On the Components of Segregation Distortion in *Drosophila melanogaster*. V. Molecular Analysis of the *Sd* Locus

Patricia A. Powers and Barry Ganetzky

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

Manuscript received February 17, 1991
Accepted for publication May 18, 1991

ABSTRACT

Segregation Distorter (*SD*) is a naturally occurring meiotic drive system comprising at least three distinct loci: *Sd*, *Rsp* and *E(SD)*. Heterozygous SD/SD^+ males transmit the *SD* chromosome in vast excess over the normal homolog. The distorted transmission involves the induced dysfunction of the spermatids that receive the SD^+ chromosome. In the 220-kb region of DNA that contains the *Sd* gene, we identified a 5-kb tandem duplication that is uniquely associated with all *SD* chromosomes, absent in SD^+ chromosomes, and detectably altered in *Sd* revertants. On northern blots, genomic probes from the tandem duplication detect an *SD*-specific 4-kb transcript in addition to several smaller transcripts present in both *SD* and SD^+. Seven classes of cDNAs derived from these transcripts have been isolated. All of these cDNAs share extensive sequence identity at their 3′ ends but differ at their 5′ ends. Sequence analysis indicates that these cDNAs potentially encode four distinct, but related, polypeptides. Introduction of the tandem duplication into SD^+ flies by germline transformation did not confer the dominant gain-of-function *Sd* phenotype. This result, taken together with our analysis of the *Sd* cDNAs, suggests that the duplication is part of a much larger gene that encodes several different polypeptides.

SEGREGATION Distorter (*SD*), a naturally occurring meiotic drive system located on the second chromosome, was first discovered by HIRAIZUMI in a population of *Drosophila melanogaster* in Madison, Wisconsin, over 30 years ago (SANDLER, HIRAIZUMI and SANDLER 1959). Males heterozygous for the *SD* chromosome and an SD^+ homolog transmit a vast excess of the *SD* chromosome to their progeny, as the result of the induced dysfunction of the SD^+-bearing gametes (HARTL, HIRAIZUMI and CROW 1967; NICOLETTI, TRIPPA and DEMARCO 1967). *SD* chromosomes, which have not been observed to arise *de novo* from SD^+ chromosomes in the laboratory, have been recovered at low frequencies (1–5%) from almost every population that has been screened for them (HIRAIZUMI and NAKAZIMA 1967; TEMIN and MARTHAS 1984). Different *SD* chromosomes can be distinguished by the set of inversions found around the centromere and in the right arm.

The molecular events that lead to the *SD*-induced sperm dysfunction are still unidentified. The first anomaly seen in sperm development in males heterozygous for *SD* and a sensitive homolog is the failure of chromatin condensation in approximately half of the developing spermatid nuclei (TOKUYASU, PEACOCK and HARDY 1977) leading to subsequent defects in maturation of the SD^+-bearing sperm.

Extensive genetic and cytogenetic studies have shown that the *SD* complex consists of a group of three tightly linked loci located near the centromere on chromosome 2 (reviewed by HARTL and HIRAIZUMI 1976; SANDLER and GOLIC 1985; TEMIN *et al.* 1991) found on all naturally occurring *SD* chromosomes. They are: Segregation distorter (*Sd*), a locus that encodes a *trans*-acting factor necessary for distortion, Enhancer of *SD*, *E(SD)*, which considerably strengthens distortion, and Responder, (*Rsp*), which acts as the target for the action of both *Sd* and *E(SD)* (TEMIN 1991).

Deleting the *Sd* locus from an *SD* chromosome results in a chromosome that segregates normally (GANETZKY 1977). Addition of extra doses of Sd^+ to heterozygous SD/SD^+ males does not reduce distortion (ULBER 1985). Thus, *Sd* behaves as a neomorph and Sd^+ behaves as an amorph with respect to segregation distortion. Deletion analysis also allowed the localization of *Sd* to polytene bands 37D2-6 (BRITTNACHER and GANETZKY 1983). A chromosome segment containing *Sd* has been translocated to the *Y* chromosome (LYTTLE 1986) where *Sd* retains its ability to distort the transmission of chromosomes carrying a sensitive *Rsp* locus. Thus *Sd* does not require its second chromosome location to function.

Enhancer of *SD*, *E(SD)*, was originally defined as an modifier that increased the strength of distortion caused by the *Sd* locus (HARTL 1974; GANETZKY 1977; BRITTNACHER and GANETZKY 1984). A deletion of *E(SD)* reduces the strength of distortion from essentially 100% to about 70% (GANETZKY 1977; BRITTNACHER and GANETZKY 1984). Subsequent analysis revealed that in certain genetic backgrounds, *E(SD)* can cause distortion even in the absence of *Sd* (SHARP,

HILLIKER and HOLM 1985; TEMIN 1991; T. W. LYTTLE, personal communication).

Current understanding of the *Rsp* locus results from the successful melding of genetic, cytogenetic and molecular approaches. *Rsp* is located in the centric heterochromatin of chromosome *2R*. A deletion of the chromosome region containing *Rsp* renders that chromosome insensitive to the action of *Sd* (GANETZKY 1977). Different alleles of *Rsp* show different degrees of sensitivity to the *Sd* locus (MARTIN and HIRAIZUMI 1979; HIRAIZUMI and THOMAS 1984; TEMIN and MARTHAS 1984). This sensitivity appears to exist in a continuum (TEMIN and MARTHAS 1984). The degree of sensitivity to *Sd* was shown to be directly correlated with the physical size of a Hoechst-bright, N-banding negative heterochromatic block in *2R* located adjacent to the centromere (PIMPINELLI and DIMITRI 1989). Molecular cloning of this region revealed a tandem array of a specific family of AT-rich repeats whose copy number displays a strong positive correlation with the sensitivity of *Rsp* (WU *et al.* 1988). Whenever genetic evidence indicates that *Rsp* has been deleted or translocated to the *Y* chromosome, cytological and molecular examination reveals a corresponding deletion or translocation of the repeat array.

GANETZKY (1977) proposed a model for segregation distortion in which the binding of the *Sd* product to the *Rsp* locus initiates the events that lead to sperm dysfunction. This model assumes that the *Sd* product binds more readily to a sensitive responder (Rsp^s) than to an insensitive responder (Rsp^i), that the *Sd* product is limited with respect to the number of binding sites available, and that Sd^+ and Rsp^s do not have any normal function with respect to segregation distortion, since the deletion of either *Sd* or Rsp^s has the effect of restoring normal chromosome transmission. BRITTNACHER and GANETZKY (1983) suggested that *Sd* might be a DNA binding protein that binds directly to the Rsp^s DNA, although the direct physical contact between the *Sd* product and Rsp^s DNA is by no means a required component of this model.

This model was recently expanded by LYTTLE (1989) and by PIMPINELLI and DIMITRI (1989) who suggested that *Sd* might produce a protein or RNA with high affinity for the *Rsp* region. According to this model, a sensitive *Rsp* region binds to the *Sd* product by some titration mechanism and carries the product through the stages of sperm development. This serves to concentrate the *Sd* product in the postmeiotic cells in which Rsp^s is present, causing the asymmetric distribution of the *Sd* product, which in turn, by some as of yet unknown mechanism, results in the asymmetric sperm dysfunction that is the hallmark of segregation distortion.

A simple view, consistent with the models of GANETZKY (1977), LYTTLE (1989) and PIMPINELLI and DIMITRI (1989) and with the current analysis of *Rsp*, is that *Sd* encodes some type of chromosomal protein that binds at Rsp^s and inhibits chromatin condensation. It is noteworthy that the failure of condensation affects all of the chromatin within the nucleus of spermatids undergoing dysfunction rather than just that in the vicinity of Rsp^s.

Although much is known about the genetics of segregation distortion, very little is known about the molecular events that actually lead to the preferential recovery of the *SD* chromosome. In order to determine the mechanism of distortion, it is essential to learn more about the product(s) of the *Sd* locus, and the nature of its interaction with *Rsp* and *E(SD)*. Toward this aim, we undertook a molecular analysis of the *Sd* locus.

MATERIALS AND METHODS

Genetic stocks: For a complete description of genetic markers, see LINDSLEY and GRELL (1968) and LINDSLEY and ZIMM (1985–1990).

***SD* chromosomes:** *SD-Mad* was isolated from a natural population in Madison, Wisconsin, by R. TEMIN in 1979. It carries the pericentric In(2LR)39D;42A and paracentric In(2R)52A2-B1;56F9–13. *SD-V017* is an inversion-free *SD* chromosome isolated from a Spanish population by R. Temin. *SD-Armidale* and *SD-Mauna* were provided by T. LYTTLE. A description of the additional *SD* chromosomes used in the study can be found in BRITTNACHER and GANETZKY (1983).

Methodology: Except where noted, recombinant DNA methods were adapted from protocols described by MANIATIS, FRITSCH and SAMBROOK (1982) and AUSUBEL *et al.* (1987).

