The Embryonically Active Gene, *unkempt*, of Drosophila Encodes a Cys_3His Finger Protein

Jym Mohler,[1] Nofit Weiss, Sumati Murli, Shahrzad Mohammadi, Kodela Vani, Georgia Vasilakis, Chan Hee Song, Alyse Epstein, Timothy Kuang, Jennifer English and Danielle Cherdak

Barnard College, New York, New York 10027

Manuscript received November 13, 1991
Accepted for publication March 3, 1992

ABSTRACT

The *unkempt* gene of Drosophila encodes a set of embryonic RNAs, which are abundant during early stages of embryogenesis and are present ubiquitously in most somatic tissues from the syncytial embryo through stage 15 of embryogenesis. Expression of *unkempt* RNAs becomes restricted predominantly to the central nervous system in stages 16 and early 17. Analysis of cDNAs from this locus reveals the presence of five Cys_3His fingers in the protein product. Isolation and analysis of mutations affecting the *unkempt* gene, including complete deletions of this gene, indicate that there is no zygotic requirement for *unkempt* during embryogenesis, presumably due to the contribution of maternally supplied RNA, although the gene is essential during post-embryonic development.

ZINC-finger proteins constitute a major type of transcription factor. These proteins are defined by motifs consisting of a specific arrangement of cysteine and histidine amino acids, that, by chelation of a Zn^{2+} ion, allows the formation of a DNA-binding structure [see reviews by EVANS and HOLLENBERG (1988) and HARRISON (1991)]. The major classes of zinc-finger proteins are defined by the nature of zinc-binding amino acids of the fingers, either as Cys_2His_2, Cys_4 or Cys_6. The Cys_2His_2 finger proteins, first characterized in the Xenopus TFIIIA and Drosophila *Krüppel* proteins (BROWN, SANDER and ARGOS 1985; ROSENBERG *et al.* 1986), generally have multiple Zn^{2+}-binding fingers ranging from 2 (ADR1; HARTSHONE, BLUMBERG and YOUNG 1986) to 14 (testis determining factor [TDF]; PAGE *et al.* 1987) each complexed with a single Zn^{2+} atom (PAVLETICH and PABO 1991). In contrast, proteins of the Cys_4 and Cys_6 classes generally contain only a single Cys_4 or Cys_6 Zn^{2+}-binding finger which binds either one (Cys_4) or two (Cys_6) Zn^{2+} atoms (see review by HARRISON 1991), although a major subclass of Cys_4 proteins, the steroid receptor DNA-binding proteins (see review by EVANS 1988), also have an additional Cys_5 Zn^{2+}-binding finger with radically different cysteine spacing, located 15 amino acids C-terminal to the Cys_4 finger.

Unlike other DNA-binding structural motifs, these zinc-finger motifs may provide a general scaffold for binding to both DNA and RNA. Three zinc-binding proteins, two Cys_2His_2 proteins (TFIIIA and p43) and one Cys_4 protein (eIF2β), function by binding to RNA. Both TFIIIA and p43 can bind to 5SRNA in Xenopus oocytes (HONDA and ROEDER 1980; PELHAM and BROWN 1980; JOHO *et al.* 1990); while TFIIIA binds the control region of the 5S RNA gene as well. The eIF2β translation initiation factor binds into a complex containing tRNA, rRNA and mRNA and contains a potential Cys_4 finger that may be involved in binding one or more of these RNAs (PATHEK *et al.* 1988). For each of these three proteins, the spacing between the coordinate amino acids is somewhat altered from the consensus DNA-binding finger motifs.

In addition to these major classes of zinc-finger proteins are a number of minor classes. One such minor class is the class of Cys_3His proteins, which includes three genes that are coinduced with the c-*fos* and c-*jun* oncogenes in cultured mammalian cell lines by serum, growth factors or tumor promoters (see review by HERSCHMAN 1991). These three proteins, TIS11/Nup475/TTP (VARNUM *et al.* 1989; DUBOIS *et al.* 1990; LAI, STUMPO and BLACKSHEAR 1990), cMG1 (GOMPERTS, PASCALL and BROWN 1990), and TIS11b (as described in the review by HERSCHMAN 1991), each contains two tandem repeats of a Cys_3His motif and are more than 70% identical within the 67 amino acid region defined by these Cys_3His repeats. One of these proteins, TIS11/nup475/TTP, has been shown to bind Zn^{2+} and to be a nuclear antigen, suggestive of a possible role as a transcription regulator (DUBOIS *et al.* 1990).

This report characterizes a fourth, more distantly related, Cys_3His protein gene, *unkempt*, in Drosophila, which was identified as encoding an abdunant embryonic RNA. Sequence analysis of the *unkempt* gene reveals the presence of five potential Cys_3His fingers. We have found *unkempt* RNAs to be ubiquitously present throughout early embryos and to become

[1] To whom correspondence should be addressed.

localized to the central nervous system during later embryonic stages. We have utilized a *w+* transposon to create mutations of the *unkempt* gene, including a complete deletion of this locus. Analysis of these mutations of the *unkempt* gene indicate that zygotic expression of the gene is not required during embryonic development, presumably due to the contribution of maternally supplied RNA from this gene.

