MOLECULAR AND CELLULAR BIOLOGY, Oct. 1988, p. 4459–4468
0270-7306/88/104459-10$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 8, No. 10

sry h-1, a New *Drosophila melanogaster* Multifingered Protein Gene Showing Maternal and Zygotic Expression

ALAIN VINCENT,[1] JANA KEJZLAROVÀ-LEPESANT,[2] LAURENT SEGALAT,[2] CONSTANTIN YANICOSTAS,[2] AND JEAN-ANTOINE LEPESANT[2]*

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

Received 25 April 1988/Accepted 18 July 1988

Low-stringency hybridization of the *Drosophila* serendipity (*sry*) finger-coding sequences revealed copies of homologous DNA sequences in the genomes of members of the family *Drosophilidae* and higher vertebrates. *sry h-1*, a new *Drosophila* finger protein-coding gene isolated on the basis of this homology, encodes a 3.2-kilobase (kb) mRNA accumulating in eggs and abundant in early embryos. The predicted *sry h-1* protein product, starting at an internal initiation site of translation, is a 868-amino-acid basic polypeptide containing eight TFIIIA-like fingers encoded by three separate exons. Links separating individual fingers in the *sry h-1* protein are variable in length and sequence, in contrast with the invariant H/C link found in most multi-fingered proteins. The similarity of the developmental pattern of transcription of *sry h-1* with that of several other *Drosophila* finger protein genes suggests the existence of a complex set of such genes encoding an information which is, at least partly, maternally provided to the embryo and required for activation of gene transcription in early embryos or maintenance of gene activity during subsequent development.

Two broad classes of genes required for proper *Drosophila* embryonic development may be distinguished on the basis of their mode of inheritance: genes whose products are accumulated during oogenesis (maternal genes) and subsequently used by the embryo and genes whose expression follows fertilization (zygotic genes). A number of these genes were identified as a result of extensive screens for maternal effect mutants (8, 27) and embryonic lethal loci yielding recognizable larval cuticle phenotypes (14, 28, 46). In these screens, however, a number of mutations may have been missed. This is the case for mutations causing developmental defects which are undetectable before the cuticle is secreted or mutations affecting genes which are expressed during both oogenesis and embryogenesis and which supply the embryo with amounts of maternal gene products sufficient to support normal morphological development. Mutations affecting multigene families where the lack of one gene product can be functionally compensated for by the product of another gene of the family would be also difficult to isolate.

An alternative approach for isolation of genes that play a critical role in early development follows the assumption that sequence similarity reflects functional homology. For example, the homeobox, a highly conserved protein motif first identified in *Drosophila* homeotic gene products and thought to mediate the control of expression of other genes because of its sequence-specific DNA-binding property, was successfully used for the isolation of additional homeobox-containing genes in *Drosophila* spp. and other eucaryotes (10).

Another DNA-binding structure, the zinc finger, first described by Miller et al. (26) in the course of their studies on the transcription factor TFIIIA of *Xenopus* spp., was later identified in other developmental gene products or transcription factors (2, 13, 15, 32, 38). The Zn^{2+} finger is typically made of a 28- to 29-amino-acid motif folded around a zinc atom coordinated by two cysteine and two histidine residues found at invariant positions (18, 26). The original model proposed by Miller and coauthors (26) postulated that direct contacts between the tip of each finger and DNA were determinant for the DNA-binding specificity. The precise correlations between finger structure, finger number, and size of the recognized DNA sequence remain as important issues to be resolved.

In *Drosophila melanogaster*, genetical and molecular studies have shown that three genes encoding Zn finger proteins, Krüppel (19), hunchback (38), and snail (2), play a critical role in the establishment of the body plan of the embryo along the anteroposterior or dorsoventral axes. In the course of a molecular characterization of a *D. melanogaster* densely transcribed region, the serendipity (*sry*) locus mapping to chromosomal position 99D and including a blastoderm-specific gene, *sryα*, we identified a few years ago the *sryβ*, and *sryδ* genes. These genes are predominantly transcribed during oogenesis, and both code for a finger protein (41, 42). They probably arose from a duplication of an ancestral gene already containing six fingers and thus represent a potentially interesting model for studying the concerted evolution of finger proteins and their regulatory properties. Recent data show that the *sryδ* gene encodes a DNA-binding protein associating with embryonic nuclei prior to the onset of zygotic transcription (F. Payre and A. Vincent, manuscript in preparation). Mutations in the *sryδ* gene lead to a late embryonic lethal phenotype with head involution defects and to a maternal effect currently under investigation. The *sryδ* defect is not compensated for by the *sryβ* gene product since none of the δ mutant alleles was rescued by transformation with a fragment that contains the *sryβ* gene (K. Kongsuwan, J. Merriam, A. Vincent, and J. Lengyel, unpublished data).

We report here the molecular structure and developmental expression of a new maternal and zygotic finger protein gene, *sry h-1*, cloned by virtue of its sequence similarity to *sryβ* and *sryδ* genes and mapping to region 98EF on the

* Corresponding author.

cytogenetic map. The pattern of expression of the *sry h-1* gene is similar to that of the *sryβ* and *sryδ* genes, suggesting that these genes belong to a complex set of multifingered proteins which play an important role in gene activation or regulation at early embryonic stages through a maximal accumulation of their transcripts (or protein product) in the mature oocyte.

MATERIALS AND METHODS

Nucleic acids. Unless otherwise noted, handling of nucleic acids followed standard protocols (23).

Southern blot hybridization. Low-stringency hybridization of Southern blots with nick-translated ^{32}P-labeled DNA fragments (10^8 cpm/μg) was carried out at 55°C in 6× SSPE (1× SSPE is 180 mM NaCl, 10 mM NaH_2PO_4 [pH 7.4], 1 mM EDTA)–2× Denhardt solution–0.1% sodium dodecyl sulfate (SDS)–100 μg of denatured calf thymus DNA per ml with 10^6 cpm of probe per ml of solution. The blots were successively washed at 60°C in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% SDS, 1.5× SSC–0.1% SDS, 1× SSC–0.1% SDS, and 0.5× SSC–0.1% SDS.

DNA cloning. The λ Dm sry h bacteriophages were isolated from the Lauer-Maniatis Canton S genomic library (24) by plaque hybridization carried out at 60°C in the conditions described for Southern blots (see above). Nitrocellulose replica filters were hybridized in parallel with the B1 and D probes. Following hybridization, the filters were successively washed at 60°C in 2× SSC–0.1% SDS, and 0.5× SSC–0.1% SDS. The h-1 E cDNA clone was isolated from a library prepared from poly(A)$^+$ RNA of embryos, kindly supplied to us by L. Kauvar (31). Subcloning of restriction fragments into pEMBL 18$^+$ and 19$^+$ phagemids and radioactive labeling of DNAs were performed as described by Vincent et al. (44).

DNA sequencing. Both genomic DNA and cDNA were sequenced by the chain termination method (34). Most of the sequence was obtained from subclones generated by digestion of large DNA fragments by using six-cutter restriction enzymes. Gaps in the sequence were filled by using synthetic oligonucleotides as primers.

Transcript mapping. Isolation of staged poly(A)$^+$ RNA, RNA analysis by Northern (RNA) blot, and S1 nuclease mapping were performed as described previously (44). S1 nuclease mapping was carried out by using 10^4 cpm of 5′-end-labeled fragment (see Fig. 6A) and 5 μg of poly(A)$^+$ RNA or 60 μg of total RNA.

In situ hybridization in tissue section. In situ hybridization to RNA in tissue sections was performed by the method of Hafen and Levine (12) with ^{35}S-labeled probes prepared by nick translation of gel-purified restriction fragments. Cryosections of *Drosophila* adult females were prepared by the method of Glaser et al. (11).

