MOLECULAR AND CELLULAR BIOLOGY, Jun 1989, p. 2526–2535
0270-7306/89/062526-10$02.00/0

Vol. 9, No. 6

Transcriptional and Posttranscriptional Regulation Contributes to the Sex-Regulated Expression of Two Sequence-Related Genes at the janus Locus of *Drosophila melanogaster*

CONSTANTIN YANICOSTAS,[1]* ALAIN VINCENT,[2] AND JEAN-ANTOINE LEPESANT[1]

Institut Jacques Monod, Centre National de la Recherche Scientifique and Université Paris 7, 2 Place Jussieu, 75251 Paris Cedex 05,[1] and Centre de Recherche de Biochimie et de Génétique Cellulaires de Centre National de la Recherche Scientifique, 31062 Toulouse Cedex,[2] France

Received 9 December 1988/Accepted 23 February 1989

We investigated the structure and developmental pattern of expression of two genes clustered at the janus locus, located at 99D.3R. Data obtained from genomic and cDNA sequencing and from a combination of S1 mapping and primer extension experiments indicated a very unusual organization of this locus, which appeared to be composed of two partially overlapping genes, designated *janA* and *janB*. These two genes were found to be transcribed in the same direction. *janA* encoded one minor and two major transcripts. The 5′ end of the *janB* mRNA mapped within the 3′ untranslated region of the *janA* transcribed sequence. The overlapping region was 118 bases long. Similarities observed between these two genes with respect to both peptidic sequence and intron position strongly suggested that this locus originated from the duplication of an ancestral transcription unit. However, each of the resulting genes has acquired its own specificity of expression linked to sex determination. The *janB* transcript was detected only in males, and its expression at the adult stage was restricted to germ line cells. The *janA* gene displayed a much more complex expression; one of the major mRNAs was found in both sexes and at all stages, whereas the two other *janA* transcripts were expressed only in males.

In *Drosophila melanogaster*, the X/A ratio is the basic signal of sex determination (8). It controls both dosage compensation and sex differentiation of germ line and somatic cells. On one hand, pole cell transplantation experiments (24, 33, 34) have demonstrated that sex differentiation of germ cells is under the control of the X/A ratio and of the *Sxl* gene, independently of the state of activity of the other sex-determining genes, *tra*, *tra-2*, *ix*, and *dsx*. On the other hand, genetic and, more recently, molecular studies have shown that at least eight genes govern the expression of sex-specific functions in somatic cells. Briefly, the X/A ratio, the zygotic products of the *sis-a* and *sis-b* genes (12, 13), and the maternal product of the *da* gene (14) determine the ON or OFF state of the key gene, *Sxl*, which in turn regulates a small cascade of sex-determining genes: *tra*, *tra-2*, *ix*, and *dsx* (for reviews, see references 1 and 26). The products of the sex-determining genes regulate the expression of somatic sex differentiation functions (realisator genes) (2, 6, 10, 11, 16, 32).

A molecular genetic approach, i.e., isolation of male- or female-specific realisator genes (16, 25, 31), should help to identify the mechanisms of control of gene expression by the X/A ratio, identify the factors involved, and assess the function of these products.

We report here on the structure and expression of two sex-regulated genes at the newly identified janus locus. These two genes, *janA* and *janB*, give rise to partially overlapping transcripts encoding proteins with strong sequence similarities. Our data suggest that *janA* and *janB* originated from the duplication of a common ancestral transcription unit and have acquired separate specificities of expression through the use of both posttranscriptional and transcriptional regulatory elements.

* Corresponding author.

MATERIALS AND METHODS

***Drosophila* stocks.** The Canton S stock of *Drosophila melanogaster* was used throughout. Germ line-less flies were obtained by crossing *tud*^{1} *bw sp*/*tud*^{1} *bw sp* (5) virgin females with wild-type males of the Canton S stock.

RNA isolation. Embryos, sex-differentiated larvae, pupae, and adult flies grown on a standard sucrose-agar-yeast-corn meal medium were frozen in dry ice and homogenized in extraction buffer (100 mM Tris hydrochloride [pH 9], 50 mM NaCl, 0.5% sodium dodecyl sulfate) in a Dounce homogenizer. The homogenate was extracted three times with water-saturated phenol at 68°C and twice with phenol-chloroform (1:1) and then ethanol precipitated. Poly(A)$^+$ mRNA was isolated by oligo(dT)-cellulose chromatography as described by Marcu et al. (23).

Northern (RNA) blot analysis. Formaldehyde-gel electrophoresis was performed by using 1 or 1.2% agarose gels as described by Maniatis et al. (22). RNA was transferred to nitrocellulose filters as described by Thomas et al. (36) in 20× SSC (SSC is 0.15 M NaCl plus 0.015 M sodium citrate) for 24 h. After baking at 80°C for 2 h, the filters were prehybridized for 4 h at 42°C, hybridized at the same temperature for 24 h, and washed as described by Wahl et al. (41). Exposure to Kodak X-Omat S or X-AR film (Eastman Kodak Co., Rochester, N.Y.) was done at −70°C with intensifying screens.

Radiolabeling of nucleic acids. Double-stranded gel-purified fragments (40 to 80 ng) were labeled by random primer labeling (17), using 30 μCi each of [^{32}P]dATP and [^{32}P]dCTP (>3,000 Ci/mmol; Amersham Corp., Arlington Heights, Ill.). Single-stranded pEMBL recombinant bacteriophage DNAs were labeled by Klenow extension, using the 17-mer M13 universal hybridization probe primer (New England BioLabs, Inc., Beverly, Mass.) as described by Hu and Messing (19), with the following modifications. The annealing mixture containing 3 μl (7.5 ng) of primer, 2 μl of single-stranded

DNA template (300 ng), and 2 μl of 10× NT buffer (500 mM Tris hydrochloride [pH 7.2], 100 mM $MgCl_2$, 1 mM dithiothreitol) in a final volume of 10 μl was incubated at 70°C for 30 min and slowly cooled at room temperature for 30 min. The probe synthesis reaction was allowed to proceed for 15 min at 20°C in a 20-μl final volume after addition of cold dGTP and dTTP (10 μM) and 5 U of Klenow enzyme.

DNA isolation and cloning. λDm janus phages were isolated from libraries prepared from poly(A)$^+$ RNA of embryos and adult males, kindly provided by L. Kauvar (28). Inserts were gel purified after digestion by suitable restriction enzymes and subcloned in both orientations in pEMBL 18^+ and 19^+ plasmids (15).

DNA sequencing. Both genomic DNA and cDNAs were sequenced by the chain termination method of Sanger et al. (30). The genomic DNA sequence was obtained from subclones generated after digestion of large DNA fragments, using six-cutter restriction enzymes.

Transcript mapping by S1 nuclease protection assay and primer extension. For S1 protection experiments (3), 10 μg of poly(A)$^+$ RNA was hybridized with purified labeled restriction fragments in 100 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES; pH 7.8)–10 mM EDTA–400 mM NaCl–70% formamide in a final volume of 20 μl. DNA was denaturated for 3 min at 75°C, and hybridization was carried out for 3 h at 50°C. The reaction mixture was then diluted with 10 volumes of ice-cold S1 buffer (280 mM NaCl, 50 mM sodium acetate [pH 4.5], 4.5 mM $ZnSO_4$, 20 mg of denaturated salmon sperm DNA per ml) and digested with S1 nuclease (1,000 U/ml) for 30 min at 30 or 37°C. Digestion products were analyzed by electrophoresis in 8% denaturing polyacrylamide gels.

