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 diagrammed 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 CHOVIK 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 (CHOVIK *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-

¹ Present address: Department of Zoology, University of Texas, Austin, Texas 78712.

The sequence data presented in this article have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00307.

FIGURE 1. Comparison of genetic and molecular maps of the *rosy* locus. The genetic map is shown on the top line, with the two sites defining the control element (*i1005* and *i409*) mapping to the left of the mutations defining the structural element. A restriction map of the 8.1-kb *SalI* fragment covering the *rosy* locus is shown in the middle. The lines between the two map lines indicate the relative positions of mutant lesions, as described in CORE *et al.* (1986). Below the molecular map is a diagram of the *rosy* mRNA, as deduced from cDNA clones (reported here and in KEITH *et al.* 1987).

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-nitrosourea (ENU, from Pfaltz and Bauer) in 2% sucrose. Following a 24-hr recovery period on standard, yeast *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*⁴¹ *kar*, *e*⁺; *MKRS* = *Tp(3)MKRS*, *M(3)S34 kar ry*² *Sb*). 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*⁵³⁸ 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 (McCARRON *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 M 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*⁺ flies in 0.25 ml of 0.1 M Tris-base pH 8.0 at 5°, and centrifuging to obtain a clear supernatant. The antigen-containing strip is placed 2 mm below the antiserum-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 antiserum, for 21 to 24 hr at 120 V. The arcs of XDH enzyme-antibody precipitate are visualized by staining for XDH activity (McCARRON *et al.* 1979).

Isolation of *rosy* genomic DNA: Recombinant DNA libraries were constructed in the bacteriophage lambda vectors EMBL4 (FRISCHAUF *et al.* 1983), Sep 6 (DAVIS, BOSTEIN and ROTH 1980), or BF101 (N. NEFF, unpublished results). DNA from 25–75 adult flies (about 6–18 µg) was partially digested with *EcoRI*, size-fractionated by sucrose gradient centrifugation, and ligated into purified lambda vector arms. The ligated DNA was packaged *in vitro* (HOHN 1979), and plated onto bacterial strain KH802. Plaques were

FIGURE 2.—Sequencing strategy. The top line represents the DNA sequence with restriction sites and mutant lesion sites indicated. Below the sequence line is a diagram of the two exons present in the I4-31 cDNA. The horizontal arrows in the middle of the diagram indicate the directions and extents of sequencing runs done on subclones of *ry*⁺ DNA. The open bars at the bottom show the extents of DNA sequence read from the indicated variant or mutant chromosomes.

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, CESARINI, 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 KH802, 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 COULSON (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 CHOVNICK (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*⁺0 and *ry*⁺5) had normal eye color, but the *ry*⁺10 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*⁺0 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 CHOVNICK 1974). In some mutants that produce unstable XDH molecules, the mutant monomers are stabilized by forming dimers with wild-type monomers (CLARK, HILLIKER, and CHOVNICK 1986). When such mutants are induced on a ry^{+5} 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^{+5} 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 retested 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 heteroallele combinations with each other, and with an array of established allele-complementing mutations (GELBART, MCCARRON and CHOVNICK 1976) to look for the recovery of normal eye color. "Leaky"

mutant alleles with normal eye color were examined for complementation in heteroallele 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 132 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 ry^{606} , ry^{406} and ry^5 , three alleles that map near the left boundary of the structural element (Figure 1) (GELBART, MCCARRON and CHOVNICK 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, ry^{5182} , ry^{5204} and ry^{5208} , mapped to the control element region to the left of ry^{606} . 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 ry^{606} , 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, ry^{523} , ry^{538} , ry^{545} and ry^{5198} , mapped close to ry^{606} 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 ry^{606} (Table 1). ry^{5198} maps close to ry^{606} , and the single conversion event demonstrates that the two lesions are separable. ry^{5198} lies clearly left of ry^5 , thus placing it near the left end of the structural region. The ry^{545} experiment gave no recombinants or convertants, suggesting a tight linkage to ry^{606} , and this mutant was not mapped further. ry^{523} gave no wild-type crossovers or conversions in large selections with each of the test alleles (Table 1). Of

TABLE 1

Number and classes of *rosy*⁺ chromosomes recovered from progeny of crosses of the type: *ry*^{5XX}/*kar*² *ry*⁺ *Ace*¹²⁶ females × *P18/MKRS* tester males

