Mutations Affecting Expression of the *rosy* Locus in *Drosophila melanogaster*

Chong Sung Lee,*1 Daniel Curtis,* Margaret McCarron,† Carol Love,† Mark Gray,* Welcome Bender* and Arthur Chovnick†

*Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115 and †Department of Molecular and Cell Biology, The University of Connecticut, Storrs, Connecticut 06268

Manuscript received October 14, 1986
Accepted February 12, 1987

ABSTRACT

The *rosy* locus in *Drosophila melanogaster* codes for the enzyme xanthine dehydrogenase (XDH). Previous studies defined a "control element" near the 5' end of the gene, where variant sites affected the amount of *rosy* mRNA and protein produced. We have determined the DNA sequence of this region from both genomic and cDNA clones, and from the *ry*−10 underproducer strain. This variant strain had many sequence differences, so that the site of the regulatory change could not be fixed. A mutagenesis was also undertaken to isolate new regulatory mutations. We induced 376 new mutations with 1-ethyl-1-nitrosourea (ENU) and screened them to isolate those that reduced the amount of XDH protein produced, but did not change the properties of the enzyme. Genetic mapping was used to find mutations located near the 5' end of the gene. DNA from each of seven mutants was cloned and sequenced through the 5' region. Mutant base changes were identified in all seven; they appear to affect splicing and translation of the *rosy* mRNA. In a related study (T. P. KEITH et al. 1987), the genomic and cDNA sequences are extended through the 3' end of the gene; the combined sequences define the processing pattern of the *rosy* transcript and predict the amino acid sequence of XDH.

The *rosy* locus in *Drosophila melanogaster* codes for the enzyme xanthine dehydrogenase (XDH). It is one of the few loci in Drosophila for which it is practical to map alleles relative to each other. *rosy* mutant larvae are killed if purine is added to the medium, so rare wild-type recombinants can be selectively recovered from crosses between mutant alleles. Such mapping by recombination has been used to generate a detailed genetic map of the locus, which is diagramed in Figure 1. The genetic map has been roughly divided into two regions, the control element and the structural element. The structural element was defined by the sites of several mutations that affect physical properties of the XDH enzyme, and which thus must be within the protein coding sequence (GELBART et al. 1974; GELBART, McCarron and CHOVNICK 1976). The control element was defined by sites in two wild-type variant strains that affect the amount of XDH protein made. The *ry*+ strain produces twice as much XDH protein and RNA, and the *ry*+10 strain half as much, as a standard wild-type stock. The sites responsible for these differences in transcriptional levels have been mapped to a region about 0.001 cM left of the structural element (CHOVNICK et al. 1976; McCARRON et al. 1979; CLARK et al. 1984). Recently, many mutations mapping to the structural region have been correlated with insertions or deletions in the DNA (COTE et al. 1986; CLARK et al. 1986; REARDON et al. 1987), and this has allowed a close alignment of the genetic and molecular maps (Figure 1). The genetic map in Figure 1 also shows the position of the adjacent lethal complementation group, called *l(3)S12*. The proximity of this neighboring gene to the nearest *rosy* structural mutations limits the leftward extent of the *rosy* control element to less than 3000 base pairs to the left of these mutant sites.

It has been our goal to describe in molecular terms cis-acting regulatory sequences in the *rosy* control element. The manageable size of the *rosy* control element makes DNA sequence analysis of several strains practical. However, we expected many DNA sequence differences in the *ry*+ and the *ry*+10 variant strains without being able to pinpoint those responsible for the differences in XDH level. Such variation was a problem, for example, in the analysis of Drosophila strains that make different amounts of the larval glue protein (MUSKAVITCH and HOGNESS 1982). In order to define sequences involved in control, it seemed essential to induce regulatory mutations on a defined background chromosome. A large mutagenesis was undertaken to induce new *rosy* alleles, and to screen these for potential regulatory mutations mapping to the control element.

The sequence differences in the mutants would not be easy to interpret without any knowledge of the mRNA sequence. Previous studies had identified the size and orientation of the *rosy* mRNA (COVINGTON, FLEENOR and DEVLIN 1984; COTE et al. 1986), but they provided only a rough position for the tran-
scribed sequence, and they gave no information on possible splicing. We were able to find rare rosy cDNA clones in cDNA libraries of high complexity, and we compared the cDNA and genomic sequences (see also KEITH et al. 1987). We were surprised to find that the region of the control element includes transcribed and translated sequences. Several new mutations in this region apparently affect splicing and translation of the rosy transcript. The designations of control element and structural element remain useful for describing the genetic mapping experiments, but we will discuss mutant sequence changes in terms of features of the transcriptional unit.

MATERIALS AND METHODS

ENU mutagenesis and HPP screen: Unmated, aged (7–10 days) $ry^+$ homozygous males were fed for 24 hr on a solution containing 0.2 mg/ml of 1-ethyl-1-nitosourea (ENU, from Pfaltz and Bauer) in 2% sucrose. Following a 24-hr recovery period on standard, yeasted Drosophila medium, the males were mated to P18/MKRS virgin females, at least 48 hr of age ($P18 = In(3L)^P + In(3R)^P18, Ubx ry^+$ $kar, e^7$; MKRS = $T(3)MKRS, M(3)s34$ $kar$ $S_b$). Matings of seven males and 20 females were carried out in half-pint cultures on 4-hydroxypyrazolo-(3,4-d)pyrimidine (allopurinol or HPP, from Sigma Chemical Co.) supplemented medium, using a concentration of 60 µg HPP/ml of medium. After 48 hr the males were removed, and the females rebrooded on fresh HPP medium each day for a total of 13 days of egg laying. Progeny were screened for flies with rosy eye color. This mutagenesis protocol gave 12–13% sex-linked lethals in preliminary trials.

