Molecular characterization of the 5' end of the *rudimentary* gene in *Drosophila* and analysis of three P element insertions

William Zerges⁺, Andor Udvardy[§] and Paul Schedl^{*}

Departments of Biology and Molecular Biology, Princeton University, Princeton, NJ 08544, USA

Received March 27, 1992; Revised and Accepted August 5, 1992

GenBank accession no. M37783

ABSTRACT

A detailed analysis of the 5' end of the *rudimentary* gene of *Drosophila melanogaster* is presented. *rudimentary* transcripts are heterogeneous at their 5' ends indicating that transcription is initiated at multiple sites within a region of ~50 bp. These transcription initiation sites are within a region that is preferentially susceptible to nuclease cleavage in isolated nuclei. Additional nuclease hypersensitive regions were found within the first exon and the first intron. Within these internal nuclease hypersensitive regions are the insertion sites for previously identified P element transposons which disrupt *rudimentary* expression. One of these P element insertions, located in the first intron, is removed from the *rudimentary* transcript with the splicing of this intron. Another P element insertion, within the first exon, is removed from the *rudimentary* transcript by novel first intron splicing involving a cryptic splice donor site, located 5' to the insertion, and either the normal acceptor site or a cryptic splice acceptor site within the second exon.

INTRODUCTION

The *rudimentary* locus (*r*) of *Drosophila melanogaster*, discovered by Morgan in 1910, is one of the classical complex loci of the fly and exhibits a complicated pattern of intragenic complementation (2). The underlying basis for this intragenic complementation is reasonably well understood. The *r* locus encodes a large multi-functional enzyme complex which catalyzes the first four steps in *de novo* pyrimidine biosynthesis: glutamine amidotransferase (GATase), carbamyl-phosphate synthetase (CPSase), aspartate transcarbamylase (ATCase), and dihydroorotase (DHOase) (11,16,25,26). Hence, as might be expected, mutations disrupting any one of these enzymatic activities can often complement mutations disrupting other activities (16). *r* mutants are pyrimidine auxotrophs; while viable on media containing pyrimidines, they exhibit obliquely truncated

wings and female sterility. These phenotypes probably reflect requirements for *de novo* pyrimidine biosynthesis in the wing imaginal discs and in the oocyte/early embryo, respectively, which can not be fully supplemented by exogenous pyrimidines.

r is located at cytological locus 15A1 on the X chromosome (17). The *r* transcription unit is slightly over 14 kb in length and encodes a poly-adenylated transcript of about 7.3 kb (32). Comparison of the nucleotide sequences of several *r* cDNAs with genomic DNA sequence showed that the *r* transcription unit contains at least seven exons and six introns (10,11). The putative translation start site is just ten nucleotides from the beginning of the first of these exons. The open reading frame extends 6,708 bp from this start codon and encodes a predicted protein of 249 kd (11). The coding region for the GATase enzyme domain is at the beginning of the message and is followed by coding sequences for the CPSase, DHOase and ATCase domains.

Previous studies described five hybrid-dysgenesis induced *r* mutants, *hd1*, *hd2*, *hd3*, *hd4*, and *hd5* (35). All correspond to P element insertions located at sites upstream of the exon that contains the putative translation start site (and, as described below, corresponds to exon 2). Thus these insertion mutations do not disrupt the protein coding sequences of the *r* message. On Northern blots, *r* mRNAs produced by these mutants were indistinguishable in size from the 7.3 kb wild type transcript (35). However, the level of *r* message in all but one of the mutants is reduced relative to that in wild type females. The exception, *hd1*, has a somewhat higher level of *r* transcript than wild type.

To learn more about the expression of the *r* gene, we have characterized the organization of its 5' end. We show that *r* has an additional 5' exon beginning upstream of the P element insertions. This exon is heterogeneous at the 5' end suggesting that there are multiple transcription initiation sites for *r* mRNAs. The 5' end of the first *r* exon is located only about 350 bp away from the probable 5' end of another transcription unit which is transcribed in opposite orientation. In addition, we have examined the chromatin structure of these two divergent transcription units and have characterized transcripts produced by two of the P element induced *r* mutants.

* To whom correspondence should be addressed

Present addresses: ⁺Department of Molecular Biology, University of Geneva, 30 Quai Ernest-Ansermet, CH1211, Geneva 4, Switzerland and [§]Institute of Biochemistry, Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary

MATERIALS AND METHODS

Drosophila strains and media

The P element insertion mutant stocks ($r^{hd3} f^1$ FM7, $r^{hd5} f^1$ FM7, $r^{hd1-5} f^1$ FM7) were obtained from Tsubota (35,36). All flies were homozygous mutant adult females and are referred to by the mutant name. All stocks were grown on yeast-glucose medium (10%(w/v) dried yeast, 10%(w/v) glucose, 3%(w/v) nipagin, 20%(w/v) agar) at 25°C.

General methods

RNA was prepared from whole flies as described previously (6). RNA and DNA samples were subjected to agarose electrophoresis and then transferred to Genescreen or Genescreen-plus membranes (NEN, Boston) by standard methods. Blots were then prehybridized and hybridized to nick-translated DNA probes with the buffer of Church (3). The strand specific RNA northern probe was synthesized as described by Melton *et al.* (24). DNA was sequenced by the method of Sanger (30) with the Sequenase system (U.S. Biochemicals).

RNAse protection and primer extension analyses

RNAse protection analyses were done with the procedure of Melton *et al.* (24). Antisense RNA probes were derived from the following genomic fragments (Fig. 1); A, 400 bp *Eco* R1-*Bam* H1; B, 1.0 kb *Bam* H1-*Eco* R1; C, 1.4 kb *Eco* R1 fragment; D, *Dra* I-*Bam* H1; E, *Dra* I to an *Exo* III deletion end point located 215 bp proximal to the left-hand *Eco* R1 site; F, a 535 bp fragment between another *Exo* III deletion end point and the *Dra* I site. Primer extensions were carried out as described by Jones (18). Approximately 5×10^4 cpm of [32 P]primer was used per hybridization. The primers were:

- 1, 5'CAACCAAGGCGGTGAATTCAGT3';
- 2, 5'GCTCCAATTAACGTCGTGA3';
- 3, 5'GCTCTCTCTTTGCTGCACTTGTA3';
- 4, 5'CGCCACCAACCAAGGCGGCTGGCC3';
- 5, 5'CTCGAGTATCAATGCTGAATTC3'.

After the extension reactions, samples were extracted once each with phenol and chloroform before the final ethanol precipitation. The cDNA extension products were resolved on denaturing acrylamide gels. Gels were exposed to preflashed X-ray film.