Computer analysis. The DNA sequence was compiled and analyzed by using the available computer programs of the Unité d'Informatique Scientifique at the Institut Pasteur and of the CITI2 (Paris, France). A search for maximal homologies between *Drosophila* finger protein sequences was performed by using a scan for a similarity to a protein subsequence by using the Dayhoff MDM78 matrix with a 30-amino-acid window (PC Gene Programs; Genofit S.A., Geneva, Switzerland).

RESULTS

Detection of *sry*-like sequences in diptera and vertebrate DNAs. To determine whether sequences similar to *sryβ* and *sryδ* finger-coding sequences were present in other eucaryotes, DNAs from various species were analyzed by Southern blotting. Blots of DNA from dipterae and vertebrates (Fig. 1B) were hybridized sequentially with the *sryδ* (D) and *sryβ* (B1 or B2) probes shown in Fig. 1A. When either finger probe was hybridized to *D. melanogaster* DNA (lanes D.m in Fig. 1B), several weak bands were observed in addition to those strongly labeled bands corresponding to fragment(s) containing the *sryβ* or *sryδ* genomic sequences (indicated by stars in Fig. 1B). With DNA from *Drosophila erecta*, a similar pattern of one major and several weaker bands was observed. A single 17-kilobase pair (kbp) *D. erecta Eco*RI fragment gives a strong hybridization signal with both the B1 and D probes (indicated by an arrow in lanes D.e). This suggests that two separate subregions (genes?) may be contained within this fragment, each one being homologous to one of the *sry* finger-coding regions. In contrast with the above result, much weaker hybridization signals were obtained with the *Drosophila virilis* DNA with the D probe only (lanes D.v). Furthermore, no hybridization band could be detected with *Calliphora* DNA with either probe under our experimental conditions (lanes Ca).

Hybridization of either *sry* probe to blots of DNA from vertebrates revealed multiple weak bands over the background. The case of chicken DNA (lanes C) is particularly striking. Some of the bands are at least as intense as those observed with *D. melanogaster* DNA in the same experiment, despite the fact that a lower number of copies of each gene is present in lanes C because of the difference in genome size. The pattern of major bands seen in each lane of vertebrate DNA is specific for either one of the two *sry* finger-coding probes, suggesting that two sets of sequences, one β related and the other δ related, have been conserved during evolution.

Identification and chromosomal location of *sry h-1*, a new finger protein gene. To verify that some of the hybridization signals correspond to bona fide finger-coding DNA sequences, we screened under low-stringency conditions a *D. melanogaster* genomic library of Canton S DNA (24) with the *sry* B1 and *sry* D probes in parallel. We purposely limited our analysis to 26 phages hybridized to both the B1 and D probes, although more than half of the strong hybridization signals were detected with only one or the other probe. Single- and double-site restriction analysis showed that two partially overlapping phages, λ Dm sry h 18 and 25, contained a 380-base-pair (bp) *Eco*RI-*Pst*I (RP) fragment giving a strong hybridization signal with each *sry* probe. Sequence analysis of this fragment identified an open reading frame with the capacity to code for several fingers (see Fig. 5). We report here the molecular analysis of the finger protein gene contained within λ Dm sry h 18, designated below as the *sry h-1* gene. In situ hybridization of biotinylated λ Dm sry h 18 phage DNA to salivary gland polytene chromosomes of the Canton S stock showed that the *sry h-1* gene is located in region 98EF of the right arm of the third chromosome (Fig. 2). No mutation has been isolated which precisely corresponds to this position.

The *sry h-1* mRNA is maternally inherited by the embryo. A physical map of the genomic region encompassing the *sry h-1* gene is shown in Fig. 3A. The λ Dm sry h 18 phage, which carries 17.5 kbp of genomic DNA, hybridizes to two polyadenylated RNA species (Fig. 4A, right panel, φ probe). One is a minor 4-kbp RNA accumulating at a uniformly low level throughout the entire *D. melanogaster* life cycle. This RNA has not been precisely mapped on the λ Dm sry h 18 phage DNA. The second is a 3.2-kb RNA, abundant in virgin