Libraries: The SD^+ genomic library was constructed by MANIATIS *et al.* (1978) using DNA from Canton S flies. SD^+ cDNA libraries were constructed by POOLE *et al.* (1985) using RNA from Oregon R flies. We used genomic DNA from *SD-Mad* flies to construct a phage *SD* library in the vector EMBL4 (FRISCHAUF *et al.* 1983) and a cosmid *SD* library in the vector pCosNeo (STELLER and PIROTTA 1985) using procedures described in MANIATIS, FRITSCH and SAMBROOK (1982). Four independent *SD* cDNA libraries were constructed using polyadenylated RNA from *SD-Mad* adult males and the vector λ GT10 (YOUNG and DAVIES 1985), three by the method of HUYNH, YOUNG and DAVIS (1985) and one by the method of GUBLER and HOFFMAN (1985).

Sequencing: DNA sequences were determined by the method of SANGER, NICKLEN and COULSEN (1977) using the Sequenase kit (U.S. Biochemical) and double-stranded plasmid DNA as a template. Nucleotide and polypeptide sequence analysis was performed using GCG software (DEVEREAUX, HAEBERLI and SMITHIES 1984) and a VAX computer.

Transformation: *P* element-mediated germline transformants were obtained essentially as described in SANTAMARIA (1986), SPRADLING (1986) and KARESS (1985). Eggs were microinjected though the chorion with a solution containing 800 μg/ml supercoiled plasmid DNA in 5 mM KCl, 0.1 mM NaH_2PO_4, pH 6.8, and 2% (v/v) Durkee green food coloring.

Genetic analysis of the transformants: The standard protocol for measuring segregation ratios in the transformed males was to mate each male of interest with two females. Each genotype was tested with 10–20 replicates. After 4 days the parents were transferred to a fresh food

FIGURE 1.—*Eco*RI restriction map of the cloned DNA from the chromosome walk in the 37D2-6 region. Vertical lines represent *Eco*RI restriction sites. p17 is the microdissected DNA fragment used to initiate the chromosome walk. The triangle marks the location of the *SD*-specific 12-kb *Eco*RI fragment. The location and orientation of the topoisomerase II gene is shown directly beneath the chromosome walk. The approximate cytological locations as determined by *in situ* hybridization, and the distal and proximal orientation of the walk are shown above the map. Underneath the map, the black bar represents the distal extent of the DNA removed in chromosomes bearing the *SD-Roma*^{rev14} deletion, an X-ray-induced deletion that deletes *Sd* and renders the *SD-Roma* chromosome unable to distort a sensitive homolog. The shaded bar represents the distal extent of the DNA removed in chromosomes bearing the *SD-72*^{pr26} deletion, an X-ray-induced deletion that does not remove *Sd* and is still capable of distortion.

vial and discarded after 4 more days. The offspring from each vial were counted through day 18 after the parents had been introduced into that vial. All experiments were carried out at 25°. In all tests, the 2^i chromosome bears a less sensitive *Rsp* locus than does the 2^j chromosome, and segregation distortion in the transformed males would be manifested by the preferential recovery of the 2^i chromosome bearing the less sensitive *Rsp* locus. The ability of the males to cause distortion is expressed in terms of k value, where k is the proportion of progeny bearing chromosome 2^i among the total progeny. Since segregation is normal in females, departures from $k = 0.5$ in the progeny of heterozygous females can be attributed to differences in relative viabilities of the segregating chromosomes. These control crosses involved the same protocol as outlined above. In an effort to keep the larval density constant, each vial contained two females and one male. If we define the number of progeny receiving the 2^i chromosome as n_i, and the number of progeny receiving the 2^j chromosome as n_j, then the relative viability effect is estimated by $W = n_i/n_j$ from the heterozygous female crosses. The corrected k value is then $n_i/(Wn_j + n_i)$ from the heterozygous male crosses.

RESULTS

The chromosome walk: V. PIROTTA performed a microdissection of the 37D2-6 region, and constructed a recombinant bacteriophage library of about 200 clones, each containing an *Eco*RI fragment from this region (as in PIROTTA 1987). Ten of these phage were assayed by *in situ* hybridization to polytene chromosomes from Drosophila larvae. One hybridized to the proximal edge of 37D2 and was chosen as the initial probe for screening a genomic library of Canton S DNA, an SD^+ strain (MANIATIS *et al.* 1978). The resulting chromosome walk is shown in Figure 1. It encompasses 220 kb of DNA extending from polytene bands 37D2 to 37D6, and spans the entire region that was defined as containing the *Sd* locus by BRITTNACHER and GANETZKY (1983). The distal endpoints of two of the X-ray-induced deletions, *SD-Roma*^{rev14} (GANETZKY 1977), a chromosome that is no longer able to distort the transmission of a sensitive homolog because it is deleted for the *Sd* locus, and *SD-72*^{pr26} (BRITTNACHER and GANETZKY 1983), a chromosome that retains the ability to distort and defines the proximal limit for the chromosome walk, are shown in Figure 1. During our analysis of the chromosome walk, the topoisomerase II gene was localized to the region 37D5,6 by NOLAN *et al.* 1988. The topoisomerase II gene structure was compared by Southern blot analysis in *SD* and SD^+ strains and no gross differences were detected.

Identification of the *Sd*-specific tandem duplication: Southern blot analysis was used to compare each DNA clone isolated from the SD^+ library to the corresponding region of genomic DNA from *SD* strains to look for *SD*-specific differences. A single restriction fragment length difference uniquely associated with all *SD* chromosomes was identified. Figure 2 shows this analysis for six *SD* strains. All nine independently isolated *SD* chromosomes that have been examined (the six shown in Figure 2 as well as *SD-Armidale*, *SD-36* and *SD-Mauna*) have the identical 12-kb *Eco*RI fragment. These chromosomes were isolated from natural populations in Asia, Europe, Australia and North America; some of these *SD* chromosomes were recently recovered from nature, while others were collected over 30 years ago. Many SD^+ chromosomes were also tested, including some from the same wild populations as the *SD* chromosomes; all contain a 7-kb *Eco*RI fragment at the corresponding location. The restriction map of cloned DNA from *SD* and SD^+ flies, shown in Figure 3, reveals that the basis for the 12-kb *Eco*RI fragment is a 5-kb tandem duplication in *SD* chromosomes of a segment of DNA that is present in a single copy in the SD^+ chromosomes.

To address the question of whether this duplication arose more than once, polymorphic restriction sites flanking the region of the duplication were examined in *SD* and in SD^+ chromosomes. All of the *SD* chromosomes have the same constellation of polymorphic restriction sites in a 50-kb region that extends 20 kb

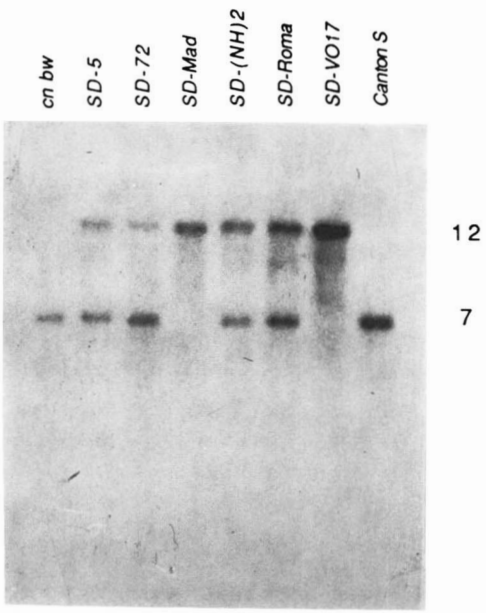

FIGURE 2.—Southern blot analysis demonstrating the presence of an *SD*-specific length difference in the region of the chromosome walk. Southern blots made with *Eco*RI digested genomic DNAs were probed with the 7-kb *Eco*RI fragment derived from SD^+ genomic DNA. The *SD-5*, *SD-72*, *SD-(NH)2* and *SD-Roma* flies were heterozygous for the *SD* chromosome and a SD^+ (*cn bw*) chromosome. *SD-Mad* and *SD-VO17* are homozygous. *Canton S* and *cn bw* are both SD^+ chromosomes.

distal and 28 kb proximal to the duplication, and have one of two conserved haplotypes in the region that is 28–43 kb proximal to the duplication (our unpublished data and reviewed in WU and HAMMER 1990) The common polymorphic haplotype and the identity of the restriction map of the tandem duplication on each of the *SD* chromosomes suggest that the duplication arose only once. Two different *SD* chromosomes, *SD-Roma* and a recently identified *SD* chromosome, *SD-VO17* (R. G. TEMIN, personal communication) both lack the pericentric and/or *2R* inversions typically associated with *SD* chromosomes and were both isolated from Mediterranean populations. If the duplication arose only once, it may have originated in the Mediterranean region on a chromosome that was free of inversions. The specific inversions associated with the different *SD* chromosomes would then have arisen subsequently after the geographic spread of these chromosomes.