Chromatin analysis

The procedures used for nuclei isolation from 0–24 hr embryos and from KC tissue culture cells, digestion with *micrococcal* nuclease and DNase I, DNA purification, and the processing of the DNA samples for hybridization were essentially as described by Udvardy and Schedl (37). Naked DNA digests were performed on total genomic DNA purified from KC cells. All digestions were performed at 25°C.

PCR amplification and cloning

cDNAs corresponding to the 5' ends *r* transcripts of a wild type strain and two P element insertion mutants (*hd3* and *hd5*) were synthesized from poly(A) RNA and then amplified by PCR. The first strands were synthesized from primer 5 (Fig. 1) hybridized to poly(A) selected RNA using the conditions described above for the primer extension experiments except that the primer was not end-labeled with 32 P. The cDNA sequences between the two *Eco* RI sites (Fig. 1) were then amplified during two rounds of

PCR using primer 1 and primer 5 (Fig. 1). The reaction conditions were; 10 mM Tris pH 8.3, 50 mM KCl, 2 mM $MgCl_2$, 0.01% gelatin, 0.2 mM dATP, 0.2 mM dCTP, 0.2 mM dGTP, 0.2 mM dTTP, 100 ng of each primer, and 2 U Taq polymerase. Samples were covered with mineral oil and amplified as follows: 2 cycles of 2 minutes at 95°C, 2 minutes at 54°C, 5 minutes at 72°C followed by 18 cycles of 40 seconds at 95°C, 90 seconds at 54°C, 90 seconds at 72°C. Samples were removed from under the mineral oil to tubes containing 2 μ l 100 mM EDTA, precipitated by adding 50 μ l of 7.5 M NH_4Ac and 150 μ l ethanol at 0°C, incubated at room temperature for 5 minutes, and DNA was pelleted in a microfuge for 15 minutes. Non-extended primers precipitate inefficiently under these conditions. The PCR products were subjected to a second round of amplification using an internal primer. Before the second amplification round, the pellet was washed with 75% ethanol. Five percent of the products of the first round were subjected to PCR under same reaction conditions. The second round was 20 cycles of 40 seconds at 95°C, 90 seconds at 54°C, 90 seconds at 72°C. After each amplification round a portion of each sample (8%) was analyzed on a 1.0% agarose gel. Each sample was digested with *Eco* RI and *Xma* I, resolved on a 1.2% agarose gel, and transferred to nitrocellulose. The fragments on this filter that hybridized to a *r* specific probe were then purified from a 1.0% low melting agarose gel and cloned into the a plasmid vector using standard cloning techniques (23).

RESULTS

The *r* transcript has a non-translated 5' exon

In previous studies (10), the 5' end of *r* was tentatively placed 410 bp upstream from the right-hand *Eco* R1 site shown in Fig. 1. This placement was based on several observations. First, a nick translated fragment containing sequences upstream (to the left) of the *Bam* H1 site (Fig. 1) did not appear to hybridize to *r* transcripts on Northern blots. Second, a single 410 nt fragment was detected in S1 protection experiments using the right-hand *Bam* H1-*Eco* R1 fragment as a probe (probe A, Fig. 1). Lastly, one primer extension product, among several, was consistent with the 410 nt S1 protection being generated by the 5' end of the *r* message.

On the other hand, there were a number of reasons to believe that this placement might not be correct. To begin with, the 5' non-translated leader would be only 15 nt, an unusually short length for invertebrate messages. Second, given the large size and relatively low abundance of the *r* message it was possible that a small exon upstream of the *Bam* H1 site may not have been detected in Northern analysis using a nick translated probe. Finally, the multiple primer extension products could be explained by premature reverse transcriptase stops. Therefore, we investigated whether sequences upstream of the putative start site are present in *r* transcripts by probing a Northern blot of RNA isolated from adult females with a more sensitive anti-sense RNA probe complementary to the upstream *Eco* R1-*Bam* H1 fragment (probe A, Fig. 1). This probe hybridized to a single RNA species (lane 2) that is indistinguishable in electrophoretic mobility from the 7.3 kb *r* transcript (lane 1) detected on a parallel lane of the same blot with a cloned *r* cDNA probe, cRUD5 (32) (Fig. 2). This observation suggests that there is at least one additional exon located upstream from the previously reported 5' end.

Figure 1. Molecular map of the 5' end of the *r* gene. The regions corresponding to the *r* exons and the first exon of the neighboring gene are indicated by the shaded and unshaded boxes, respectively. The inset map shows the organization of the locus. The positions of the multiple transcription start sites of both genes are indicated. The nuclease hypersensitive regions are indicated by the black bars above the genomic map. The sites of the five P-element insertions (36) are indicated. The RNase protection probes (A–F) and the oligonucleotide primers (1–5) are indicated by the arrows. The strand used for each probe and primer is indicated by the direction of its arrow (which points 5' to 3').

RNase protection analysis of the first exon of *r*

To further localize this additional 5' exon, the anti-sense RNA probes (A–E) shown in Fig. 1 were used for RNase protection analysis of the *r* transcripts of wild type females. Probe C, which corresponds to the 1.4 kb *Eco* R1 fragment (Fig. 1), gave two different protection products (Fig. 3A, lane 1). The first is 410 nt. Lane 3 shows that a protection product of the same size was detected with probe B. Hence this protection product maps to the right hand side of the *Eco* R1 fragment and corresponds to exon 2 (Fig. 1). The second protection product observed with probe C was a smaller doublet of 230–235 nt. As shown in Fig. 3A, this doublet was not observed with probe B, but was detected with probe A (lane 2). This difference maps the new 5' exon, the ~230 nt doublet, to the left of the *Bam* H1 site in the 1.4 kb *Eco* R1 fragment. Further evidence localizing exons 1 and 2 came from analysis of the probe C protection products in RNA prepared from the deletion mutant *hd1-5* which does not produce any detectable 7.3 kb *r* message (36). In *hd1-5* a ~6.5 kb deletion extends from just downstream of the left *Eco* R1 site, through exon 2, and well into the gene. Lane 6 of Fig. 3 shows that the 410 nt exon 2 protection product was absent, while the level of the exon 1 protection product was reduced by approximately 10 fold relative to the *adh* control and appeared to be slightly truncated.