Primer extension experiments were carried out essentially as described by Calzone et al. (9), with minor modifications. p331, a 22-mer oligonucleotide (5′-TTTTCATTTTATTTAA CCAAAG-3′) complementary to both janus transcription units between positions +862 and +883, and p429, a 21-mer oligonucleotide (5′-TTGATCATGACATACTTGAAG-3′) complementary to the *janA* transcript between positions +186 and +206, were used as primers. ^{32}P-5′-end-labeled oligonucleotide (10 nmol; 5×10^6 cpm/pmol of ends) was coprecipitated with 5 μg of poly(A)$^+$ RNA. Pellets were dissolved in 20 μl of 15 mM Tris hydrochloride (pH 8)–1 mM EDTA. Mixtures were heated for 3 min at 85°C, incubated for an additional 3 h at 65°C after addition of NaCl to 400 mM, and then diluted to 100 μl in reverse transcriptase buffer containing 200 U of Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories, Inc., Bethesda, Md.). Extension products were analyzed by electrophoresis in 8% denaturating polyacrylamide gels.

Determination of mRNA poly(A) tail length with RNase H. For RNase H digestion, 25 μg of total RNA was incubated for 20 min at 37°C in a 30-μl reaction mixture containing 100 mM KCl, 20 mM Tris hydrochloride (pH 7.5), 10 mM $MgCl_2$, 1 mM dithiothreitol, 1.4 U of RNase H (Bethesda Research Laboratories), and 200 μg of oligo(dT) (Boehringer Mannheim Biochemicals, Indianapolis, Ind.) per ml. Control experiments were performed by omitting RNase H and oligo(dT) or oligo(dT) alone. RNAs were then phenol extracted and subjected to Northern blot analysis.

RESULTS

Developmental Northern blot analysis of *Drosophila* poly(A)$^+$ RNAs, using probes adjacent to the serendipity (*sry*) locus, initially identified a transcription unit showing

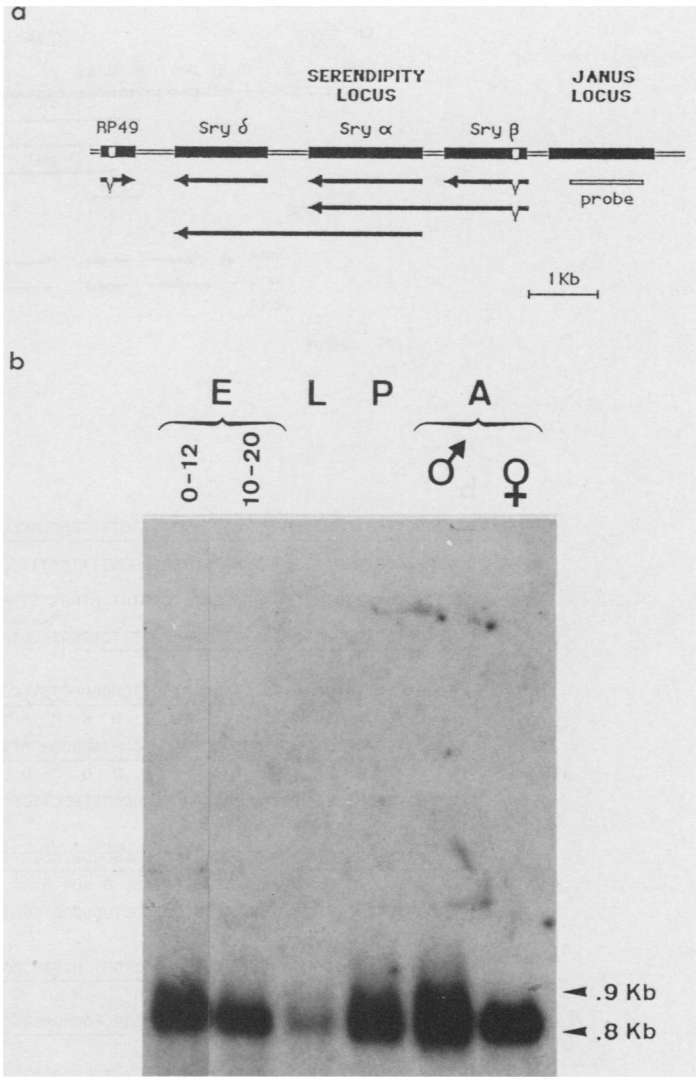
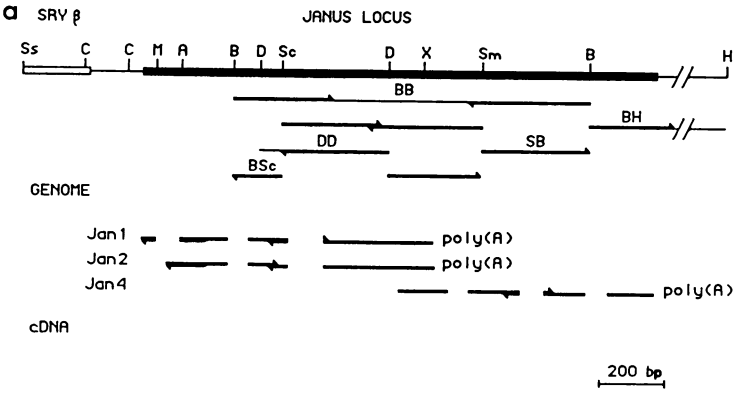


FIG. 1. (a) Partial transcriptional map of the 99D region. Symbols: ▬, genes (the corresponding mRNAs are indicated by arrows pointing in the direction of transcription); ▭, introns. Position of the BB genomic subclone used as the hybridization probe in panel b is shown (see precise position in Fig. 2a). (b) Transcriptional profile of the janus locus throughout development. Samples (1.5 μg) of poly(A)$^+$ RNA from different developmental stages (0- to 12- and 10- to 20-h-old embryos [E], third-instar larvae [L], brown pupae [P], and adult males and females [A]) were electrophoresed, transferred to nitrocellulose, and hybridized to single-stranded BB DNA. *Eco*RI-*Hin*dIII-digested λ phage DNA was run in parallel as a molecular weight marker (not shown).

sex specific expression at 99D (38, 40). Figure 1a shows the approximate position, relative to the *sry* locus, of this transcription unit, previously termed Y and redesignated here as janus.

To delineate more precisely the janus transcription unit(s) and direction of transcription, Northern blots of mRNA extracted from various stages were hybridized with a single-stranded probe generated from a 1-kilobase (kb) genomic *Bam*HI-*Bam*HI (BB) fragment (see position of Fig. 2a). The results showed that the janus region is transcribed throughout development in an orientation opposite to that of the *sry* genes. The reproducible widespread appearance of the band detected in males compared with that observed in females