Parental alleles		Crossovers		Convertant <i>ry</i> ^{5XX}	Convertant <i>ry</i> ⁺	Sample size (× 10 ⁻⁶)
<i>ry</i> ^{5XX}	<i>ry</i> ⁺	<i>kar ry</i> ⁺	<i>ry</i> ⁺ <i>Ace</i> ¹²⁶	<i>ry</i> ⁺	<i>kar ry</i> ⁺ <i>Ace</i> ¹²⁶	
5208	606	5 (1.05) ^a	0	2 (1.05)	3 (1.00)	0.51
5204	606	5 (1.05)	0	1 (1.05)	3 (1.00)	0.30
5182	606	3 (1.05)	0	2 (1.05)	2 (1.00)	0.61
5198	606	0	0	1 (1.05)	0	0.49
5198	5	3 (1.05)	0	4 (1.05)	5 (1.00)	0.37
					1 (1.03) ^b	
545	606	0	0	0	0	0.36
538	606	0	0	2 (1.05)	3 (1.00)	0.44
523	606	0	0	0	1 (1.00)	0.98
					1 (1.03) ^b	
523	406 ^c	0	0	0	1 (1.02)	0.34
523	5	0	0	1 (1.05) ^d	3 (1.00)	0.31

^a The numbers in parentheses indicate the relative electrophoretic mobility of the native XDH homodimer in the given fly. The mobilities of the *ry*⁶⁰⁶, *ry*⁴⁰⁶ and *ry*⁵ enzymes (the background stocks from which the *ry*⁶⁰⁶, *ry*⁴⁰⁶ and *ry*⁵ mutants were respectively derived) are 1.00, 1.02 and 1.00, respectively. The mobility of the *ry*⁵ enzyme is 1.05.

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

^c This chromosome carries *kar*² *ry*⁴⁰⁶ *pic*¹⁶²³. The lethal *pic* allele serves as a substitute marker in place of *Ace*¹²⁶.

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

particular interest is the *ry*⁵²³ × *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*⁴⁰⁶ and *ry*⁵ 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*⁵³⁸ is an allele from a previous screen using DEB as mutagen. Since it was known to map near *ry*⁶⁰⁶, it was included in this study. Several conversion events separate the *ry*⁵³⁸ 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*⁵ 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*⁵. Electrophoresis of enzyme from *ry*⁵¹⁸²/

FIGURE 3.—XDH activity and CRM from first exon mutants. A, Data from fluorometric assays on extracts from the upstream mutants, *ry*⁵¹⁸² (◇), *ry*⁵²⁰⁴ (○), and *ry*⁵²⁰⁸ (●). The *ry*⁵ (□) extract serves as a control and the underproducer variant from the *ry*¹⁰ (■) strain serves as a reference standard for comparison. B, Rocket-line immunoelectrophoresis on extracts of *ry*⁵, *ry*¹⁰ and the upstream mutants. The ratio of CRM to enzyme activity is the same for *ry*⁵, *ry*¹⁰ and *ry*⁵¹⁸²; the other two mutants do not make detectable XDH CRM.

*ry*¹³ heterozygotes gives a triple banded XDH pattern. The lower *ry*⁵¹⁸² band is less intense than the upper *ry*¹³ band, which appears normal (not shown), reflecting the *cis*-acting nature of the *ry*⁵¹⁸² lesion.

*ry*⁵²⁰⁴ 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 ry^{+13} shows only the slow ry^{+13} band with the expected intensity.

ry^{5198} has about 40% of the XDH activity of ry^{+5} and 40% of ry^{+5} CRM levels in adults. As in ry^{5182} , the eyes of this mutant appear wild type. The enzyme has a normal ry^{+5} mobility of 1.05 and is not heat labile, and ry^{5198} does not complement other *rosy* alleles. By these analyses it appeared that both ry^{5198} and ry^{5182} produce normal XDH molecules, but at levels of only 35–40% of wild type.

ry^{545} , ry^{538} and ry^{523} all have a completely null *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 *rosy* genomic DNA and cDNAs: The DNA of the *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 ry^{+5} stock. Thus it was essential to isolate *rosy* DNA from the ry^{+5} stock for sequence comparisons between wild-type and mutants.

The upstream region was sequenced first from the ry^{+5} DNA. To be sure that the sequence included the entire 5' end of the locus, the sequence was initiated at the *Pst*I site near the left end of the 8.1-kb *Sal*I fragment. Insertions into this site are known to inactivate the adjacent *IS12* locus, but to have no effect on *rosy* expression (CLARK and CHOVIK 1986). The sequence was continued for approximately 3 kb through the *Eco*RI site near the middle of the 8.1-kb *Sal*I fragment; we predicted that this *Eco*RI site would be within the *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 *Eco*RI restriction cut was arbitrarily designated as position 0 in numbering the sequence.