The limits of the 5' *r* exon were more precisely defined with probes D and E. Probe D has the same right-hand end point as probe A (the *Bam* H1 site) but extends 100 nt further upstream to the *Dra* I site. Probe D gave a protection pattern which is different from that obtained with probe A (Fig. 3, lane 4). Instead of two bands, five major fragments were observed. Two of these were the same size as the fragments observed with probe A. The

Figure 2. Parallel lanes of a northern blot of RNA isolated from adult females were probed either with a *r* cDNA probe (lane 1) or Probe A (lane 2).

other three were between 30 and 45 nt longer and, therefore, must have been derived from exons which have different 5' ends located between the *Dra* I and *Eco* R1 restriction sites. These findings indicate that the newly identified *r* exon is heterogeneous

Figure 3. RNase protection analysis of the first and second exons of the *r* locus and the first exon of the distal neighboring gene. All lanes of the denaturing acrylamide gels resolve protection products of RNA isolated from adult wild type females except for lane 6 of panel a which contains the negative control in which RNA isolated from hd1-5 was used. The probes used were as follows (Fig. 1). Panel a, lane 1, probe C; lane 2, probe A; lane 3, probe B; lane 4, probe D; lanes 5 and 6, probes C and a probe to the second of the *adh* transcript. Panel b, lane 1, probe E; lane 2, probe D; lane 3, probe F. Lanes with molecular weight markers are labeled 'M'.

Figure 4. Primer extension analysis of multiple 5' ends of the *r* transcript. Five primer extension products from the primers 2 and 3 (Fig. 1) are shown in the first lanes of panels a and b, respectively. Sequencing ladders in panels of the corresponding *r* sequence from primers 2 and 3 are also shown in panels a and b, respectively. cDNAs extended from primer 2 hybridized to RNA isolated from the *r* deletion mutation hd1-5 are shown in lane 2 of panel b.

in length due, at least in part, to different 5' ends. This conclusion was confirmed with probe E which has the same distal end point as probe D (the *Dra* I site) but extends less proximally, to an Exo III deletion end point located 208 bp downstream of the left *Eco* R1 site (Fig. 1). As shown in Fig. 3b lane 1, probe E gave five protection products, each of which was 25 nt shorter than the corresponding protection product for probe D (Fig. 3b, lane 2). The structure of this 5' exon is summarized in Fig. 1.

The relative levels of the five protection products generated by probe D were examined at different stages of development: 0–24 hr embryos, first, second, and third instar larvae, early and late pupae, and adult males and females. We found that the relative levels of the five products was constant throughout development (unpublished data).

Primer extension analysis revealed heterogeneous 5' ends of *r* messages

Results of primer extension experiments confirmed that the different 5' ends of exon 1 correspond to 5' termini of *r* transcripts. Fig. 4 shows cDNAs extended from two antisense [³²P]oligonucleotide primers (#2 and #3) which are complementary to different sites in the *r* transcript (Fig. 1). The cDNAs extended from each primer were resolved next to the corresponding sequencing ladders so that the size of each extension product could be correlated with a specific position on the genomic DNA sequence. Both primers gave five major cDNA extension products. Each of these products corresponds to one of the 5' ends of exon 1 defined by RNase protection analysis. Moreover, the level of these products is reduced on RNA from hd1-5, Fig. 4B, lane 2. Thus, we conclude that exon 1 (Fig. 1) is the first exon of the *r* transcript and that the *r* transcripts initiate at five distinct sites within a 50 bp region.

DNA sequence analysis

The sequence of a ~4.0 kb DNA segment spanning the 5' end of the *r* transcription unit is shown in Fig. 5. The sequence downstream of the *Eco* R1 site was reported previously (10,11). Position +1 corresponds to the most distal of the five 5' ends of the *r* transcript. The 5' ends of *r* transcripts are indicated by arrows. Like the promoters of many other genes encoding steps of intermediary metabolism (7), the *r* promoter lacks any sequence resembling a TATA box. Though unlike other such genes the *r* 5' flanking region is not particularly GC rich. Also, none of the transcription start sites fit the consensus sequence for transcription initiation of *Drosophila* genes transcribed by RNA polymerase II (15).

The results presented above, and the nucleotide sequence of a PCR amplified *r* cDNA (Fig. 7), indicate that the first exon of *r* extends from the five discrete 5' ends (Fig. 5) to a single splice donor site at position +273. The nucleotide sequence at this site (TC/GTAAGT) is consistent with the consensus sequence of splice donor sites, AG/GTAAGT (34). The first intron extends to position +1045 which matches the splice acceptor consensus sequence (Py(n)NCAG/GG). Upstream from the acceptor site, between positions +1031 and +1035, is a sequence that could be a branch site for splicing of the first intron (19).

Fig. 5 also shows ~2.6 kb of upstream flanking sequences. In a previous study (27) a transcription unit immediately distal to *r* was identified. The orientation and probable 5' end of this nearby transcription unit was determined by RNase protection. As shown in Fig. 3B (lane 3), protection products were observed when probe F was hybridized to RNA isolated from adult