MATERIALS AND METHODS

Isolation of cDNA clones: The cDNA library of M. NOLL, constructed of cDNAs derived from 0–4-hr embryonic RNA inserted into λgt10 (FRIGERIO *et al.* 1988) was screened twice by hybridization with ^{32}P-labeled Drosophila DNA in Bluescript KS$^-$ according to the procedure of SAMBROOK, FRITSCH and MANIATIS (1989). Two different probes were used: three clones (SM1, SM4 and SM7) were isolated using genomic *unk* DNA (−32 to −37 region from the K15 phage clone) and 32 clones (NW1–32) were isolated when probed with the SM1 cDNA, which was isolated in the first screen. A plasmid cDNA library of N. BROWN, constructed of cDNAs derived from 4–8-hr embryonic RNA inserted into pNB40 (BROWN and KAFATOS 1988), was screened with the ^{32}P-labeled SM1 cDNA restriction fragment recovered from an agarose gel according to the procedures of SAMBROOK, FRITSCH and MANIATIS (1989); a single cDNA clone (15C) was isolated in this screen.

Sequence analysis: The cDNA clones and genomic DNA (2.4-kb *Xho*I-*Xho*I and 1.5-kb *Xho*I-*Eco*RV fragments [pX7 and pXV1] from the −35 region of the K15 phage clone of the walk of MOHLER *et al.*, 1991) were recloned into Bluescript KS$^-$ for sequencing. The ends of the SM4, NW1 and NW29 clones were sequenced, following rescue of single-stranded DNA by f1R408 helper phage, using oligonucleotide primers corresponding to the T7 and T3 promoters in Bluescript KS$^-$ with the Sequenase Kit (USB) according to procedures of the kit. The ends of the 15C cDNA clone and the upstream genomic sequence were sequenced from T3 and Sp6 promoter primers (15C) or a synthetic primer (5′-GTTCGTTTCATTTG-3′, Operon Technologies) derived from the 5′ end of the 15C sequence (pX7) using the double-stranded sequencing procedures of AUSUBEL *et al.* (1987).

Sets of nested deletions were prepared in both directions for SM4, in the 3′ to 5′ direction for NW29, and in the 5′ to 3′ direction for 15C and pXV1 according to the procedure of HENIKOFF (1984) following double digestion of the plasmids with *Kpn*I and *Hin*dIII (SM4, NW29 and 15C) or *Sac*I and *Eco*RI (pX7). Deletion constructs were sequenced using the T3 promoter primer (SM4, NW29 and 15C) or the T7 promoter primer (pXV1) of either rescued single-strand DNA (SM4 and NW29) or double-strand DNA (15C and pXV1). The sequence was analyzed using HIBIO DNASIS and PROSIS software (Hitachi).

Isolation of *unkempt* alleles: *unk* alleles were generated in one of three screens: two involved hybrid-dysgenesis induced loss of a P(w,ry)A (HAZELRIGG, LEVIS and RUBIN 1984) insertion in *unk*, the third isolated radiation-induced mutations that failed to complement an *unk* mutation for viability. All stocks in these schemes carried *w* or w^{1118} on the *X* chromosome. The first hybrid-dysgenesis screen (KIDWELL 1986) is shown in Figure 1a. unk^{OR20}/*TM3* females were mated to *Ly/TM3 P*-element bearing males at 25°. unk^{OR20}/*TM3* F_1 males were crossed to *Ly/TM3* sibling females at 18°. 38 w^- flies were recovered from approximately 1500 F_2 flies screened.

The second dysgenesis screen utilized Δ2–3 (ROBERTSON *et al.* 1988) as shown in Figure 1b. *ry* Δ*2-3* males were crossed to e^s *bar-3* unk^{OR20}/*TM3* females at 25°. Approximately 400 F_1 e^s *bar-3* unk^{OR20}/*ry D2-3* males were individually mated to *bar-3* females. F_2 e^s *bar-3* $unk^{*(OR20)}$/*bar-3* males which had lost the w^+ marker were mated to e^s *bar-3* unk^{OR20}/*TM3* or unk^{OR20}/*TM3* females. At most three reversion chromosome were retained from each dysgenic male to limit multiple recovery of the same event. Of the recovered 275 F_2 males which had lost the w^+ marker, 82 failed to complement for viability with unk^{OR20} in the subsequent generation and were retained as stocks.

In the γ-irradiation mutagenesis screen (Figure 1c), *w* males were irradiated with 4000R from a ^{137}Cs-radiation source and mated to e^s *bar-3* unk^{r93b}/*TM3* females. Each of 2948 F_1 */*TM3* males was individually mated back to e^s *bar-3* unk^{r93b}/*TM3* females. For each F_1 male, F_2 */e^s bar-3* unk^{r93b} flies were scored for viability. The F_2 */*TM3* progeny of the one male displaying loss of *unk* function were crossed *inter se* to establish a stock (unk^{CHS20}).

Southern and Northern analysis: Drosophila DNA was prepared by homogenization of adult flies in 100 mM Tris, pH 9, 20 mM EDTA pH 8, 5% Sucrose, 1% SDS. The homogenate was extracted with equal volumes of phenol twice, and of $CHCl_3$ twice. The DNA was precipitated from the extracted homogenate with ethanol. Phage and plasmid DNAs were isolated as described by SAMBROOK, FRITSCH and MANIATIS (1989). Drosophila RNA was prepared by homogenization of 0.5 g of Drosophila embryos in a mixture of 25% phenol, 25% $CHCl_3$, 50% Homogenization Buffer (0.15 M Na-acetate, 5 mM EDTA, 1% SDS, 50 mM Tris pH 9). The aqueous homogenate was recovered and extracted once with a 50:50 mixture of phenol and $CHCl_3$, then extracted twice with $CHCl_3$. Nucleic acid was recovered by precipitation with ethanol. Poly(A^+) and poly(A^-) RNA was prepared by fractionation over oligo(dT) cellulose columns.