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

In order to define the transcribed sequences, we searched *Drosophila* cDNA libraries for clones homologous to *rosy* genomic DNA. We found *rosy* cDNAs in the libraries from L. KAUVAR and T. KORN-

BERG, made from poly(A)⁺ RNA from various stages of development (POOLE *et al.* 1985). We isolated 7 cDNA clones homologous to the *rosy* locus, from the third instar libraries I4, H2, and H3. Six of these had homology only to sequences to the right of the *Eco*RI 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 *Eco*RI 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 *rosy* mRNA, so the transcription start site remains to be determined. The right end of the clone is exactly at the *Eco*RI site, which is probably a construction artifact (if the double-stranded cDNA was not completely methylated at internal *Eco*RI 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 ry^{+10} control variant. The extent of the DNA sequenced from each genotype is indicated in Figure 2. Sequencing reactions on subclones of 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 ry^{+10} and the 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 ry^{+10} are shown in Figure 4. The three mutations which mapped furthest left, ry^{5208} , ry^{5182} and ry^{5204} , were all found within the upstream exon defined by the I4-31 cDNA clone. ry^{5208} and ry^{5182} each have a single guanine to adenine transition. 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 ry^{+5} with the three mutants in the first exon is shown in Figure 5.

The remaining four mutations, which mapped close to ry^{606} , were found within or adjacent to the second exon (Figure 4). Two mutations, ry^{523} and ry^{545} , both change the last guanine of intron 1 to an adenine.