2615 TGTTCCTCTGGCGCGAGAAGAACTTCACCGATACCGCGGGCCTGACACGCGCCTTCGTCCAGGAGCAGCACTCCACGCCGCTCTTACGCTGCGTAGCGA
-2515 GCTCTTCAAGTGGTACTACTTCGTGATCCACGCCAGGGCAATCGTCTGGCCGCCCTGCGCTCCGGCAGTCCCAAGCTCTCTGCACAGCACTGGACCATC
-2415 GCCTACGCGGGCGACGCACATCATCTTCGGGACGCCGAATCAGCTCTGTGCTGCCGAAAAGTTCGACGACGATCGCTCGTGGTGAAGATGCTGCCCG
-2315 GCCAGATGCACACGCCGCCATCGGAGGCTTGCCAGTGGCTACCGATCGCCTGTGCTCCGCCAAAATGAGCGGGATGGCTGTACCCTGGAATGTTTGAAC
-2215 AAATAAACTTATGCATTAGTGATAAGTGTTTTATTCTTAGCCAGTTCATGAATAAATAATATATAATTTGATTAAATATGTGTGTTTCGGTTCAGTCT
-2115 GCAAGTGATTATACAAATATTAGAAAGGTAACCAAAACAAACAAATTTTACAGAACGGTACAGCATAGAACCGTGGTGGTCAATGGTTTCAATTAGCCCTA
-2015 ATCAGTGGTTCCTTGTGTGGTGAAGCTGTTCGTCCGTATTGTACTACTTTCGCCGAAGTCCACCTCCTGCTCGAAGAACCTCTCTCTCGCTGTCAT
-1915 CGGACTCCGACTCCGGCTCCGGCTCCGGCTCCGAGTTGCGGGCGGGCCTGTTGATAGCGGAACTGCGAGTTGGGCAGCGGATTTGACACGGTGGCGCGA
-1815 GCGATTGCCCCAATAGGCTACGACGATGACTGCTGTCATGCTGAGTAACCTCCGCTGGTTTATCTCTCAGTTCCTCTCAGTTCCTCTCTCTCTCG
-1715 GGGCGCGCATCCATGGACTCCTCATGGTGGCGGAGTGATCTCGATCTCTTCGTCTCGTCCAGATGCGGTGCTCCATTGGCGAACGAGGCGAAGCGCAAAG
-1615 GTACACGCGAGTTTGCCCATACCCGCCACTGCTTGCATCCGGACTGTTGACGTAATTCGCGAGGTCGCCGAGCGATTTCGAGGTCTCCACGGCCGCTGCT
-1515 CTGTTCCGATCTCTCTGGGTGAGCATGCGCCTCTTTGGCGGAAAACGGCGGCATACATGAACCTTGTCTGTTCTGCTCGCAAAATGGGGAAGTAGTGG
-1415 GCCACAAAGACGACGACCATGATGAAGATCACAGCGCTGGTGTGCCAGCTGCTGAAGTAAGCCATAACGGCGCCAATGGCCTGACGACCCACATGACGA
-1315 TGGTTTGGCATCGAGGATTCTTGGGGCGACCAATGCGATACGAGATCTGTAGAATAATCGAGGCAATTAGCTGGGGTTATCTAATCAAAGGGAATATCA
-1215 AGTGATTACGCACCAAGGAACGAGATCAGTCCGGTGATCACCAAGTAGCCAGCAGCATGGTTCGCGGTAGTGATCAGAAATCAAACGCGAGTTTGTCCCGCAG
-1115 CTGTTTGTATCATAGAAAGCAATGGTCCAGCGCCCAATTAGGACGCCATACATGATGTACGTGCGCGAAAGAGTTTGGCGGCCAAATAGATGACGCCAC
-1015 AGCAGGGAGGCGCATATGCGGATGACAATGCCGACCATAGTAGAACAACACTGTTCTTTGCTAGCGCGCGGCGCTGAAGAGGATCAGTATGCCGACG
-915 GGAACCTGTGCGGACGCGCCAGATGCTACGCGGAATGCTATTCAAGATGACCGTGTACCGTTCGCGGGAGGAATCCCGATACGACTGCTGCTGAAGGCGCA
-815 GAGCTGACGAGCTGGCGCTTCTGCGGCAACAGGGTGATGCTGAACAGCGATTGCTTGTCTTGTAGTGCTGCAACACCTCCTCCGCGCTTCTGCCGCGG
-715 TACTGGGAATAATCATGCTCCGATGCGGACGAGCACTCCAGCTCTCGAAGAGACTGACACCTCGACGCGTTTACCAGGATAGCAGAACGCTGCGCA
-615 GAGTCTGAGCTCCCAAGATTGCTTCCGACAAATTTGATGTATGCTGCCGCGCTCAGGTAGGCGACGCTCGAGGAAGCCGATTTGCCAGGAATACCCGATGT
-515 GAAATATCCAGCCACGGTTAGCATCAGTAGCTCAGCCGAATGCATTATGAACGGTTGCAATGGCGTTAGTCTTATCAATCAGTGCCTTTTCACTCTGTTG

-415 CTCTCGCAGGACGTTGCTTGTAAATGCTGACTTGCTTATGCTTGATGTTTTGATGTTTACTTTACTTTAAGGAATCGAATGAAAAGCAACGCCGCTT

-315 TGGTTTATGCTCTACTCAGCAGCGGACGGGCAGTATATCGTAACGTAAGAAAATGTATCGATAGGTGTGTGTGACTACACTATTGTGATTTCAAAT

-215 TTGTGACTAACCATTAAATGTGGATTATTTATTTCCCTTTGGAATACTTTGCAGAACACCCCATTTAAACATGTTTGTCTAATATTATACAAATATCTTA

-115 TCTTATTACATAACAATAAATGAATAAATACGTTCTTGGCAAGTTCCCAACCTGCAATATTACTTCTAGTAAAGTTCCACACCGCCGAGCTAAGCTTTG **A**

-15 CTGTGGCATTTTTACGACATAATGCACGAAGATTCCGATTCAACCAAGGCGGTGAATTCAGTTGTTGACTTTGATTTTCCAGACGTTAAATTGGAGCCA **EcoRI**

85 ATTTGCAAGCAAAACACAGTTAAATGCTCAAACATAACCATCCGAGCAGTGCGAAGTTAACCAAGCTCTAAAAGCGGTGCGAAAGTAACAAATCAAGT

185 TGTCAATCGAATTAAAGTCGAACCAAAACGTTTCACTACAAGTGACAGCAAGAGAGAGAGCTTCGCTTGAAAGTACACGGAATCGTAAGTGAGCTGT **C** **D** **B** **hd1** **hd3** **E1** **I1**

285 AGTTGTAACGGGTGCAGATGGGAAAAGGATAACTTGATGGCGATACTAATGCATCGTAGCTTAAAGGATTCTCAGTTGTTGTTAATCCTCAATTCTCTGCCAA

385 AGTCCAATTCAAAGTTAAAGTCAATGTTATTGTTATCTCATTGTGTCAGCCAACCGGTTTCCAATTGGATCCAAACTTCTCACACGTGCTTACGCACAC

485 ACAAACGCGACCTCTCACACTCGCTCTTCTCTTTCTTCTCCGCGCTAGTCGTCTGTGTAACAAACCGTGAATGGCAATATCAATGACAACAGCAAC

585 TTGAAACGCTTGATAAGCAGTGTCCATTCCCAATCAGAACCGTGTGTGCCCCGCTCAACATGTTGCCAGACAGTTAAATACATGATACATTATCTAT

685 CGCGGCTTCTTATTGGATTCTTTGTATATTTTCTGGGCGGCAACTGTGTGCGTGTGACGGTGTGTTTATAAACACATGTGTTTTTTGAACAAGATAA

785 CCTGTGAATTATATTACATTACATTTAAATCGAGGTATGCAACGGCCTCTGCGTATACATATAGCATAGTCATAGCATATACTCATTAAAGTAAAGTTC

885 CCAGAGTCGCCGAATATTGTTATCAGATTTTCGACTCAATATTGTGTCTACTTGCCCTTCGTAGAGATATGATCTTTTAATTAATTAACCTTTGCTCGTTTA

985 TTTAAGACGGTTTTTCGAGTGACAGGTACTAAAGAATACCCCTATGGTTTCCAATCCTCCCCGCAACTCGTCCAATATGGCCTCTACGGACTGTTTACCTCG **I1** **E2** **M** **A** **S** **T** **D** **C** **Y** **L** **A**