For Drosophila DNA, either 1 μg or the equivalent of 1 fly was digested with an appropriate restriction enzyme overnight in the buffer supplied by the enzyme manufacturer. For phage or plasmid clone DNAs, 0.1 μg of DNA was similarly digested. Digested DNA was electrophoresed at 25 V overnight in 0.7% agarose gels in TAE buffer and blotted to Nylon 66 membrane (Micron Separation) as described by SAMBROOK, FRITSCH and MANIATIS (1989). For Drosophila RNA, 5 μg of poly(A^+) and 50 μg poly(A^-) RNA was loaded onto formaldehyde denaturing gels and run at 150 V for 2 hr. The northern gel was blotted overnight to Nylon 66 membrane as described by SAMBROOK, FRITSCH and MANIATIS (1989). Filters were hybridized to plasmid or phage cloned DNAs labelled with ^{32}P by random priming (SAMBROOK, FRITSCH and MANNIATIS, 1989) at 70° in 0.3 M NaCl overnight, and nonhybridized label was removed with four half-hour washes at 70° in 0.1 M NaCl. Hybridized filters were autoradiographed for 15 min to 2 weeks depending on the intensity.

***In situ* hybridization and α-HRP staining of embryos:** *In situ* hybridization to embryos was performed according to the procedure of TAUTZ and PFEIFLE (1989) using digoxygenin labeled probes. The SM1 cDNA clone, used for *unkempt* hybridizations, is a slightly smaller cDNA clone than the SM4 cDNA clone and its sequences are entirely deleted in *Df(3R)r94a* and unk^{r98b}. Immunological staining of Drosophila embryos was performed according to the procedure of BIER *et al.* (1989) for α-β-galactosidase staining of embryos, except rabbit α-HRP (Sigma, diluted 1:50) was used as a primary antibody. Staging of embryos is according to CAMPOS-ORTEGA and HARTENSTEIN (1985).