Analysis of *SD* revertants: To extend the correlation between *Sd* activity and the duplication, several *SD* revertants (chromosomes that are derived from *SD* chromosomes and that have lost the ability to distort a sensitive homolog) were assayed for the presence of alterations involving the 12-kb *Eco*RI fragment. Three X-ray-induced *SD* revertants with no cytologically detectable lesions (GANETZKY 1977) were analyzed by restriction mapping of genomic DNA. One of these ($SD\text{-}5^{rev19}$) retained the 12-kb *Eco*RI fragment, with no apparent alterations either in the 12-kb *Eco*RI fragment or elsewhere in the region of the chromosome walk. The other two ($SD\text{-}5^{rev2}$ and $SD\text{-}5^{rev7}$) have the 7-kb *Eco*RI fragment that is characteristic of SD^+ chromosomes. Throughout the entire length of the chromosome walk, the latter two revertant chromosomes resembled the *cn bw* chromosome at each of the locations where the *cn bw* and *SD-5* chromosomes differ. The simplest explanation for this arrangement is that during the mutagenesis experiment, both of these revertants arose from rare recombination events in *SD/cn bw* males. In addition, a spontaneous *SD* revertant (T. W. LYTTLE, personal communication), two putative *P* element-induced revertants (SHIMIKAWA 1987) and three "suppressors of *SD*" (GOLIC 1990), that are no longer able to distort a sensitive homolog were assayed for the presence of the 12-kb *Eco*RI fragment. All contained the 7-kb *Eco*RI fragment at that location, despite the presence of the 12-kb *Eco*RI fragment in their parent *SD* chromosomes. In addition, several *SD* revertants, generated by EMS mutagenesis (J. G. BRITTNACHER and B. GANETZKY, unpublished results) were analyzed in the region of the duplication by restriction mapping of genomic DNA. All of these revertants retain the 12-kb *Eco*RI fragment. The putative lesion in one of these, *ER8*, was detected because it generated a novel *Eco*RI restriction site within the duplication (indicated by an arrow in Figure 3). The lesions of the other EMS-induced revertants were not detected in this analysis, and may be extremely small deletions or base substitutions or reside outside the *Sd* locus itself.

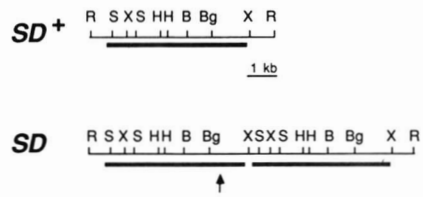

FIGURE 3.—Restriction map of the *SD*-specific length difference indicating that it is a tandem duplication. The upper line shows the restriction map of the 7-kb *Eco*RI fragment from SD^+ chromosomes. The lower line shows the restriction map of the corresponding region found in all *SD* chromosomes examined. The *Eco*RI fragment is 12 kb and the restriction map reveals that it contains a region that is tandemly duplicated. The approximate extent of the duplication as judged by the restriction map is indicated by the hatched lines shown below the map. The arrow indicates the location of the novel *Eco*RI site in an EMS-induced revertant (see text). Enzymes used were *Bam*HI (B), *Bgl*II (Bg), *Hin*dIII (H), *Eco*RI (R), *Sst*I (S) and *Xba*I (X).

Analysis of transcripts encoded within this region: The region surrounding the duplication was used to probe northern blots made using polyadenylated RNA from homozygous *SD* and SD^+ strains. Probes consisting of the entire duplication, or portions of the duplication detected two bands in *SD* flies: a transcript of 4 kb that is only found in *SD* RNA from both males and females, and a broad band of hybridization at 2 kb found in both *SD* and SD^+ RNA from both males and females (Figure 4). Developmental

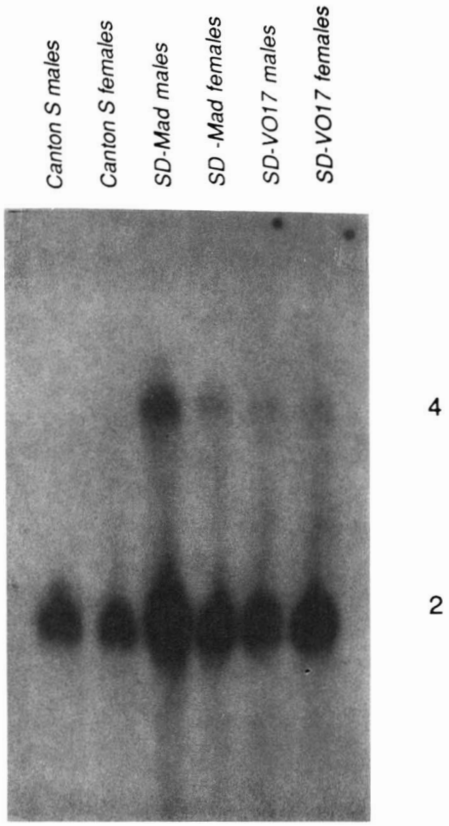

FIGURE 4.—Northern blot analysis of transcripts detected with genomic probes from the *Sd*-duplication. An autoradiograph of a northern blot made with polyadenylated RNA isolated from *Canton S* (*SD*$^+$) males (lane 1), *Canton S* (*SD*$^+$) females (lane 2), *SD-Mad* homozygous males (lane 3), *SD-Mad* homozygous females (lane 4), *SD-VO17* homozygous males (lane 5) and *SD-VO17* homozygous females (lane 6), hybridized with the 3.5-kb *Bam*HI-*Eco*RI genomic fragment from the proximal side of the *Sd*-duplication is shown. Lane 3 contains more RNA than the other lanes.

northern blots indicate that these transcripts are expressed only in the pupal and adult stages (data not shown). All of the genomic probes from the duplication detect both the 2 kb and the 4-kb transcripts. In addition, *SD*$^+$ cDNA probes that correspond to the 2-kb transcript detect both the 2-kb and the 4-kb transcripts on northern blots of *SD* RNA. We conclude that the 2-kb and the 4-kb transcripts share extensive sequence identity. We examined the pattern of transcripts in two of the *SD* revertants described above. We detected both the 4-kb and the 2-kb transcripts in *SD-5*^{rev19} and in *ER8*. Other transcripts encoded by the regions that flank the duplication have identical patterns in *SD* and *SD*$^+$ flies. In particular, we detected no difference in the size or abundance of the nearby topoisomerase II transcript in polyadenylated RNA from *SD*$^+$ *vs. SD-Mad* homozygotes.

Analysis of cDNAs: Several cDNA libraries were constructed using polyadenylated RNA from *SD-Mad* male flies. These libraries were screened with genomic DNA probes from the duplication. Portions of these cDNAs were also used to rescreen the cDNA libraries. Several existing cDNA libraries constructed from *SD*$^+$ strains (POOLE *et al.* 1985) were also screened. Approximately full-length cDNAs corresponding to the 2-kb transcript that is present in both *SD* and *SD*$^+$ RNA were recovered from the *SD* and *SD*$^+$ cDNA libraries. These two cDNAs (S1 from the *SD* cDNA library, and N1 from the *SD*$^+$ cDNA library) are essentially identical. Additional cDNAs that hybridize to the region of the duplication and that share extensive sequence identity to cDNAs S1 and N1 were also recovered from the *SD* cDNA libraries. The cDNAs from the *SD* library fall into seven classes as summarized in Figure 5.

All of the cDNAs that have been isolated share extensive identity in the 3′ region, and six of the seven classes terminate at the same 3′ position. S2 cDNAs contain the 3′ sequences found in the other classes, but are about 1 kb longer at the 3′ end. This longer 3′ end is encoded in part by genomic sequences located within the duplication and in part by genomic sequences located 40 kb distal to the duplication.

All of the cDNA classes differ at their 5′ ends. Several of the cDNAs, such as S3 and S7, must be specific to *SD* chromosomes because they contain sequences from both halves of the duplication and so can only be produced by *SD* chromosomes bearing the duplication. Three classes of cDNAs extend 5′ proximal to the duplicated region. The 5′ ends of cDNAs S2, S4 and S5 are derived from sequences located about 50 kb proximal to the duplication. Although the sequences encoding these 5′ ends appear to be clustered, the 5′ ends of cDNAs S2, S4 and S5 do not share any sequence identity. The S6 cDNA is colinear with the S1 and N1 cDNAs, but extends an additional 300 bp 5′ to cDNAs S1 and N1. Since additional members of the S1 class were identified, each sharing the same 5′ end, we think that it is unlikely that cDNAs S1 and N1 are truncated cloning artifacts of the longer S6 cDNA. None of the cDNAs was found to hybridize to the genomic fragment that contains the new *Eco*RI site in the *SD* revertant *ER8*. However, we cannot exclude that the region of identity was so small that we failed to detect it by hybridization, or that this region has a regulatory role, but is removed in the completely spliced cDNA.

None of the cDNAs that we have isolated so far, or any combinations of cDNAs, can be arranged to create a cDNA corresponding to the entire 4-kb transcript. However since the 4-kb transcript is relatively abundant and since all of the independent cDNAs that we obtained (over 50 from four independent *SD* cDNA libraries) fall into only seven classes, it seems likely that one of these classes will correspond to a truncated cDNA representing the 4 kb transcript.