1085 CACTGGAGGATGGCACAGTGTGCGCGGGTACTCCTTTGGCTACGTTCCCTCGGAGAATGAATCCAAAGTGGGTGTTGGTGGCGAGGTGGTCTTCCAAAC **L** **E** **D** **G** **T** **V** **L** **P** **G** **Y** **S** **F** **G** **Y** **V** **P** **S** **E** **N** **E** **S** **K** **V** **G** **V** **G** **E** **V** **F** **Q** **T**

1185 GGGCATGGTTGGCTACGAGGCGCTCAGGATCGATCGTACAGTCTGATGCTGCTGCTTACCTATCCACTGATCGGTAACATATGGGTTACGAGG **G** **M** **V** **G** **Y** **T** **E** **A** **L** **T** **D** **R** **S** **Y** **S** **A** **Q** **I** **L** **V** **L** **T** **Y** **P** **L** **I** **G** **N** **Y** **G** **V** **P** **A**

1285 CCGGATGAGGACGACGAGCTGCCGCTGCATCTCGAGTGGACTGAAGGACGTCGTCAGGCGCCGCTTGGTGTGGGCGAGGCTGGCTAAGAGGC **P** **D** **E** **D** **E** **H** **G** **L** **P** **L** **H** **F** **E** **W** **T** **E** **G** **R** **R** **P** **G** **Q** **P** **P** **W** **W** **A** **R** **L** **A** **K** **R** **Q**

Figure 5. DNA sequence of the 5' end and the upstream flanking region of the *r* gene. The sequence downstream of the *Eco* R1 site was previously reported (10,11). The arrows clustered around the *Eco* R1 site show the transcription start sites of the *r* gene. The position of the distal most transcription start site is designated as +1. The putative transcription start sites of the neighboring gene are indicated by the arrows between positions -367 and -400. The junctions of the first intron (I1) with the first exon (E1) and the second exon (E2) are indicated. A sequence matching the criteria for branch formation, located just upstream from the splice acceptor site of the first intron, is underlined. The 8 bp duplications at the sites of the r^{hd1} and r^{hd3} P-element insertions (36) are underlined. The short sequence identities between *r* and the Hamster *CAD* gene (described in the text) are underlined and labeled A-E.

females. A probe complementary to probe F was not protected (unpublished data). These findings indicate that the adjacent gene is transcribed in the opposite direction from *r*. Indeed, two consecutive open reading frames (of approximately 800 bp) are present on the opposite strand from *r* beginning at position -471 (Fig. 5). Like the first exon of the *r* gene, multiple RNase protection products were observed with probe F as though this neighboring transcription unit also has multiple transcription start sites (indicated by the second set of arrows between positions -399 and -367 in Fig. 1). If these protection products

correspond to authentic 5' ends of this adjacent transcription unit, then the spacer region separating the two genes would only be 367 base-pairs.

Chromatin structure analysis

Promoters and other *cis*-acting regulatory sequences of eukaryotic genes are often located in nuclease hypersensitive regions of chromatin (8,13). The *r* gene is expressed in *D. melanogaster* embryos and tissue culture cells (27, 32), and in both cases the gene should be in a 'active' (or potentially active) chromatin

Figure 6. The chromatin structure of the genomic region flanking the 5' end of the *r* gene is punctuated by four nuclease hypersensitive regions. Hypersensitive regions are indicated above the molecular map in panel a. The probe, exons 1 and 2 of *r* (shaded boxes), and the exon 1 of the neighboring gene (unshaded box) are indicated. (Although the neighboring gene is shown as lacking introns, its organization has not been determined.) Panels b, c and d show the results of the indirect-end-labeling experiments described in the text. The guide next to the lanes is similar to the map in panel a, except that it portrays the logarithmic proportions of the gel. The clustered transcription start sites of each gene are marked by asterisks. So that the positions of the hypersensitive regions could be precisely aligned on the molecular map, purified genomic DNA was digested to completion with *Sac* I and partially digested with either *Mbo* I (Mbo) or *Eco* R1(R1) (see panel b). In each panel the source of nuclei is indicated as follows: KC: KC tissue culture cells, and EM: 0–24 hr embryos. In cases where samples are duplicated in a panel, the DNA sample are from different preparations. Digests of naked genomic DNA from KC cells are shown for micrococcal nuclease in panel C: the lane labeled DNA MN, and DNase I in panel D; the lane labeled DNA DN. The second KC MN digest in panel C is from heat shocked KC cells. Note the pattern is essentially unchanged by heat shock even though the *r* gene is likely to be turned off following the shift to the elevated temperature (37°C).

configuration. To identify specific DNA segments that might be involved in transcriptional regulation of the *r* gene, the indirect end labeling technique (39) was used to examine the nuclease accessibility of the DNA flanking the 5' end of the *r* transcription unit in chromatin. We prepared nuclei prepared from tissue culture cells or embryos and incubated with micrococcal nuclease or DNase I (37). The DNA was isolated and digested to completion with the restriction enzyme *Sac* I and analyzed by

agarose gel electrophoresis and blotting. The nitrocellulose filters were then probed with a *Sac* I-*Bgl* II restriction fragment located near the 3' end of the distal transcription unit (see diagram Fig. 6). The autoradiograms in Fig. 6b, c and d show the cleavage pattern obtained with the two nucleases in nuclei prepared from tissue culture cells and embryos. As a control we also digested naked genomic DNA with micrococcal nuclease (DNA MN in Fig. 6c) and DNase I (DNA DN in Fig. 6d).

85	ATTGCAAGCAACACAGTTAAATGCTCAAACATAACATATCCGAGCAGTGCGAAGTTAACAGCTCTAAAAGCGGTGTCGCAAGTAAACAAATCAAGT	genomic
	ATTGCAAGCAACACAGTTAAATGCTCAAACATAACATATCCGAGCAGTGCGAAGTTAACAGCTCTAAAAGCGGTGTCGCAAGTAAACAAATCAAGT	wt
	ATTGCAAGCAACACAGTTAAATGCTCAAACATAACATATCCGAGCAGTGCGAAGTTAACAGCTCTAAAAGCGGTGTCGCAAA.....	hd3-a
	ATTGCAAGCAACACAGTTAAATGCTCAAACATAACATATCCGAGCAGTGCGAAGTTAACAGCTCTAAAAGCGGTGTCGCAAA.....	hd3-b
	(AAAGTAAACAAATCAAGT	hd5
	hd3	
185	TGTCAATCGAATTAAAGTCGAAACAAAACGTGTTCACTACAAGTGCAGCAAGAGAGAGCTTCGCTTGAAAGTACACGGAATCGTAAGTGAGCTGT	genomic
	TGTCAATCGAATTAAAGTCGAAACAAAACGTGTTCACTACAAGTGCAGCAAGAGAGAGCTTCGCTTGAAAGTACACGGAATC.....	wt
		hd3-a
		hd3-b
	TGTCAATCGAATTAAAGTCGAAACAAAACGTGTTCACTACAAGTGCAGCAAGAGAGAGCTTCGCTTGAAAGTACACGGAATC.....	hd5
285	AGTTGTAACGGGTGCAGATGGGAAAAGGATAACTTGATGGCGATA	genomic
		wt
		hd3-a
		hd3-a
		hd5
	Intron 1	
		