<PstI> -2901
 CTGCAGGTAA GTGACTGTGA
 -2801
 TTGAACCGT GCGTCGCCCT TATCACTTAA TTATCAATGT GGCATCCGCA GGTATCAATG GTCGTCAGG AGATTGAGC AGCAGCAATT GCGTGAGGCG
 <Bgl 2> -2701
 ATGAGACGAC AAGCCCTTTA TGAGGGCACC CAAAGGGAGA GAGACTTAGA TCTGAAGTCG GCGTAGCATA AGTCAGATCC TTCATTACTA CTTGATGTT
 -2601
 TGTTTTCACT CTAGGATACC AAGTAATCCT GGCATTGTCA AATCCCTTGT GTTGTAACCT ACCTTTTGTG ACCAAACTTT ATAGTAGCCA ATATATTATA
 -2501
 TATATGTGT AAGTGAATC ATAGCTTGGG GTTGTCTGCG CACTCGTAGG GAATTAGTTA TGCCACCATT TTTGATTGGG GATTGACAGA AATCAATCCA
 -2401
 TCATTTTAAAT GATTGTGGTT CCTAAACCT GTTGGGCTTT ATTGTGCTTT CGATTGTATT TAAGCTGATA GAGCTAGTTG TTTACTTTTG AAGAACATTG
 -2301
 AACCCATTTT GAGCAGCTTT TAACCTTTTA ATAGTTTTTT TTTACCATTT ACATAAAAGC AAAATTTACA GTCTAAAGT TGTATATCA ATACGTTCTG
 -2201
 GTACGTAAAA GGGTTATTTA ATGCAACATT TTTACCACTT TAGATAGTAT ACTTTTACTT TGAGTGGGAG CCATATCACA AAAATGAAAT GAAAGGTTAA
 <Bgl2>
 AAAAGTATCT AAAAGTAAGT TTCAATTAAT ATTCTGAAA ATAATTGCAT TTCATGTACG TTTCTCTGAA ATACCCCTAA TTCGATACTA TTTATAGATT
 > C -2001
 TGGAAGATAT ATAGCTAGGA AATAGAATTA TATATTGGC TAGTTTGAAC ACCTTGTGAA CATTTTGTGTT AAAAAAAG CGTATGGTGG TACTAGAAAA
 -1901
 CTTAAGATA CAGATAAATA ACTTCACAT ACCGAGGTTT TGGTGTCTT TGAGACTGAA GGACAATCT CAAATGGTC TTAATTCTTG AGATAGAAAA
 -1801
 C GCTGAAGTTC TCATGCTGAA TAGTGATGAC TACGATTITA TCTGACAGAT TGCTTCTCTA GCTGATCATA CCGATCTCGG ATGGGTATCT GCTCTAATCC
 <BclI>
 GAGATAGCAG GTAGTAGAAC TGCTGCTCG TAAOCCCTAG ACATGGCCAT TTCCACGCAC ATGTAGGCGA TCCTGCACCT GTCGTTCTCT CGTCTTATCT
 G C
 CACAAAAATA AAAGTCGGCG TTTGGATGGA TTGCTCGGCT GGCTGTGGC ACATTCATTA CGTTTTCGTT GAGCGCGCGA GAGCGTTGTG TTGTTTTTTA
 <Hinc2> CG -1601
 ATCGCGCGGC ACATGAGGTT GTCGGTCTC TAAITTTGAA ATAAAAAAT TGTTTTGGCT TTACATTCAA GGAACGCACC TCCTCCCTCC CATATATGTA
 [<-5' end of 14-31 cDNA clone
 A in *ry*⁵¹⁸² -1401
 AACGAGCTTC GTAATTGCAG GCTAACACTT ATATAGGCAT TGGTACTGAG ATACAGCGAG ATACCGTGAG CTACAGGTAC ATAGCACTTC ACGATGCTA
 MetSerA
 G [-] both in *ry*⁵²⁰⁴ A in *ry*⁵²⁰⁸ T <C1a
 ACTCGGTTTT GGTGTTTTTC GTGAATGAA AGAAGGTACT TAAAGATACT CGGATGCOCT GCGTGTCTTT CCTTAAAGTC GGCCCACTCC TACTATATCG
 snSerValle uValPhePhe ValAsnGlyL ysLys<- intron 1
 1> -1201
 ATTCCCTTTA TTGTTTTATT GCCCTAATTG CGTTTAACTG GCATTGATG GAATTTGCOCT TCTTGCTCAA TTGAATGTGT CGAATGTATA TAGTATTGCT
 -1101
 TGTGCTATA GTGATAACAA TTGACCCCA TGTGAAGCOCT CGTGTGTCG TGGTGGCCCG TTATCGCCCG ATCAAGATGG TGCACCTGCC AATCGCGCAT
 -1001
 GGGAGCAGTG CTACTCGGAT GTGGGTGAGC TGGGATCGTC ATGGGCAITTT GTGTCTGCCA GATAGCTCAG GCCCTCTAAT CCGCACATTC CCACCGGCTG
 T <SstI> -901
 GCGGTGTGTT ACAATCAGT GAGCTCGGCT GGATTATCA GATAAAGCTG GCCATATTTA GCATCTGGTG TTGGTGGAGT GCAATTGGAG TTCTTTACTT
 -801
 TGTGCTACA GCTTGCACAC AAGTGAAGTA AATAATAATA AATAACCAAT TGCTCGTGGC CCCGGCATTT GATAGTGATA GATAGCAATC TCTGTTGCT
 -701
 CAGGGCTGGC AGAGCAGATA ACACCTTTCG TGTTTTAACT TTAACAAG TCACAGTTTT GACAAACAT TGCTTTAAAC CATATAACT TATAATTAAA
 [-] <Bgl 2>TA -601
 GCAATTATAA TGGGTTAAGT AAGCACAGAT CTAGAAGCTA CAAGATACAA GATACAAAT TTCCAATTAA TGGATAAAT GCATTAGATA CTTTGGGAA
 <Xba 1>
 A in *ry*⁵²³ and *ry*⁵⁴⁵ -501
 ATTGAITTCG TTTTTCAT TTTATTATAT TTATACATTT CTTATTCAG GTGACCGAAG TGTGCGCTGA TCCGGAGTGC ACGCTCTCTCA CATCTCTCGG
 intron 1 -> ValThrGluV alSerProAs pProGluCys ThrLeuLeuT hrPheLeuAr
 G -401
 CGAAAGCTG CCGCTGTGCG GAAGCAAGTT GGGATGTGCG GAAGCGGAT GCGGCGCTG CACCGTGATG GTGTCCCGCC TAGACCGTCG GGCCAAACAG
 gGluLysLeu ArgLeuCysG lyThrLysLe uGlyCysAla GluGlyGlyC ysGlyAlaCy sThrValMet ValSerArgL euAspArgAr gAlaAsnLys
 -301
 ATACGTCACC TGGCGGTCAA CGCCTGCTG ACGCCAGTGT GCTCGATGCA CGGATGTGCG GTGACCACTG TGGAGGGTAT AGGAGACACC AAAACGCGTC
 IleArgHisL euAlaValAs nAlaCysLeu ThrProValC ysSerMetHi sGlyCysAla ValThrThrV alGluGlyIl eGlySerThr LysThrArgL
 <BamHI> T in *ry*⁵¹⁹⁸ -201
 TGCATCCGCT GCAGGAGCGA CTGGCCAAAG CGCACGGATC CCACTGCGGC TTCTGCACGC CGGCAATTTG GATGTCCATG TACGCACTTC TCGGGAACGC
 euHisProVa lGlnGluArg LeuAlaLysA laHisGlySe rGlnCysGly PheCysThrP roGlyIleVa lMetSerMet TyrAlaLeuL euArgAsnAl
 [- 56 bp deletion in *ry*⁵³⁸ <Xho 1> -101
 GGAGCAACCC TCTATGCGAG ACTTGGAGGT GGCATTCCAA GGTAACTGCT GCCGCTGCAC CGGCTATGCA CCCATTCTCG AGGGCTACAA GAAGTTCAAC
 aGluGlnPro SerMetArgA spLeuGluVa lAlaPheGln GlyAsnLeuC ysArgCystH rGlyTyrArg ProIleLeuG luGlyTyrLy sThrPheThr
 -1
 AAGGAGTTTG CTTGCGGAAG GGGGAGAGAG TGCTGCAAG TTATGGGAA AGGATGTGGA ACCGATCGCG AGACCGATGA CAAGCTCTTC GAGCGCAGCG
 LysGluPheA laCysGlyMe tGlyGluLys CysCysLysV alSerGlyLy sGlyCysGly SerAspAlaG luthrAspAs pLysLeuPhe GluArgSerG
 EcoRI> <BamHI> <Pvu2> +100
 AATTCCAGCC CTGCGATCCC AGCCAGGAAC CCATCTTCCC ACCGGAACCT CAGCTGAGTG ACGCCTTCGA TTCTCAGAGT TTGATCTTTA GTTCGGATAG
 luPheGlnPr oLeuAspPro SerGlnGluP roIlePhePr oProGluLeu GlnLeuSerA spAlaPheAs pSerGlnSer LeuIlePheS erSerAsp