b

```
ATTCGGCTTAAACAATTTAATTTGTGTATATTTTGTTGTGAACGCCAGAGCTGTGCCGATAGTGCCGATAGTATCGACTGCGTGCTGTCG -181
Sryβ ↔ Start transcription
GCGTAATCGATAAATTTGCTGTCACTGATAACAACGTTTTCTTTTTGAGTTTATTAATTATTACTAAAATATACTGAGTAGTACAAAAAA -91

CGTTTTCCCAAATGTACTAAAGAAATAACTGAATATTATAATTTTAATAGTATCGATACATAAGGTGAACGAGAATAAAAGTATCTGGTC -1
5'End Jan1cDNA                                    5'End Jan1intron1                    5'End Jan2cDNA
ACATTGCTGGACTAAAGCAGCGTTTTTGGAAAATTTGCCGGTTGGTAAGACATTAAATTCTGTTTTCAAACACTTTTCCACAATGAATCG 90
        A                                                                                 M N R
                3'End Jan1intron1
CCTCCAACTGCTTTCCAAAGGACTACGACTGATTCACAAAATGTCCGAGGAAGCACTTGCCGGCGTGCCACTGGTGCACATCAGTCCAGA 180
  L Q L L S K G L R L I H K M S E E A L A G V P L V H I S P E

GGGCATCTTCAAGTATGTCATGATCAATGTCTTCGATGGAGGAGATGCTTCAAAGGCGGTGATCCGCGGATTTGCGGACTGCACATGGCA 270
  G I F K Y V M I N V F D G G D A S K A V I R G F A D C T W H

TGGTAAGTCGGATCCTCATCACCCATCAAGTGCCCACTTAGCTTGGTTACTGTCCCACAGCCGACATCTTCGAGCGCGAGGAGGAGGTCT 360
                                                    C
                                                A D I F E R E E E V

TTAAAAAACTGGGGCTGCGGGCCGAGTGTCCTGGCGGCGGTCGCATTGAACACAATCCCGAGAAGAAGTACTTGAAGGTCTACGGATACT 450
F K K L G L R A E C P G G G R I E H N P E K K Y M K V Y G Y

CGCAGGTGGGTCTATTCCTTGAGTAAAGGGGTCGCTGGGCAGTGGATGGACTGATGTATCTAACTACTTGAAAAATTTTCTGGAGTTCTA 540
S Q

ATCAAGTCTTTCTATTTAAGGGCTTTGGAAAAGCTGATCACGCGCAGACCAAACGCATCCTGGCCACCAAATACCCGGACTACACGATCG 630
                  G F G K A D H A Q T K R I L A T K Y P D Y T I

AAATCTCCGATGAGGGATATTAGCTGCAATCAACGAGAGAAGACTCCACATAAGCACACTGACATGAATTTATACCATTGGCTTCGATCC 720
E I S D E G Y END
                                                    5'End Jan4cDNA
TGTGTGCCATGATTTTATTGGAAATGGCATTTAAAATTGAGAAATACTCTGAAAGGCAGTTAGTCTGTAGCTTTGCAACTGCTCGCACTA 810
                                                                Poly(A) Jan1 Jan2
AACCTTTTCGGATCTAAATTAATCAGTTTGTACACAAATTTCGTTTCTTTTCCTTTGGTTAAATAAAATGAAAATGTTCAAGTCATTGCG 900
                                                          M K M F K S L R

TCTGCTTCCTCATATTGTTTCTCCGTTTCGTAAGGCTTAGGAATATTCAATATTAAGATTTACAAGCCCTAATATACTTGGTTTTAGAAA 990
  L L P H I V S P F                                                                   Q

AATGTTACTCAACCGATTTGATAAGTTTGGTAGGCGTTCCCCGGGTCAAGATAACCAAGGGTCAGAATCGTTATTTGTTGGTGAATATTC 1080
K C Y S T D L I S L V G V P R V K I T K G Q N R Y L L V N I

ATACGCATGGCTTCACGAAGTATGGAAGAGTTATTGTCCGTGGCGCCGATGTTGACAATCACTGTGAGTTTCCACTGCTGGACGCTTAAC 1170
H T H G F T K Y G R V I V R G A D V D N H

CTTGAGCAGTCTTACAAATCCTTCTTTCAGTGGCGGTCTTCGACTCGATTTTGGAGGAGCTGGAACCCGAGGGCATATGTGCCAAAATCC 1260
                  L A V F D S I L E E L E P E G I C A K I

TCGGTGGTGGAAGGATTCTCAACGAGGCAGAAAATAAAAAAATTAAGATCTATGGCACCTCCAGGGTAAGTAGAGGATCCTTGGTCCTTG 1350
L G G G R I L N E A E N K K I K I Y G T S R

AAGCACCGGCTAATGGTTCTTGATGGGTCTCCCTAGACTTTCGGCGGTGCTGATCACACAAGGACAAGGAATATACTTCAAGCGTGGACC 1440
                                    T F G G A D H T R T R N I L Q A W T

ACTTATAAGGACTTTAAGATAACCGTTAAACAATAAAGTTGCATAAATTTCGAAAATGGAAATTCAGTACTAATAAAAAGAAAATAGAAT 1530
T Y K D F K I T V K Q END
Poly(A) Jan4
ATAAAACTAGCGCTCTTTCAATATTATTAAGGGGTAATCGACAGGCGATTGTAATTTGGCTTCGATCCTGTGTGCCATCATTTTATTGGA 1620
```

FIG. 2. Genomic and cDNA sequences of the janus locus. (a) Simplified restriction map of the janus locus (Ss, *Sst*I; C, *Cla*I; M, *Msp*I; A, *Apa*LI; B, *Bam*HI; D, *Dra*I; Sc, *Sca*I; X, *Xmn*I; Sm, *Sma*I; H, *Hin*dIII) and positions of genomic subclones and cDNA clones. Arrows indicate extent and direction of sequencing. (b) Nucleotide sequence of the janus locus and deduced amino acid sequences starting from the first AUG in each of the three cDNA clones Jan 1, Jan 2, and Jan 4. Symbols: —, cDNA sequences identical to those of genomic DNA except for two polymorphic nucleotides at positions +9 and +350; - - -, 5′- and 3′-end region of the Jan 2 cDNA. Putative TATA boxes are boxed; AUG initiating codons are underlined. The genomic sequence from positions −270 to +285 was previously determined by Vincent et al. (38).

(Fig. 1b) suggested the existence of more than one transcript in males. Further analysis with subclones of the BB fragment (results not shown) suggested that these male transcripts originated from two closely linked but separate transcription units.

***janA* and *janB*, two genes with partly overlapping transcripts.** To confirm the existence of separate classes of *jan* mRNAs, cDNA clones were isolated by screening cDNA libraries made from poly(A)$^+$ RNA from 0- to 12-h-old embryos or adult males (28), using the 1-kb BB fragment as a probe. Three independent embryonic (referred to as Jan 1, 2, and 3) and one male (Jan 4) cDNA clones were isolated and sequenced. A 1,500-base-pair (bp) genomic DNA fragment encompassing the janus cDNAs was also sequenced by the strategy summarized in Fig. 2a.

Comparison of the genomic and cDNA sequences shows that Jan 1 cDNA consists of four exons with respective sizes of 44, 162, 125, and 335 bp; the corresponding introns are 66, 58, and 106 bp long. The Jan 2 cDNA starts at position +72 on the genomic sequence (with position +1 corresponding to the Jan 1 cDNA 5′ end), i.e., inside the first intron of Jan 1, and is composed of three exons of 201, 125, and 340 bp, respectively; the corresponding introns are 58 and 106 bp long. Thus, the Jan 1 and Jan 2 cDNAs differ in their 5′ ends. The poly(A) addition sites are at position +891 for Jan 1 and position +895 for Jan 2. The cDNA clone Jan 3 is only 402 bp

