology-dependent pathways of ligation (template-directed and post-repair ligation), and if those pathways required greater terminal homologies in *E. coli* than in mammalian cells, then, as observed, direct sequence repeats would be a regular feature of *E. coli* deletion junctions.

ACKNOWLEDGMENTS

We thank B. Engels for assistance with the quantitative aspects of our data and our colleagues Amy Bejsovec, Suzanne Sprunger, Brian Carr, and John Collins for critically reading the manuscript.

This work was supported by Public Health Service individual research grant GM30132 from the National Institutes of Health.

LITERATURE CITED

1. **Albertini, A. M., M. Hofer, M. P. Calos, and J. H. Miller.** 1982. On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. Cell **29**:319–328.
2. **Albertson, D. G., J. E. Sulston, and J. G. White.** 1978. Cell cycling and DNA replication in a mutant blocked in cell division in the nematode *C. elegans*. Dev. Biol. **63**:165–178.
3. **Anderson, R. A., S. Kato, and R. D. Camerini-Ortero.** 1984. A pattern of partially homologous recombination in mouse L cells. Proc. Natl. Acad. Sci. USA **81**:206–210.
4. **Been, M. D., R. R. Burgess, and J. J. Champoux.** 1984. Nucleotide sequence preference at rat liver and what germ type 1 DNA topoisomerase breakage sites in duplex SV40 DNA. Nucleic Acids Res. **12**:3097–3114.
5. **Benton, W. D., and R. W. Davis.** 1977. Screening lambda-gt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
6. **Bonven, B. J., E. Gocke, and O. Westergaard.** 1985. A high affinity topoisomerase I binding sequence is clustered at DNAse I hypersensitive sites in Tetrahymena R-chromatin. Cell **41**:541–551.
7. **Brenner, S.** 1974. The genetics of *Caenorhabditis elegans*. Genetics **77**:71–94.
8. **Bullock, P., J. J. Champoux, and M. Botchan.** 1985. Association of crossover points with topoisomerase I cleavage sites: a model for nonhomologous recombination. Science **230**:954–958.
9. **Bullock, P., W. Forrester, and M. Botchan.** 1984. DNA sequence studies of simian virus 40 chromosomal excision and integration in rat cells. J. Mol. Biol. **174**:55–84.
10. **Chang, X.-B., and J. H. Wilson.** 1987. Modification of DNA ends can decrease end joining relative to homologous recombination in mammalian cells. Proc. Natl. Acad. Sci. USA **84**:4959–4963.
11. **Dibb, N. J., D. M. Brown, J. Karn, D. G. Moerman, S. L. Bolten, and R. H. Waterston.** 1985. Sequence analysis of mutations that affect the synthesis, assembly and enzymatic activity of *unc-54* myosin heavy chain of *C. elegans*. J. Mol. Biol. **183**:543–551.
12. **Eide, D., and P. Anderson.** 1985. The gene structures of spontaneous mutations affecting a *Caenorhabditis elegans* myosin heavy chain gene. Genetics **109**:67–79.
13. **Eide, D., and P. Anderson.** 1985. Transposition of Tc1 in the nematode *C. elegans*. Proc. Natl. Acad. Sci. USA **82**:1756–1760.
14. **Farabaugh, P. J., U. Schmeissner, M. Hofer, and J. H. Miller.** 1978. Genetic studies of the *lac* repressor. VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. J. Mol. Biol. **126**:847–857.
15. **Foster, T. J., V. Lundblad, S. Hanley-Way, S. M. Halling, and N. Kleckner.** 1981. Three Tn*10*-associated excision events: relationship to transposition and role of direct and inverted repeats. Cell **23**:215–227.
16. **Franklin, N. C.** 1967. Extraordinary recombinational events in *Escherichia coli*. Their independence of the *rec*$^+$ function. Genetics **55**:699–707.
17. **Franklin, N. C.** 1971. Illegitimate recombination, p. 175–194. *In* A. D. Hershey (ed.), The bacteriophage lambda. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Frischauf, A.-M., H. Lehrach, A. Poustka, and N. Murray.** 1983. Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. **170**:827–842.
19. **Glickman, B. W., and L. S. Ripley.** 1981. Structural intermediates of deletion mutagenesis: a role for palindromic DNA. Proc. Natl. Acad. Sci. USA **81**:512–516.
20. **Goldberg, S. Z., D. Kuebbing, D. Trauber, M. P. Schafer, S. E.**