		
		
985	TTTAAGACGGTTTTTCGAGTGACAGGTACTAAAGAATACCCTATGGTTTCCAATCCTCCCGAGCAACTCGTCCAATATGGCCTCTACGGACTGTTACCTCG	genomic
	CAACTCGTCCAATATGGCCTCTACGGACTGTTACCTCG	wt
	CAACTCGTCCAATATGGCCTCTACGGACTGTTACCTCG	hd3-a
	CAACTCGTCCAATATGGCCTCTACGGACTGTTACCTCG	hd3-b
	CAACTCGTCCAATATGGCCTCTACGGACTGTTACCTCG	hd5
1085	CACTGGAGGATGGCAGAGTGTGCCCGGTACTCCTTTGGCTACGTTCCCTCGGAGAATGAATCCAAAGTGGGTGTTGGTGGCGAGGTGGTCTTCCAAAC	genomic
		hd3-b
	CACTGGAGGATGGCAGAGTGTGCCCGGTACTCCTTTGGCTACGTTCCCTCGGAGAATGAATCCAAAGTGGGTGTTGGTGGCGAGGTGGTCTTCCAAAC)	hd5
1185	GGGCATGGTTGGCTACACGGAGCGCTACGGATCGATCGTACAGTGCTCAGATCCTGGTGCTTACCTATCCACTGATCGGTAACCTATGGGGTACCAGCG	genomic
	CG	hd3-b
1285	CCGGATGAGGACGAGCAGGACTGCCGCTGCACTTCGAGT	genomic
	CCGGATGAGGACGAGCAGGACTGCCGCTGCACTTCGAGT	hd3-b

Figure 7. Nucleotide sequence of cloned *r* cDNAs. PCR amplified *r* cDNAs generated from RNA from wild type, r^{hd3} homozygous, and r^{hd5} homozygous females are shown below the wild type genomic sequence.

As evident from inspection of the autoradiograms in Fig. 6, the DNA segment spanning the 5' end of the *r* transcription unit contains several major nuclease hypersensitive regions. These hypersensitive regions are cleaved by both micrococcal nuclease and DNase I and are observed in nuclei prepared from tissue culture cells and from 0–24 hr embryos. Moreover, with both enzymes, the cleavage pattern in chromatin is clearly distinct from that obtained in the corresponding naked DNA digests. This result would indicate that these four hypersensitive regions are chromatin specific.

The nuclease hypersensitive regions were aligned with respect to the underlying sequence using molecular weight markers and partial restriction enzyme digests of genomic DNA (in Fig. 6a, *Mbo*I and *Eco* R1). The precision of the localization of these regions is approximately twenty-five base-pairs. The most distal hypersensitive region (region 1) exposes the putative transcription start sites of the neighboring gene and extends approximately 100 bp into the intergenic spacer. This is followed by a nuclease resistant segment of nearly 200 bp which covers much of the short intergenic spacer separating the two transcription units. The length of this nuclease resistant segment is sufficient to accommodate a nucleosome core particle. Hypersensitive region 2 is about 150 bp in length and exposes the *r* promoter. The *r* transcription unit contains two internal hypersensitive regions. The first of these, region 3, maps ~200 bp (or one nucleosomal unit) downstream of region 2 and exposes ~50 bp at the 3' end of the non-translated exon 1. The second internal hypersensitive

region, region 4, begins ~400 bp from the *r* promoter and is approximately 225 bp in length. It has a more complex DNase I and micrococcal nuclease cleavage pattern than the other regions. We also found that the four DNase I and micrococcal nuclease hypersensitive regions are preferentially cleaved by S1 nuclease and *N. crassa* nuclease (not shown), enzymes that are thought to recognize either single-stranded DNA or double-stranded DNA in an altered non-B-form conformation (21,31).

Transcripts in two P element induced *r* mutants

Of the five P element insertion *r* mutations reported previously (35), three are within the first *r* exon (r^{hd1} , r^{hd2} and r^{hd3}). The two others are within the first intron (r^{hd4} and r^{hd5} , Fig. 1). Northern analysis of mRNA produced by these mutants (36) revealed that *r* transcript levels are reduced relative to wild type in all except $hd1$, which has about a two fold higher *r* transcript level than wild type. Considering that three of the mutants have P elements of 0.7–1.0 kb inserted in the first exon, it was surprising that the transcripts produced by the mutants have the same apparent electrophoretic mobility as the 7.3 kb wild type transcript.

The r^{hd5} allele, which contains an incomplete P element insertion of approximately 1.1 kb (35) within the first *r* intron, has a weak *r* mutant wing phenotype. An r^{hd5} cDNA generated by reverse transcription and PCR amplification (Fig. 7) was found to be identical to a cDNA of the wild type *r* mRNA across the junction of exons 1 and 2. Thus the r^{hd5} P element insertion

sequences are spliced out of the *r* transcript with the first intron. This conclusion is supported by RNase protection and primer extension analysis of the *r*^{hd5} transcript which revealed that the locations of the 5' ends, and the organization of the first and second exons, are indistinguishable from those of the wild-type *r* transcript (unpublished data).

In contrast to hd5, the *r* transcripts produced by hd1 and hd3 clearly differ in structure from *r* transcripts of a wild type stock. The *r*^{hd1} P element insertion is located about 60 bp upstream of the 5' splice junction of exon 1 (Fig. 5) and this mutant has a moderate *r* wing phenotype (35). Although wild-type levels of the second *r* exon were observed in RNase protection experiments using RNA from hd1, no protection products corresponding to the first exon were detected (unpublished data). Moreover, we were also unable to generate hd1 cDNAs by reverse transcription and PCR amplification using oligonucleotide primers in exons 1 and 2. The findings would indicate that the hd1 *r* transcripts lack the normal first exon and hence might be initiated from within the P element insertion.