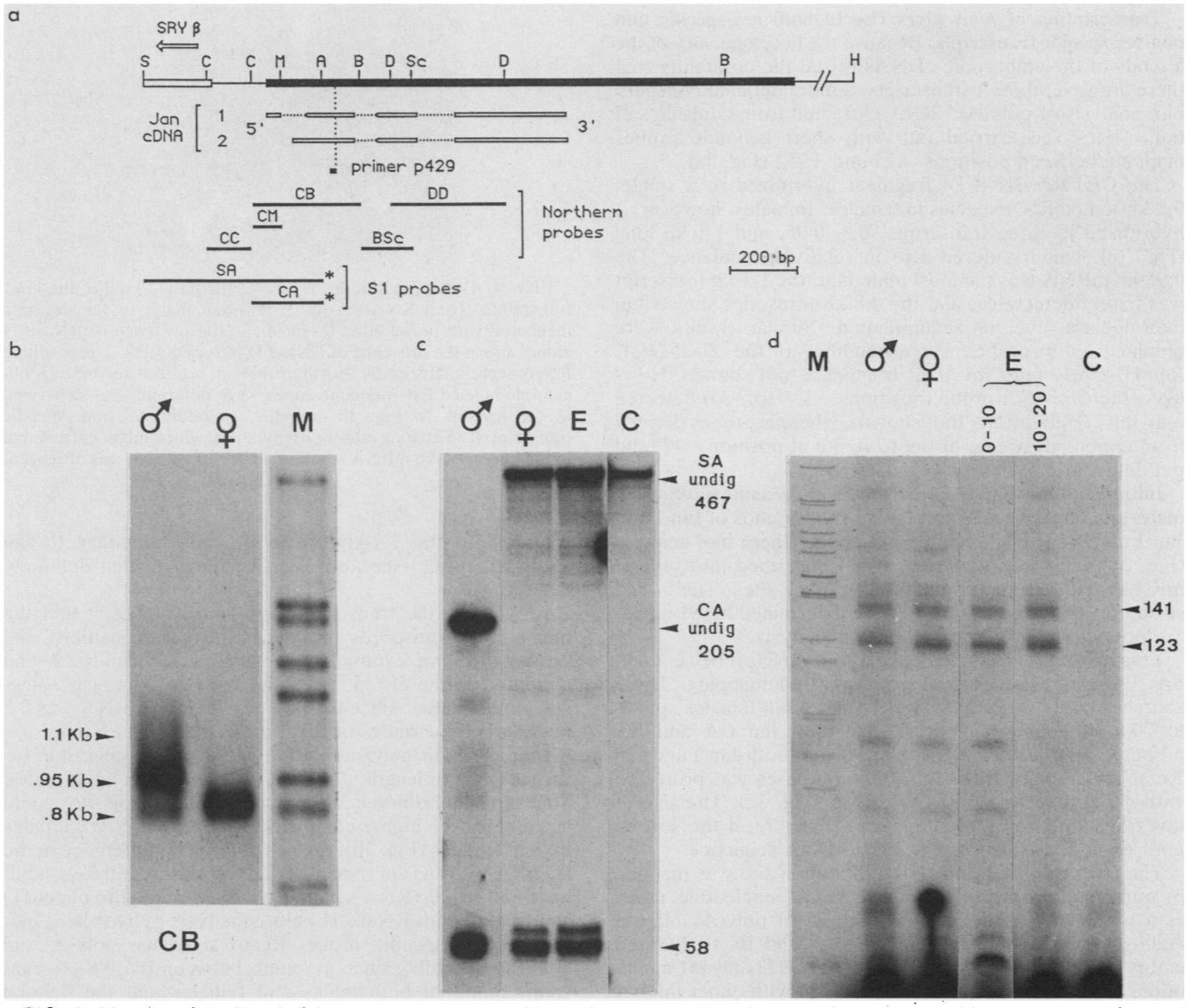


FIG. 3. Mapping of the 5′ end of the *janA* mRNAs. (a) Positions of the Jan 1 and Jan 2 cDNA clones (▭) and introns (·····), of several genomic subclones or DNA fragments used as probes for Northern blot experiments or S1 nuclease protection assay (▬) and of the primer p429 with respect to the genomic DNA sequence shown at the top. *, ^{32}P-labeled 5′ ends of the *Sst*I-*Apa*LI (SA) and *Cla*I-*Apa*LI (CA) probes. (b) Northern blot analysis of *janA* transcripts, using the oligonucleotide-labeled *Cla*I-*Bam*HI (CB) fragment as a probe. A 10-μg sample of poly(A)$^+$ RNA from adult males and females was loaded per slot; the molecular weight marker (M) was as in Fig. 1b. (c) S1 nuclease protection pattern obtained with the *Sst*I-*Apa*LI (SA) or *Cla*I-*Apa*LI (CA) ^{32}P-5′-labeled fragment used as a probe. Sizes (in nucleotides) were determined by reference to *Hpa*II-digested pBR322 (panel d). Yeast tRNA was used as a control (C). Reproducibly observed were additional faint bands corresponding to fragments of approximately 100, 120, and 150 bases, which could correspond to minor *janA* mRNA species, as also suggested by the sequence of the Jan 2 cDNA. (d) cDNA extension products with the p429 primer and poly(A)$^+$ RNA from 0- to 20-h-old embryos (E) and adult males and females. Control RNA (C) and size markers (M) were as in panel c.

long, and its sequence confirms the 3'-end structure of the Jan 1 and Jan 2 cDNAs between positions +367 and +827.

The Jan 4 cDNA is completely different from the other cDNAs; it is composed of four exons with respective sizes of 152, 157, 125, and 152 bp; the corresponding introns are 58, 57, and 60 bp long. Moreover, Jan 4 overlaps with Jan 1 and Jan 2 at its 5' end. Indeed and surprisingly, the 5' end of Jan 4 maps within the 3' region of the embryonic cDNA clones, at position +778. It possesses a poly(A) tail starting at position +1536.

These sequencing data show that the janus locus encompasses two separate but partly overlapping transcription units. One, corresponding to the embryonic cDNAs, is referred to as *janA*; the other, corresponding to male cDNA, is referred to as *janB*.

Transcription of *janA* gives rise to both sex-specific and non-sex-specific transcripts. Because the heterogeneity of the 5' ends of the embryonic cDNAs raised the possibility that there are several *janA* transcripts, a more detailed Northern blot analysis of poly(A)$^+$ RNA extracted from adult flies of both sexes was carried out with short genomic probes mapping between positions –175 and +752 (Fig. 3a).

The *Cla*I-*Bam*HI (CB) fragment hybridized to a single, 0.8-kb-long mRNA species in females. In males, however, it hybridized to three transcripts, 0.8, 0.95, and 1.1 kb long (Fig. 3b), which differed also in relative abundance. The 0.95-kb mRNA was the most abundant, the 1.1-kb transcript was barely detectable, and the 0.8-kb transcript showed an intermediate level of accumulation. Similar results were obtained with probes corresponding to the *Dra*I-*Dra*I, *Bam*HI-*Dra*I, and *Cla*I-*Msp*I fragments (not shown). However, the *Cla*I-*Cla*I probe (positions −175 to −33) detected only the 1.1-kb mRNA (not shown), whereas probes derived from regions upstream of the *Cla*I site at position −175 did not detect any of the *janA* transcripts.

Initiation of the major *janA* mRNAs at the same positions in males and females. The structures of the 5' ends of Jan 1 and Jan 2 cDNAs (Fig. 2a and b) and the Northern blot analysis (Fig. 3b) suggested that *janA* was transcribed into several mRNAs with alternate initiation or splice sites. The 5'-end structures of *janA* transcripts were determined by S1 nuclease protection and primer extension analysis.

Fragments 5' end labeled at position +168 (probes CA and SA; Fig. 3a) were used as probes for S1 mapping. Three sources of poly(A)$^+$ RNA, adult males, adult females, and 0- to 20-h-old embryos, were used. With the CA and SA probes, 5' end labeled at a site internal to both Jan 1 and Jan 2 cDNAs, a major fragment of 58 ± 2 bases was protected with either source of poly(A)$^+$ RNA (Fig. 3c). The size of this fragment confirmed the 5'-end structure of the second *janA* exon deduced from the Jan 1 cDNA sequence.

The *janA* mRNA transcription initiation site was mapped by primer extension, using a synthetic oligonucleotide, p429, as a primer (Fig. 3a) and four sources of poly(A)$^+$ RNA: adult male, adult female, and 0- to 10- and 10- to 20-h-old embryos. Two major extension products, 123 and 141 nucleotides long, were reproducibly observed with either mRNA template (Fig. 3d). The 141-nucleotide product indicated a 5' end at an adenine residue at position +1, which was the expected site with Jan 1 cDNA being a full-length copy of the *janA* messenger. At 34 bp upstream from this A residue is found the sequence TACATAA, which is reminiscent of a TATA box element. However, the presence of the 123-nucleotide-long fragment together with the sequence of the Jan 2 cDNA suggested the existence of alternative structures of *janA* transcripts. These structures, including that corresponding to the 1.1-kb transcript complementary to the *Cla*I-*Cla*I probe (see above), have not yet been definitely determined.