Northern blot analyses using as probes portions of the various classes of cDNAs or genomic sequences that correspond to the cDNAs, show that the broad band of hybridization at 2 kb is actually a collection

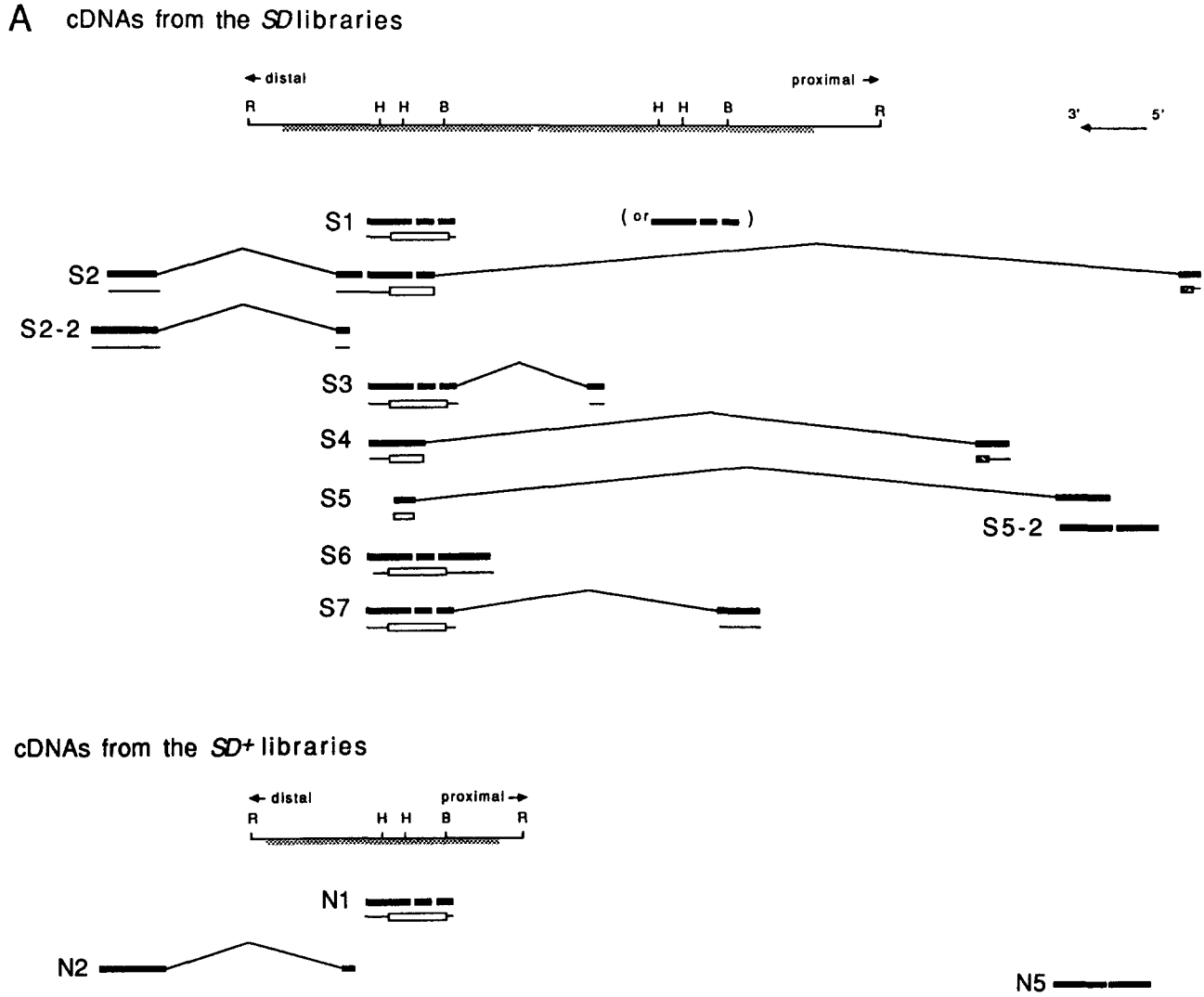

FIGURE 5.—cDNAs corresponding to transcripts from the region of the *Sd* duplication. A, The upper line shows a restriction map of the 12-kb *Eco*RI fragment from an *SD* chromosome. The seven classes of *Sd* cDNAs identified are shown under the map. cDNA S2-2 overlaps and extends cDNA S2. cDNA S5-2 overlaps and extends cDNA S5. For each cDNA, the upper line represents the organization of the cDNA relative to genomic DNA. Solid bars represent exons and thin lines represent intervening sequences. The minimum number of exons that are necessary to account for the observed restriction sites and hybridization patterns are shown. However the organization of exons may be more complex than is actually represented here. To simplify the figure, we have arbitrarily drawn cDNAs S1, S2, S4, S5 and S6 as being derived from the left hand portion of the duplication. The direction of transcription is from right to left. The portion of each cDNA that has been sequenced is indicated below the exons in each case. Open bars represent apparent coding regions; thin black lines represent noncoding regions. The open reading frames of the various cDNAs are identical in those regions where the open bars overlap. The filled boxes for cDNAs S2 and S4 represent coding sequences at the 5′ ends that differ from that of cDNA S1.

of several transcripts of similar size (data not shown). All members of the 2-kb class of transcripts were found in both *SD* and SD^+ polyadenylated RNA.

The cDNAs recovered from SD^+ libraries are shown in Figure 5B. These libraries were not extensively screened, and so SD^+ cDNAs corresponding to all seven of the classes of *SD* cDNAs have not yet been identified. However we predict that many of the cDNAs that were isolated from the *SD* libraries will also be present in the SD^+ libraries because as described above, northern blots of RNA from both *SD* and SD^+ flies probed with *SD* cDNAs give complex patterns of transcripts that differ only by the presence of the 4-kb *SD*-specific transcript.

DNA sequence analysis: The complete DNA sequences of the S1 and N1 cDNAs were determined, and the sequence of S1 is shown in Figure 6. The S1 and N1 sequences are essentially identical, differing only by a few nucleotides at the 5′ and 3′ ends. Both of these cDNAs are incomplete at the 3′ end since they lack poly(A) tracts. This DNA sequence has no significant similarity with any sequence in the GenBank, NFBR or EMBL DNA databases.

The S1 (and N1) sequence contains a single open reading frame. Conceptual translation of the open reading frame predicts a polypeptide of 363 amino acids, as shown in Figure 6. The predicted polypeptide sequence has no significant similarity to any protein sequence in the GenBank, NFBR or Swissprot protein databases, or to any translation product (in any read-

```
   1  ctgactgctcgacgttggttgcagaaaaaaaaacaaggaaaacaacgacctgagattatgctgaaataaatacgctgatatgagctaccgtgctgacctg  100

 101  tagggcaaaccgacacgaggcATGTTTAGGAAACTGCTTAAGATGTGGATTCTGCTCCGACCCACCCACTGGCTGATCCTGATAGCCCTGTGCGCGGTCA  200
                          MetPheArgLysLeuLeuLysMetTrpIleLeuLeuArgProThrHisTrpLeuIleLeuIleAlaLeuCysAlaValT

 201  CCAGTGCAGGTTACTGGTTGCTTTGGTCAGAGATTCGCCTGGAACATGCCTTCAAGCCGCTTTCCAAGCTGGGCGATTCTCTGAGCCCAGATCAGCATGC  300
      hrSerAlaGlyTyrTrpLeuLeuTrpSerGluIleArgLeuGluHisAlaPheLysProLeuSerLysLeuGlyAspSerLeuSerProAspGlnHisAl

 301  CTCGTCCACCACCGATGACTTTGACTTCGAGGAGCACCTAGTGGTGCTTTACAATCGCGTACCGAAAACGGGATCCACCAGCTTTGTTAACATAGCATAC  400
      aSerSerThrThrAspAspPheAspPheGluGluHisLeuValValLeuTyrAsnArgValProLysThrGlySerThrSerPheValAsnIleAlaTyr

 401  GATCTGTGCAAGCCCAATAAATTCCATGTGCTGCACATCAATGTCACTGCCAACATGCACGTCCTCTCGCTTCCCAATCAAATCCAATTCGTGCGCAATG  500
      AspLeuCysLysProAsnLysPheHisValLeuHisIleAsnValThrAlaAsnMetHisValLeuSerLeuProAsnGlnIleGlnPheValArgAsnV

 501  TTTCCAGGTGGCACGAGATGAAGCCAGCTCTTTATCACGGCCACATGGCCTTCCTAGACTTCTCAAAATTCCAAATCGCCCACAAGCCCATCTACATCAA  600
      alSerArgTrpHisGluMetLysProAlaLeuTyrHisGlyHisMetAlaPheLeuAspPheSerLysPheGlnIleAlaHisLysProIleTyrIleAs

 601  TTTGGTGCGCAAACCGCTCGACAGACTTGTCTCCTACTATTACTTTCTACGCTTTGGCGAACACTACCGACCGAATTTAGTCCGCAAGAAGGCGGGCAAT  700
      nLeuValArgLysProLeuAspArgLeuValSerTyrTyrTyrPheLeuArgPheGlyGluHisTyrArgProAsnLeuValArgLysLysAlaGlyAsn

 701  AAAATTACCTTCGATGAGTGCGTGGTGCAGAAACAACCCGACTGTGATCCCAAAAACATGTGGCTGCAGATACCCTTCTTCTGTGGCCATGCAGCCGAGT  800
      LysIleThrPheAspGluCysValValGlnLysGlnProAspCysAspProLysAsnMetTrpLeuGlnIleProPhePheCysGlyHisAlaAlaGluC

 801  GCTGGGAACCCGGCAGCAGTTGGGCCTTGGACCAGGCCAAGCGCAATCTAGTCAACGAATACTTCCTAGTCGGAGTCACAGAGCAGATGTACGAGTTTGT  900
      ysTrpGluProGlySerSerTrpAlaLeuAspGlnAlaLysArgAsnLeuValAsnGluTyrPheLeuValGlyValThrGluGlnMetTyrGluPheVa

 901  GGATCTGCTCGAGAGATCCCTTCCAAGAATTTTTCACGGCTTTCGTGAGCACTATCATAATTCAAACAAATCTCATCTGCGGGTGACATCTTCCAAGCTT  1000
      lAspLeuLeuGluArgSerLeuProArgIlePheHisGlyPheArgGluHisTyrHisAsnSerAsnLysSerHisLeuArgValThrSerSerLysLeu

1001  CCGCCCAGGAATCGACAATTAAATCCATTCAAAAGACAAAAAATCTGGCAAATGGAAAACGATCTGTACGATTTCGCGCTGGCCCAATTCGAGTTCAACAA  1100
      ProProArgAsnArgGlnLeuAsnProPheLysArgGlnLysSerGlyLysTrpLysThrIleCysThrIleSerArgTrpProAsnSerSerSerThrA

1101  GAAGAAGCTCATGCAGCCGGACAACAAGCACGTTCAGAAGTTCATGTACGAGAAGATTCGGCCCAAATGATTGCAGAGGAGGCCGGACGGCAGAGCGAGG  1200
      rgArgSerSerCysSerArgThrThrSerThrPheArgSerSerCysThrArgArgPheGlyProAsnAspCysArgGlyGlyArgThrAlaGluArgGl

1201  ATTCGGCCAGTGAtctaaccaggatcgtccgacaattgtattgtaaatacgtaattctccttcgcatagattgatctgtacttaagagtaactcatccgt  1300
      yPheGlyGlnEnd

1301  acacttaagcgctacctaagactttagcattgaatatcgttgacatattactaacccaaacctttgtacgtagtctaagctttaaacgtataaaacagc  1400

1401  ggagaacgcttttgtagttcatcgtaagcaatatctagcaacaattgtctgcacaatatctctatatgattagtcaaatgtacgttttatactcgacaat  1500

1501  acaacggaatgaatggatgaattgaa  1526
```