The *r*^{hd3} P element is inserted just upstream of the 5' splice donor of the first exon and thus might be expected to disrupt the normal splicing of *r* intron 1. Two *r*^{hd3} cDNAs, hd3-a and hd3-b (Fig. 7), were obtained which have exon 2 joined to exon 1 at a cryptic splice donor site located 102 nt 5' to the normal donor site. The sequence of this donor site (AA/GTAAAC) is consistent with the consensus for invertebrate splice donor sites (AG/GTAAGT) (34). This cryptic donor site was joined either to the intron 1 acceptor site (hd3-a, Fig. 7) or to a cryptic acceptor site within exon 2 (hd3-b). The *r* transcripts produced by this mutant most probably initiate at the *r* promoter because the set of primer extension products synthesized with hd3 RNA (using primer 3) were same as those observed with wild type RNA (unpublished data). Results of RNase protection analysis of hd3 RNA were consistent with the hd3-a transcript being the predominant *r* transcript species expressed from this allele (unpublished results). However, other minor species, which were not accounted for by cDNAs, are also expressed from this allele.

DISCUSSION

This report describes the characterization of the 5' end of the *r* gene of *D. melanogaster*. Our results place the *r* promoter only about 350 bp from the 5' end of another transcription unit in the opposite orientation and having two consecutive extended open reading frames of unknown function. The *r* promoter region does not appear to contain a TATA box and transcription initiation occurs at 5 discrete sites within a fifty bp interval. While we have not ruled out the possibility that this heterogeneity arises from the developmental or tissue specific use of different transcription initiation sites, this scenario seems unlikely since the ratios of the different 5' ends were found to be constant throughout development and between males and females. The adjacent gene also appears to produce transcripts with several different 5' ends. The first *r* exon extends downstream to a single splice donor site located at position +273 bp. The 3' splice acceptor of intron 1 is located 13 nt upstream from the AUG thought to correspond to the translation initiation codon for the 249 kd *r* polypeptide. Two additional AUG codons are found in the 5' *r* leader encoded by exon 1. The first is at +8 and hence is present only in transcripts initiating at the most distal of the five *r* transcription start sites. This AUG is followed by a short

17 codon open reading frame terminating with UGA. The second AUG codon is at position +107 and is present in all five exon 1 variants. Like the upstream AUG, it is also followed by a short 15 codon open reading frame which terminates with a UAA instead of UGA.

As has been observed for many polymerase II transcribed genes (13), the promoter regions of both *r* and its distal neighboring transcription unit are highly sensitive to cleavage by both DNase I and micrococcal nuclease in isolated nuclei. The *r* 5' hypersensitive region spans just over 150 bp and includes all five of the transcription start sites. In addition to promoters, enhancer elements are often hypersensitive to nuclease cleavage in chromatin. Such hypersensitive sites are not evident in the very short 5' flanking spacer which separates *r* from the adjacent gene. Instead, this 5' spacer segment is most probably associated with a nucleosome core particle. There are, however, two quite prominent hypersensitive regions within the *r* transcription unit. The first of these, region #3, is about 200 bp downstream from the *r* promoter, close to the 3' end of exon 1. The second, region #4, is 400 bp from the *r* promoter and contains sequences within the first intron. Given the fact that the *r* gene has only a very short and nuclease resistant 5' flanking segment, these findings suggest that expression may be regulated by internal control elements located in hypersensitive regions #3 and #4. Consistent with this possibility, the internal control elements of a variety of other genes are found to be hypersensitive to nuclease attack in isolated nuclei (1,5,13,22).

The characterization of the 5' end of the *r* homolog from the Syrian hamster, the *CAD* gene, was reported recently (9). As might be expected for non-protein coding regions of genes from such highly diverged species, there is little apparent overall sequence similarity between the 5' regions of *r* and *CAD*. Yet there are some short elements which have an identical or very similar nucleotide sequence. The most interesting of these elements are the ones located in nuclease hypersensitive regions at the 5' of the *r* gene. The first element ('A' in Fig. 5) is an 8-mer located within the 5' hypersensitive region just upstream from the *r* and *CAD* transcription start sites. The second ('B' in Fig. 5), is also an 8-mer but is located in the middle of the hypersensitive region near the end of *r* exon 1. This same 8-mer is found in exon 1 of *CAD*. The third similarity ('C' in Fig. 5) is a 12 bp sequence (with 11 out of 12 residues identical). And the fourth ('D' in Fig. 5) has 13 out of 14 residues identical. The latter two sequences are located in intron 1 of both *r* and *CAD* and, in the *r* gene, are within the large nuclease hypersensitive region #4. That these sequences have similar locations within *r* and *CAD*, and are located in hypersensitive regions of *r*, suggest that they have a functional role in the expression or regulation of *r* and *CAD* expression.

A number of findings of interest have come from our analysis of the P element induced *r* mutations. As indicated in the diagram in Fig. 1, all five of the P element insertion mutations described previously (35) are within (or at the edge of) the nuclease hypersensitive regions at the 5' end of the transcription unit. The *r*^{hd1}, *r*^{hd2}, and *r*^{hd3} insertions are located either in, or at the edge of, the short hypersensitive region #3 near the end of exon 1. *r*^{hd4} and *r*^{hd5} are located in the large hypersensitive region, #4, of the first *r* intron. This preference for nuclease hypersensitive regions in the *r* locus would suggest that potential insertion sites for P elements might be restricted to sequences located in DNA segments which are accessible in chromatin to the various

enzymes involved in the insertion process. This view is supported by recent studies on 5' P element insertions in the suppressor of sable (su(s)) locus (38). As in *r*, all five of the P element insertions are located in DNase I hypersensitive regions in chromatin. Moreover, a comparison of the DNA sequence for the various P insertion sites in these two genes reveals little similarity (35, 38). This would also support the suggestion that chromatin structure, rather than the presence of a specific consensus insertion sequence, may be the critical determinant in site selection. The importance of chromatin structure in P element site selection is further supported by the finding that P elements preferentially insert around the 5' ends of a wide variety of other *D. melanogaster* genes. These include *RpII215* (33), *Notch* (20), *Sex-lethal* (Sxl) (29), *yellow* (12), *singed* (28), *daughterless* (4). Moreover, in the cases of *Notch* and *Sxl*, chromatin structure analysis indicates that P element insertions in these loci are also in hypersensitive DNA segments (Schedl, unpublished data).

Analysis of some of the P element induced *r* mutations revealed that each of these insertions have a different effect on the expression of *r* transcripts. The P element insertion of *r*^{hd5} is in the first intron and the *r* transcripts it produces appear to be spliced at the normal *r* first intron junctions. Thus, as proposed previously (36), the slight *r* mutant phenotype of this allele is likely to result from an effect of this insertion on the level or perhaps the tissue specificity of *r* transcription.