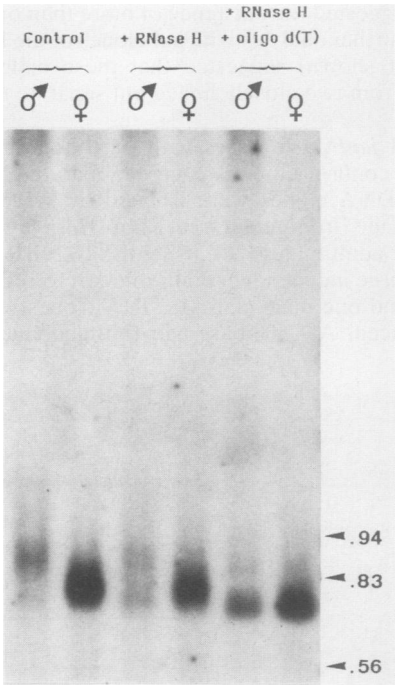


FIG. 4. Determination of the size of the poly(A) tail of the *janA* transcripts. Total RNA (25 μg) from adult males or females was incubated with no addition (control), in the presence of RNase H alone, and in the presence of RNase H plus oligo(dT). The resulting RNAs were analyzed by Northern blot hybridized to the oligonucleotide-labeled CB fragment. Sizes were determined as described in the legends to Fig. 1a and 3b. An additional male-specific transcript, 0.85 kb long after deadenylation, which might correspond to the minor 1.1-kb mRNA revealed by the CC probe, was observed.

Altogether, the transcript-mapping data suggest that the major *janA* transcripts initiate at the same positions and display the same exon-intron structures (see also the 3'-end structure in Fig. 5b) in males and females. This was somewhat unexpected, since the non-sex-specific mRNA was 0.8 kb long and the male-specific mRNA was 0.95 kb long.

Differences in polyadenylation are mainly responsible for the variation in length of the *janA* transcripts in adult males. Since no major differences in initiation or splicing sites could account for the higher complexity of *janA* mRNAs in males than in females (Fig. 3b), the possibility of a difference in the length of poly(A) tail remained. To test this hypothesis, adult male and female RNA samples were hybridized to oligo(dT) and treated with RNase H before analysis by Northern blot with a *janA*-specific probe. Removal of the poly(A) tail abolished the difference in length between the 0.8-kb transcript found in both males and females and the 0.95-kb male-specific transcript (Fig. 4), which indicated that these mRNAs differ only in the discrete sizes of their poly(A) tails.

Initiation of *janB* mRNAs within the 3' transcribed region of *janA*. The *janB* transcription initiation site was mapped by S1 nuclease protection analysis of poly(A)$^+$ RNA extracted from either male or female adults, using uniformly labeled single-stranded DNA fragments (see genomic position in Fig. 5a) as probes.

With the *Sca*I-*Sma*I probe, four mRNA protected frag-

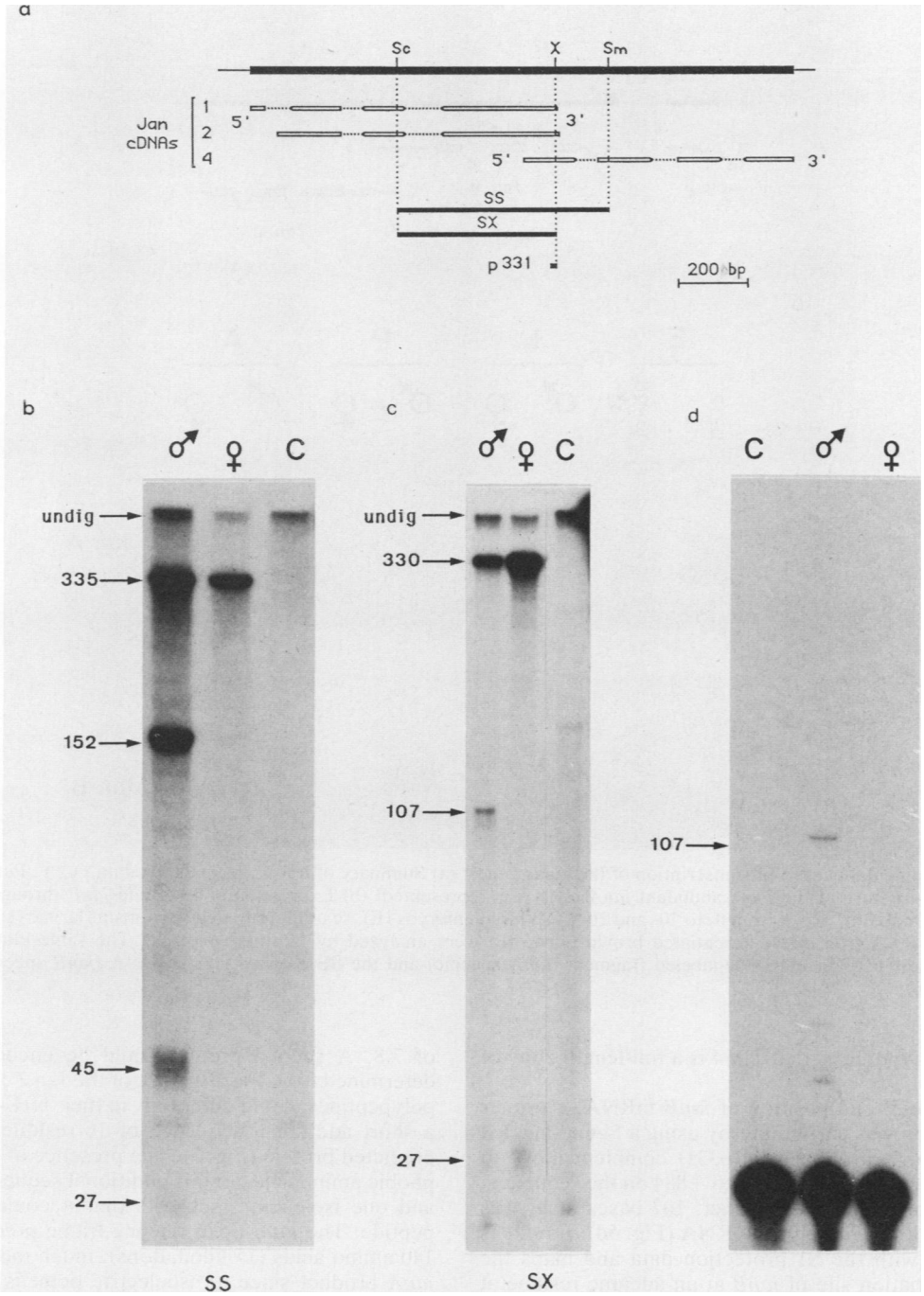


FIG. 5. Mapping of the 5′ end of the *janB* mRNA. (a) Strategy for mapping the 5′ termini of the *janB* transcripts. ▭, Locations of cloned janus cDNAs with respect to their genomic positions (top). Positions of the uniformly labeled probes *Sca*I-*Sma*I (SS) and *Sca*I-*Xmn*I (SX) (see Materials and Methods) and of primer p331 are shown below. (b and c) S1 nuclease-protected fragments detected after electrophoresis in denaturating 6% polyacrylamide gels, using the SS (b) or SX (c) fragment as a probe. (d) cDNA extension products primed by the ^{32}P-5′-labeled p331 oligonucleotide. Sizes (in nucleotides) were determined as in Fig. 3c and d (S1 protection assay) or according to sequencing ladders generated by using the same primer and an appropriate DNA template (primer extension assay). Poly(A)$^+$ RNA was from adult males and females. Yeast tRNA was used as a control (C).

ments with respective sizes of 335, 152, 45, and 27 bases were detected (Fig. 5b); the 152- and the 45-base fragments were detected only with RNA from adult males. With the *Sca*I-*Xmn*I probe, three fragments of 330, 107, and 27 bases, respectively, were observed (Fig. 5c). The 107-base fragment was detected only with adult male RNA. A protection pattern identical to that obtained with adult female RNA was observed with embryonic RNA (data not shown). These S1 mapping data confirmed the exon-intron structure of the *janA* 3′ end and the *janB* 5′ end as deduced from genomic and cDNA sequencing (see Fig. 3a). The sizes of the male-specific fragments (152 and 107 nucleotides) are con-