FIGURE 6.—Nucleotide sequence and the predicted polypeptide of cDNA S1. The location of four leucines that form a heptad repeat are underscored.

ing frame) of any nucleotide sequence in the GenBank or EMBL DNA databases. Of possible interest is the observation that the region from amino acids 33–54, contains four leucine residues spaced 7 amino acids apart. This arrangement is the hallmark of leucine zippers (LANDSCHULZ, JOHNSON and MCKNIGHT 1988) or leucine coils (O'SHEA, RUTKOWSKI and KIM 1989), sequences that are capable of forming amphipathic helices and that are associated with regions involved in protein-protein interactions. However, an examination of the leucine repeat region of the S1 (and N1) polypeptide shows that it contains amino acids (such as the proline at position 46) that might prevent it from forming a strongly amphipathic helix, and thus the significance of the leucine heptad is still unclear. No other motifs characteristic of DNA binding proteins were identified in this polypeptide sequence.

In order to understand the structural relationships among the various *Sd* cDNAs, nucleotide sequences were obtained for portions of the remaining classes of *Sd* cDNAs. cDNAs S1, S3, S6 and S7 all share the same open reading frame (referred to as the "common" open reading frame), but each differ at the 5′ noncoding end. cDNAs S3 and S7 are particularly intriguing because despite having the same open reading frame as the others, they also contain noncoding sequences derived from both halves of the duplication. Thus S3 and S7 can only be encoded by chromosomes bearing the *SD*-specific duplication.

S2 cDNAs span a region of genomic DNA of about 100 kb. The 3′ end of the untranslated region is derived from genomic sequences 40 kb distal to the *SD*-specific duplication and the 5′ end is transcribed from sequences located about 50 kb proximal to the duplication. The deduced translation product is a polypeptide 376 amino acids in length that shares 339 amino acids in common with the S1 polypeptide, from position 25 to the carboxy terminus. The amino terminal 37 amino acids are not present in the S1 polypeptide but are encoded by one or more exons transcribed from the region 50 kb proximal to the duplication and spliced to the rest of the open reading frame encoded within one part of the duplication. Figure 7 shows the deduced amino acid sequences encoded by cDNAs S1, S2, S4 and S5.

S4 cDNAs span a region of genomic DNA of at least 50 kb. The 3′ portion of the cDNA is identical to that of S1 containing sequences corresponding to codons 251–363 of the common open reading frame. The 5′ portion of this cDNA is derived from a region 50 kb proximal to the duplication. The junction region between the 5′ sequences and the common por-

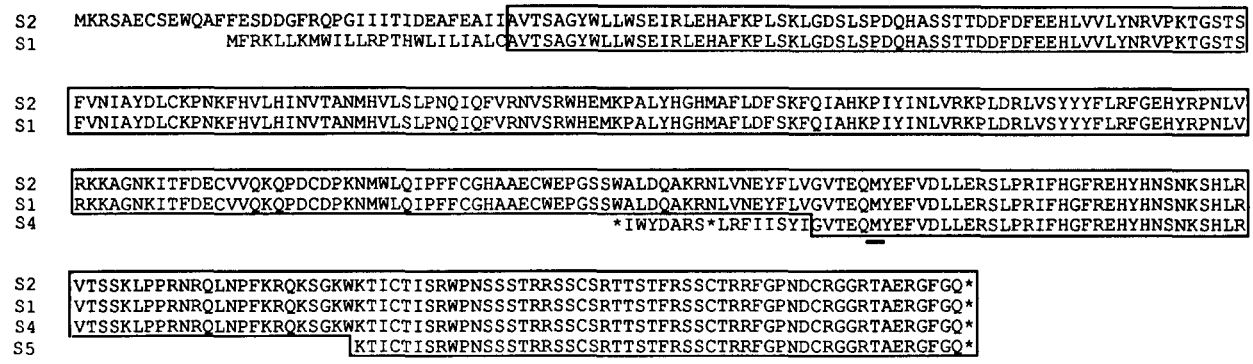

FIGURE 7.—Comparison of the predicted polypeptides encoded by the *Sd* cDNAs. The predicted polypeptides from cDNAs S1, S2, S4 and S5 are shown. Identical sequences are boxed. *Sd* cDNAs S3, S6 and S7 encode the same polypeptide as S1. The S4 polypeptide has two in-frame termination codons, and the 5′-most methionine is underscored. The portion of the S5 polypeptide sequence that has been determined is shown.

tion of the cDNA has been sequenced from three independent cDNA clones. It contains multiple stop codons in all frames. If this cDNA represents a transcript that is actually translated, it would initiate at the methionine located at position 256 in the S1 polypeptide, resulting in a truncated 107 amino acid polypeptide (Figure 7).

The S5 cDNA also spans a region of over 50 kb. About 400 bp of sequence from each end of this cDNA were determined. The 3′ end sequence is identical to the 3′ ends of cDNAs S1, S3, S4, S6 and S7 and includes a portion of the common open reading frame. The 5′ end sequence is unique to S5 cDNAs. Oligonucleotide probes from cDNA S1 were used to analyze the internal portion of cDNA S5. No homology was detected using probes derived from the central portion of cDNA S1, demonstrating that the complete common open reading frame is not present in S5. The portion of cDNA S5 that has been sequenced is shown in Figure 7.

Transformation with the 12-kb *Eco*RI fragment: Although the 5′ and 3′ ends of some of the cDNAs lie outside of the 12-kb *Eco*RI fragment that contains the duplication, the entire coding region of cDNAs S1, S3, S6 and S7 lie within the 12 kb *Eco*RI fragment. Therefore in an effort to demonstrate directly that this region encodes *Sd* activity, *P* element-mediated transformation experiments were undertaken using a construct containing the 12-kb *Eco*RI fragment inserted in the CaSpeR vector (PIROTTA 1988). From 626 injected SD^+ embryos, we recovered four independent germline transformants, three with insertions of the 12-kb *Eco*RI fragment on chromosome *2* and another on the *X*. *P* element-mediated mobilization of some of the inserts resulted in secondary jumps, one from the *X* to chromosome *2* and the other from chromosome *2* to chromosome *3*. We confirmed by Southern blot analysis that all of these transformants did in fact contain the 12-kb *Eco*RI fragment.