Determination of lethal stage: Embryonic viability of

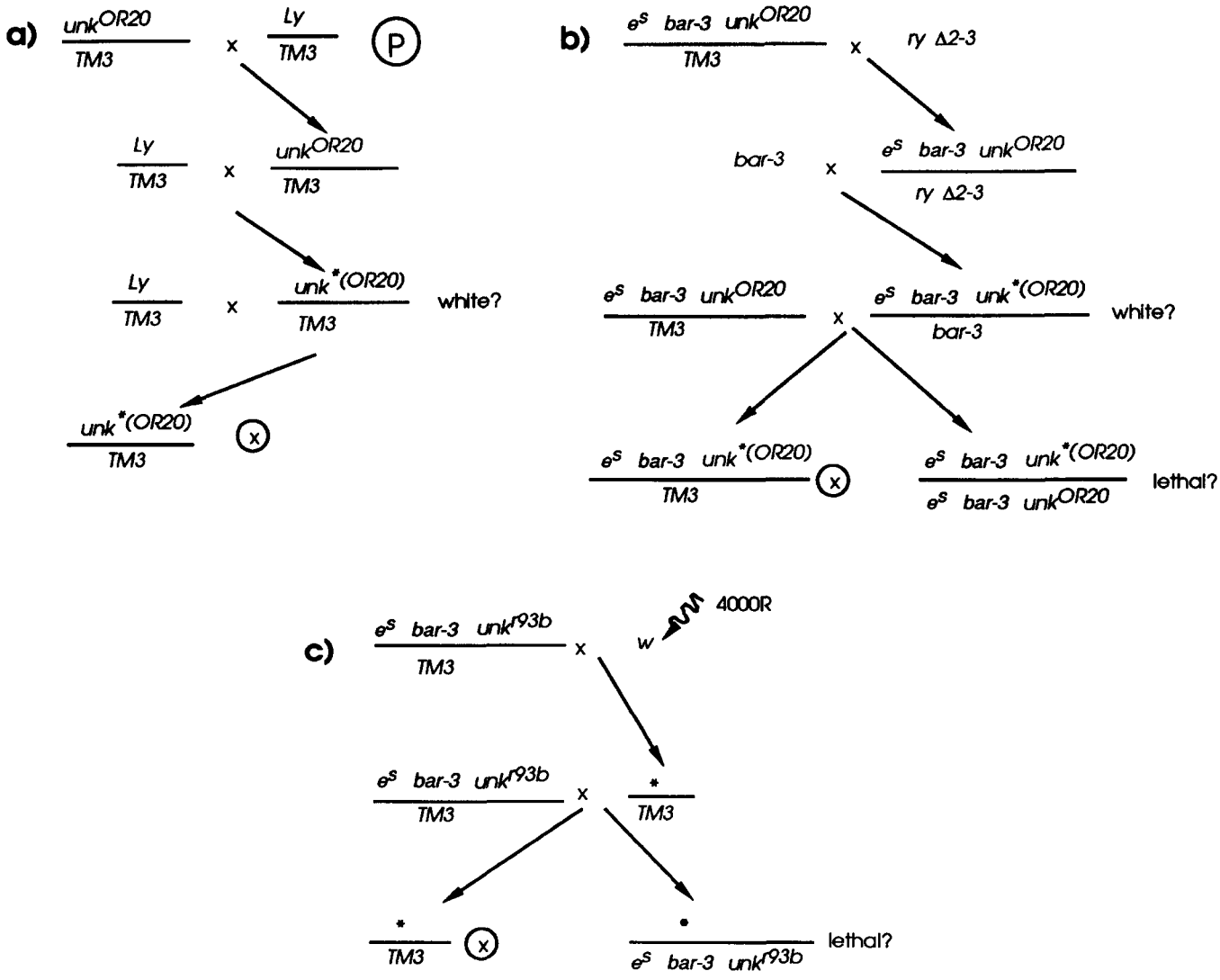

FIGURE 1.—Schemes used to generate *unk* mutations. (a) Hybrid dysgenesis eye color reversion of unk^{R14}; (b) Δ2-3 induced eye color reversion of unk^{R14}; and (c) γ-irradiation mutagenesis.

three *unk* alleles (unk^{r98b}, unk^{R14} and unk^{CHS20}) was determined by egg hatch rates of progeny of the F_1 cross of the *unk* stock and wild-type flies. Embryonic viability for other *unk* mutations was determined by phenotypic analysis of cuticle preparations of progeny of the *unk* mutant stock balanced with *TM3, daeh*, screening for phenotypes distinct from that of homozygous *TM3, daeh* embryos. Postembryonic lethal phases of unk^{r98b}, unk^{CHS20} and unk^{R14} homozygous embryos and unk^{r98b}/unk^{R14}, unk^{r98b}/unk^{r98b}, and unk^{r98b}/unk^{CHS20} larvae were determined by analysis of development of Tb^+ larvae from crosses between *unk/TM6B, Tb* stocks.

RESULTS

The *unkempt* locus was initially identified as a site of a P(w,ry)A transposon insertion in 94D10-E1 (P((w,ry)A^R)O20(1C), HAZELRIGG, LEVIS and RUBIN 1984; LEVIS, HAZELRIGG and RUBIN 1985) in a chromosomal region homologous to an abundant embryonic RNA identified by reverse northern analysis. The transposon was found to be inserted approximately 30 kb proximal to the CNC (*cap'n'collar*) gene at 94E3-4 and 15 kb distal to the *hedgehog* (*hh*) locus at position −36 on the chromosome walk of MOHLER *et al.* (1991), and to map to position 3–79.2 on the genetic map, 0.1 cM distal to the *bar-3* allele of *hh* (3–79.1) and 0.3 cM proximal to *l(1)5G83* (Figure 2). The chromosome bearing this transposon insertion was homozygous lethal and the transposon insertion allele in the *unkempt* locus has been designated unk^{OR20}.

Expression of *unk* RNAs and isolation of cDNAs: Northern analysis of RNAs from the *unkempt* locus identified three prominent poly(A^+) RNA species expressed during embryogenesis of approximately 2.5, 4.1 and 6.5 kb in size (Figure 3). The *unkempt* RNAs are most abundant during the first 5 hr of embryogenesis, when the 2.5 and 4.1 are major RNA species and the 6.5 kb is a minor RNA species from this locus. After 5 hr of embryogenesis, the 2.5-kb RNA is

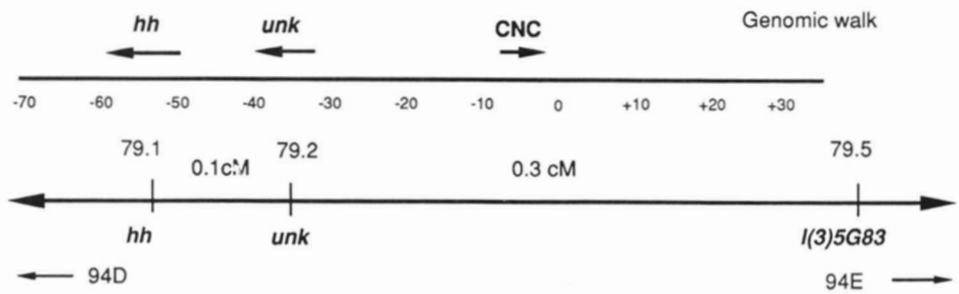

FIGURE 2.—Map of the 94E region. The top line shows the chromosome walk of MOHLER *et al.* (1991) with the positions of the genomic loci of the CNC gene (MOHLER *et al.* 1991), *unk* (this study) and *hh* (J. MOHLER and K. VANI, manuscript submitted). The bottom line shows the map position of *unk*, *hh* and *l(3)5G83* on the recombination map.

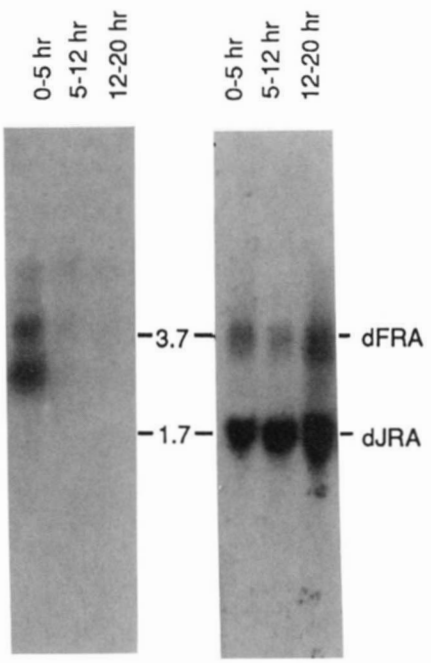

FIGURE 3.—Northern Analysis of embryonic RNA. RNAs from 0–5-, 5–12- and 12–20-hr-old embryos were analyzed for *unk* RNA. The filter was probed with the SM1 *unk* cDNA (left). The filter was reprobed with dJRA and dFRA cDNAs (right) as size and loading controls (dJRA is expressed continuously in embryogenesis and dFRA is expressed through 18 hr of development, PERKINS *et al.* 1990).

essentially missing, and the amount of the 4.1-kb RNA is reduced to about the same level as the 6.5-kb RNA, which is roughly constant in amount throughout embryogenesis.