In contrast, the *r*^{hd1} and *r*^{hd3} P element insertions are within exon 1 sequences and the *r* transcripts produced by these mutants are structurally different from wild type. In RNA from hd1, we did not detect any transcripts having wild type 5' ends. Yet adult hd1 flies were observed previously to produce relatively high levels of normal size *r* transcripts (35) and we have found that exon 2 has an apparently wild type structure. A plausible explanation for these results is that the *r* transcripts produced by *r*^{hd1} are initiated from some site (or sites) within the P element. These transcripts might then be spliced from a 5' junction somewhere further 3', but still within the P element, to the normal *r* first intron acceptor site. Thus, the *r* mutant phenotype of this allele could well result from the absence of the wild type 5' non-translated leader sequence or an altered pattern of expression.

Unlike hd1, most of the *r* transcripts produced by hd3 are initiated at the normal *r* transcription start sites. However, the splicing of the *r* first intron is effected. In this allele, the P element insertion is just eleven nucleotides upstream of the first intron splice donor site, and our results indicate that the element affects the utilization of this junction. *r*^{hd3} transcripts were detected from which the P element had been spliced out by the use of a cryptic splice donor site in the first *r* exon. This cryptic donor site was either joined to the splice acceptor site of intron 1 or to a cryptic splice acceptor site within exon 2. The mutant phenotype of hd3, therefore, likely results from the aberrant processing of its *r* transcripts.

ACKNOWLEDGEMENTS

We thank D.Bopp, B.Bucher, R.Doreig, B.Kellum, G.Horabin, K.Louis, M.Samuels, E.Schmidt, and B.Suter for helpful comments and discussions. We also thank S.Tsubota for providing the mutant strains. This work was supported by NIH grant GM25976. W.Z. was supported by NIH Predoctoral Training Grant GM07312.

REFERENCES

- Aronow, B., Lattier, D., Silbiger, R., Dusing, M., Hutton, J., Jones, G., Stock, J., McNeish, J., Potter, S., Witte, D. and Wiginton, D. (1989) *Genes Devel.*, 3, 1384–1400.
- Carlson, P.S. (1971) *Genet. Res. Camb.*, 17, 53–81.
- Church, G.M. and Gilbert, W. (1984) *Proc. Natl. Acad. Sci., USA* 81, 1991–1995.
- Cronmiller, C., Schedl, P. and Cline, T.W. (1988) *Genes Devel.*, 2, 1666–1676.
- DeLotto, R. and Schedl, P. (1984) *J. Mol. Biol.*, 179, 607–628.
- Doerig, R.E., Suter, B., Gray, M. and Kubli, E. (1988) *EMBO J.*, 7, 2579–2584.
- Dynan, W.S. (1986) *Trends Genet.*, 2, 196–197.
- Elgin, S.C.R. (1988) *J. Biol. Chem.*, 263, 19259–19262.
- Farnham, P.J. and Kollmar, R. (1990) *Cell Growth and Diff.*, 1, 179–189.
- Freund, J.N., Vergis, W., Schedl, P. and Jarry, B.P. (1986) *J. Mol. Biol.*, 189, 25–36.
- Freund, J.N. and Jarry, B.P. (1987) *J. Mol. Biol.*, 193, 1–13.
- Geyer, P.K. and Corces, V. (1987) *Genes Devel.*, 1, 996–1004.
- Gross, D.S. and Garrard, W.T. (1988) *Ann. Rev. Biochem.*, 57, 159–197.
- Han, S., Udvardy, A. and Schedl, P. (1984) *J. Mol. Biol.*, 179, 469–496.
- Hultmark, D., Klimenz, R. and Gehring, W.J. (1986) *Cell*, 44, 429–438.
- Jarry, B.P. and Falk, D. (1974) *Mol. Gen. Genet.*, 135, 113–122.
- Jarry, B.P. (1979) *Mol. Gen. Genet.*, 172, 199–202.
- Jones, K.A., Yamamoto, K.R. and Tjian, R. (1985) *Cell*, 42, 559–572.
- Keller, E.B. and Noon, W.A. (1985) *Nucl. Acids Res.*, 13, 4971–4981.
- Kelly, M.R., Kidd, S.R., Berg, L. and Young, M.W. (1987) *Mol. Cell. Biol.*, 7, 1545–1548.
- Larsen, A. and Weintraub, H. (1982) *Cell*, 29, 609–622.
- Louis, C., Schedl, P., Samal, B. and Worcel, A. (1980) *Cell*, 22:387–392.
- Maniatis, T., Fritsch, E.R. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor University Press, Cold Spring Harbor.
- Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) *Nucl. Acids Res.*, 12, 7035–7056.
- Norby, S. (1973) *Hereditas*, 73, 11–16.
- Rawls, J. and Fristrom, J. (1975) *Nature (London)*, 255, 738–740.
- Rawls, J.M., Freund, J.N., Jarry, B.P., Louis, C.M., Seagraves, W.A. and Schedl, P. (1986) *Mol. Gen. Genet.*, 202, 493–499.
- Roiha, H., Rubin, G.M. and O'Hare, K. (1988) *Genetics*, 119, 75–83.
- Salz, H.K., Cline, T.W. and Schedl, P. (1987) *Genetics*, 117, 221–231.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, 74, 546.
- Schon, E., Evans, T., Welsh, J. and Efstratiadis, A. (1983) *Cell*, 35, 837–848.
- Seagraves, W.A., Loius, C.M., Tsubota, S., Schedl, P.M. and Jarry, B.P. (1984) *J. Mol. Biol.*, 175, 1–17.
- Searles, L.L., Greenleaf, A., Kemp, W.E. and Voelker, R.A. (1986) *Mol. Cell. Biol.*, 6, 3312–3319.
- Shapiro, M.B. and Senapathy, P. (1987) *Nucl. Acids Res.*, 15, 7155–7156.
- Tsubota, S.I., Ashburner, M. and Schedl, P. (1985) *Mol. Cell Biol.*, 5, 2567–2574.
- Tsubota, S.I. and Schedl, P. (1986) *Genetics*, 114, 165–182.
- Udvardy, A. and Schedl, P. (1984) *J. Mol. Biol.*, 172, 385–403.
- Voelker, R.A., Graves, J., Gibson, J. and Eisenberg, M. (1990) *Genetics*, 126, 1071–1092.
- Wu, C. (1978) *Cell*, 16, 797–806.