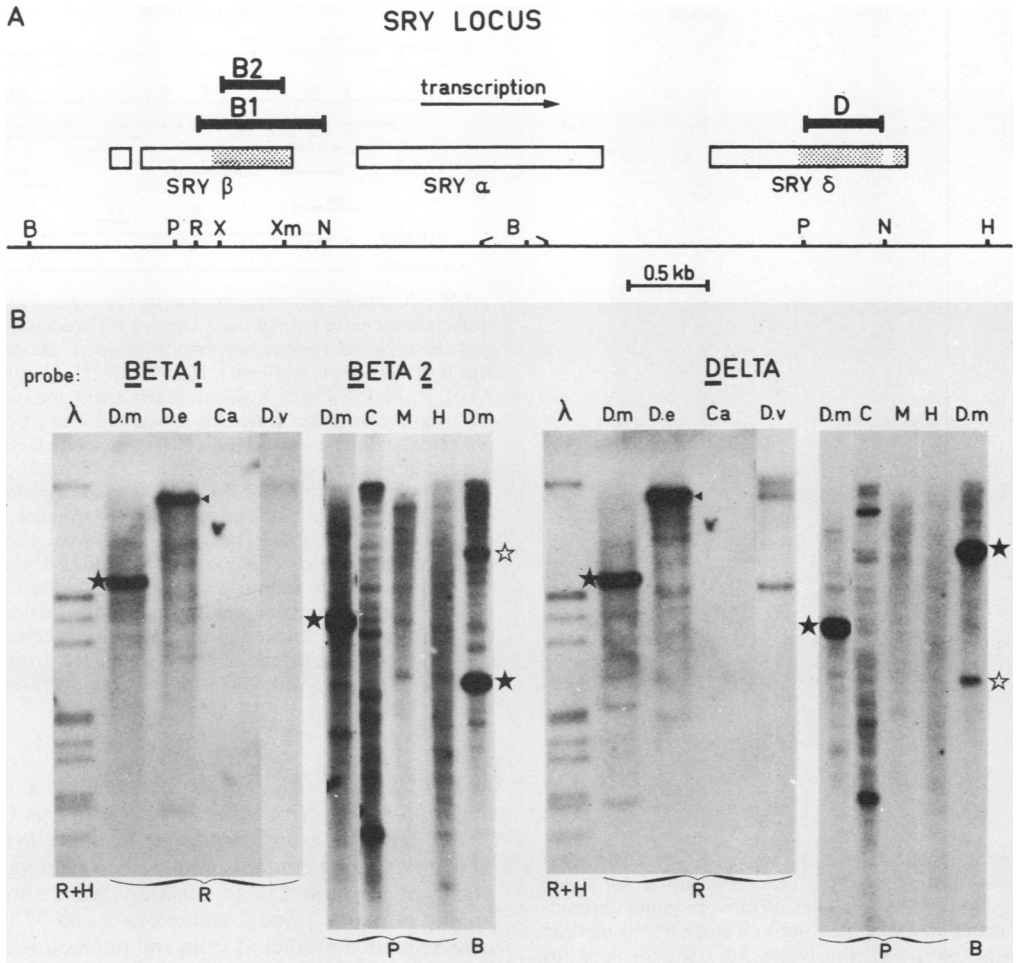

FIG. 1. Detection of DNA sequences homologous to the *sry* finger-coding regions in the genomes of various eucaryotes. (A) Restriction endonuclease map of the *sry* locus showing the *sry*β (B1 and B2) and *sry*δ (D) fragments used as probes. Boxes correspond to the *sry*β-, *sry*α-, and *sry*δ-coding regions, with the stippled areas indicating the finger-coding domains. (B) Southern blot of genomic DNAs of *D. melanogaster* (D.m), *D. erecta* (D.e), *D. virilis* (D.v), *C. vicina* (Ca), chicken (C), mouse (M), and human (H) cleaved to completion with various enzymes as indicated below each lane (B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; N, *Nru*I; P, *Pst*I; X, *Xho*I; Xm, *Xmn*I). 5′-end-labeled fragments from λ phage DNA cleaved with *Eco*RI and *Hin*dIII were used as molecular weight standards (from top to bottom: 21.7, 5.15, 4.27, 3.48, 1.98, 1.90, 1.59, 1.37, 0.94, 0.83, and 0.56 kbp). Hybridization was carried out as described in Materials and Methods with the B1 (Beta 1), B2 (Beta 2), and D (Delta) probes. Filters were washed in 0.5× SSC–0.1% SDS (B1 and D probes with diptera DNAs) or in 1.5× SSC–1% SDS (B2 and D probes with vertebrate DNAs) at 60°C. ★, Position of restriction fragments containing sequences of the *sry*β and/or *sry*δ genes hybridized to their homologous B1, B2, or D probes; ☆, position of restriction fragments containing sequences of the *sry*β or *sry*δ genes detected by cross-hybridization with either probe; ▶, 17-kbp restriction fragment hybridized to the B1 or D probes in *D. erecta* DNA.

females (lanes F) and early (E1, 0- to 2-h and E2, 2- to 4-h) embryos and detectable but to a much lower level during other developmental stages (E3, 10- to 20-h-old embryos; L, larvae; P, pupae; M, adult males). Only the 3.2-kb RNA is complementary to the *Eco*RI-*Pst*I fragment (RP, Fig. 3B), carrying the similarity to the *sry*β and *sry*δ finger-coding regions (Fig. 4A, left panel, RP probe), and to the *Xho*I-*Pst*I (XP; Fig. 3B) fragment (results not shown). From this result, we deduced that the 3.2-kb RNA is transcribed from the *sry h-1* finger-coding region. This RNA is, at least partly, maternally inherited by the embryo as shown by in situ hybridization of a cDNA probe to sections of ovarian chambers and embryos (Fig. 4B). *sry h-1* transcripts are first seen over nurse cells up to stage 10A of oogenesis. Between stages 10A and 10B, they are transferred to the growing oocytes, where they accumulate (Fig. 4B, left panel). This confirms a maternal origin of the *sry h-1* mRNA present in early embryos. After egg deposition and before the cellular blastoderm stage, there is a net decrease of *sry h-1* transcripts which appear evenly distributed within the embryo during gastrulation. Later during embryogenesis, a nonuniform distribution of the residual transcripts (or transcriptional activity) between different tissues of the embryo is noticeable but has not been thoroughly examined (results not shown).

Structure and sequence of the *sry h-1* gene. A 3.2-kbp-long cDNA (h-1E) complementary to the XP and RP fragments (Fig. 3B) was isolated by screening a λ gt10 cDNA library made from poly(A)$^+$ RNA from 0- to 12-h-old embryos (31). Comparison of the restriction maps of the *sry h-1* genomic DNA and h-1E cDNA suggested the existence of two short introns within the *sry h-1*-transcribed region (Fig. 3B). The complete nucleotide sequence of this region is shown in Fig. 5. Partial sequencing of the h-1E cDNA allowed a precise positioning of the intron-exon boundaries and 3′ polyadenylation site. Intron 1 is located between nucleotides 1263 and

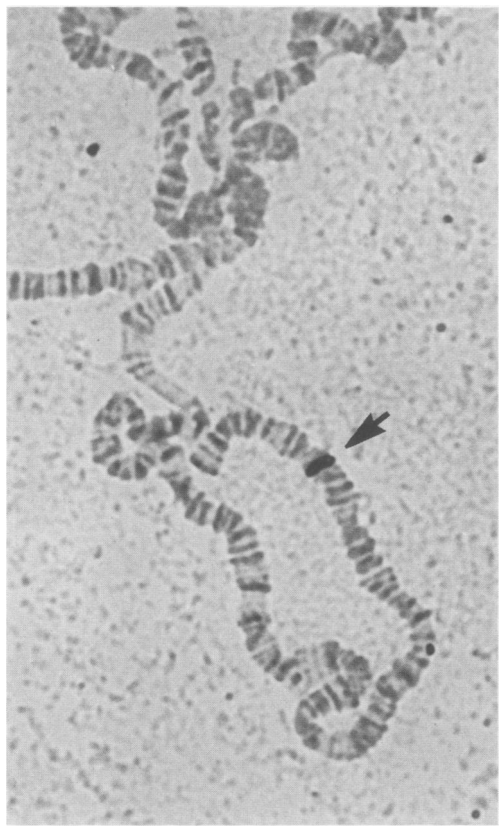

FIG. 2. Cytogenetic localization of the *sry h-1* gene. λ Dm sry h 18 phage DNA was biotinylated by nick translation by using Bio-11-dUTP and hybridized to squashes of salivary gland chromosomes of late third instar larvae of the Canton S stock by the method of Langer-Safer et al. (20), with the Detek kit I (Enzo) used for staining. A single hybridization signal was observed at position 98 EF on the right arm of the third chromosome (→).

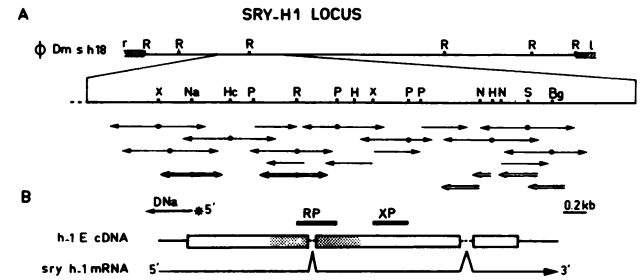

FIG. 3. Molecular organization of the *sry h-1* locus. (A) Restriction endonuclease map of the λ Dm sry h 18 recombinant phage (top) and its enlarged transcribed region (below). Restriction enzymes: Bg, *Bgl*II; D, *Dra*I; R, *Eco*RI; Hc, *Hin*cII; H, *Hin*dIII; Na, *Nar*I; N, *Nco*I; P, *Pst*I; S, *Sal*I; X, *Xho*I. Each arrow indicates the sequencing from a restriction fragment subcloned into pEMBL 18^+ or 19^+ phagemids; double arrows indicate sequenced cDNA subclones. (B) Transcription map of the *sry h-1* gene. The 5'-to-3' orientation of the transcript was established by single-stranded probe mapping. The entire λ Dm sry h 18 phage and single-stranded DNA subclones containing the XP and RP fragments were used for Northern blot analysis. The position of the introns and of the 3' terminus of the *sry h-1* transcript were deduced from partial sequencing of the h-1E cDNA. The 5' terminus was S1 mapped by using the *Dra*I-*Nar*I (DNa) fragment 5' labeled at the *Nar*I site. The boxed area corresponds to the open reading frame deduced from the nucleotide sequence (Fig. 5), with the stippled area representing the finger-coding domain.