Genomic DNA flanking the insertion site of the P(w,ry)A transposon in unk^{OR20} was used to screen two cDNA libraries. Thirty-five cDNA clones were isolated from the 0–4 hr cDNA library of M. NOLL (FIGERIO *et al.* 1986) and one from the 4–8-hr library of N. BROWN (BROWN and KAFATOS 1988). The largest three cDNAs from the Noll library (SM4 [2.2 kb], NW1 [2.4 kb] and NW29 [2.4 kb]) and the cDNA clone from the Brown library (15C [4.2 kb]) were analyzed further. The sizes of these cDNA clones correspond roughly to the sizes of the two major classes of RNA (4.1 and 2.5 kb) and probably reflect nearly full length cDNA clones. Mapping of these cDNAs onto the genomic region by hybridizaton and alignment of restriction sites yields the map of the transcription unit shown in Figure 4.

Sequence analysis of cDNA clones: Sequence analysis of the four cDNAs shows that the cDNA clones are colinear. Figure 5 shows the sequence of the largest cDNA, 15C, preceeded by upstream genomic sequence, and the alignment of the other three cDNAs. Three cDNA clones, NW1, NW29 and 15C have 5′ ends within 100 bp of each other; the fourth cDNA, SM4, appears to have a cloning artifact at the 5′ end with portions of the more 5′ sequences appended in reverse orientation upstream to nucleotide 498. Each of the four cDNAs has a unique poly(A^+) addition site: NW1, NW29 and SM4 are terminated in a 150-bp region approximately 2.5 kb from the 5′ end of the cDNA region, and 15C is terminated approximately 4.2 kb from the 5′ end. Thus, the structure of these cDNAs implies that the 2.5- and 4.1-kb RNAs from which they are derived encode an identical protein product and differ primarily in the length of the 3′ trailer region. This sequence encodes a predicted 614 amino acid protein starting from a methionine codon at nucleotide −34 to a stop codon at nucleotide 1809. A second possible start codon is located at nucleotide 15 (within the known transcribed region, as judged from the cloned cDNAs), from which a predicted 599 amino acid product would be generated. Regardless of the correct start position, this predicted protein sequence can be divided into three structural domains: 1) an N-terminal domain with five Cys_3His repeats, 2) a long C-terminal α-helical domain, and 3) an intervening region consisting of 22% serine.

The five Cys_3His repeats are shown aligned in Figure 6. The repeats are approximately 40 amino acids long and contain 11 positions with identical or similar (D = E, F = Y, T = S, K = R) amino acids present in at least four of the five repeats. The consensus for this repeat, **YK** $^{T}/_{S}$ xx**C** x_{4-12}**G**x**C**xx**G**xx**C**x $^{F}/_{Y}$ x**H**xxx $^{E}/_{D}$ x_{7-9}, is remarkable for its alignment of cysteine, histidine and glycine within the repeat. A search of