The mutant $ry^{58}$ was derived in an earlier mutagenesis using 0.01% 1:2:3:4-diepoxybutane (DEB, from Aldrich Chemical Co.) in 2% sucrose as the mutagen.

Genetic mapping: Procedures for selective recovery of $ry^+$ flies in genetic recombination experiments on purine-supplemented food have been described elsewhere (MC-CARRON et al. 1979). Since mutants with different XDH activity levels required different levels of purine for effective killing, pilot experiments were done to titrate the purine for each mutant before large-scale mapping experiments were undertaken.

XDH enzyme and protein assays: XDH activity was measured in extracts of 2-day-old adult flies by measuring the change in fluorescence associated with the conversion of 2-amino-4-hydroxypteridine into isoxanthopterin (MCCARRON et al. 1979). Isoelectric forms of the enzyme were resolved on nondenaturing polyacrylamide gels and stained by an enzyme-linked reduction of nitroblue tetrazolium (MCCARRON et al. 1979). XDH protein was measured by crossed-line rocket immunoelectrophoresis in agarose gels, using rabbit anti-XDH antiserum. The procedure is based on that of MCCARRON et al. (1979), and is more sensitive than the earlier procedure. A 1% agarose gel is poured in three segments of the same width. In all three segments the agarose is buffered by 0.22 mM Tris-base, 0.15 M boric acid, 0.006 M EDTA (pH 8.4). The upper segment of the gel contains anti-XDH serum at a concentration of 0.9% and is 75 mm long. The middle segment contains 5% XDH antigen extract and is 1.5 mm long. Antigen extract is prepared by grinding 25 homozygous adult $ry^{58}$ flies in 0.25 ml of 0.1 M Tris-base pH 8.0 at $5^\circ$, and centrifuging to obtain a clear supernatant. The antigen-containing strip is placed 2 mm below the antisemur-containing slab. The third segment is a contact gel formed by pouring buffered agarose on both sides of the antigen strip. 4 mm diameter wells are cut in this segment 1 mm below the antigen strip. Samples (12 µl) of mutant or wild type extract (prepared as described above, using either 25 homozygous or 50 hemizygous flies) are placed in the wells and the enzyme is electrophoresed towards the anode, that is through the antigen strip and into the antisemur, for 21 to 24 hr at 120 V. The arcs of XDH enzyme-antibody precipitate are visualized by staining for XDH activity (MC-CARRON et al. 1979).

Isolation of rosy genomic DNA: Recombinant DNA libraries were constructed in the bacteriophage lambda vectors EMBL4 (FRISCHAUF et al. 1983), Sep6 (DAVIS, BOTTSTEIN and ROETH 1980), or BFI01 (N. NEFF, unpublished results). DNA from 25–75 adult flies (about 6–18 µg) was partially digested with EcoRI, size-fraccionated by sucrose gradient centrifugation, and ligated into purified lambda vector arms. The ligated DNA was packaged in vitro (HOMMA 1979), and plated onto bacterial strain KH802. Plaques were
screened by the method of Benton and Davis (1977), using a nick-translated pBR322 plasmid probe containing a 7.7-kb EcoRI fragment which extends leftward from the EcoRI site at coordinate 0 in the rosy gene (Figure 1). Phage clones containing rosy DNA were recovered from each of these libraries, and SalI or EcoRI fragments from the rosy locus (Figure 1) were subcloned from each into the plasmid vectors pBR322 or pEMBL9 (Dente, Cesaroni, and Cortese 1983). Methods for phage stock growth, preparation of phage DNA and genomic fly DNA, and Southern blot analysis were as described in Bender, Spierer, and Hogness (1983).

cDNA isolation: cDNA libraries made from D. melanogaster early and late third instar larval RNA in the vector lambda gt10 were kindly provided by L. Kauvar, B. Drees, S. Poole, and T. Kornberg (Poole et al. 1985). cDNA phages were plated onto bacterial strain K802, and plaques were screened using a nick-translated rosy fragment extending from the BclI site at -1837 to the EcoRI site at 0 (Figure 4). After plaque purification of the rosy cDNA containing phage, the cDNA insert was subcloned into the vector pEMBL9 for sequence analysis.

DNA sequencing: Fragments of the rosy upstream region generated by restriction enzyme digestion were subcloned into M13 (Messing 1983), or pEMBL vectors, and the DNA sequence determined by the dideoxy method of Sanger, Nicklen, and Couhon (1977). A few start sites for sequencing were generated by random cutting with DNase I (Hong 1982). We obtained dideoxy nucleotide mixes, M13 sequencing primer and Klenow fragment enzyme from New England Biolabs. Reactions were performed with [32P]dATP (New England Nuclear) as the radioactive nucleotide and electrophoresed on straight 6% or 8% polyacrylamide gels or on step gradient gels as described (Biggin, Gibson, and Hong 1983).

The ry+ sequence has been determined completely on both DNA strands, as indicated by the sequencing strategy in Figure 2. To facilitate the location of single base changes in the mutants, sequencing ladders of the same cloned fragments from several mutants and from ry+ were typically run side-by-side on the same gel.