1325, and intron 2 is located between nucleotides 2528 and 2639. Intron sequences at the 5' and 3' splice junctions closely match the consensus (36). 3' splice signals C/TT A/G AC/T (17) are observed within each of the two introns at positions −20 and −21, respectively. The 3' polyadenylation and pre-mRNA cleavage site at position 3353 is preceded by a canonical AAUAAA polyadenylation site located 24 bp upstream. The 5' transcription initiation site of the *sry h-1* gene was mapped on the DNA sequence by S1 nuclease analysis performed with poly(A)$^+$ RNA from embryos and virgin female ovaries (Fig. 6) with the *Nar*I-*Dra*I fragment (DNa probe, Fig. 6A) labeled at its 5' end. The S1 protected fragments were separated by electrophoresis in either denaturing acrylamide urea gels (Fig. 6B and C) or acrylamide neutral gels (Fig. 6D). The results indicate two major 5' ends at the level of adenine nucleotides (referred to as positions + 1 and + 21), 297 and 276 nucleotides upstream of the *Nar*I site, respectively. Position +1 corresponds to the most 5' nucleotide on the h-1E cDNA sequence, indicating that this cDNA is full length. A third minor band comigrates in high resolution denaturing polyacrymalmide gels (lane 2, Fig. 6C) with the 286-bp 5'-end-labeled *Nar*I-*Xho*I fragment (Fig. 6A) used as a size standard (lane S', Fig. 6C). This band corresponds to a 5' end at position +12. No TATA box-like sequence is observed at a consensus position. The total length of the exons mapped in the above experiments comes to 3.18 kbp, consistent with the 3.2 kb size determined for the *sry h-1* transcript by gel electrophoresis (Fig. 4A).

Two separate exons encode the eight *sry h-1* Zn^{2+} fingers. The *sry h-1* gene contains a long open reading frame extending from position 251 to position 3024, showing in-frame joining of exons 1 and 2 and exons 2 and 3. The 868-amino-acid sequence predicted from the nucleotide sequence (Fig. 5) indicates a large (calculated molecular weight, 96,000) fairly basic (calculated pI, 8.6) protein with two proline-rich domains (residues 141 to 213 and 593 to 669). The 28- to 29-residue DNA-binding finger motif Y/F-X-Cys-$X_{2/4}$-Cys-X_{12}-His-X_3-His-X_5 (26, 41) is tandemly repeated eight times in the central, basic portion of the *sry h-1* protein. The third finger is interrupted at the genomic DNA level by the first intron. Apart from the TFIIIA gene (39), this is so far the only other instance in which an intron is located within a finger-coding region.

Other structural features of the *sry h-1* polypeptide include a polyalanine repeat, 11 residues long, located 28 amino acids from the finger domain towards the carboxy end. A similar polyalanine repeat has been found in the proteins coded by several *Drosophila* developmental genes (see reference 21). The glutamine, serine, and threonine residues are unevenly distributed with short stretches rich in either glutamine (residues 169 to 186 and 632 to 647) or serine plus threonine (residues 52 to 85 and 115 to 162), a distribution which is reminiscent of that found in the Sp1 protein (15). The glutamine residues are not present as polyglutamine tracts however, as it has been found in proteins containing the Opa (or M) repeats (7, 45). There are five potential N-linked glycosylation sites, three of which are clustered between residues 97 and 130.

The 5' untranslated region of the *sry h-1* mRNA contains five AUG codons. S1 mapping and cDNA sequencing data indicate that the h-1E cDNA is full length and derives from a mRNA molecule initiated at position +1, 250 nucleotides upstream of the *sry h-1* protein translation initiation site. Of

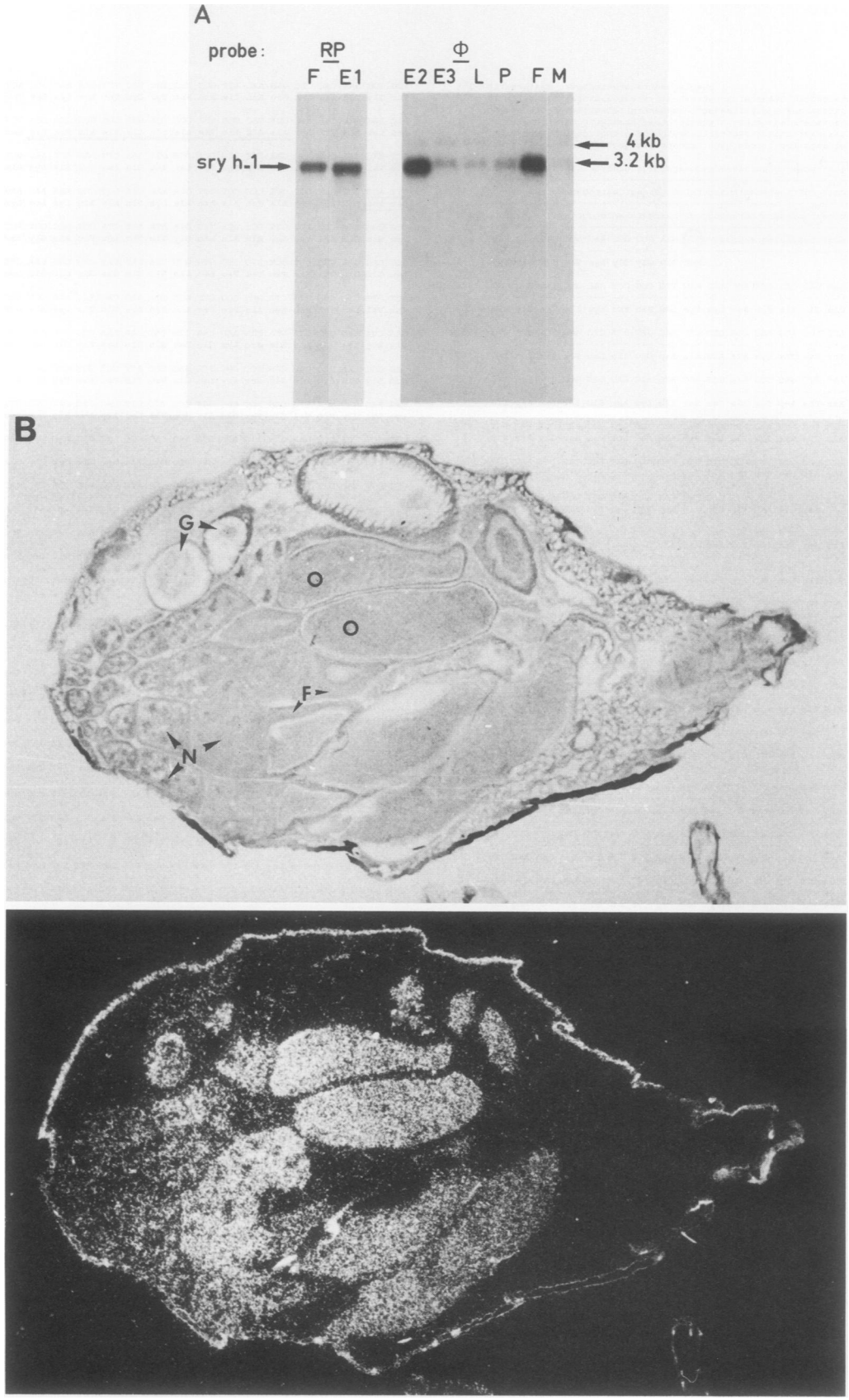

FIG. 4. Analysis of the transcription of the *sry h-1* gene. (A) Developmental profile of *sry h-1* transcript accumulation. Samples of poly(A)$^+$ RNA prepared from virgin females (F); 0- to 2-h (E1), 2- to 4-h (E2), and 10- to 20-h (E3)-old embryos; larvae (L); pupae (P); and adult males (M) were electrophoresed and transferred to nitrocellulose filters as described by Vincent et al. (44). Blots were probed either with a single-stranded probe corresponding to the RP (left panel) fragment (Fig. 3) or with nick-translated λ Dm sry h 18 phage DNA (ϕ) (right panel). (B) Detection of the *sry h-1* transcript by in situ hybridization. A 2-kbp *Eco*RI-*Eco*RI double-stranded DNA fragment containing sequences complementary to the *sry h-1* mRNA between the *Eco*RI site at position 1166 and the 3′ end of the mRNA was cut out of the h-1E cDNA, gel purified, labeled by nick translation to a specific activity of 4×10^8 cpm/μg, and used as a hybridization probe to detect *sry h-1* transcript on a longitudinal cryosection of an adult female abdomen. Top panel, Bright-field illumination; bottom panel, dark-field illumination. Specific labeling is detected over the nurse cells (N) and the oocytes (O). Note that follicle cells (F) are not labeled. The strong labeling over the lumen of the gut (G) is due to a non-specific trapping of the probe by the gut content.