FIGURE 4. The DNA sequence of the upstream region of *ry*⁺. The sequence is numbered with the restriction site of *EcoRI* as the 0 point. Sequence changes for the various mutations are shown above the *ry*⁺ sequence line. The *ry*¹⁰ wild-type was sequenced between -2105 and -260; the differences relative to *ry*⁺ are shown in italics above the *ry*⁺ line. The predicted XDH protein sequence is shown below the DNA sequence. The horizontal arrows indicate tandem repeats in the DNA sequence. Selected restriction sites with six-base recognition sequences are indicated. The *ry*⁺ sequence is continued for 100 bases to the right of the *EcoRI* site at 0. This sequence matches the sequence of the Canton S wild type reported in this issue by KEITH *et al.* (1987) with the exception of a polymorphism at +74 (a G in Canton S-no change in the protein sequence). This sequence, together with that of KEITH *et al.* (1987), has been submitted to the GenBank and EMBL databases under the accession number Y00307.

*ry*⁵¹⁹⁸ has a single cytosine to thymidine base change, which results in an amino acid substitution (serine to phenylalanine) in the XDH protein. *ry*⁵³⁸ has a 56

base pair deletion within the protein coding sequences of the second exon. This deletion would cause a translation frameshift.

FIGURE 5.—Sequencing gel of the three first exon mutants. Sequencing reactions through the first exon from ry^{+5} and the three upstream mutants were run in parallel. The mutant changes are indicated next to their positions on the gel. The ry^{5204} sequence is not in exact register because it is derived from a vector with a polylinker of different length.

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 ry^{+4} and ry^{+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 ry^{5182} , ry^{5204} and ry^{5208} each

map clearly to the left of the structural element marker ry^{606} , with the genetic distance between the mutants and ry^{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 ry^{606} and these mutants to be between 0.7 and 2.1 kb. The recombination map predicts that the ry^{606} lesion is near the left end of the second exon (Figure 1). Indeed, preliminary molecular mapping suggests that the ry^{606} lesion is about 0.5 kb inside the exon, or about 1.3 kb from the base changes in ry^{5182} , ry^{5204} and ry^{5208} (M. GRAY and W. BENDER, unpublished observations). Since we have sequenced nearly 2 kb of the upstream DNA from ry^{5208} and ry^{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 ry^{5204} , but the mapping data agrees with its position near ry^{5208} and ry^{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 ry^{5198} , ry^{545} and ry^{538} each mapped as single site mutations located close to ry^{606} . The mapping data from ry^{523} , however, suggested that this allele might be a double mutant. The ry^{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 ry^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 ry^{523} parental flanking markers, indicating a partial conversion of ry^{523} . We interpret these data to mean that one of two separable ry^{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 ry^{523} is indeed a double mutant. In addition to the transition at -561, 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 ry^{545} chromosome does not carry this second mutation. This establishes that the ry^{523} and ry^{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.