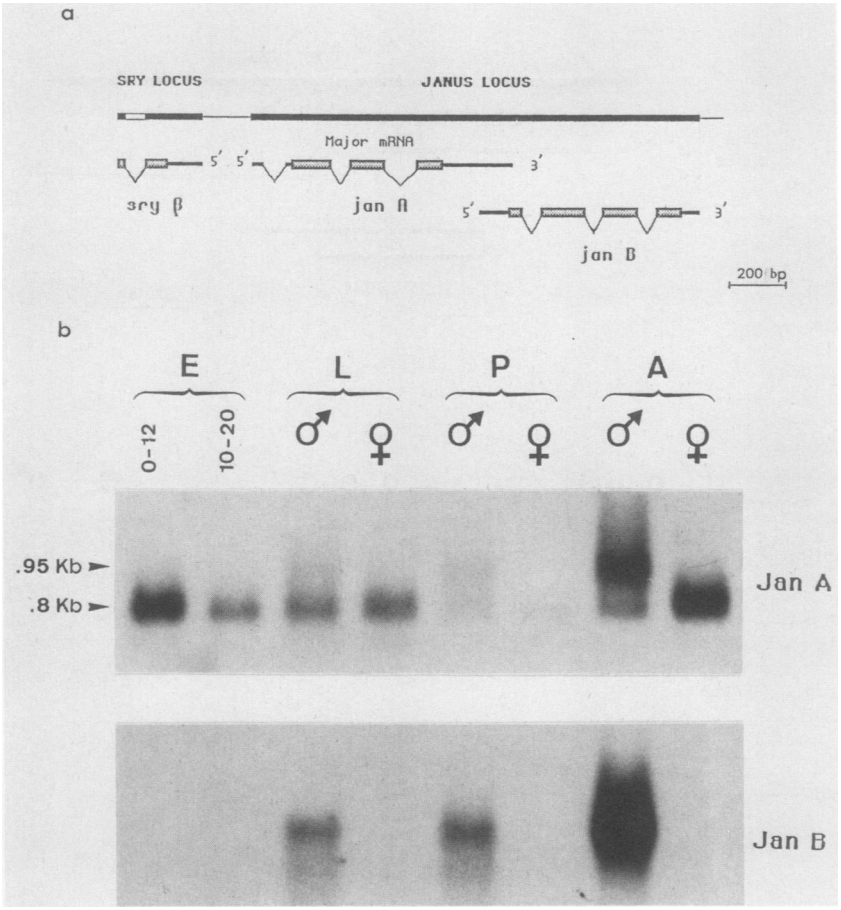


FIG. 6. Developmental pattern of transcription of the janus genes. (a) Summary of transcript-mapping data. ▭, Putative open reading frames. Only the structure of the most abundant *janA* mRNA is represented. (b) Expression of *janA* and *janB* throughout development. Samples (10 μg) of poly(A)$^+$ RNA from 0- to 10- and 10- to 20-h-old embryos (E), sex-differentiated third-instar larvae (L), and adults (A) or 2 μg of poly(A)$^+$ RNA from sex-differentiated brown pupae (P) were analyzed by Northern blotting. The same blot was successively hybridized with the CB oligonucleotide-labeled fragment (*janA* specific) and the SB Klenow-labeled DNA (*janB* specific) as probes (see positions in Fig. 2b).

sistent with the hypothesis that Jan 4 is a full-length copy of *janB* mRNA.

To confirm the 5′-end position of *janB* mRNA, a primer extension reaction was carried out by using a 5′-end-labeled synthetic oligonucleotide primer (p331) complementary to the *janB* transcript (positions +863 to +884 on the sequence; Fig. 5a). A major elongation product, 107 bases long, was observed only with male poly(A)$^+$ RNA (Fig. 5d). Its size is fully consistent with the S1 protection data and maps the transcription initiation site of *janB* at an adenine residue at position +778 of the sequence (Fig. 2b). Moreover, 28 bp upstream of this A residue is found the sequence TTTA AAA, which is reminiscent of a TATA box element.

Both the Jan 4 cDNA sequence and the transcript mapping data reported above support the conclusion that *janB* mRNA initiates within the *janA* 3′ transcribed but untranslated region. The *janA-janB* overlapping region would thus extend over 118 bp. The structures of these two transcription units are summarized in Fig. 6a.

***janA* and *janB* encode sequence-related polypeptides.** The sequences of the janus genomic region and cDNAs showed that *janA* and *janB* each contain a single long open reading frame. The *janA* open reading frame predicts a polypeptide of 115 amino acids (12 kilodaltons) with an isoelectric point of 7.8. A second protein could be encoded by *janA*, as determined from the structure of the Jan 2 cDNA. These two polypeptides would differ only in their NH_2-terminal ends by a short additional sequence of 16 residues for the second predicted protein (Fig. 7). The presence of six highly hydrophobic amino acids in this additional sequence, five leucines and one isoleucine, suggests that it could act as a signal peptide. The *janB* open reading frame predicts a protein of 140 amino acids (15 kilodaltons), much more basic than the *janA* product since its isoelectric point is 9.8. Searches in NBRF and GENEBANK data bases did not reveal any significant sequence similarity to any known protein.

On the other hand, computer comparison of the predicted *janA* and *janB* polypeptides revealed a large sequence similarity between these two proteins, with 37% of the amino acid residues found to be identical at homologous positions. The significance of this sequence similarity is reinforced by the fact that it is not interrupted by the last two 3′ introns, which are located at identical positions (Fig. 7) with respect to the *janA* and *janB* protein-coding sequences. This result suggests that these two genes resulted from a tandem duplication of an ancestral transcription unit already containing at least the two end-most 3′ introns.

Sex-specific expression of janus transcripts throughout de-

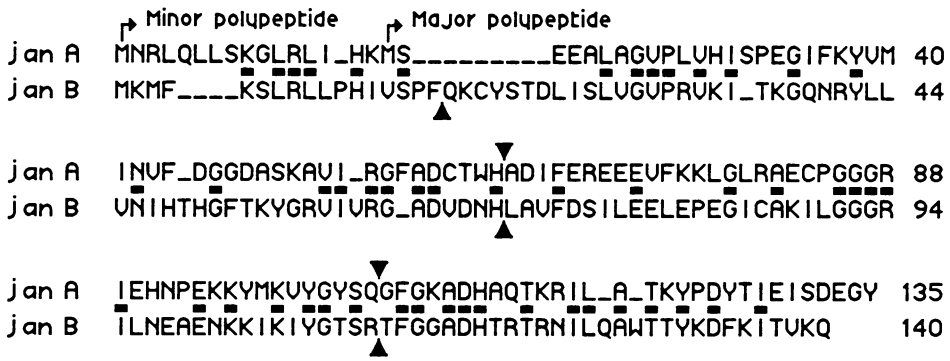


FIG. 7. Sequence alignment of the predicted *janA* and *janB* polypeptides. Positions of identical amino acids (■) and introns (▲, ▼) are indicated. Amino acid numbers are indicated on the right.

velopment. Expression of the two janus transcription units throughout development and in each sex was studied by Northern analysis of RNA samples from 0- to 12- and 10- to 20-h-old embryos, sex-differentiated larvae and pupae, and adult males and females by using successively probes CB and SB (Fig. 2a and 3a), each of which is specific for one of the two genes. This analysis confirmed both the abundance of the 0.8-kb *janA* mRNA in females and early embryos and its presence at a lower level at all other stages studied in both sexes. In contrast, the 0.95-kb *janA* messenger was detected only in males from the third-instar larval stage (Fig. 6b). Similarly, the *janB* transcript was present only in males, from third-instar larval to adult stage, with a peak of accumulation in adults. In first- and second-instar larval stages, none of the male-specific janus transcripts were detected (not shown).

The male-specific janus mRNAs are germ line specific in adults. To determine whether the sex-specific expression of *janA* and *janB* transcripts is restricted to either germinal or somatic tissues, the presence of the transcripts was examined in germ line-less adult progeny of homozygous tudor females. Neither the 0.95-kb *janA* mRNAs nor the *janB* transcript was detected in germ line-less males (Fig. 8). We can therefore conclude that expression of these two male-specific transcripts is also strictly germ line specific. Conversely, the non-sex-specific 0.8-kb *janA* transcript was detected in germ line-less adults of either sex and must therefore be expressed in somatic tissues. The level of accumulation of this transcript was nevertheless considerably lower in germ line-less than in wild-type females, whereas in adult males the level appeared to be less affected by the lack of germinal tissues. These results were also independently confirmed by in situ hybridization to sections of adult abdomen of both sexes (not shown).