**RESULTS**

**Mutagenesis:** We wished to generate regulatory mutations mapping near the 5′ end of the rosy locus. Thus, a mutagenesis and screening procedure was designed to find mutations that reduce the quantity of XDH produced without changing the physical properties of the enzyme. Such mutations were then mapped by recombination, and mutations mapping to the 5′ region were reserved for analysis. Flies with reduced levels of XDH were recognized by their high sensitivity to allopurinol (HPP) added to standard Drosophila food. HPP is an inhibitor of XDH (Pomales et al. 1963), and Keller and Glassman (1965) showed that large amounts of HPP in the food could produce phenocopies of the rosy eye color in wild-type flies. Subsequently, Clark, Hilliker, and Chovich (1979) were able to determine the minimum HPP concentrations which would give rosy eye color phenocopies in wild-type strains that have different levels of XDH enzyme. Pilot experiments established concentrations where most wild-type strains (including ry and ry+) had normal eye color, but the ry" strain (the control variant with about 50% of normal activity) had rosy eye color.

The ry+ wild-type chromosome was chosen as the target of the mutagenesis. It has a normal control region, in that the ry+ level of XDH approximates that of ry+ and most other wild-type chromosomes (McCarron et al. 1979). However, ry+ has an XDH enzyme that moves faster in electrophoresis than that
of any other wild-type allele tested (McCarron, Gelbart and Chovnicky 1974). In some mutants that produce unstable XDH molecules, the mutant monomers are stabilized by forming dimers with wild-type monomers (Clark, Hilliker, and Chovnicky 1986). When such mutants are induced on a ry+ background, they can be recognized in heterozygotes with a slow wild-type electromorph, since the XDH hybrid dimers are electrophoretically separated from the homodimers.

We used ENU as a mutagen. Relative to other alkylating agents like ethyl methanesulfonate (EMS), ENU is more efficient, produces fewer chromosomal rearrangements, and gives fewer mosaic offspring of treated flies (Vogel and Natarajan 1979a,b). ENU treated ry+ males were crossed to rosy mutant females, and their offspring were screened directly on HPP food for exceptional flies with rosy eye color.

**Initial tests of mutants:** Each fly with a rosy eye color recovered in the HPP screen was crossed back to the P18/MKRS tester stock, and the progeny were restested on normal food and on HPP food. From a total of 0.9 × 10^6 screened progeny, 376 confirmed rosy mutations were recovered. Of these, 136 displayed rosy eye color on standard food, and these were assumed to be null mutations. The other 240 showed partially or completely wild-type eye pigmentation; these apparently had partial XDH activity. Confirmed rosy mutants were crossed to flies carrying a large deletion for the rosy region. If the new mutation proved lethal over the deletion, it was crossed to smaller deletions and lethal point mutations in the neighboring complementation groups (Hilliker et al., 1980) in order to define the genetic basis of the lethality. Nine mutant chromosomes carried such lethals; five were new deletions extending into adjacent lethal complementation groups, and four were double mutants.

All of the partial activity mutants were then raised on food with varying levels of HPP to titrate the XDH level. A group of 81 mutants were set aside at this point because the partial XDH levels were so high that the purine levels required to kill mutant larvae would also select against wild type. Thus, we could not do fine structure recombination experiments on these mutants to determine if they mapped to the control element. We excluded this group of 81 as well as the five deletions, and examined further a total of 132 rosy null mutants and 158 partial activity mutants.

**Mutations affecting the XDH protein:** Several tests were used to identify and exclude mutants with altered XDH protein: (1) allele complementation. XDH-, rosy eye color mutants were examined in mutant heteroalleles with each other, and with an array of established allele-complementing mutations (Gelbart, McCarron and Chovnicky 1976) to look for the recovery of normal eye color. “Leaky” mutant alleles with normal eye color were examined for complementation in heteroallelic combinations by HPP titration; (2) production of cross-reacting material (CRM) in quantities greater than that expected from XDH activity levels. CRM was assayed with anti-XDH antibody in rocket-line electrophoresis tests (McCarron et al. 1979); (3) electrophoretic mobility shift of the mutant XDH; (4) thermolability of the mutant XDH; and (5) analysis of hybrid XDH dimers in heterozygotes with ry+13, our wild-type strain with XDH molecules of the slowest electrophoretic mobility. In some cases, negative complementation was seen in the gel pattern of mutant/ry+13 heterozygotes. The intensity of the ry+13 homodimer band was reduced, with no obvious hybrid dimer band. In these cases, we infer that the hybrid dimer is formed but is enzymatically inactive. Of the 152 new apparent null mutations, 76 were classified as protein alterations on the basis of these tests. Likewise, 123 of the 158 partial activity mutants were classified as protein alterations. Thus, 56 nulls and 35 partial activity alleles remained for further study as potential regulatory mutations.

**Fine structure mapping:** Mutations that possibly affected the control element were next mapped in fine structure recombination experiments. The mutations were mapped relative to ry606, ry406 and ry5, three alleles that map near the left boundary of the structural element (Figure 1) (Gelbart, McCarron and Chovnicky 1976). A total of 61 mutations have been mapped in experiments involving selections on 2.6 × 10^7 progeny.

Only three of these 61 mutants, ry519, ry538 and ry545, mapped to the control element region to the left of ry606. All three were partial activity alleles. The results of the mapping experiments with these alleles are shown in Table 1. All of the crossovers were characterized by the production of normal levels of XDH, and this enzyme had the electrophoretic mobility of the ry+5 enzyme. This mobility is expected since these mutations lie to the left of ry606, and the bulk of the protein-coding sequence in the recombinants is derived from the ry+5 chromosome. Conversions of all three mutations were recovered, as noted in Table 1.