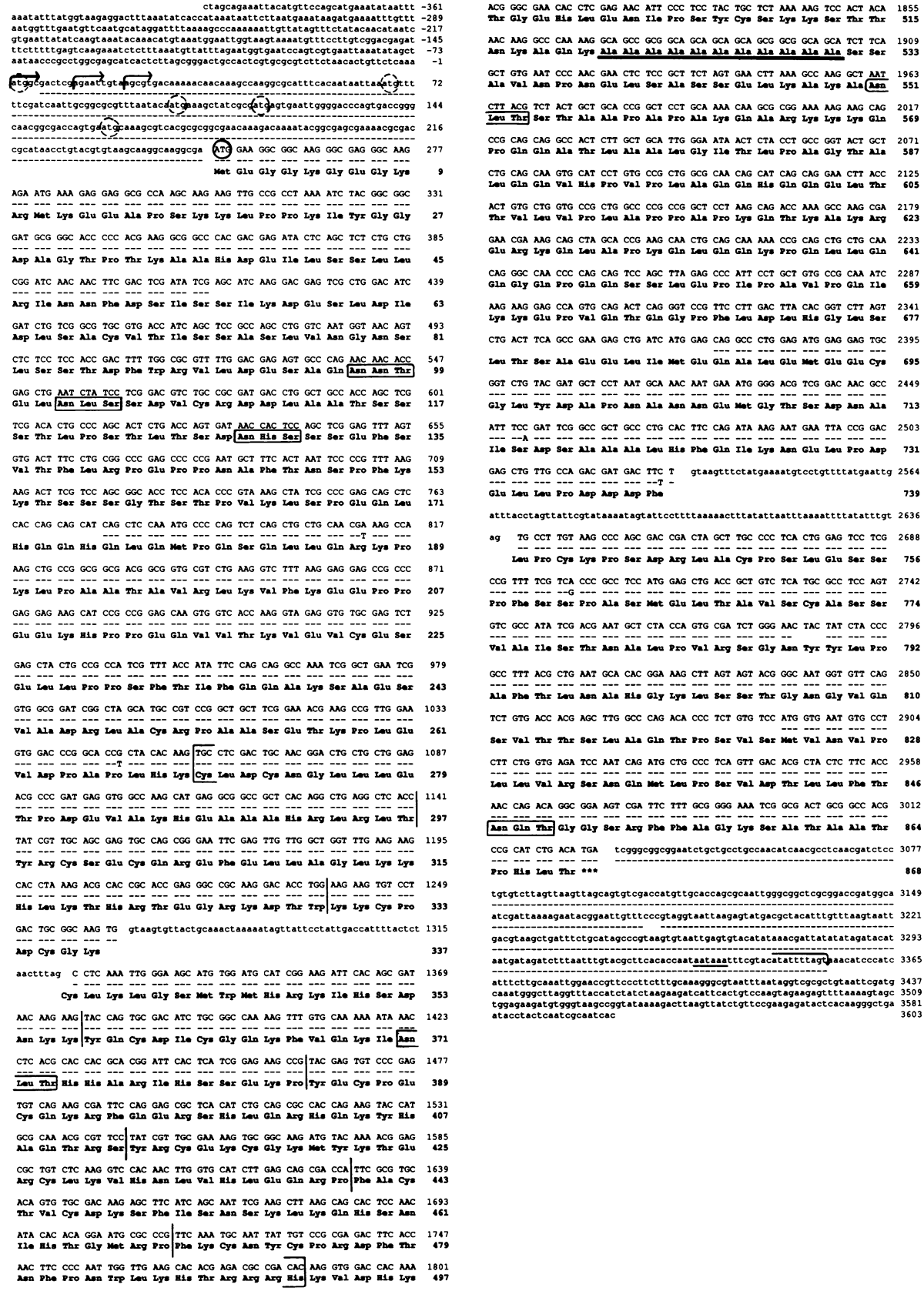

FIG. 5. DNA and amino acid sequences from the *sry h-1* transcribed region. The 5′-to-3′ genomic sequence of the mRNA-like strand is presented. The A nucleotide at the most upstream transcription initiation site is referred to as nucleotide 1. The determined cDNA sequences are underlined by dotted lines except for nucleotide substitutions which are indicated. The predicted amino acid sequence of the *sry h-1* open reading frame is indicated under the nucleotide sequence. The amino acid positions are designated by boldface numerals. The finger-coding domain is bracketed. Vertical lines separate individual fingers. Potential glycosylation sites are boxed. The polyalanine stretch is underlined. The putative initiation codon and polyadenylation signal are circled and underlined, respectively. AUG codons in the 5′-untranslated region are circled with a dotted line. The 5′ and 3′ ends of the *sry h-1* transcripts are indicated by bent arrows.

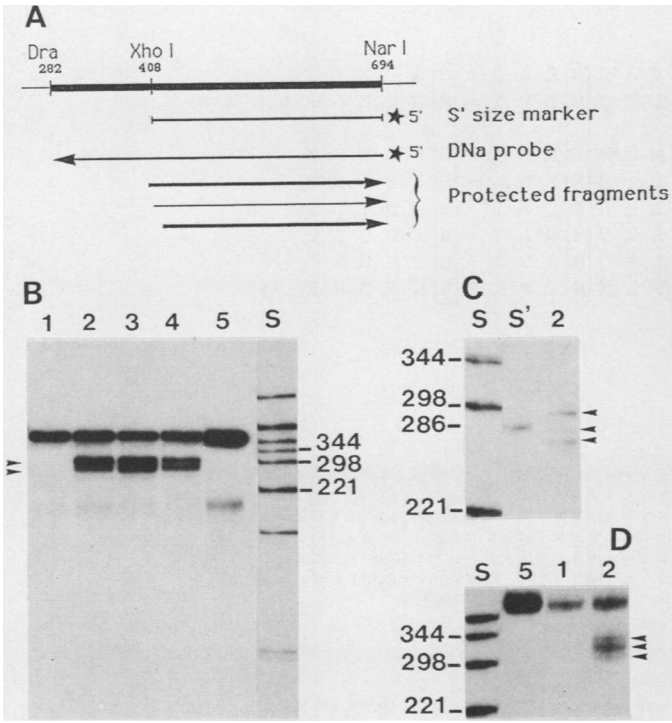

FIG. 6. S1 nuclease mapping of the 5' ends of the *sry h-1* mRNAs. (A) Diagram of the labeled probe and protected fragments. ★5', ^{32}P-labeled 5' end of the 414-bp *Nar*I-*Dra*I (DNa) fragment (Fig. 3B) (positions 282 to 694 on the sequence shown in Fig. 5) used as a probe for mapping the 5' end of the *sry h-1* mRNA. The 5'-end-labeled (★5') 286-bp *Nar*I-*Xho*I fragment was used as a size marker (S'). Fragments protected by the *sry h-1* transcripts are indicated by the three arrows pointing rightward. (B, C, D) Analysis of protected DNA fragments on 6% (panel B) or 4.5% (panel C) polyacrylamide gels containing 7.5 M urea or on 5% polyacrylamide gels without urea (panel D). The positions of the 344-, 298-, and 221-nucleotide DNA fragments from *Hin*fI-digested pBR322 DNA used as a size standard (lanes S) are indicated. The 5'-end-labeled *Nar*I-*Dra*I (lanes 5) probe fragment and the 5'-end-labeled *Nar*I-*Xho*I (lane S') fragment were also run as size standards. Origin of the RNA used for hybridization with the probe: lane 1, mouse rRNA; lane 2, RNA from virgin adult females; lane 3, poly(A)$^+$ RNA from 0- to 12-h-old embryos; lane 4, poly(A)$^+$ RNA from 12- to 20-h-old embryos. ▶, Positions of the protected fragments.

the five AUG initiator codons preceding the predicted start of translation of the *sry h-1* protein, none starts an open reading frame longer than 19 amino acids, the corresponding stop codons being all located upstream of the major open reading frame. Therefore, *sry h-1* adds to the growing list of genes in which translation termination-reinitiation presumably occurs, allowing initiation to take place at an internal AUG (30).