DISCUSSION

The 99D region of *D. melanogaster* is a densely transcribed region that has been shown to contain the *sry* α, *sry* β, and *sry* δ (38) and rp49 (37) genes within a span of 8 kb of genomic DNA. Expression of each of these genes appears autonomous with respect to transcript accumulation at different periods of development (38, 39). Besides these four well-characterized genes, another transcription unit, previously designated Y, has been shown to map in the same region, upstream of and very close the the *sry* β gene (38, 40).

Organization of the janus locus. We show here that the locus designated janus displays an unusual organization. It encompasses two partly overlapping genes, *janA* and *janB*, transcribed in the same direction. The 5′ end of the *janB* mRNA maps within the 3′ untranslated region of *janA*, and the overlap region between the *janA* and *janB* transcribed regions is 118 nucleotides long. Few examples of overlapping transcription units encoding separate protein products have been reported in *D. melanogaster*. In the *Ddc* (35) and *tra* (4) genes, the overlapping region encompasses short segments of the 3′ region of each of the two genes. In contrast, *janB* and the Actin 5C overlapping gene (29) remain the only examples in which transcription initiates within the transcribed region of an upstream gene.

Similarities between the *janA* and *janB* genes include a significant identity of the predicted peptidic products and of the positions of the introns; this finding strongly suggests that these two genes originated from the tandem duplication of an ancestral transcription unit. A similar duplication event is probably at the origin of the *sry* β and *sry* δ genes located

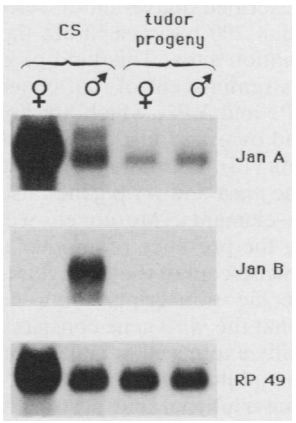


FIG. 8. Expression of janus in germ line-less flies. Northern blots of poly(A)$^+$ RNA from adult males and females of the wild-type Canton S stock (CS) and adult male and female progeny of a cross between *tud*^{1} *bw sp*/*tud*^{1} *bw sp* females and wild-type CS males (tudor progeny) were probed sequentially as described in the legend to Fig. 6b. To confirm that all lanes contained similar amounts of RNA, the same blot was further hybridized to an oligonucleotide-labeled probe specific for the rp49 gene. Since ribosomal proteins are accumulated during oogenesis, the rp49 transcript is more abundant in adult females than in adult males. In males, the intensity of the 0.95-kb mRNA on Northern blots was usually greater than that of the 0.8-kb transcript, although the 0.95-kb transcript occasionally gave a weaker signal, as on this RNA blot, possibly as an effect of aging.

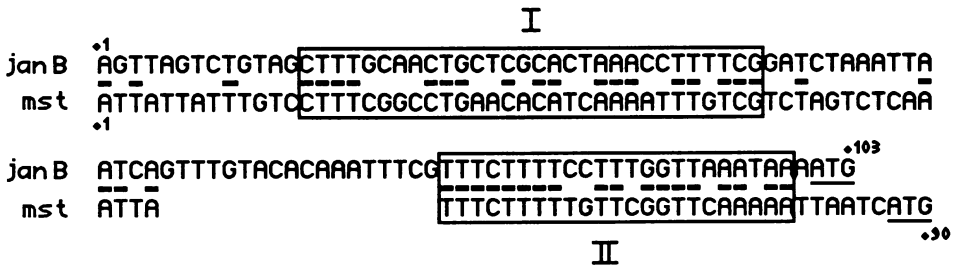


FIG. 9. Sequence alignment of the 5′ untranslated region of the *janB* (positions +778 to +886 in Fig. 2a) and *mst(3)gl-9* (20) mRNAs. +1 indicates the start of transcription. The AUG initiating codons are underlined. Positions of identical nucleotides are indicated (■). Boxes I and II delimit regions of higher sequence similarity which display 60 and 78%, respectively, identical nucleotides at homologous positions.

at the adjacent *sry* locus (40). Unlike what is observed for the *sry* genes, however, *janA* and *janB* have diverged not only in their protein-coding sequences but also through the acquisition of a specific pattern of developmental expression. This raises the interesting questions of the structure and specificity of expression of the putative common ancestor and the identity of the regulatory elements, if any, that are still shared by *janA* and *janB*. Whereas the *janA* gene is expressed in both sexes and in somatic and germ line cells, expression of *janB* is restricted to germ line tissues of adult males. We recently showed, by P-element-mediated transformation, that no more than 175 bases upstream from the *janB* transcription initiation site are required to drive the testis-specific expression of this gene (C. Yanicostas, manuscript in preparation). This result indicates that the 5′-flanking sequence and most of the coding sequence of the *janA* gene are dispensable for correct regulation of the *janB* gene.

In adult females, the *janA* gene is predominantly transcribed in ovaries, as is the *sry* β gene located immediately adjacent on the chromosome and transcribed in the opposite orientation (as described above and in reference 38). Moreover, no more than 200 bases separate the *janA* and *sry* β transcription initiation sites. This kind of gene organization and expression is reminiscent of that observed for the yolk protein genes YP1 and YP2, which are both transcribed in adult fat body and ovarian follicle cells and share transcriptional *cis* elements (18). By analogy, it raises the interesting possibility that the *janA* and *sry* β genes also share transcriptional regulatory elements. Moreover, we have some evidence supporting the presence of another transcription unit less than 1 kb downstream of the *janB* gene, which gives rise to two male-specific transcripts (data not shown). This finding suggests that the *janB* gene consists of a larger cluster of genes specifically expressed in males from the third-instar larval stage to the adult stage.

Sex-specific transcriptional and posttranscriptional regulation of the janus genes. As an approach to investigate the mechanisms of sex-specific gene expression, genes encoding either male- or female-specific products have been identified by screening of genomic libraries, using sex-specific cDNAs as probes (16, 25, 31). One of the best studied of these genes is the testis-specific gene *mst(3)gl-9* (32). Kuhn et al. have shown that a *cis*-acting element mediating spermatid-specific translational control of *mst(3)gl-9* is located within the first 120 bases of the 5′ region of the mRNA, including the entire untranslated leader plus 10 nucleotides following the AUG codon (20). A striking sequence similarity is observed between the *janB* untranslated leader and this 120-base-long region (Fig. 9). That the region of homology between *janB* and *mst(3)gl-9* might be a consensus sequence involved in testis-specific translational regulation is an exciting possibility.

Precise sizing of *janA* transcripts in adult males and females showed that the poly(A) tail of the male-specific transcript is, on average, 180 residues longer than that of the non-sex-specific transcript. The involvement of the poly(A) tail length in posttranscriptional regulation of mRNAs is well documented. Variation in poly(A) tail length may affect both mRNA translational efficiency (27) and stability (7, 42). That a long poly(A) tail serves to increase mRNA stability in the *Drosophila* testis is a reasonable hypothesis, since mRNAs synthesized in premeiotic spermatocytes must be stored for at least 3 days before translation during the spermatocyte maturation process (21).

Together, our data show that both transcriptional and posttranscriptional elements contribute to the differences in expression of the janus genes in males and females. Because it contains two duplicated transcription units with overlapping sequences, the janus locus represents a unique model system for the study of sex-specific regulation of gene expression.

ACKNOWLEDGMENTS

We are grateful to Paul Berreur for helpful suggestions and discussions during the course of this work and to Jana Kejzlarovà-Lepesant for help with cDNA libraries. We thank Larry Kauvar of the gift of cDNA libraries and Roger Ollo for the gift of the tudor mutant strain.