Four alleles, ry523, ry538, ry545 and ry519, mapped close to ry606 near the left end of the structural element (Figure 1). However, the absence of crossover events leaves ambiguous the order of the mutant sites with respect to ry606 (Table 1). ry519 maps close to ry606, and the single conversion event demonstrates that the two lesions are separable. ry519 lies clearly left of ry5, thus placing it near the left end of the structural region. The ry545 experiment gave no recombinants or convertants, suggesting a tight linkage to ry606, and this mutant was not mapped further. ry523 gave no wild-type crossovers or conversions in large selections with each of the test alleles (Table 1). Of
particular interest is the $ry^{523} \times ry^{606}$ cross, which gave two conversions of $ry^{606}$. One of these conversions was a co-conversion of the electrophoretic site, $e^{507}$, which is heterozygous in the cross. Since $e^{507}$ is located to the right of $ry^{606}$ (Fig. 1), this implies that $ry^{523}$ is located either to the left of $ry^{606}$ or to the right of $e^{507}$. The failure to recover crossovers in all recombination tests of $ry^{523}$ with $ry^{606}$, $ry^{606}$ and $ry^{1}$ favors the location of $ry^{523}$ just to the left of $ry^{606}$. One conversion of $ry^{523}$ was recovered, but it had a purine-sensitive phenotype (Table 1). This suggested that the $ry^{523}$ chromosome carries two mutations in the rosy gene (see DISCUSSION). $ry^{523}$ is an allele from a previous screen using DEB as a mutagen. Since it was known to map near $ry^{606}$, it was included in this study. Several conversion events separate the $ry^{523}$ site from that of $ry^{606}$, but the absence of crossover events place the two sites close together (Table 1). The remaining 54 mutants which have been mapped fall to the right of $ry^{606}$ (data not shown), and have not been examined further.

**Mutant phenotypes:** The mutant $ry^{5182}$ fails to complement with other rosy alleles, and produces XDH enzyme with the heat stability and electrophoretic mobility characteristic of the $ry^{+}$ wild-type. Adult flies have wild-type red eye color, and the XDH activity level (Figure 3A) and anti-XDH CRM level (Figure 3B) are both about 35% of that associated with extracts of $ry^{+}$. Electrophoresis of enzyme from $ry^{5182}$/

### Table 1

<table>
<thead>
<tr>
<th>Parental alleles</th>
<th>Crossovers</th>
<th>Conver-</th>
<th>Conver-</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>tant $ry^{mx}$</td>
<td>tant $ry^{1}$</td>
<td>size</td>
</tr>
<tr>
<td>$ry^{mx} \times ry^{1}$</td>
<td>$kar \ ry^{1}$</td>
<td>$ry^{+} \ Ace^{mx}$</td>
<td>$ry^{+} \ Ace^{1}$</td>
<td>($\times 10^{4}$)</td>
</tr>
<tr>
<td>$ry^{mx}$</td>
<td>$ry^{1}$</td>
<td>$kar \ ry^{1}$</td>
<td>$kar \ Ace^{1}$</td>
<td></td>
</tr>
<tr>
<td>5208</td>
<td>606</td>
<td>5 (1.05)</td>
<td>0</td>
<td>2 (1.05)</td>
</tr>
<tr>
<td>5204</td>
<td>606</td>
<td>5 (1.05)</td>
<td>0</td>
<td>1 (1.05)</td>
</tr>
<tr>
<td>5182</td>
<td>606</td>
<td>3 (1.05)</td>
<td>0</td>
<td>2 (1.05)</td>
</tr>
<tr>
<td>5198</td>
<td>606</td>
<td>0</td>
<td>0</td>
<td>1 (1.05)</td>
</tr>
<tr>
<td>5198</td>
<td>5</td>
<td>3 (1.05)</td>
<td>0</td>
<td>4 (1.05)</td>
</tr>
<tr>
<td>545</td>
<td>606</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>558</td>
<td>606</td>
<td>0</td>
<td>0</td>
<td>2 (1.05)</td>
</tr>
<tr>
<td>523</td>
<td>606</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>525</td>
<td>406</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>523</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1 (1.05)</td>
</tr>
</tbody>
</table>

* The numbers in parentheses indicate the relative electrophoretic mobility of the native XDH homodimer in the given fly. The mobilities of the $ry^{+}$, $ry^{1}$ and $ry^{mx}$ enzymes (the background stocks from which the $ry^{606}$, $ry^{523}$ and $ry^{1}$ mutants were respectively derived) are 1.00, 1.02 and 1.00, respectively. The mobility of the $ry^{+}$ enzyme is 1.05.

* The 1.03 mobility of the XDH produced in this fly indicates that the $e^{507}$ site (see Figure 1) has been converted to $e^{606}$.

* This chromosome carries $kar^{2} \ ry^{606} \ Ace^{6232}$. The lethal $pic$ allele serves as a substitue marker in place of $Ace^{1}$.

* This convertant has a purine sensitive rosy phenotype (see text).

**Figure 3.**—XDH activity and CRM from first exon mutants. A, Data from fluorometric assays on extracts from the upstream mutants, $ry^{182}$ (O), $ry^{206}$ (O), and $ry^{204}$ (O). The $ry^{+}$ control extract serves as a control and the underproducer variant from the $ry^{+}$ strain serves as a reference standard for comparison. B, Rocket-line immuno-electrophoresis on extracts of $ry^{+}$, $ry^{+}$ and the upstream mutants. The ratio of CRM to enzyme activity is the same for $ry^{+}$, $ry^{+}$ and $ry^{182}$; the other two mutants do not make detectable XDH CRM.