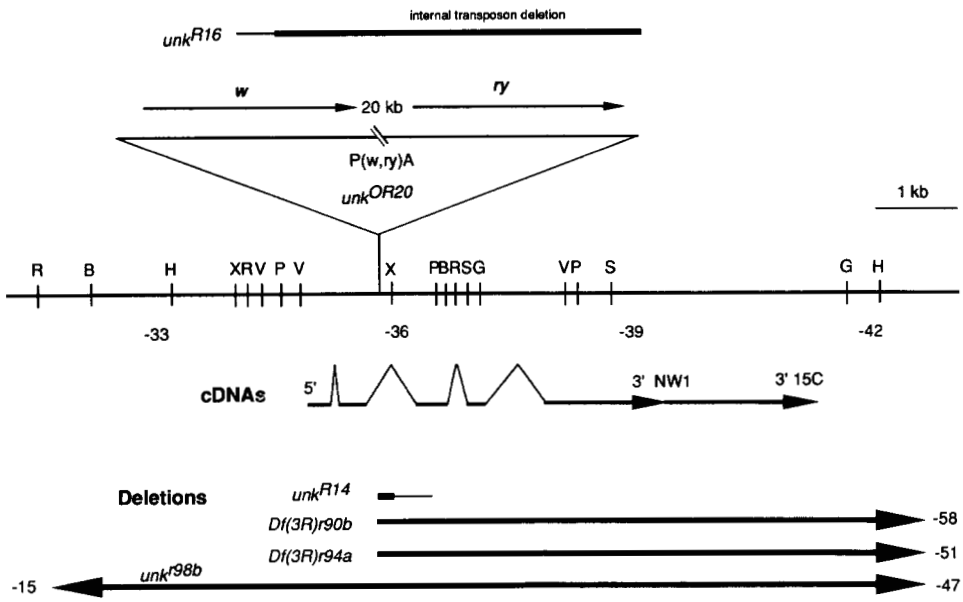

FIGURE 4.—Restriction map of the K15 genomic clone containing the *unk* gene. The transcript map of *unk* was determined by hybridization of cDNAs to restriction digests of K15 and the alignment of restriction sites. The positions of first two exons were further aligned by direct sequence comparison of the cDNAs with the genomic sequence of the pXV1 subclone. The position of the P(w,ry)A insertion was mapped by analysis of altered restriction fragments determined by southern analysis to genomic DNA. [The original published position of −31 (MOHLER *et al.* 1991) was incorrectly determined due to confusion over rearranged restriction fragments in a subclone generated during the original chromosome walk.] Chromosomal DNA deleted in imprecise revertants of unk^{OR20} is indicated by thick lines for each mutant; limits of breakpoint uncertainty is indicated by thin lines. (Restriction sites: B = *Bam*HI, G = *Bgl*II, H = *Hin*dIII, *P* = *Pst*I, R = *Eco*RI, S = *Sal*I, V = *Eco*RV, X = *Xho*I.)

the PROT-SWISS database (R15.0) resulted in the identification of two other proteins containing these repeats: cMG1 (GOMPERTS, PASCALL and BROWN 1990) and TIS11/Nup475/TTP (VARNUM *et al.* 1989; DUBOIS *et al.* 1990; LAI, STUMPO and BLACKSHEAR 1990). Both of these proteins are rapidly coinduced with c-*fos* in mammalian fibroblasts by growth factors, insulin or other mitogens. These proteins both contain two Cys_3His repeats and are highly related in the region of these repeats (72% identity over a 67 amino acid region, GOMPERTS, PASCALL and BROWN 1990). Each repeat of CMG1 and TIS11 contains all of the conserved amino acids of the repeat consensus derived for the *unk* protein repeats. Although the function of these repeats is not known, TIS11/Nup475/TTP has been shown to be a zinc-binding nuclear protein (DUBOIS *et al.* 1990), suggesting that the Cys_3His structure may bind with a metal ion in a manner analogous to zinc-finger structures.

N-terminal to the first Cys_3His repeat is a potential nuclear targeting signal, namely a string of basic amino acids flanked by helix breakers (GERACE and BURKE 1988): **RRRRPVRKRDG**, amino acids (aa) 72–82. However, because similar domains can be found in a number of non-nuclear proteins (*e.g.*, the extracellular matrix retention domain of PDGF-B, LA ROCHELLE *et al.* 1991), the existence of a potential nuclear targeting signal in the *unk* protein is not indicative of a nuclear localization of this protein. C-terminal to the Cys_3His fingers of *unkempt* is a serine-rich domain (aa 315–474), consisting of 22% serine (35 out of 159 aa). Similar serine-rich domains are also present C-terminal to the Cys_3His fingers in TIS11/Nup475/TTP and cMG1 proteins (TIS11/Nup475/TTP: 33 serine out 157 amino acids; cMG1: 31 serine out of 157 amino acids).

Theoretical prediction (CHOU and FASMAN 1978; ROSE 1978) of the secondary structure of *unkempt* protein suggests that the C-terminal third (from aa 474–575) consists primarily of a long α-helical domain, with a single predicted turn around aa 494. This α-helical domain is unique to *unkempt*, and is not found in the TIS11/Nup475/TTP or cMG1 Cys_3His finger proteins. This helix contains two regions of strong amphipathicity: aa 474–504 and aa 542–564 (Figure 7, a and b), in which one face of the helix is hydrophobic and another face contains charged residues. Such amphipathic helices have been shown to have a role in formation of protein-protein dimers in the case of myosin (MCLACHLAN and KARN 1983) and a variety of transcription factors (MURRE *et al.*, 1989; O'SHEA *et al.* 1989; VINSON, SIGLER and MCKNIGHT 1989).

Isolation of *unkempt* mutations: To determine whether the lethality of the chromosome bearing the w^+ in the *unk* (unk^{OR20}) was due to the insertion at 94E, we induced loss of the w^+ activity of the P(w,ry)A insertion by hybrid dysgenesis. Of the 38 w^- revertants of unk^{OR20} isolated in the first screen, 33 complemented the lethality of unk^{OR20}. Southern analysis of