This work was supported by grants from the Centre National de la Recherche Scientifique, the Ligue Nationale Française Contre le Cancer, and the Fondation pour la Recherche Médicale.

LITERATURE CITED

1. **Baker, B. S., R. N. Nagoshi, and K. C. Burtis.** 1987. Molecular genetic aspects of sex determination in Drosophila. Bioessays **6**:66–70.
2. **Belote, J. M., A. M. Handler, M. F. Wolfner, K. J. Livak, and B. S. Baker.** 1985. Sex-specific regulation of yolk protein gene expression in Drosophila. Cell **40**:339–348.
3. **Berk, A. J., and P. A. Sharp.** 1978. Spliced early mRNAs of simian virus 40. Proc. Natl. Acad. Sci. USA **75**:1274–1278.
4. **Boggs, R. T., P. Gregor, S. Idriss, J. M. Belote, and M. McKeown.** 1987. Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the *transformer* gene. Cell **50**:739–747.
5. **Boswell, R. E., and A. P. Mahowald.** 1985. *tudor*, a gene required for assembly of the germ plasm in *Drosophila melanogaster*. Cell **43**:97–104.

6. **Bownes, M., and R. Nöthiger.** 1981. Sex determining genes and vitellogenin synthesis in *Drosophila melanogaster*. Mol. Gen. Genet. **182:**222–228.
7. **Brewer, G., and J. Ross.** 1988. Poly(A) shortening and degradation of the A+U-rich sequences of human c-*myc* mRNA in a cell-free system. Mol. Cell. Biol. **8:**1697–1708.
8. **Bridges, C. B.** 1925. Sex in relation to chromosomes and genes. Am. Nat. **59:**127–137.
9. **Calzone, F. J., R. J. Britten, and E. H. Davidson.** 1987. Mapping of gene transcripts by nuclease protection assays and cDNA primer extension, p. 611–632. *In* S. L. Berger and A. R. Kimmel (ed.), Guide to molecular cloning techniques. Academic Press, Inc. (London), Ltd., London.
10. **Chapman, K. B., and M. F. Wolfner.** 1988. Determination of male-specific gene expression in *Drosophila* accessory glands. Dev. Biol. **126:**195–202.
11. **Chen, P. S., E. Stumm-Zollinger, T. Aigaki, J. Balmer, M. Bienz, and P. Böhlen.** 1988. A male accessory gland peptide that regulates reproductive behavior of female *D. melanogaster*. Cell **54:**291–298.
12. **Cline, T. W.** 1986. A female-specific lethal lesion in an X-linked positive regulator of the Drosophila sex determination gene, *Sex-lethal*. Genetics **113:**641–663.
13. **Cline, T. W.** 1988. Evidence that *sisterless-a* and *sisterless-b* are two of several discrete "numerator elements" of the X/A sex determination signal in Drosophila that switch *Sxl* between two alternative stable expression states. Genetics **119:**829–862.
14. **Cronmiller, C., and T. W. Cline.** 1987. The Drosophila sex determination gene *daughterless* has different functions in the germ line versus the soma. Cell **48:**479–487.
15. **Dente, L., G. Cesarini, and R. Cortese.** 1983. pEMBL: a new family of single stranded plasmids. Nucleic Acids Res. **11:** 1645–1655.
16. **DiBenedetto, A. J., D. M. Lakich, W. D. Kruger, J. M. Belote, B. S. Baker, and M. F. Wolfner.** 1987. Sequences expressed sex-specifically in *Drosophila melanogaster* adults. Dev. Biol. **119:**242–251.
17. **Feinberg, P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137:**6–13.
18. **Garabedian, M. J., B. M. Shepherd, and P. C. Wensink.** 1986. A tissue-specific transcription enhancer from the Drosophila yolk protein 1 gene. Cell **45:**859–867.
19. **Hu, N. T., and J. Messing.** 1982. The making of strand-specific M13 probes. Gene **17:**271–277.
20. **Kuhn, R., U. Schäfer, and M. Schäfer.** 1988. Cis-acting regions sufficient for spermatocyte-specific transcriptional and spermatid-specific translational control of the *Drosophila melanogaster* gene *mst(3)gl-9*. EMBO J. **7:**447–454.
21. **Lindsley, D. L., and K. T. Tokuyasu.** 1980. Spermatogenesis, p. 225–294. *In* M. Ashburner and T. R. F. Wright (ed.), The genetics and biology of *Drosophila*, vol. 2. Academic Press, Inc. (London), Ltd., London.
22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 202–203. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Marcu, K., O. Valbuena, and R. P. Perry.** 1978. Isolation, purification and properties of mouse heavy-chain immunoglobulin mRNAs. Biochemistry **17:**1723–1733.
24. **Marsh, J. L., and E. Wieschaus.** 1978. Is sex determination in germ line and soma controlled by separate genetic mechanisms? Nature (London) **272:**249–251.
25. **Monsma, S. A., and M. F. Wolfner.** 1988. Structure and expression of a Drosophila male accessory gland gene whose product resembles a peptide pheromone precursor. Genes Dev. **2:** 1063–1073.
26. **Nöthiger, R., and M. Steinmann-Zwicky.** 1985. Sex determination in *Drosophila*. Trends Genet. **1:**209–215.
27. **Palatnik, C. M., C. Wilkins, and A. Jacobson.** 1984. Translational control during early *Dictyostelium* development: possible involvement of poly(A) sequences. Cell **36:**1017–1025.
28. **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.
29. **Rao, J. P., R. S. Zafar, and A. Sojda.** 1988. Transcriptional activity at the 3′ end of the actin gene at 5C on the X chromosome of *Drosophila melanogaster*. Biochim. Biophys. Acta **950:**30–44.
30. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
31. **Schäfer, U.** 1986. Genes for male-specific transcripts in *Drosophila melanogaster*. Mol. Gen. Genet. **202:**219–225.
32. **Schäfer, U.** 1986. The regulation of male-specific transcripts by sex determining genes in *Drosophila melanogaster*. EMBO J. **5:**3579–3582.
33. **Schüpbach, T.** 1982. Autosomal mutations that interfere with sex determination in somatic cells of *Drosophila* have no direct effect on the germline. Dev. Biol. **89:**117–127.
34. **Schüpbach, T.** 1985. Normal female germ cell differentiation requires the female X chromosome to autosome ratio and expression of Sex-lethal. Genetics **109:**529–548.
35. **Spencer, C. A., R. D. Gietz, and R. B. Hodgetts.** 1986. Overlapping transcription units in the Dopa decarboxylase region of *Drosophila*. Nature (London) **322:**279–281.
36. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77:**5201–5205.
37. **Vaslet, C. A., P. O'Connell, M. P. Izquierdo, and M. Rosbash.** 1980. Isolation and mapping of a cloned ribosomal protein gene of *Drosophila melanogaster*. Nature (London) **285:**674–676.
38. **Vincent, A., H. V. Colot, and M. Rosbash.** 1985. Sequence and structure of the *Serendipity* locus of *Drosophila melanogaster*. J. Mol. Biol. **186:**149–166.
39. **Vincent, A., H. V. Colot, and M. Rosbach.** 1986. Blastoderm specific and read-through transcription of the *sry* alpha gene transformed into the *Drosophila* genome. Dev. Biol. **118:**480–487.
40. **Vincent, A., P. O'Connell, M. R. Gray, and M. Rosbash.** 1984. *Drosophila* maternal and embryo mRNAs transcribed from a single transcription unit use alternate combinations of exons. EMBO J. **3:**1003–1013.
41. **Wahl, G. M., M. Stern, and G. R. Stark.** 1979. Efficient transfer of large DNA fragments from agarose gels to diazo-benzyloxymethyl paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. USA **76:**3683–3687.
42. **Zeevi, M., J. R. Nevins, and J. E. Darnell, Jr.** 1982. Newly formed mRNA lacking polyadenylic acid enters the cytoplasm and the polyribosomes but has a shorter half-life in the absence of polyadenylic acid. Mol. Cell. Biol. **2:**517–525.