$ry^{+13}$ heterozygotes gives a triple banded XDH pattern. The lower $ry^{182}$ band is less intense than the upper $ry^{+13}$ band, which appears normal (not shown), reflecting the cis-acting nature of the $ry^{182}$ lesion.

$ry^{204}$ and $ry^{5208}$ mutant flies, when grown on normal food, have an eye color intermediate between rosy and wild-type, and thus have some XDH activity. Both mutants showed barely detectable XDH activity in the fluorometric assay, and no detectable CRM in the rocket immunoelectrophoresis test (Figure 3). We estimate that both mutants have less than 5% of the wild-type activity. Both fail to complement with all
other alleles tested. Heat lability and electrophoretic mobility could not be tested. For both mutants, electrophoresis of extracts from heterozygotes with \( \text{ry}^{+13} \) shows only the slow \( \text{ry}^{+13} \) band with the expected intensity. 

\( \text{ry}^{598} \) has about 40% of the XDH activity of \( \text{ry}^{+5} \) and 40% of \( \text{ry}^{+5} \) CRM levels in adults. As in \( \text{ry}^{5182} \), the eyes of this mutant appear wild type. The enzyme has a normal \( \text{ry}^{+5} \) mobility of 1.05 and is not heat labile, and \( \text{ry}^{5182} \) does not complement other \( \text{rosy} \) alleles. By these analyses it appeared that both \( \text{ry}^{598} \) and \( \text{ry}^{5182} \) produce normal XDH molecules, but at levels of only 35–40% of wild type.

\( \text{ry}^{545} \), \( \text{ry}^{536} \) and \( \text{ry}^{523} \) all have a completely null \( \text{rosy} \) phenotype. They produce no XDH enzyme activity or CRM, and are noncomplementers (data not shown). At this stage in our analysis, the biochemical and genetic data for each of these seven alleles was consistent with their being regulatory mutants.

**Isolation of \( \text{rosy} \) genomic DNA and cDNAs:** The DNA of the \( \text{rosy} \) locus was cloned initially from recombinant bacteriophage libraries made from the Canton S and Oregon R wild-type stocks (BENDER, SPIERER and HOGNESS 1983). Neither stock was necessarily isogenic for the third chromosome, and neither would likely be identical in DNA sequence to the single third chromosome of the \( \text{ry}^{+5} \) stock. Thus it was essential to isolate \( \text{rosy} \) DNA from the \( \text{ry}^{+5} \) stock for sequence comparisons between wild-type and mutants.

The upstream region was sequenced first from the \( \text{ry}^{+5} \) DNA. To be sure that the sequence included the entire 5′ end of the locus, the sequence was initiated at the \( \text{PvuII} \) site near the left end of the 8.1-kb \( \text{SalI} \) fragment. Insertions into this site are known to inactivate the adjacent \( \text{ls12} \) locus, but to have no effect on \( \text{rosy} \) expression (CLARK and CHOVNICK 1986). The sequence was continued for approximately 3 kb through the \( \text{EcoRI} \) site near the middle of the 8.1-kb \( \text{SalI} \) fragment; we predicted that this \( \text{EcoRI} \) site would be within the \( \text{rosy} \) protein coding sequence from the correlation of genetic and molecular maps (Figure 1). The sequence was determined completely in both directions as outlined in Figure 2, and it is shown in Figure 4. The site of the \( \text{EcoRI} \) restriction cut was arbitrarily designated as position 0 in numbering the sequence.

The sequence around the \( \text{EcoRI} \) site has one continuous open reading frame from left to right (Fig. 4). This open frame continues for about 3.4 kb beyond the \( \text{EcoRI} \) site [excluding introns—see KEITH et al. (1987) in this issue]. This reading frame spans mutation sites known to be within the \( \text{rosy} \) structural element (COTE et al. 1986).

In order to define the transcribed sequences, we searched Drosophila cDNA libraries for clones homologous to \( \text{rosy} \) genomic DNA. We found \( \text{rosy} \) cDNAs in the libraries from L. KAUVAR and T. KORNBERG, made from poly(A)+ RNA from various stages of development (POOLE et al. 1985). We isolated 7 cDNA clones homologous to the \( \text{rosy} \) locus, from the third instar libraries I4, H2, and H3. Six of these had homology only to sequences to the right of the \( \text{EcoRI} \) site (position 0); these are described in the accompanying paper (KEITH et al. 1987). The remaining clone, designated I4-31, was homologous to two separate regions to the left of the \( \text{EcoRI} \) site. The comparison of cDNA and genomic sequences revealed two exons separated by an 815 bp intron, as diagramed in Figure 2, and as shown in detail in Figure 4. We suspect the left end of the clone falls short of the 5′ end of the \( \text{rosy} \) mRNA, so the transcription start site remains to be determined. The right end of the clone is exactly at the \( \text{EcoRI} \) site, which is probably a construction artifact (if the double-stranded cDNA was not completely methylated at internal \( \text{EcoRI} \) sites, or if restriction enzyme contaminated the methylase preparation, these sites would be cut).

A continuous open reading frame in the cDNA sequence begins with the ATG codon at −1407 in the first exon. The open frame continues across the splice junction (−1366 to −551) to join with the large open reading frame encoding the XDH protein. The left edge of exon 2 apparently coincides with the left end of the genetically defined structural element (Figure 1). The previous mapping of the structural element did not detect the protein-coding sequences in exon 1.