We analyzed the *Rsp* locus of the four second chromosome transformant lines by comparing the recovery of the transformant second chromosome when heterozygous with an *SD* chromosome in males [***Sd?***,Sd^+ *E(SD)*$^+$ *Rsp?/Sd E(SD) Rsp*^{i}], and by hybridization of DNA from these flies with the cloned *Rsp*^{s} probe (WU *et al.* 1988). All four transformed second chromosomes carry *Rsp*^{s} as measured by both criteria so that each second chromosome was ***Sd?***,Sd^+ *E(SD)*$^+$ *Rsp*^{s}.

Each of the six insertion lines was tested for *Sd* activity. For inserts on the second chromosome the tested males were: ***Sd?***,Sd^+ *E(SD)*$^+$ *Rsp*^{s}/Sd^+ *E(SD) Rsp*^{i} and were measured for self-distortion of the *Rsp*^{s} homolog in the presence of the *Rsp*^{i} homolog. For the inserts on the *X* and the third chromosomes the tested males were: ***Sd?***,Sd^+ *E(SD)*$^+$ *Rsp*^{i}/Sd^+ *E(SD)*$^+$ *Rsp*^{ss} and were measured for the distortion of the supersensitive second chromosome relative to its insensitive homolog. The results of these tests are shown in Table 1. Control crosses were carried out to measure the corresponding segregation ratios in females. No significant distortion relative to control crosses was observed for any of the insertion lines. Although transcript analysis has not yet been carried out on these transformants, we would predict that they do not express the *Sd*-specific 4-kb transcript. Several of the 2-kb transcripts contain sequences from regions that lie outside the 12-kb *Eco*RI fragment and we expect that this will be true for the 4-kb transcript as well. We conclude that the 12-kb *Eco*RI fragment from *SD* chromosomes is not sufficient to produce detectable distortion.

Conservation of the duplication in other *Drosophila* species: The cloned 12-kb *Eco*RI fragment containing the duplication was used as a probe to hybridize to *Eco*RI digested genomic DNA from several *Drosophila* species under various hybridization stringency conditions. At a stringency that would allow hybridization between sequences that share at least 70% nucleotide identity, no specific hybridization was observed with *D. virilis* or *D. pseudoobscura* DNA. A single 7-kb *Eco*RI band was seen with the more closely

TABLE 1

Measuring distortion by *SD*$^+$ chromosomes bearing the 12-kb *Eco*RI fragment containing the *Sd* duplication

Chromosome	Relevant loci on each chromosome				k
	x	2^l	2^r	3	
a.	**Sd?**	*Rsp^i*	*Rsp^{ss}*		
1-91					0.502 ± 0.053
b.		**Sd?** *Rsp^s*	*Rsp^{ss}*		
2-52					0.507 ± 0.112
2-91					0.650 ± 0.113
2-102					0.504 ± 0.153
c.		*Sd$^+$ E(SD) Rsp^i*	**Sd?** *Rsp^s*		
2-52					0.575 ± 0.105
2-58					0.530 ± 0.062
2-91					0.492 ± 0.076
d.		*Rsp^i*	*Rsp^{ss}*	**Sd?**	
3-102					0.533 ± 0.038
e.		**Sd** *E(SD) Rsp^i*	*Rsp^s*		
SD-Mad					0.989 ± 0.005

Sd? represents the 12-kb *Eco*RI fragment containing the *Sd* duplication inserted by *P* element-mediated transposition as is shown in its respective chromosomal location by column. **SD** *E(SD) Rsp^i* represents the native *SD-Mad* chromosome, whose *k* value is presented for comparison. *k* is the proportion of progeny bearing the 2^l chromosome among the total progeny, measured as described in the MATERIALS AND METHODS. The specific crosses were: a, 1-**Sd?**, *CyO/Gla* males × *cn bw/cn bw* females; b, 2-**Sd?**/*RCy40* bw males × *lt pk cn bw/lt pk cn bw* females; c, 2-**Sd?**/*lt pk cn bw* males × *lt pk cn bw/lt pk cn bw* females; d, 3-**Sd?**, *CyO/lt pk cn bw* males × *cn bw/cn bw* females; and e, *SD-Mad/cn bw* males × *cn bw/cn bw* females (taken from BRITTNACHER and GANETZKY 1983).

related *D. simulans*, *D. mauritiana* and *D. sechellia* DNA. Conservation of the 7-kb *Eco*RI fragment in *D. melanogaster*, *D. simulans*, *D. mauritiana* and *D. sechellia* suggests that there has been no rearrangement in this region by insertion or deletion and that the sequences have not diverged by more than 30% since the last common ancestor of these four species.

DISCUSSION

Previous analysis of radiation-induced deletions of the *Sd* locus demonstrated that *Sd* is a neomorphic mutation, perhaps associated with the presence of some foreign DNA (GANETZKY 1977; BRITTNACHER and GANETZKY 1983). The deletion analysis also mapped the *Sd* locus to within several bands at 37D2-6. Those results provided the foundation for undertaking the present molecular analysis of *Sd*. A chromosome walk in the 37D2-6 region was carried out with the expectation that there would be a recognizable molecular difference between *SD* and *SD*$^+$ chromosomes in this region that would define the *Sd* locus. This expectation was realized in the present studies with the discovery of a length difference that uniquely distinguishes *SD* and *SD*$^+$ chromosomes. However, contrary to the suggestion that *Sd* might represent an insertion of some foreign piece of DNA, its origin appears endogenous to the Drosophila genome as a tandem duplication of a 5 kb genomic fragment that is normally present in a single copy in *SD*$^+$.

We propose that this tandem duplication corresponds to at least a portion of the *Sd* locus itself, on the basis of the following results. First, the duplication is present on every *SD* chromosome examined. Some of these chromosomes were recently obtained from nature and some have been in laboratory stocks for more than 30 years. They include *SD* chromosomes isolated from natural populations in North America, Australia, Europe and Asia. There is no reason to expect that these *SD* chromosomes would, without exception, share an identical "polymorphism" if it were unrelated to *Sd* itself. Second, the duplication has not been found on *SD*$^+$ chromosomes including those obtained from the same natural populations as the *SD* chromosomes. Third, the majority of radiation-induced *SD* revertants, a spontaneous revertant, two hybrid dysgenesis-induced revertants, and an EMS-induced revertant all have alterations involving this duplication. Taken together, these data provide a strong circumstantial case for the connection between *Sd* and the duplication. If the duplication does not define the *Sd* locus, then some other difference must be found in the 37D2-6 region whose pattern of presence or absence shows the same kind of correlation with *Sd* that we have found for the duplication. In screening virtually the entire region that contains *Sd* by Southern blot analysis, we have not detected any such *SD*-specific difference besides the duplication. Although we cannot eliminate the possibility that *Sd* is associated with an alteration too subtle to detect on Southern blots, the fact that *Sd* is a dominant neomorph that appears to have arisen only in nature leads us to believe that the alteration involves something more than a single nucleotide substitution.

If the duplication does correspond to *Sd* itself, the presence of the same duplication on *SD* chromosomes from around the world argues in favor of a single common origin for these chromosomes. Otherwise one must propose that there is something special about this region that promotes the occurrence of a particular breakage and reunion event. Sequence analysis of the junction fragments of the duplication should reveal the existence of any unusual features that could engender such an occurrence. In addition, sequence analysis of related portions of the duplication will indicate the evolutionary age of the duplication.

Assuming that the tandem duplication is causally associated with distortion, its phenotypic consequence is not due simply to a dosage effect of *SD*$^+$ because addition of *Dp(2;Y)G*, which includes the entire 37D region from *SD*$^+$, to otherwise normal flies does not produce distortion. Instead, it seems likely that it is the particular location of the duplication endpoints that produce the neomorphic activity of *Sd*. In that case, the gain-of-function phenotype of *Sd* is most likely to originate in one of two ways: either (1) the

production of a novel fusion protein, which could involve coding sequences contained entirely within the duplication or fusion of some coding information from the duplication with another unrelated gene; or (2) the ectopic expression or gross overexpression of some wild-type gene product. The dominant activation of some mammalian oncogenes by certain rearrangements (HALUSKA, TSUJIMOTO and CROCE 1987) and aberrant expression of the *Antp* gene product in the eye-antennal disk in Drosophila as the result of an inversion breakpoint (FRISCHER, HAGER and GARBER 1986) serve as examples for this possibility. Similar explanations have been proposed to explain the dominant neomorphic activity of mutant alleles of the *Bar* locus in Drosophila (TSUBOTA *et al.* 1989) and the *Knotted* locus in maize (VEIT *et al.* 1990).

Information bearing on these possibilities is provided by the results from northern blot and cDNA analysis. We found that in the region of the duplication, *SD* and SD^+ chromosomes express a similar, perhaps identical, family of 2-kb transcripts. In addition, we have identified at least one transcript, 4 kb in length, that is exclusively expressed by *SD* chromosomes. Because the 4-kb transcript is expressed only in *SD* flies, it is tempting to speculate that this transcript is responsible for distortion. For example, this transcript could encode the putative fusion protein. Alternatively, the coding information of the 4-kb transcript might be identical with that of the smaller transcripts also present in SD^+, in which case it could represent the transcript whose misexpression leads to the aberrant phenotype. It is also possible that the functional *Sd* protein is instead encoded by one of the smaller transcripts.