Lewis, R. A. Popp, and W. F. Anderson. 1986. A 66-basepair insert bridges the deletion responsible for a mouse model of beta-thalassemia. J. Biol. Chem. **261**:12368–12374.
21. **Henthorn, P. S., D. L. Mager, T. H. Huisman, and O. Smithies.** 1986. A gene deletion ending within a complex array of repeated sequences 3′ to the human beat-globin gene cluster. Proc. Natl. Acad. Sci. USA **83**:5194–5198.
22. **Hohn, B.** 1979. In-vitro packaging of lambda and cosmid DNA. Methods Enzymol. **68**:299–309.
23. **Horsthemke, B., U. Beisiegel, A. Dunning, J. R. Havinga, R. Williamson, and S. Humphries.** 1987. Unequal crossing-over between two Alu-repetitive DNA sequences in the low-density-lipoprotein-receptor gene. A possible mechanism for the defect in a patient with familial hypercholesterolaemia. Eur. J. Biochem. **164**:77–81.
24. **Ikeda, H.** 1986. Bacteriophage T4 DNA topoisomerase mediates illegitimate recombination *in vitro*. Proc. Natl. Acad. Sci. USA **83**:922–926.
25. **Ikeda, H., K. Aoki, and A. Naito.** 1982. Illegitimate recombination mediated *in vitro* by DNA gyrase of *Escherichia coli*: structure of recombinant DNA molecules. Proc. Natl. Acad. Sci. USA **79**:3724–3728.
26. **Ikeda, H., K. Moriya, and T. Matsumoto.** 1981. *In vitro* study of illegitimate recombination: involvement of DNA gyrase. Cold Spring Harbor Symp. Quant. Biol. **45**:399–408.
27. **Inselberg, J.** 1967. Formation of deletion mutations in recombination-deficient mutants of *Escherichia coli*. J. Bacteriol. **94**: 1266–1267.
28. **Jennings, M. W., R. W. Jones, W. G. Wood, and D. J. Weatherall.** 1985. Analysis of an inversion within the human beta-globin gene cluster. Nucleic Acids Res. **13**:2897–2906.
29. **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. USA **80**:4253–4257.
30. **Kopchick, J. J., and D. W. Stacey.** 1984. Differences in intracellular DNA ligation after microinjection and transfection. Mol. Cell. Biol. **4**:240–246.
31. **Lehrman, M. A., D. W. Russell, J. L. Goldstein, and M. S. Brown.** 1987. Alu-Alu recombination deletes splice acceptor sites and produces secreted low density lipoprotein receptor in a subject with familial hypercholesterolemia. J. Biol. Chem. **262**:3354–3361.
32. **Lehrman, M. A., W. J. Schneider, T. C. Sudhof, M. S. Brown, J. L. Goldstein, and D. W. Russell.** 1985. Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. Science **227**:140–146.
33. **Mager, D. L., P. S. Henthorn, and O. Smithies.** 1985. A chinese $^{G}\psi^{+}(^{A}\psi\phi\beta)^{0}$ thalassemia deletion: comparison to other deletions in the human β-globin gene cluster and sequence analysis of the breakpoints. Nucleic Acids Res. **13**:6559–6575.
34. **Marmur, J.** 1961. A procedure for the isolation of deoxyribonucleic acid from micro-organisms. J. Mol. Biol. **3**:208–218.
35. **Marvo, S. L., S. R. King, and S. R. Jaskunas.** 1983. Role of short regions of homology in intermolecular illegitimate recombination events. Proc. Natl. Acad. Sci. USA **80**:2452–2456.
36. **McClintock B.** 1941. The stability of broken ends of chromosomes in *Zea mays*. Genetics **26**:234–282.
37. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
38. **Michel, B., and S. D. Ehrlich.** 1986. Illegitimate recombination at the replication origin of bacteriophage M13. Proc. Natl. Acad. Sci. USA **83**:3386–3390.
39. **Miller, C. K., and H. M. Temin.** 1983. High efficiency ligation and recombination of DNA fragments by vertebrate cells. Science **220**:606–609.
40. **Molineaux, S. M., H. Engh, F. De Ferra, L. Hudson, and R. A. Lazzarini.** 1986. Recombination within the myelin basic protein gene created the dysmyelinating shiverer mouse mutation. Proc. Natl. Acad. Sci. USA **83**:7542–7546.