We also cloned and sequenced DNA from the seven upstream mutants and the \( \text{ry}^{+10} \) control variant. The extent of the DNA sequenced from each genotype is indicated in Figure 2. Sequencing reactions on subclones of \( \text{ry}^{+5} \) and one or more other alleles were usually run in parallel, in adjacent lanes of a sequencing gel, so that sequence differences were readily apparent. As expected, there were numerous differences between the \( \text{ry}^{+10} \) and the \( \text{ry}^{+5} \) wild types (25 differences in 1845 bp, or about 1.4% variation). Most of the mutants had only a single change. The sequence changes of each of these mutations and of \( \text{ry}^{+10} \) are shown in Figure 4. The three mutations which mapped furthest left, \( \text{ry}^{5208} \), \( \text{ry}^{5182} \) and \( \text{ry}^{5204} \), were all found within the upstream exon defined by the I4-31 cDNA clone. \( \text{ry}^{5208} \) and \( \text{ry}^{5182} \) each have a single guanine to adenine transition. \( \text{ry}^{5204} \) has a single base change and a single base deletion within a seven base stretch of the upstream exon. The exact site of the deleted base is uncertain since one base is lost from a run of five Ts. A sequencing gel comparing \( \text{ry}^{+5} \) with the three mutants in the first exon is shown in Figure 5.

The remaining four mutations, which mapped close to \( \text{ry}^{606} \), were found within or adjacent to the second exon (Figure 4). Two mutations, \( \text{ry}^{522} \) and \( \text{ry}^{545} \), both change the last guanine of intron 1 to an adenine.
phenylalanine) in the XDH protein. This results in an amino acid substitution (serine to threonine) in the full-length protein sequence. The single nucleotide change ATG to GTA at position 5198 in the ry+5 sequence has a single cytosine to thymidine base change, which results in an amino acid substitution (serine to phenylalanine) in the XDH protein. ry+5 has a 56 base pair deletion within the protein coding sequences of the second exon. This deletion would cause a translation frameshift.
A major goal of our studies on the rosy locus has been to determine the DNA sequence of the genetically defined control element, and to determine its position relative to the rosy mRNA. This has now been accomplished, and we find that the first protein coding exon of the mRNA lies in the middle of the control element region. We have not yet defined the exact molecular distance between the two exons (Figure 1). Indeed, preliminary molecular mapping suggests that the $\text{rosy}^{5208}$ lesion is about 0.5 kb inside the exon, or about 1.3 kb from the base change in $\text{rosy}^{5182}$, $\text{rosy}^{5204}$ and $\text{rosy}^{5208}$ (M. Gray and W. Bender, unpublished observations). Since we have sequenced nearly 2 kb of the upstream DNA from $\text{rosy}^{5204}$ and $\text{rosy}^{5182}$, there is little doubt that the single alteration found in each case is the mutation responsible for the phenotype. Less DNA was sequenced from $\text{rosy}^{5208}$, but the mapping data agrees with its position near $\text{rosy}^{5204}$ and $\text{rosy}^{5182}$. The location of all three of these mutations within the first exon and their predicted in vivo effects (discussed below) suggest that no additional lesions contribute to their phenotypes.

As described earlier, the three mutants $\text{rosy}^{5198}$, $\text{rosy}^{545}$ and $\text{rosy}^{538}$ each mapped as single site mutations located close to $\text{rosy}^{606}$. The mapping data from $\text{rosy}^{523}$, however, suggested that this allele might be a double mutant. The $\text{rosy}^{523}$ lesion was separable by gene conversion from each of the three markers it was mapped against, yet it gave no crossovers and was not itself converted to wild type in the three experiments. One exceptional chromosome arose in the experiment with $\text{rosy}^{5}$ (Table 1). The fly carrying this chromosome had a purine-sensitive rosy eye color phenotype, unlike the rosy null phenotype of either parent. The chromosome still carried the $\text{rosy}^{523}$ parental flanking markers, indicating a partial conversion of $\text{rosy}^{523}$. We interpret these data to mean that one of two separable $\text{rosy}^{523}$ mutant sites was converted to wild type, with the remaining unconverted site causing the purine sensitive phenotype. We have also found molecular evidence that $\text{rosy}^{523}$ is indeed a double mutant. In addition to the transition at $\text{rosy}^{523}$, there is another change, not yet sequenced, in the second exon to the right of the EcoRI site (M. Gray and W. Bender, unpublished observations). The $\text{rosy}^{545}$ chromosome does not carry this second mutation. This establishes that the $\text{rosy}^{523}$ and $\text{rosy}^{545}$ chromosomes represent independent isolations of an identical lesion at the splice junction site, and furthermore argues that this one lesion they have in common is sufficient to cause the XDH null phenotype seen in both mutants.

DISCUSSION

A major goal of our studies on the rosy locus has been to determine the DNA sequence of the genetically defined control element, and to determine its position relative to the rosy mRNA. This has now been accomplished, and we find that the first protein coding exon of the mRNA lies in the middle of the control element region. We have not yet defined the exact DNA sequence differences responsible for the altered rosy expression in the $\text{rosy}^{+4}$ and $\text{rosy}^{+10}$ control variant strains, but we have induced, mapped, and sequenced new point mutations in the region. These mutations fall in the exon sequences or adjacent splice junction sequences. The mutant sites, together with the cDNA structure, allow us to identify sequences necessary for RNA processing and translation initiation. None of the new point mutations analyzed to date appear to affect the initiation of transcription.