DISCUSSION

Since the description of the TFIIIA zinc/DNA-binding unit (26), several genes encoding putative fingered proteins have been independently isolated from various eucaryotes. They include *Drosophila sry*β and *sry*δ (42), Krüppel, (32), hunchback (38), snail (2), terminus (1), *Saccharomyces cerevisiae ADR1* (13), human *Sp1* (15), and a sex-determining region of the human Y chromosome (29). The *ADR1* and *Sp1* gene products were already known as transcription factors binding to specific DNA sequences (16, 37). Krüppel and hunchback are two of the three members of the gap class of segmentation genes required for the normal establishment of adjacent and partially overlapping segment regions of the embryo (19, 38). snail is a zygotic gene whose expression is essential for the correct specification of the embryonic dorsoventral pattern (2). Additional sequences encoding putative Zn^{2+} finger proteins were found in genomic or cDNA clones isolated by cross-hybridization with a finger-coding DNA probe made from Krüppel (3, 4, 33, 35). Sequence similarity between individual Cys_2/His_2 proteins appears mainly, if not exclusively, restricted to the finger domain, suggesting that the proteins share some DNA-binding properties but carry specific regulatory functions. The *sry*β and *sry*δ genes, which map at a single locus, most probably arose from the evolutionarily recent duplication of an ancestral finger protein gene. Southern blot analysis of DNAs from several dipterae and vertebrates strengthen this hypothesis, inferred previously from DNA sequence comparison (42). A similar result was recently reported by Mc Keown et al. (25), concerning a 7.5- to 8-kbp tandem duplication located close to the *tra* gene. In that case, *Drosophila mauritiana*, like *D. melanogaster*, has the tandem duplication, whereas *Drosophila simulans* has a single copy. Because they also show similar patterns of transcription during development, they represent an interesting, and so far unique, model for addressing questions about the evolution of the finger protein gene structure and function (42).

We report here the characterization of a new *D. melanogaster* multifingered protein gene, *sry h-1*, cloned by virtue of its similarity to *sry*β and *sry*δ probes and showing a similar pattern of transcript accumulation during development. Sequences of the eight *sry h-1* finger motifs are shown in Fig. 7A. As their consensus sequence is Y/F-X-Cys-X_2-Cys-X_3-Phe-X_5-Leu-X_2-His-X_3-His, they are referred to as Cys_2/His_2 fingers as opposed to the Cys_2/Cys_2 finger characteristic of steroid hormone receptors and of some yeast activators (6, 15, 29a).

The conservation of DNA sequences hybridizing to the *sry* probes is not restricted to the *Drosophila* genome. In fact, the signal corresponding to the *sry h-1* gene (Fig. 1, *Drosophila* DNA lanes) is less intense than some other signals seen in the *Drosophila* or vertebrate DNA lanes (not shown). One cannot conclude that all the observed bands correspond to finger-coding sequences on the sole basis of Southern hybridization signals obtained under low stringency conditions, but the structure of the *sry h-1* gene suggests that at least some do. The most intense signals are either *sry*β or *sry*δ specific. This may be due to both a significant sequence divergence of the two probes and the absence in the proteins of invariant links joining adjacent fingers (42). Indeed, sequence comparisons reveal that the invariant H/C link TGEKPY/F characteristic of Krüppel is the major motif common to Krüppel-related genes in *Drosophila*, mouse (3, 4), or *Xenopus* (34) genomes, as well as to the genes for the human SP1 or yeast ADR1 proteins (13, 15) at the nucleotide and amino acid sequence levels. Cross-hybridization between *sry*β or *sry*δ and *sry h-1* genes is limited to a fraction of the finger-coding domains. This cross-hybridization precisely corresponds to the region of maximal homology between the seven predicted *Drosophila* multifingered proteins known to date, as found by computer search (Fig. 7B). A link motif of the general sequence XDXKKY is found in the sry h-1, sryβ, sryδ, and snail proteins (Fig. 7B). This conserved motif with no proline residue might have some specific properties. It is interesting that some finger proteins,