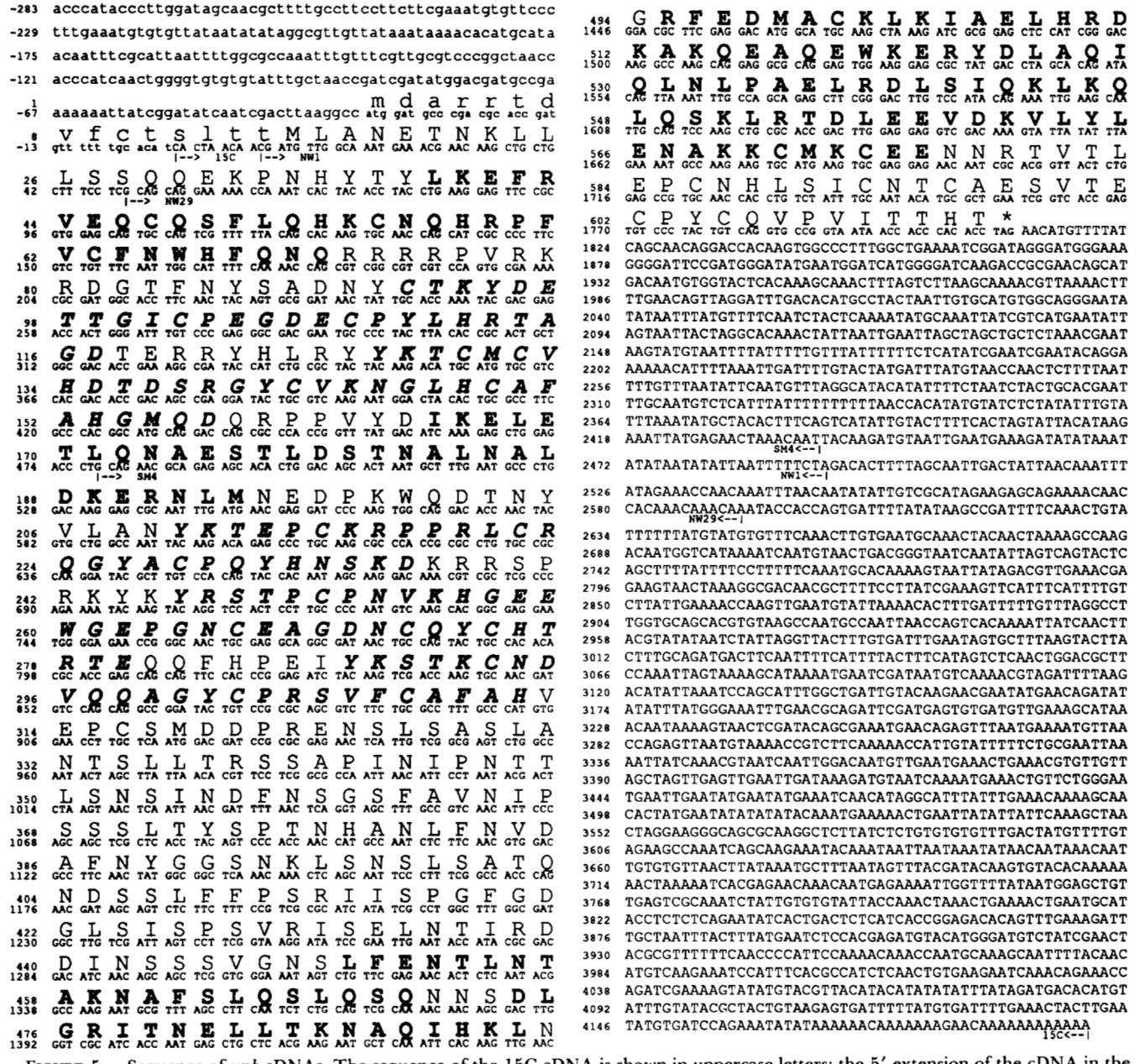

FIGURE 5.—Sequence of *unk* cDNAs. The sequence of the 15C cDNA is shown in uppercase letters; the 5′ extension of the cDNA in the genomic DNA is shown in lowercase. The DNA sequence is numbered starting with +1 from the beginning of the 15C cDNA. The ends of other sequenced cDNAs aligned with the 15C cDNA are indicated below the sequence. The predicted protein sequence is shown above the DNA sequence, from two possible methionine initiation codons. The protein sequence beginning from the methionine codon within the cDNA is shown in uppercase, the protein sequence from the preceding methionine codon in the genomic sequence is shown in lowercase. The protein sequence is numbered starting from the first methionine codon as +1. Amino acids of the Cys_3His fingers are indicated by boldface italics. Predicted regions of α-helix longer than seven amino acids are highlighted in boldface. The DNA sequence has been submitted to the EMBL/GenBank database (accession number Z11527).

the DNA of several of these complementing revertants showed the complete loss of the P(w,ry)A insertion, confirming that the insertion at the 94E site is the cause of the lethal phenotype. Four of the noncomplementing alleles, unk^{R11}, unk^{R14}, unk^{R23} and unk^{R31}, failed to complement unk^{OR20} for adult viability. The fifth noncomplementing allele, unk^{R16}, complemented unk^{OR20} for viability, but displayed with incomplete penetrance a variety of adult cuticular defects including roughened eyes, splayed wings and crossed scutellar bristles (the "unkempt" phenotype) when heterozygous with unk^{OR20}. In contrast, the unk^{R16}/unk^{R14} heterozygotes were not viable, suggesting that the unk^{R14} revertant may be a stronger allele of *unk* than the unk^{OR20} insertion. Comparison of the DNAs of three derivatives, unk^{R14}, unk^{R16} and unk^{R31}, by Southern analysis with wild-type and unk^{OR20} DNA indicated that these three alleles were due to imprecise excision of the P(w,ry)A transposon at the 94E locus. (The two other noncomplementing derivatives, unk^{R11}