The demonstration that SD^+ chromosomes express a family of 2-kb transcripts related or identical to those expressed by *SD* and sharing common sequences with the *SD*-specific 4-kb transcript provides the first evidence for the existence of an Sd^+ product. Previously, Sd^+ could be defined only as the absence of *Sd*. We now presume that the Sd^+ locus is contained, at least partially, within the 5-kb fragment that is duplicated in *SD* and expresses a family of 2-kb transcripts. The biological function of Sd^+ remains unknown. Is its activity completely distinct from *Sd* or is it involved at some level with chromatin structure or chromosome segregation? Although answers to these questions are not yet forthcoming, they could provide essential clues concerning how *Sd* is able to exert its effect and how its activity is derived from Sd^+. A loss-of-function mutation of Sd^+ should be especially informative in this regard. The fact that the *Sd*-associated duplication appears to express the same family of 2-kb transcripts that is found in Sd^+, in addition to specifying at least one *SD*-specific transcript, suggests that the *Sd* locus has not lost the original Sd^+ activity in acquiring its new activity.

To learn more about the protein(s) encoded by the various transcripts derived from the *Sd* region, cDNAs were isolated from *SD* and SD^+ libraries. The *SD* cDNAs fall into seven classes that share a common 3′ region but differ at their 5′ ends. Some of these transcripts contain sequences derived from both parts of the duplication and must, therefore, be *Sd*-specific. Other cDNAs appear to correspond to transcripts found in SD^+ as well as *SD* flies. The seven classes of cDNAs represent four different open reading frames which are identical at the 3′ end and differ at their 5′ ends. The most frequent open reading frame encodes a 363 amino acid polypeptide that is identical to one encoded by an SD^+ cDNA. Searches of nucleotide sequence and amino acid sequence databases revealed no significant similarity of this polypeptide with any other known sequence. This polypeptide was found to contain four leucines arranged in heptad repeat motif similar to a leucine zipper. However, this region also contains amino acids that would prevent it from forming a strongly amphipathic helix so the potential significance of this motif is still uncertain.

The remaining open reading frames are encoded by cDNAs which include sequences derived from regions 5′ to the duplication. One class of cDNAs encodes a polypeptide 376 amino acids in length of which the final 339 are identical with residues 25–363 of the common polypeptide. The remaining 37 amino acids at the amino terminus are encoded by sequences located about 50 kb proximal to the duplication. Still another class a 5′ end derived from sequences located about 46 kb proximal to the duplication and spliced to a portion of the common sequence. The 5′ end of this transcript has no open reading frame before the junction with the common sequence. If translated, this transcript would encode a truncated protein consisting of the carboxy terminal 107 amino acids of the common polypeptide. These particular splice variants have not yet been isolated from SD^+ cDNA libraries, which we have screened less exhaustively than the *SD* libraries. However, we are inclined to believe that they are also present in SD^+ because there is an equally complex pattern of small transcripts seen on northern blots of *SD* and SD^+ flies and because it is hard to envision how the duplication could affect the choice of transcriptional start sites located 40–50 kb away.

We have not yet isolated a full-length cDNA that corresponds to the *SD*-specific 4-kb transcript. Because of the relative abundance of this transcript it seems very likely that some of the shorter incomplete cDNAs that were isolated represent portions of this transcript. However, we are not able to assemble a composite 4-kb sequence from any combination of these partial cDNAs. Thus, the nature of the polypeptide encoded by the 4-kb transcript and exactly how

it is related to the common polypeptide are still open questions.

The complexity of the transcripts derived from the region of the tandem duplication in *SD* and SD^+ makes it difficult to identify with certainty which among them, if any, is causally responsible for distortion. Because most models of segregation distortion have postulated a direct binding of an *Sd* polypeptide with the *Rsp* locus, we carried out gel shift experiments, but did not detect specific binding of a fusion protein derived from the S1 polypeptide to cloned Rsp^s DNA (our unpublished results). Either the expressed polypeptide does not bind directly to the Rsp^s repeat, or some aspect of this *in vitro* assay was inappropriate to detect such binding. Further experiments will be necessary to distinguish these possiblities.

Germline transformation experiments were undertaken in an attempt to make a functional connection between the tandem duplication and the *Sd* locus. Because the S1 cDNA is contained entirely within the tandem duplication, as are the complete coding regions of several of the other related cDNAs (S3, S6 and S7), we tested whether the dominant gain-of-function *Sd* phenotype could be conferred by this duplication. Six transformants in which the 12 kb *Eco*RI fragment containing the complete duplication had been introduced into SD^+ flies were examined for their ability to cause distorted segregation of a Rsp^s-bearing chromosome. Since significant distortion was not observed in any of these transformants, we conclude that the 12-kb *Eco*RI fragment alone is not sufficient to cause distortion. It is possible that all the relevant coding information is contained within this segment but it lacks some essential *cis*-acting regulatory sequences needed to produce the expression pattern responsible for distortion. Alternatively, distortion may require one of the spliced transcripts containing coding information derived from genomic sequences located far outside the duplication or some other transcript not yet identified. In the absence of compelling positive results from germline transformation, our identification of the tandem duplication as the *Sd* locus rests on circumstantial evidence. Thus, although we still believe that the tandem duplication is in some way causally related to distortion, final confirmation of this interpretation must await further investigation.

The topoisomerase II gene is located immediately adjacent to the tandem duplication in a region that is actually spanned by at least the transcripts corresponding to cDNAs S2, S4 and S5. This discovery raises intriguing questions about the potential involvement of topoisomerase II in segregation distortion. Topoisomerase II is the major structural component of the scaffolding that maintains DNA in an organized loop structure in metaphase chromosomes (GASSER *et al.* 1986) and is required for the condensation of eukaryotic chromosomes (ADACHI, LUKE and LAEMMLI 1991). Topoisomerase II binds primarily at the scaffold associated regions (ADACHI, KAS and LAEMMLI 1990), regions that are AT-rich and contain consensus topoisomerase II cleavage sites (GASSER *et al.* 1986). It is possible that topoisomerase II also physically interacts with the AT-rich Rsp^s region, although the outcome of such a binding is unknown since Rsp^s lacks sequences that resemble the consensus topoisomerase II cleavage site. Northern blots of RNA isolated from *SD* flies revealed no obvious alterations in the size or abundance of the topoisomerase II transcript compared with SD^+. Nor did we detect any physical alteration of the topoisomerase II gene in *SD* flies by Southern blot analysis. The possibility that *Sd* is associated, either directly or indirectly, with a more subtle alteration of topoisomerase II expression or function cannot be ruled out. However, if this is the case then this alteration of topoisomerase II has no overt somatic or germline effects in homozygotes such as *SD-Mad* or *SD-V017*, which are homozygous viable and fertile. Furthermore, any interaction between the *Rsp* locus and topoisomerase II must not be a prerequisite for normal sperm development because Rsp^s- or Rsp^{ss}-bearing chromosomes are not differentially recovered from a chromosome whose entire *Rsp* region has been deleted when both are segregating in an SD^+ male, and because *Rsp*-deletion homozygotes. are still viable and fertile.

Clearly, many questions still remain concerning the underlying basis of segregation distortion. The present analysis of the *Sd* region now provides the molecular tools that will be necessary to obtain the answers.

We are indebted to V. PIROTTA for constructing the microdissected library which provided us with a molecular entry into the 37D2-6 region. We also thank RAYLA TEMIN, KATE LOUGHNEY and JAMES CROW and our colleagues in the Ganetzky laboratory for helpful discussions, and ROBERT KREBER and CHRISTINA PETERSEN for their skillful technical assistance. This work was supported by grant DMB 8709888 from the National Science Foundation. This is publication 3188 from the Laboratory of Genetics, University of Wisconsin, Madison. The sequence of the S1 cDNA has been submitted to GenBank under the accession number X60218.

LITERATURE CITED

ADACHI, Y., E. KAS and U. K. LAEMMLI, 1990 Preferential cooperative binding of DNA topoisomerase II to scaffold-associated regions. EMBO J. **8:** 3997–4006.

ADACHI, Y., M. LUKE and U. K. LAEMMLI, 1991 Chromosome assembly *in vitro*: topoisomerase II is required for condensation. Cell **64:** 137–148.

AUSUBEL, F. M., R. BRENT, R. E. KINGSTON, D. D. MOORE, J. A. SMITH, J. G. SEIDMAN and K. STRUHL (Editors), 1987 *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.

BRITTNACHER, J. G., and B. GANETZKY, 1983 On the components of segregation distortion in *Drosophila melanogaster*. II. Deletion mapping and dosage analysis of the *Sd* locus. Genetics **103:** 659–673.

BRITTNACHER, J. G., and B. GANETZKY, 1984 On the components of segregation distortion in *Drosophila melanogaster*. III. Nature of the *Enhancer of SD*. Genetics **197:** 423–434.