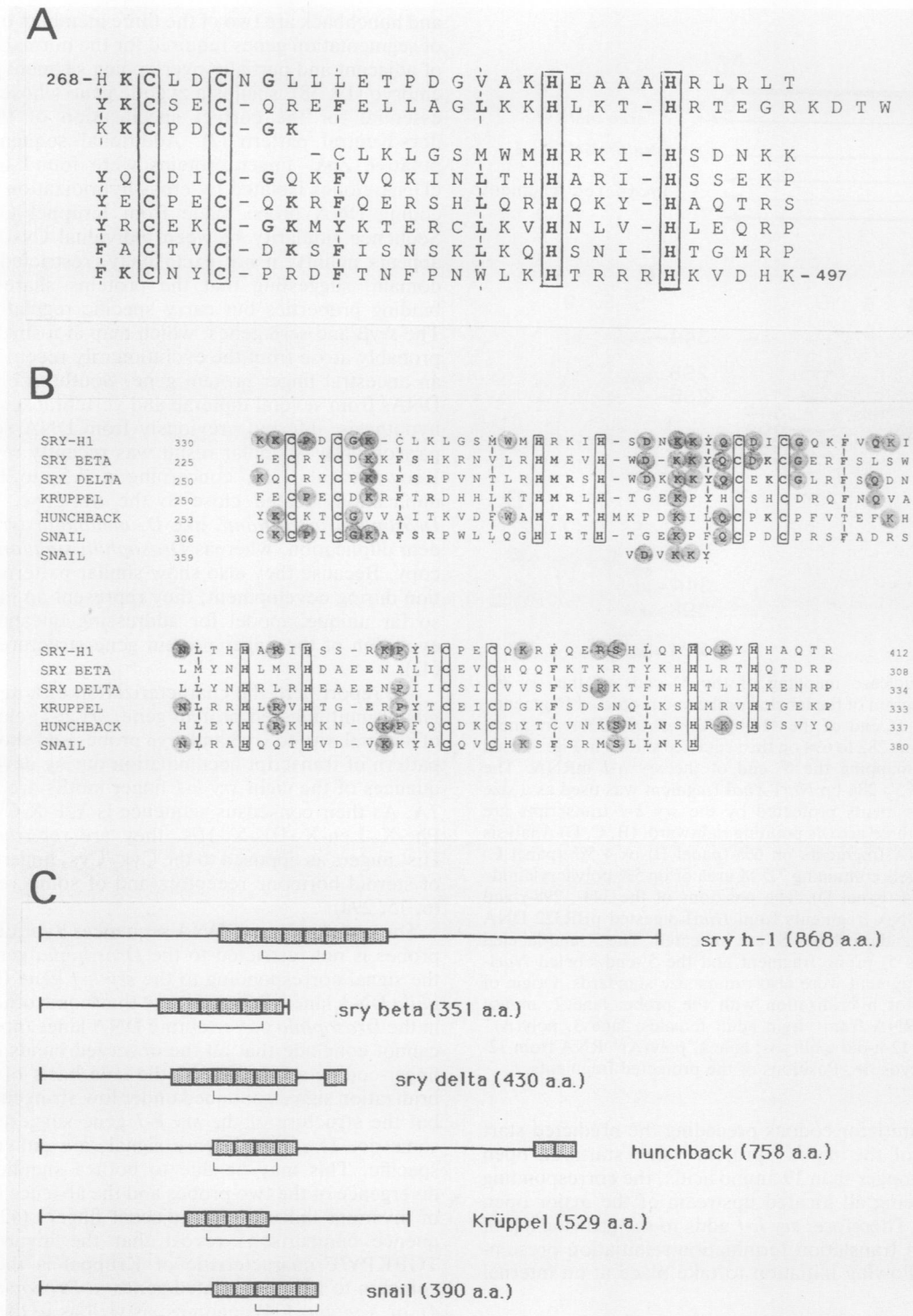

FIG. 7. Structure and sequence comparison of *Drosophila* finger proteins. (A) Alignment of finger motifs repeated in the *sry h-1* polypeptide. The first cysteine of each Cys-X_2-Cys doublet is placed at the third position. The positions of the NH2 and COOH terminal amino acids are indicated. The amino acids which are identical at the same position in at least 4 different repeats are shown in boldface, with the invariant amino acids at finger consensus positions enclosed in boxes or connected by vertical lines. (B) Alignment of part of the finger domains of all six predicted *Drosophila* finger proteins: *sry h-1*, *sry*β (42), *sry*δ (42), Krüppel (32), hunchback (38), and snail (2). The polypeptide sequences were aligned from a computer scan for similarity to a protein subsequence by using a 30-amino-acid window (see Materials and Methods). Only the best alignment score is shown. Amino acids identical at the same position in *sry h-1* and any other protein are circled and shaded. Amino acids identical at the same position in at least three proteins apart from *sry h-1* are shown in boldface. (C) Schematic arrangement of fingers in the *Drosophila* sry h-1, sryβ, sryδ, Krüppel, hunchback, and snail proteins drawn to scale. Orientation of each polypeptide is amino (left) to carboxy (right) terminus. The total number of amino acids of each protein is indicated. Each complete finger is drawn as a square. The position of maximal sequence similarity between the different proteins (Fig. 7B) is underlined with brackets. a.a., Amino acids.

such as TFIIIA or snail, possess one or two copies of the TGEKPF link in addition to unrelated link sequences (2, 26); furthermore, six of the seven introns interrupting the TFIIIA finger-coding region are located within the interfinger links irrespective of their sequence, i.e., TGEKPF or another sequence (39). These results suggest that the isolation of finger proteins by cross-hybridization with variable interfinger sequences may be problematic.

sry h-1 is the only *Drosophila* finger protein gene (out of seven) which shows an intron interrupting the finger domain at the genomic DNA level. This intron is located within the tip of the third finger, a domain of the finger proposed to interact specifically with DNA (26). The exon 1-exon 2 in-frame joining results not only in an additional cysteine residue creating a Cys-X_2-Cys motif within the finger but also, more unexpectedly, in an atypical finger structure with only 11 residues instead of 12 between the consensual second cysteine and first histidine residues. So far, 12 amino acids have invariably been found in finger tips (consensus: C-X_{12}-H) (6, 18; Payre and Vincent, in press). Since no intron has yet been located within fingertips except for one in *Xenopus* TFIIIA, we do not know whether the structural variation we observe in the third *sry h-1* finger is a consequence of the intron location or has some functional significance. It must, however, be noted that the first *sry h-1* finger also deviates from the consensus with 13 amino acids in place of 12.

Like the *sryβ* (43), *sryδ* (F. Payre and A. Vincent, unpublished data), and hunchback (38) genes, the *sry h-1* gene shows both maternal and zygotic transcription. The *sry h-1* mRNA is maximally accumulated in oocytes during late stages of oogenesis and maternally inherited by the embryo. We have not determined yet at which stage the zygotic transcription of *sry h-1* is initiated (*sryβ*, *sryδ*, and hunchback are zygotically transcribed prior to the cellular blastoderm stage [38, 43; Payre and Vincent, in preparation]). Recent data obtained with an anti-*sryδ* antibody show that at least one of the sry finger proteins is synthesized in the mature egg and is associated with embryo nuclei at the time of zygotic transcription and gene activation (Payre and Vincent, in preparation). In situ hybridization to tissue sections does not reveal any tissue-specific localization of the *sry h-1* mRNA during embryogenesis, although differences in staining intensities of different tissues are reproducibly observed. This situation is similar to that observed for *sryβ* and *sryδ* (C. Yanicostas, unpublished results) and contrasts with the pattern of zygotic transcription of hunchback (38), Krüppel (19), or terminus (1), with transcripts spatially localized in specific regions of the embryo.

Posttranscriptional regulation of *sry h-1* expression cannot, however, be excluded in view of the structure of the *sry h-1* mRNA which shows a 5′ untranslated region with 5 AUG initiation codons preceding the protein-predicted start of translation. Translational regulation through several small open reading frames in the mRNA leader sequence has been proposed to participate in generating the anteroposterior concentration gradient of the caudal protein (22) and the pattern of Krüppel protein at stages later than gastrulation (9) and has been shown to modulate the expression of yeast GCN4 mRNA (40).

Developmental patterns of transcription and in situ hybridization data suggest that the *sry h-1*, *sryβ*, and *sryδ* (Yanicostas, unpublished) genes may not be involved directly in embryonic pattern formation but rather in regulation of general cell functions such as cell metabolism, cell division, or cell differentiation. In view of the similarities of structure and expression of these genes, we propose that they belong to a complex set of genes maternally and zygotically expressed and coding for regulatory DNA-binding proteins with a role in gene activation in early embryos. Whether the acquisition by the zygotic genome of competence for transcription at cycle 10 (5) or the general transcriptional activation of this genome at cycles 11 or 12 depends upon the accumulation of maternal finger proteins in the egg or early embryo is an important issue to be resolved. Detailed functional analysis of the *sry* genes should help to elucidate some of the mechanisms involved in this activation, as well as in selective maintenance of gene expression (or repression) paralleling cell determination and tissue specification during development.

ACKNOWLEDGMENTS

We thank Larry Kauvar for a gift of an embryonic cDNA library and Judith Lengyel and James Kadonaga for sharing unpublished results.

This work was supported by grants to A.V. and J.-A.L. from the Centre National de la Recherche Scientifique, the Fondation pour la Recherche Médicale, and the Ligue Nationale Française contre le Cancer.