The *rosy* open reading frame: Comparison of the sequences of the *rosy* genomic and cDNAs reveals a long open reading frame initiating at the ATG methionine codon at -1407 and extending across three

FIGURE 6.—Summary diagram of mutant sequence changes and their proposed effects.

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 CHOVIK 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 occur in two situations, where there is an upstream ATG with a poor sequence context, such that it is not recognized efficiently by ribosomes, or where there is translational initiation at an early ATG, termination, and then reinitiation at a downstream ATG. The sequence context of a given ATG affects the efficiency of its use in initiation (KOZAK 1986). The optimal context for the rat preproinsulin gene, the best studied case, is CCACCATGG, with the A at -3 and the G at +4 (in bold face) the most critical positions (KOZAK 1986). This sequence agrees with a consensus derived from a large survey of eukaryotic genes (KOZAK 1984a). A consensus sequence from 63 *Drosophila* genes is similar: N(C/A)A(A/C)(A/C)ATGN (D. CAVENER, personal communication). While the -1505 ATG is embedded in a poor initiation context (CATATATGT), the -1407 context is good (TCACGATGT). In addition, the -1505 ATG is followed by three in-frame stop codons in the next 51 bp. It has been shown for the rat preproinsulin gene (KOZAK 1984b, 1986) and for other genes (LIU, SIMONSEN and LEVINSON 1984; HUGHES *et al.* 1984; PERSING, VARUMUS and GANEM 1985), that an upstream ATG will not prevent internal initiation if there is a stop codon in frame with the first ATG and preceding the inter-

nal ATG. It is likely, then, that translation of the XDH polypeptide normally begins at the -1407 ATG. This conclusion is confirmed by an analysis of the mutant sites (see below).

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 *ry*⁵¹⁸² 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 *ry*⁵¹⁸² mutation is in a sequence context very similar to the consensus (ATACCATGG). In *ry*⁵¹⁸², 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 *ry*⁵¹⁸² 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 *ry*⁵²⁰⁴ 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 *ry*⁵²⁰⁴ 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 *ry*⁵²⁰⁴ and *ry*⁵¹⁸² mutations support the assignment of the -1407 ATG as the XDH translational initiation codon.