41. **Muller-Hill, B., and J. Kania.** 1974. *lac* repressor can be fused to beta-galactosidase. Nature (London) **249**:561–563.
42. **Myerowitz, R., and N. D. Hogikyan.** 1987. A deletion involving Alu sequences in the beta-hexosaminidase alpha-chain gene of French Canadians with Tay-Sachs disease. J. Biol. Chem. **262**: 15396–15399.
43. **Nalbantoglu, J., D. Hartley, G. Phear, G. Tear, and M. Meuth.** 1986. Spontaneous deletion formation at the *aprt* locus of hamster cells: the presence of short sequence homologies and dyad symmetries at deletion termini. EMBO J. **5**:1199–1204.
44. **Nalbantoglu, J., G. Phear, and M. Meuth.** 1987. DNA sequence analysis of spontaneous mutations at the *aprt* locus of hamster cells. Mol. Cell. Biol. **7**:1445–1449.
45. **Nicholls, R. D., N. Fischel-Ghodsian, and D. R. Higgs.** 1987. Recombination at the human α-globin gene cluster: sequence features and topological constraints. Cell **49**:369–378.
46. **Orr-Weaver, T. L., and J. W. Szostak.** 1983. Yeast recombination: the association between double-strand gap repair and crossing-over. Proc. Natl. Acad. Sci. USA **80**:4417–4421.
47. **Park, E.-C., and H. R. Horvitz.** 1986. *C. elegans unc-105* mutations affect muscle and are suppressed by other mutations that affect muscle. Genetics **113**:853–867.
48. **Roth, D. B., T. N. Porter, and J. H. Wilson.** 1985. Mechanisms of nonhomologous recombination in mammalian cells. Mol. Cell. Biol. **5**:2599–2607.
49. **Roth, D. B., and J. H. Wilson.** 1985. Relative rates of homologous and nonhomologous recombination in transfected DNA. Proc. Natl. Acad. Sci. USA **82**:3355–3359.
50. **Roth, D. B., and J. H. Wilson.** 1986. Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. Mol. Cell. Biol. **6**:4295–4304.
51. **Rothstein, R., C. Helms, and N. Rosenberg.** 1987. Concerted deletions and inversions are caused by mitotic recombination between delta sequences in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **7**:1198–1207.
52. **Sander, M., and T. Hsieh.** 1985. *Drosophila* topoisomerase II double-strand DNA cleavage: analysis of DNA sequence homology at the cleavage site. Nucleic Acids Res. **13**:1057–1071.
53. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
54. **Schaaper, R. M., B. N. Danforth, and B. W. Glickman.** 1986. Mechanisms of spontaneous mutagenesis: an analysis of the spectrum of spontaneous mutation in the *Escherichia coli lacI* gene. J. Mol. Biol. **189**:273–284.
55. **Slightom, J. L., A. E. Blechl, and O. Smithies.** 1980. Human fetal $^{G}\psi$- and $^{A}\psi$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21**:627–638.
56. **Spritz, R. A., and S. H. Orkin.** 1982. Duplication followed by deletion accounts for the structure of an Indian deletion β^{0}-thalassemia gene. Nucleic Acids Res. **10**:8025–8029.
57. **Spudich, J. A., V. Horn, and C. Yanofsky.** 1970. On the production of deletions in the chromosome of *E. coli*. J. Mol. Biol. **53**:49–67.
58. **Stringer, J. R.** 1985. Recombination between poly[d(GT) · d (CA)] sequences in simian virus 40-infected cultured cells. Mol. Cell. Biol. **5**:1247–1259.
59. **Sulston, J. E., and S. Brenner.** 1974. The DNA of *C. elegans*. Genetics **77**:95–104.
60. **Vanin, E. F., P. S. Henthorn, D. Kioussis, F. Grosveld, and O. Smithies.** 1983. Unexpected relationships between four large deletions in the human β-globin gene cluster. Cell **35**:701–709.
61. **Williams, T. J., and M. Fried.** 1986. Inverted duplication-transposition event in mammalian cells at an illegitimate recombination join. Mol. Cell. Biol. **6**:2179–2184.
62. **Wilson, J. H., P. Berger, and J. Pipas.** 1982. Somatic cells efficiently join unrelated DNA segments end-to-end. Mol. Cell. Biol. **2**:1258–1269.