LITERATURE CITED

1. **Baldarelli, R. M., P. A. Mahoney, F. Salas, E. Gustavson, P. D. Boyer, M. F. Chang, M. Roark, and J. A. Lengyel.** 1988. Transcripts of the *Drosophila* blastoderm-specific locus, *terminus*, are concentrated posteriorly and encode a potential DNA-binding finger. Dev. Biol. **125**:85–95.
2. **Boulay, J. L., C. Dennefeld, and A. Alberga.** 1987. The *Drosophila* developmental gene *snail* encodes a protein with nucleic acid binding fingers. Nature (London) **330**:395–398.
3. **Chavrier, P., P. Lemaire, O. Revelant, R. Bravo, and P. Charnay.** 1988. Characterization of a mouse multigene family that encodes zinc finger structures. Mol. Cell. Biol. **8**:1319–1326.
4. **Chowdhury, K., U. Deutsch, and P. Gruss.** 1987. A multigene family encoding several "finger" structures is present and differentially active in mammalian genomes. Cell **48**:771–778.
5. **Edgar, B. A., and G. Schubiger.** 1986. Parameters controlling transcriptional activation during early Drosophila development. Cell **44**:871–877.
6. **Evans, R. M., and S. M. Hollenberg.** 1988. Zinc fingers: gilt by association. Cell **52**:1–3.
7. **Frigerio, G., M. Burri, D. Bopp, S. Baumgartner, and M. Noll.** 1986. Structure of the segmentation gene *paired* and the Drosophila PRD gene set as part of a gene network. Cell **47**:735–746.
8. **Gans, M., C. Audit, and M. Masson.** 1975. Isolation and characterization of sex-linked female-sterile mutants in *Drosophila melanogaster*. Genetics **81**:683–704.
9. **Gaul, V., E. Seifert, R. Schuch, and H. Jäckle.** 1987. Analysis of *Krüppel* protein distribution during early Drosophila development reveals posttranscriptional regulation. Cell **50**:639–647.
10. **Gehring, W. J.** 1987. Homeo boxes in the study of development. Science **236**:1245–1252.
11. **Glaser, R. L., M. F. Wolfner, and J. T. Lis.** 1986. Spatial and temporal pattern of *hsp26* expression during normal development. EMBO J. **5**:747–754.
12. **Hafen, E., and M. Levine.** 1986. The localization of RNAs in Drosophila tissue sections by in situ hybridization, p. 139–157. *In* D. B. Roberts (ed.), Drosophila, a practical approach. IRL Press, Oxford.
13. **Hartshorne, T. A., H. Blumenberg, and E. T. Young.** 1986. Sequence homology of the yeast regulatory protein ADR1 with Xenopus transcription factor TFIIIA. Nature (London) **320**: 283–287.
14. **Jurgens, G., E. Wieschaus, C. Nüsslein-Volhard, and H. Kluding.** 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. II. Zygotic loci on the third chromosome. Roux's Arch. Dev. Biol. **193**:283–295.
15. **Kadonaga, J. T., K. R. Carner, F. R. Masiarz, and R. Tjian.**

1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell **51**:1079–1090.

16. **Kadonaga, J. T., K. A. Jones, and R. Tjian.** 1986. Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends Biochem. Sci. **11**:20–23.
17. **Keller, E. B., and W. A. Noon.** 1985. A conserved internal signal in introns of *Drosophila* pre-mRNAs. Nucleic Acids Res. **13**: 4971–4981.
18. **Klug, A., and D. Rhodes.** 1987. "Zinc fingers": a novel protein motif for nucleic acid recognition. Trends Biochem. Sci. **12**:464–469.
19. **Knipple, D. C., E. Seifert, U. B. Rosenberg, A. Preiss, and H. Jäckle.** 1985. Spatial and temporal patterns of *Krüppel* gene expression in early Drosophila embryos. Nature (London) **317**: 40–44.
20. **Langer-Safer, P. R., M. Levine, and D. C. Ward.** 1982. Immunological method for mapping genes on *Drosophila* polytene chromosomes. Proc. Natl. Acad. Sci. USA **79**:4381–4385.
21. **Macdonald, P. M., P. Ingham, and G. Struhl.** 1986. Isolation, structure, and expression of *even-skipped*: a second pair-rule gene of Drosophila containing a homeo box. Cell **47**:721–734.
22. **Macdonald, P. M., and G. Struhl.** 1986. A molecular gradient in early *Drosophila* embryos and its role in specifying the body pattern. Nature (London) **324**:537–545.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connel, D. Quon, G. K. Sim, and A. Efstratiadis.** 1978. The isolation of structural genes from libraries of eucaryotic DNA. Cell **15**:687–701.
25. **McKeown, M., J. M. Belote, and B. S. Baker.** 1987. A molecular analysis of *transformer*, a gene in *Drosophila melanogaster* that controls female sexual differentiation. Cell **48**:489–499.
26. **Miller, J., A. D. McLachlan, and A. Klug.** 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. **4**:1609–1614.
27. **Mohler, J. D.** 1977. Developmental genetics of the Drosophila egg. I. Identification of 59 sex-linked cistrons with maternal effects on embryonic development. Genetics **85**:259–272.
28. **Nüsslein-Volhard, C., E. Wieschaus, and H. Kluding.** 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. I. Zygotic loci on the second chromosome. Roux's Arch. Dev. Biol. **193**:267–282.
29. **Page, D. C., R. Mosher, E. M. Simpson, E. M. C. Fisher, G. Mardon, J. Pollack, B. McGillivray, A. De la Chapelle, and L. G. Brown.** 1987. The sex-determining region of the human Y chromosome encodes a finger protein. Cell **51**:1091–1104.

29a. **Payre, F., and A. Vincent.** 1988. Finger proteins and DNA specific recognition: distinct patterns of conserved aminoacids suggest different evolutionary modes. FEBS Lett. **234**:245–250.

30. **Peabody, D. S., and P. Berg.** 1986. Termination-reinitiation occurs in the translation of mammalian cell mRNAs. Mol. Cell. Biol. **6**:2695–2703.
31. **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.
32. **Rosenberg, U. B., C. Schröder, A. Preiss, A. Kienlin, S. Côté, J. Riede, and H. Jäckle.** 1986. Structural homology of the product of the *Drosophila Krüppel* gene with *Xenopus* transcription factor IIIA. Nature (London) **319**:336–339.
33. **Ruiz i Altaba, A., H. Perry O'Keefe, and D. A. Melton.** 1987. *Xfin*: an embryonic gene encoding a multifingered protein in *Xenopus*. EMBO J. **6**:3065–3070.
34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
35. **Schuh, R., W. Aicher, U. Gaul, S. Côté, A. Preiss, D. Maier, E. Seifert, U. Nauber, C. Schröder, R. Kemler, and H. Jäckle.** 1986. A conserved family of nuclear proteins containing structural elements of the finger protein encoded by *Krüppel*, a Drosophila segmentation gene. Cell **47**:1025–1032.
36. **Shapiro, M. B., and P. Senapathy.** 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. **15**: 7155–7174.
37. **Shuster, J., J. Yu, D. Cox, R. V. L. Chan, M. Smith, and E. T. Young.** 1986. ADR1-mediated regulation of ADH2 requires an inverted repeat sequence. Mol. Cell. Biol. **6**:1894–1902.
38. **Tautz, D., R. Lehmann, H. Schnürch, R. Schuh, E. Seifert, A. Kienlin, K. Jones, and H. Jäckle.** 1987. Finger protein of novel structure encoded by *hunchback*, a second member of the gap class of *Drosophila* segmentation genes. Nature (London) **327**: 383–389.
39. **Tso, J. Y., D. J. Van Den Berg, and L. J. Korn.** 1986. Structure of the gene for *Xenopus* transcription factor TFIIIA. Nucleic Acids Res. **14**:2187–2200.
40. **Tzamarias, D., D. Alexandraki, and G. Thireos.** 1986. Multiple *cis*-acting elements modulate the translational efficiency of GCN4 mRNA in yeast. Proc. Natl. Acad. Sci. USA **83**:4849–4853.
41. **Vincent, A.** 1986. TFIIIA and homologous genes. The "finger" proteins. Nucleic Acids Res. **14**:4385–4391.
42. **Vincent, A., H. V. Colot, and M. Rosbash.** 1985. Sequence and structure of the *Serendipity* locus of *Drosophila melanogaster*. A densely transcribed region including a blastoderm-specific gene. J. Mol. Biol. **186**:149–166.
43. **Vincent, A., H. V. Colot, and M. Rosbash.** 1986. Blastoderm-specific and read-through transcription of the *sry* alpha gene transformed into the *Drosophila* genome. Dev. Biol. **118**:480–487.
44. **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.
45. **Warthon, K. A., B. Yedbovnick, V. G. Finnerty, and S. Artavanis-Tsakonas.** 1985. Opa: a novel family of transcribed repeats shared by the *Notch* locus and other developmentally regulated loci in *Drosophila melanogaster*. Cell **40**:55–62.
46. **Wieschaus, E., C, Nüsslein-Volhard, and C. Jurgens.** 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. III. Zygotic loci on the X-chromosome and fourth chromosome. Roux's Arch. Dev. Biol. **193**:296–307.