The *ry*⁵²⁰⁸ 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 ry^{5208} , has been synthesized *in vivo* in a cloned rabbit β -globin gene (WEIRINGA *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 ry^{5208} 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 ry^{5208} is shorter by 150–200 bp than the wild-type message (D. CURTIS and W. BENDER, unpublished results). A shorter ry^{5208} 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 ry^{5204} and ry^{5208} 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 ry^{523} and ry^{545} 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 ry^{545} and ry^{523} 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 ry^{5198} 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 ry^{538} 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 ry^{538} 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 ry^{+10} sequence might be responsible for the two-fold reduction of *rosy* RNA levels in the ry^{+10} 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 ry^{+10} *rosy* RNA. Genetic mapping data places the ry^{+10} controlling element between the adjacent *l(3)S12* gene and the ry^{+4} control site (CLARK *et al.* 1984) (see Figure 1). The ry^{+4} 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 ry^{+4} transcriptional control element to lie within the transcribed sequences of the first intron, and the ry^{+10} 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 ry^{+4} and ry^{+10} , exceptional flies were recovered carrying crossovers between the ry^{+10} and ry^{+4} 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. CHOVIK, unpublished data). This will allow us to place limits on which of the many differences in the ry^{+10} sequence (relative to ry^{+5}) 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. SPIERER 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.
- BENTON, W. D. and R. W. DAVIS, 1977 Screening lambda-gt recombinant genes by *in-situ* hybridization to single plaques *in situ*. *Science* **196**: 180-182.
- BIGGIN, M. D., T. J. GIBSON and G. F. HONG, 1983 Buffer gradient gels and ³⁵S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* **80**: 3963-3965.
- BIRD, M. J. and O. G. FAHMY, 1953 Cytogenetic analysis of the action of carcinogens and tumor inhibitors in *Drosophila melanogaster*. I. 1:2,3:4-Diepoxybutane. *Proc. R. Soc. Lond. (Biol.)* **140**: 556-578.
- CHOVNICK, A., W. GELBART, M. MCCARRON, B. OSMOND, E. P. M. CANDIDO and D. L. BAILLIE, 1976 Organization of the rosy locus in *Drosophila melanogaster*: evidence for a control element adjacent to the xanthine dehydrogenase structural element. *Genetics* **84**: 233-255.
- CLARK, S. H. and A. CHOVNICK, 1986 Studies of normal and position-affected expression of *rosy* region genes in *Drosophila melanogaster*. *Genetics* **114**: 819-840.
- CLARK, S. H., S. J. HILLIKER and A. CHOVNICK, 1979 Genetic organization of the *rosy* locus in *Drosophila melanogaster*: pilot studies on the induction and analysis of low XDH activity mutants of the rosy locus as putative control mutants. pp. 117-121. In: *Eukaryotic Gene Regulation* (ICN-UCLA Symposia on Molecular and Cellular Biology, Vol. XIV), Edited by R. AXEL, T. MANIATIS and C. F. FOX. Academic Press, New York.
- CLARK, S. H., A. J. HILLIKER and A. CHOVNICK, 1986 Genetic analysis of the right (3') end of the *rosy* locus in *Drosophila melanogaster*. *Genet. Res.* **47**: 109-116.
- CLARK, S. H., S. DANIELS, C. A. RUSHLOW, A. J. HILLIKER and A. CHOVNICK, 1984 Tissue-specific and pretranslational character of variants of the rosy locus control element in *Drosophila melanogaster*. *Genetics* **108**: 953-969.
- CLARK, S. H., M. MCCARRON, C. LOVE and A. CHOVNICK, 1986 On the identification of the rosy locus DNA in *Drosophila melanogaster*: intragenic recombination mapping of mutations associated with insertions and deletions. *Genetics* **112**: 755-767.
- COTE, B., W. BENDER, D. CURTIS and A. CHOVNICK, 1986 Molecular mapping of the rosy locus in *Drosophila melanogaster*. *Genetics* **112**: 769-783.
- COVINGTON, M., D. FLEENOR and R. B. DEVLIN, 1984 Analysis of xanthine dehydrogenase mRNA levels in mutants affecting the expression of the rosy locus. *Nucleic Acids Res.* **12**: 4559-4573.
- DAVIS, R. W., D. BOTSTEIN and J. R. ROTH, 1980 *Advanced Bacterial Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- DENTE, L., G. CESARINI and R. CORTESE, 1983 pEMBL: a new family of single stranded plasmids. *Nucleic Acids Res.* **11**: 1645-1655.
- EDWARDS, T. C. R., E. P. M. CANDIDO and A. CHOVNICK, 1977 Xanthine dehydrogenase from *Drosophila melanogaster*. *Mol. Gen. Genet.* **154**: 1-6.
- FRISCHAUF, A. M., H. LEHRACH, A. POUSTKA and N. MURRAY, 1983 Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* **170**: 827-842.
- GELBART, W., M. MCCARRON and A. CHOVNICK, 1976 Extension of the limits of the XDH structural element in *Drosophila melanogaster*. *Genetics* **84**: 211-232.
- GELBART, W. M., M. MCCARRON, J. PANDEY and A. CHOVNICK, 1974 Genetic limits of the xanthine dehydrogenase structural element within the rosy locus in *Drosophila melanogaster*. *Genetics* **78**: 869-886.
- HILLIKER, A. J., S. H. CLARK, A. CHOVNICK and W. M. GELBART, 1980 Cytological analysis of the chromosomal region immediately adjacent to the rosy locus in *Drosophila melanogaster*. *Genetics* **55**: 95-110.
- HOHN, B., 1979 *In vitro* packaging of lambda and cosmid DNA. *Methods Enzymol.* **8**: 299-309.
- HONG, G. F., 1982 A systematic DNA sequencing strategy. *J. Mol. Biol.* **158**: 539-549.
- HUGHES, S. K., E. MELLSTROM, E. KOSIK, F. TAMANOI and J. BRUGGE, 1984 Mutation of a termination codon affects *src* initiation. *Mol. Cell Biol.* **4**: 1738-1746.
- KELLER, E. B. and W. A. NOON, 1985 Intron splicing: a conserved internal signal in introns of *Drosophila* pre-mRNAs. *Nucleic Acids Res.* **13**: 4971-4981.
- KELLER, E. C. and E. GLASSMAN, 1965 Phenocopies of the *ma-l* and *ry* mutants of *Drosophila melanogaster*: inhibition *in vivo* of xanthine dehydrogenase by 4-hydroxypyrazolo (3,4-d) pyrimidine. *Nature* **208**: 202-203.
- KEITH, T. P., M. A. RILEY, M. KREITMAN, R. C. LEWONTIN, D. CURTIS and G. CHAMBERS, 1987 Sequence of the structural gene for xanthine dehydrogenase (*rosy* locus) in *Drosophila melanogaster*. *Genetics* **116**: 67-73.
- KOZAK, M., 1978 How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**: 1109-1123.
- KOZAK, M., 1984a Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* **12**: 857-872.
- KOZAK, M., 1984b Selection of initiation sites by eucaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. *Nucleic Acids Res.* **12**: 3873-3893.
- KOZAK, M., 1986 Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**: 283-292.
- LIU, D., C. C. SIMONSEN and A. D. LEVINSON, 1984 Initiation of translation at internal AUG codons in mammalian cells. *Nature* **309**: 82-85.
- MCCARRON, M., W. M. GELBART and A. CHOVNICK, 1974 Intracistronic mapping of electrophoretic sites in *Drosophila*