Mutant positions: The sequence differences found in each of the seven mutations are summarized in Figure 6. Based on sequence coverage and the genetic mapping data, we are confident that the identified DNA sequence changes are in fact the lesions responsible for the rosy mutant phenotypes.

The three mutations $\text{rosy}^{5182}$, $\text{rosy}^{5204}$ and $\text{rosy}^{5208}$ each map clearly to the left of the structural element marker $\text{rub}^{606}$, with the genetic distance between the mutants and $\text{rub}^{606}$ ranging from 0.001 to 0.003 cM. Based on previous correlations between genetic and molecular distance (Bender, Spierer and Hogness 1983; Cote et al. 1986), we would estimate the molecular distance between $\text{rosy}^{606}$ and these mutants to be between 0.7 and 2.1 kb. The recombination map predicts that the $\text{rosy}^{606}$ lesion is near the left end of the second exon (Figure 1). Indeed, preliminary molecular mapping suggests that the $\text{rosy}^{606}$ lesion is about 0.5 kb inside the exon, or about 1.3 kb from the base changes in $\text{rosy}^{5182}$, $\text{rosy}^{5204}$ and $\text{rosy}^{5208}$ (M. Gray and W. Bender, unpublished observations). Since we have sequenced nearly 2 kb of the upstream DNA from $\text{rosy}^{5204}$ and $\text{rosy}^{5182}$, there is little doubt that the single alteration found in each case is the mutation responsible for the phenotype. Less DNA was sequenced from $\text{rosy}^{5208}$, but the mapping data agrees with its position near $\text{rosy}^{5204}$ and $\text{rosy}^{5182}$. The location of all three of these mutations within the first exon and their predicted in vivo effects (discussed below) suggest that no additional lesions contribute to their phenotypes.
introns to terminate at +3762 (KEITH et al. 1987). The open reading frame could encode a polypeptide of 1335 amino acids with a molecular weight of approximately 147,000. This is in close agreement with the 150-kD size of XDH monomers predicted from polyacrylamide gel electrophoresis (EDWARDS, CANDIDO and CHOVNICK 1977). There is one ATG codon upstream of the −1407 ATG within the cDNA sequence (at −1505), and since we have not yet fixed the start site for transcription, there may be more upstream ATGs. Preliminary S1 and primer extension data suggest that the 5′ end of the rosy mRNA lies approximately 100 bp upstream of the 5′ end of the cDNA. In any case, there are stop codons in all frames in the sequence following the −1505 ATG and preceding the −1407 ATG. Typically, eukaryotic ribosomes will initiate translation at the first methionine codon in an mRNA, but there are many examples of internal initiation [for discussion, see PEABODY and BERG (1986)]. The use of an internal ATG codon can reduce the efficiency of proper initiation.

**Effects of the mutations:** The mutations which we have characterized affect splicing and translation of the rosy message or the sequence of the XDH polypeptide, and are summarized in Figure 6. We will discuss them in their sequence order from left to right.

The **ry5182** lesion may hinder proper translation of the message. The G to A transition at −1435 creates a new ATG initiation codon upstream of the −1407 ATG. The ATG produced by the **ry5182** mutation is in a sequence context very similar to the consensus (ATACATGG). In **ry5182**, translation of the message should preferentially begin at the new ATG upstream of the −1407 start and proceed out of frame with respect to the long open reading frame. Termination of translation would occur at the stop codon at −1402. Under the ribosomal scanning model of KOZAK (1978, 1984b), ribosomal reinitiation at the proper start site should not occur since the −1402 stop site is downstream of the −1407 start. However, recent results show that reinitiation may occur at an internal ATG if translational termination occurs as much as 50–90 bp downstream of the ATG (PEABODY and BERG 1986; PEABODY, SUBRAMANI and BERG 1986). Since **ry5182** flies have 35% of normal XDH activity and CRM, ribosomes could be reinitiating at the correct ATG initiation codon, but the mutation significantly reduces the efficiency of proper initiation.

The **ry5204** mutation causes a translation frameshift near the amino terminal end of the XDH protein. The frameshift occurs only 8 codons past the −1407 ATG, with a TGA stop codon following in the new frame only 2 codons downstream. We speculate that the trace amount of XDH produced in **ry5204** flies is a result of translational reinitiation events further downstream that produce a small amount of an enzymatically active amino-truncated product. The effects of the **ry5204** and **ry5182** mutations support the assignment of the −1407 ATG as the XDH translational initiation codon.

The **ry5208** mutation at −1366 falls within the XDH open reading frame in the last base of the first exon.
It is at the third position of a codon and does not change the amino acid sequence of the protein product. The mutation might, however, affect the efficiency or accuracy of splicing of the adjacent intron. Although the sequence of the exon at the exon/intron border is not invariant in eukaryotic transcripts, there is a strong consensus for AG as the last two bases of an exon (Mount 1982). In a survey of 35 splice sites in 22 Drosophila genes [references in Keller and Noon (1985) and our compilation], we find no strong conservation at the −2 position, but a strong preference for guanine (63%) over adenine (9%) at −1.