DEVEREAUX, J., P. HAEBERLI and O. SMITHIES, 1984 A compre-

hensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

FRISCHAUF, A.-M., H. LEHRACH, A. POUSTKA and N. MURRAY, 1983 Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. **170:** 827–842.

FRITSCHER, L. E., F. S. HAGER and R. L. GARBER, 1986 An inversion that disrupts the *Antennapedia* gene causes abnormal structure and localization of RNAs. Cell **47:** 1017–1023.

GANETZKY, B., 1977 On the components of segregation distortion in *Drosophila melanogaster*. Genetics **86:** 321–355.

GASSER, S. M., T. LAROCHE, J. FALQUET, E. BOY DE LA TOUR and U. K. LAEMMLI, 1986 Metaphase chromosome structure, involvement of topoisomerase II. J. Mol. Biol. **188:** 613–629.

GOLIC, K., 1990 Examination of a mutable system affecting the components of the segregation distorter meiotic drive system of *Drosophila melanogaster*. Genetics **125:** 51–76.

GUBLER, U., and B. J. HOFFMAN, 1983 A simple method and very efficient method for generating cDNA libraries. Gene **25:** 263–269.

HALUSKA, F., Y. TSUJIMOTO and C. M. CROCE, 1987 Oncogene activation by chromosome translocation in human malignancy. Annu. Rev. Genet. **21:** 321–345.

HARTL, D. L., and Y. HIRAIZUMI, 1976 Segregation Distortion, pp. 615–666 in *Genetics and Biology of Drosophila*, Vol 1b, edited by M. ASHBURNER and E. NOVITSKI. Academic Press, New York.

HARTL, D. L., Y. HIRAIZUMI and J. F. CROW, 1967 Evidence for sperm dysfunction as the mechanism of segregation distortion in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **58:** 2240–2245.

HIRAIZUMI, Y., and K. NAKAZIMA, 1967 Deviant sex ratio associated with segregation distortion in *Drosophila melanogaster*. Genetics **55:** 681–697.

HIRAIZUMI, Y., and A. M. THOMAS, 1984 Supressor systems of *Segregation Distorter* (*SD*) chromosomes in natural populations of *Drosophila melanogaster*. Genetics **106:** 279–292.

HUYNH, T. V., R. A. YOUNG and R. W. DAVIS, 1985 Constructing and screening cDNA libraries in λgt10 and λgt11, pp. 49–78 in *DNA Cloning, a Practical Approach*, Vol.1, edited by D. M. GLOVER. IRL Press, Oxford.

KARESS, R., 1985 P element mediated germ line transformation of Drosophila, pp. 121–141 in *DNA Cloning, a Practical Approach*, Vol. 2, edited by D. M. GLOVER. IRL Press, Oxford.

LANDSCHULZ, W. H., P. F. JOHNSON and S. L. MCKNIGHT, 1988 The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240:** 1759–1764.

LINDSLEY, D., and E. H. GRELL, 1968 *Genetic Variation in Drosophila melanogaster*. Carnegie Inst. Wash. Publ. 627.

LINDSLEY, D., and G. ZIMM, 1985–1990 The genome of *Drosophila melanogaster*. Drosophila Inform. Serv. **62, 64, 65, 67** and **68.**

LYTTLE, T. W., 1986 Additive effects of multiple *Segregation Distorter* (*SD*) chromosomes on sperm dysfunction in *Drosophila melanogaster*. Genetics **114:** 203–216.

LYTTLE, T. W., 1989 The effect of novel chromosome position and variable dose on the genetic behavior of the Responder (*Rsp*) element of the *Segregation Distorter* (*SD*) system of *Drosophila melanogaster*. Genetics **121:** 203–214.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MANIATIS, T., R. C. HARDISON, E. LACY, J. LAUER, C. O'CONNELL, D. QUOM, D. K. SIM and A. ESTRADIATIS, 1978 The isolation of structural genes from libraries of eukaryotic DNA. Cell **15:** 687–701.

MARTIN, D. W., and Y. HIRAIZUMI, 1979 On the models of segregation distortion in *Drosophila melanogaster*. Genetics **93:** 423–435.

NICOLETTI, B., G. TRIPPA and A. DEMARCO, 1967 Reduced fertility in *SD* males and its bearing on segregation distortion in *Drosophila melanogaster*. Atti Acad. Naz. Lincei **43:** 383–392.

NOLAN, J. M., M. P. LEE, E. WYCKOFF and T.-S. HSIEH, 1988 Isolation and characterization of the gene encoding Drosophila DNA topoisomerase II. Proc. Natl. Acad. Sci. USA **83:** 3664–3668.

O'SHEA, E. K., R. RUTKOWSKI and P. KIM, 1989 Preferential heterodimer formation by isolated leucine zippers from *fos* and *jun*. Science **243:** 538–542.

PIMPINELLI, S., and P. DIMITRI, 1989 Cytogenetic analysis of segregation distortion in *Drosophila melanogaster*: the cytological organization of the *Responder* (*Rsp*) locus. Genetics **121:** 765–772.

PIROTTA, V., 1988 Vectors for P-mediated transformation in Drosophila, pp. 437–456 in *Vectors, a Survey of Molecular Cloning Vectors and Their Uses*, edited by R. L. RODRIGUEZ and D. T. DENHARDT. Butterworths, Boston.

POOLE, S. J., L. M. KAUVAR, B. DREES and T. KORNBERG, 1985 The *engrailed* locus of Drosophila: structural analysis of an embryonic transcript. Cell **40:** 37–43.

SANDLER, L. Y., and K. GOLIC, 1985 Segregation distortion in Drosophila. Trends Genet. **1:** 181–185.

SANDLER, L. Y., Y. HIRAIZUMI and I. SANDLER, 1959 Meiotic drive in natural populations of *Drosophila melanogaster*. I. The cytogenetic basis of segregation-distortion. Genetics **44:** 232–250.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SANTAMARIA, P., 1986 Injecting eggs, pp. 159–173 in *Drosophila, a Practical Approach*, edited by D. B. ROBERTS. IRL Press, Oxford.

SHARP, C. B., A. J. HILLIKER and D. C. HOLM, 1985 Further characterization of genetic elements associated with the Segregation Distorter phenomenon in *Drosophila melanogaster*. Genetics **110:** 671–688.

SHIMAKAWA, E. S., 1987 A Molecular and population genetic study of Segregation distorter in *Drosophila melanogaster*. Ph.D. dissertation, University of Hawaii, Manoa.

SPRADLING, A. C., 1986 P element mediated transformation, pp. 175–196, in *Drosophila, a Practical Approach*, edited by D. B. ROBERTS. IRL Press, Oxford.

STELLER, H., and V. PIROTTA, 1985 A transposable P vector that confers selectable G418 resistance to Drosophila larvae. EMBO J. **4:** 167–171.

TEMIN, R. G., 1991 The independent distorting ability of the *Enhancer of Segregation Distortion, E(SD)*, in *Drosophila melanogaster*. Genetics. **128:** 339–356.

TEMIN, R. G., and M. MARTHAS, 1984 Factors influencing the effect of segregation distortion in natural populations of *Drosophila melanogaster*. Genetics **107:** 375–393.

TEMIN, R. G., B. GANETZKY, P. A. POWERS, T. W. LYTTLE, S. PIMPINELLI, P. DIMITRI, C.-I. WU and Y. HIRAIZUMI, 1991 Segregation Distortion in *Drosophila melanogaster*: genetic and molecular analyses. Am. Nat. **137:** 287–331.

TOKUAYSU, K. T., W. J. PEACOCK and R. W. HARDY, 1977 Dynamics of spermiogenesis in *Drosophila melanogaster*. VII. Effects of *Segregation Distorter* (*SD*) chromosome. J. Ultrastruct. Res. **58:** 96–107.

TSUBOTA, S., D. ROSENBERG, H. SZOSTAK, D. RUBIN and P. SCHEDL, 1989 The cloning of the *Bar* region and the *B* breakpoint in *Drosophila melanogaster*: evidence for a transposon-induced rearrangement. Genetics **122:** 881–890.

ULBER, M. E., 1983 Segregation Distortion in *Drosophila melanogaster*: Spermatid degeneration and dosage analysis of Sd^+. M.S. Thesis, University of Wisconsin, Madison.

VEIT, B., E. VOLLBRECHT, J. MATHERN and S. HAKE, 1990 A tandem duplication causes the *Kn1-0* allele of *Knotted*, a dominant morphological mutant of maize. Genetics **125:** 623–631.

WU, C.-I., and M. F. HAMMER, 1990 Molecular evolution of ultraselfish genes, in *Evolution at the Molecular Level*, edited by R. K. SELANDER, A. G. CLARK and T. S. WHITTAM. Sinuar Press, Sunderland, Mass.

WU, C.-I., T. W. LYTTLE, M. L. WU and G.-F. LIN, 1988 Association between a satellite DNA sequence and the *Responder* of Segregation Distorter in *D. melanogaster*. Cell **54:** 179–189.

YOUNG, R. A., and R. W. DAVIS, 1983 Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. USA **80:** 1194–1198.

Communicating editor: T. W. CLINE