- melanogaster*: fidelity of information transfer by gene conversion. *Genetics* **76**: 289–299.
- MCCARRON, M., J. O'DONNELL, A. CHOVNICK, B. S. BHULLAR, J. HEWITT and E. P. M. CANDIDO, 1979 Organization of the rosy locus in *Drosophila melanogaster*: further evidence in support of a cis-acting control element adjacent to the xanthine dehydrogenase structural element. *Genetics* **91**: 275–293.
- MESSING, J., 1983 New M13 vectors for cloning. *Methods Enzymol.* **101**: 20–79.
- MITCHELL, P. J., G. URLAUB and L. CHASIN, 1986 Spontaneous splicing mutations at the dihydrofolate reductase locus in chinese hamster ovary cells. *Mol. Cell. Biol.* **6**: 1926–1935.
- MOUNT, S. M., 1982 A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**: 459–472.
- MUSKAVITCH, M. A. T. and D. S. HOGNESS, 1982 An expandable gene that encodes a *Drosophila* glue protein is not expressed in variants lacking remote upstream sequences. *Cell* **29**: 1041–1051.
- PEABODY, D. S. and P. BERG, 1986 Termination-reinitiation occurs in the translation of mammalian cell mRNAs. *Mol. Cell. Biol.* **6**: 2695–2703.
- PEABODY, D. S., S. SUBRAMANI and P. BERG, 1986 Effect of upstream reading frames on translation efficiency in Simian Virus 40 recombinants. *Mol. Cell. Biol.* **6**: 2704–2711.
- PERSING, D. H., H. E. VARMUS and D. GANEM, 1985 A frameshift mutation in the pre-S region of the human hepatitis B virus genome allows production of surface antigen particles but eliminates binding to polymerized albumin. *Proc. Natl. Acad. Sci. USA* **82**: 3440–3444.
- POMALES, R., S. BIEBER, R. FRIEDMAN and G. H. HITCHINGS, 1963 Augmentation of the incorporation of hypoxanthine into nucleic acids by the administration of an inhibitor of xanthine oxidase. *Biochim. Biophys. Acta* **72**: 119–120.
- 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.
- REARDON, J. T., C. A. LILJESTRAND-GOLDEN, R. L. DUSENBERY and P. D. SMITH, 1987 Molecular analysis of diepoxybutane-induced mutations at the rosy locus of *Drosophila melanogaster*. *Genetics* **115**: 323–331.
- SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463–5467.
- TREISMAN, R., S. ORKIN and T. MANIATIS, 1983 Specific transcription and RNA splicing defects in five cloned β -thalassaemia genes. *Nature* **302**: 591–596.
- VOGEL, E. and A. T. NATARAJAN, 1979a The relation between reaction kinetics and mutagenic action of monofunctional alkylating agents in higher eukaryotic systems. I. Recessive lethal mutations and translocations in *Drosophila*. *Mutat. Res.* **62**: 51–100.
- VOGEL, E. and A. T. NATARAJAN, 1979b The relation between reaction kinetics and mutagenic action of monofunctional alkylating agents in higher eukaryotic systems. II. Total and partial sex-chromosome loss in *Drosophila*. *Mutat. Res.* **62**: 101–123.
- WEIRINGA, B., F. MEYER, J. REISER and C. WEISSMANN, 1983 Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit β -globin gene. *Nature* **301**: 38–43.

Communicating editor: A. SPRADLING