A guanine to adenine transition mutation in the last base of an exon, identical to that occurring in ry5208, has been synthesized in vivo in a cloned rabbit β-globin gene (Weiranga et al. 1983), but no effect on splicing was observed in a transfection assay in HeLa cells. In contrast, mutations in the conserved splice donor sequences within the intron, in human globin genes, result in the activation of cryptic splice junction sequences and improper splicing of the globin message (Treisman, Orkin and Maniatis 1983). Likewise, a mutation in the first base of an exon next to a splice acceptor site can block proper splicing in vivo (Mitchell, Urlaub and Chasin 1986). We suspect that the ry5208 mutation at the last base within the Drosophila exon might have an effect similar to the splice donor mutations in the β-globin introns. Preliminary evidence suggests that the rosy mRNA produced in ry5208 is shorter by 150–200 bp than the wild-type message (D. Curtis and W. Bender, unpublished results). A shorter ry5208 message is consistent with the use of a cryptic 5′ splice junction to the left (5′) of the normal site, or the selection of a different 3′ splice acceptor site. The trace amount of XDH produced in this mutant might be synthesized from rare, properly processed messages, or from translational initiation within the open reading frame downstream from the normal start and first intron splice sites. Since ry5204 and ry5208 both make only trace amounts of XDH, we think that there are no major alternate forms of XDH mRNA which lack the exon including these mutant sites.

The G to A transition in ry523 and ry545 occurs at the last base (−551) of the intron. Eukaryotic introns usually end with the dinucleotide sequence AG (Mount 1982; Keller and Noon 1985). These bases are apparently essential for proper excision of the intron; alteration of the guanine nucleotide to a cytosine completely prevents in vitro splicing of the first intron in the human β-globin gene (Reed and Maniatis 1985). We expect that proper splicing of the large intron in the rosy message is eliminated in the ry545 and ry523 null mutants. Since these alleles make no detectable XDH, there must not be any alternate forms of XDH mRNA which do not use this splice acceptor site.

The theoretical translation of the ry5198 sequence indicates that one serine amino acid residue in the wild type is changed to a phenylalanine in the mutant. Since the mutant retains a wild-type phenotype on normal food and has 40% of the normal level of XDH CRM, it is possible that this single amino acid change partially destabilizes the enzyme, and causes abnormally rapid turnover of the molecule.

The effect of the ry538 lesion on rosy expression is straightforward: a deletion of 56 bp within the protein coding region results in both a loss of part of the coding sequence and a translational frame shift. The null phenotype of this mutant indicates that any translational product that might be produced from the aberrant message has no XDH enzymatic activity. The ry538 allele was induced by diepoxybutane, a mutagen which frequently generates small deletions in other systems (Bird and Fahmy 1953; Reardon et al. 1987).

In our preliminary northern blot analysis, none of these seven alleles shows an obvious drop in the amount of rosy mRNA. This confirms our belief that the loss of XDH activity seen in each mutant is not due to reduced message levels.

**Transcriptional control:** We are unable to tell which of the 25 variant sites in the ry510 sequence might be responsible for the two-fold reduction of rosy RNA levels in the ry510 strain. None of the variant sites correspond to any of the mutant lesions, and none would cause a predictable change in processing or translation of ry510 rosy RNA. Genetic mapping data places the ry510 controlling element between the adjacent l(3)S12 gene and the ry54 control site (Clark et al. 1984) (see Figure 1). The ry54 site has been seen to co-convert with a polymorphic restriction site within the large first intron (S. Clark, unpublished data). Thus, we expect the ry54 transcriptional control element to lie within the transcribed sequences of the first intron, and the ry510 control site to lie in the nontranscribed DNA upstream of the first exon. The genetically defined “control region” therefore overlaps rosy transcribed sequences. In half-tetrad mapping experiments between the control element variants ry54 and ry510, exceptional flies were recovered carrying crossovers between the ry510 and ry54 control element sites, and others were recovered carrying conversions of each of the sites (Clark et al. 1984). Work is in progress to locate on the DNA sequence the approximate sites of crossing over in these genetic events (S. Clark and A. Chovnick, unpublished data). This will allow us to place limits on which of the many differences in the ry510 sequence (relative to ry54) are functionally significant in the control of XDH expression.

The point mutations we have sequenced all fall within or adjacent to rosy exons, and affect message splicing, translation, or the amino acid sequence of the protein. None are likely to affect transcriptional
control. It is possible that there are few, if any, regulatory sequences for the rosy gene which can be altered by a single base change to result in a large reduction of expression. We chose initially to work with a subset of our point mutants with at least a threefold reduction in XDH levels in order to permit genetic mapping experiments with the purine selection. Although some mutants were recovered with smaller effects, this selection constraint biased our sequencing studies towards the stronger mutants. The mutants we found have helped define the limits of the rosy transcript, and have pointed out structural features of the rosy message. In order to define sequences responsible for transcriptional control of the rosy gene, our future work with point mutations will focus on alleles with relatively subtle effects. We also hope to generate small deletions in the region surrounding the first exon, which may have more dramatic effects on transcription.

We thank Jessica Moore for assistance in refining the DNA sequence. This work was supported by a fellowship to M.G. from the Damon Runyon-Walter Winchell Cancer Fund, by grants from the National Institutes of Health to W.B. and A.C., and by a gift to W.B. from the C. Marran Foundation.

LITERATURE CITED

BENDER, W., P. SPiRER and D. S. HOGNESS, 1983 Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the Bithorax Complex in Drosophila melanogaster. J. Mol. Biol. 168: 17–33.


KOZAK, M., 1986 Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eucaryotic ribosomes. Cell 44: 283–292.


MCCARBERT, M. W. M. GELBART and A. CHOVNICK, 1974 Intracistronic mapping of electrophoretic sites in Drosophila