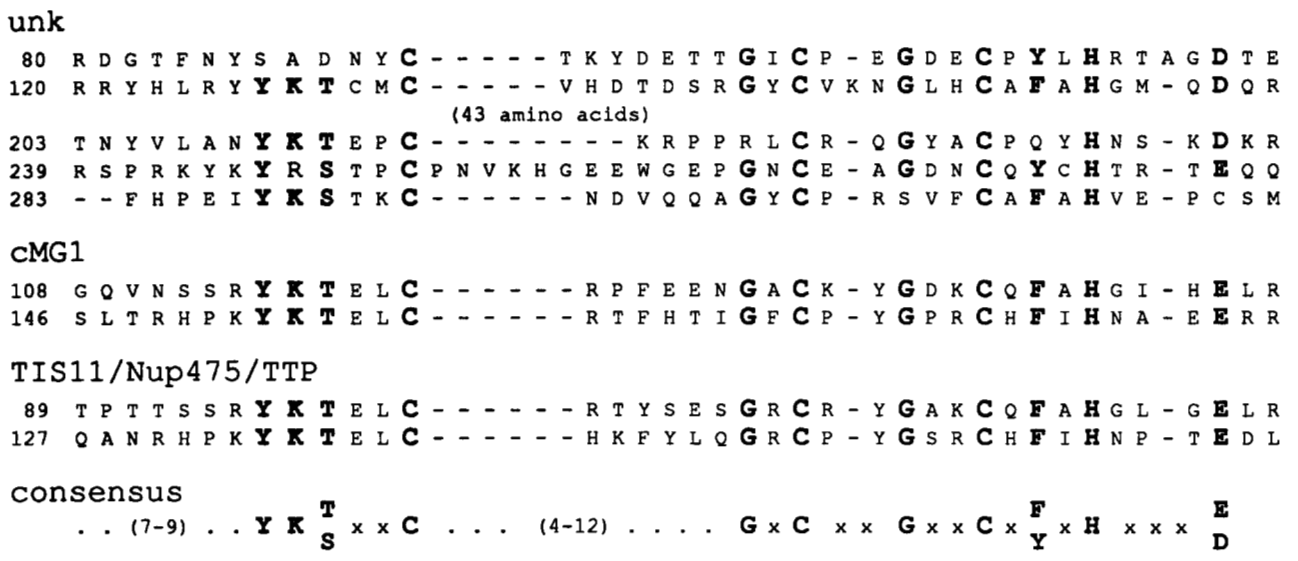

FIGURE 6.—Alignment of the Cys_3His repeats. The five Cys_3His repeats are shown aligned and compared with the aligned Cys_3His motifs from the cMG1 (GOMPERTS, PASCALL and BROWN 1990) and the TIS11/Nup475/TTP (numbered as from LAI, STUMPO and BLACKSHEAR 1990). The consensus sequence of the *unk* repeats is shown below.

and unk^{R23}, were lost before they could be completely analyzed.) The unk^{R16} revertant appeared to be associated with internal deletion of the P(w,ry)A element (Figure 4). The unk^{R31} revertant retained both *Hin*dIII sites associated with the ends of the P(w,ry)A element and is, therefore, presumed to also be an internal alteration of the P(w,ry)A transposon. In contrast, the unk^{R14} derivative removed flanking genomic DNA, including an *Xho*I site, as well as part of the P(w,ry)A element (Figure 4).

An additional screen for hybrid-dysgenesis induced P(w,ry)A transposon loss was conducted to screen for deficiencies extending from *unk* to either side, possibly including the two flanking loci, *hh* and CNC. In these screens, a total of 82 w^- derivatives of unk^{OR20} which retained the *unk* lethality were recovered and assayed for *hh* activity and for the presence of an *Eco*RI polymorphic site (differing between the unk^{OR20} and *TM3* chromosomes) located at position −18 between *unk* and CNC in the chromosome walk of MOHLER *et al.* (1991). In general, the nature of the alterations at the *unkempt* locus were not determined. Two w^- derivatives, *Df(3R)r90b* and *Df(3R)r94a*, failed to complement *hh* mutations and possess deficiency breakpoints at the P(w,ry)A insertion site and within the *hh* locus, at about −51 and −58 (Figure 4; J. MOHLER and K. VANI, manuscript submitted). Two other derivatives, unk^{r93b} and unk^{r98b}, were missing the *Eco*RI restriction fragment characteristic of the unk^{OR20}-linked alleles of the −18 polymorphism. However, unk^{r93b} retains another unk^{OR20}-linked polymorphism at position −22 and appears to have an altered set of *Eco*RI restriction fragments at the −18 position; this allele appears to contain a complex rearrangement of DNA around the *unk* locus. unk^{r98b} is missing both polymorphic sites at positions −22 and −18, as well as an additional polymorphic site at −44. Further analysis of the unk^{r98b} chromosome has shown that this mutation is a deletion from approximately −47 to −15 (Figure 4). The effects of the proximal breakpoints of unk^{r98b}, *Df(3R)r90b* and *Df(3R)r94a* on expression of the *hedgehog* locus are discussed elsewhere (J. MOHLER and K. VANI, manuscript submitted).

One additional, radiation-induced allele, unk^{CHS20}, was independently obtained by failure to complement unk^{r93b} for viability. The independent origin of the unk^{CHS20} allele demonstrates that the observed lethality is the effect of a disruption of the *unkempt* gene and not peculiar to derivatives of the unk^{OR20} chromosome.

***unkempt* embryonic spatial expression patterns:** *In situ* hybridization of cDNA clones from this region to embryos revealed ubiquitous expression of this RNA starting in early preblastoderm stages, when staining is predominantly localized to the cytoplasmic islands (Figure 8a). During blastoderm formation, this staining segregates with the cytoplasm to the periphery of the embryo (Figure 8b). This early preblastoderm staining suggests a maternal contribution of this RNA to the embryo. Upon gastrulation, an increase in RNA level can be observed in the invaginating tissues (the ventral and cephalic folds and both the posterior and anterior mid-gut invaginations) in comparison to noninvaginating tissues (Figure 8c). This increased RNA level is present in mesodermal cells during the early extended germband-stage embryos (Figure 8d). During the extended germband stage, this difference in staining between the mesoderm and

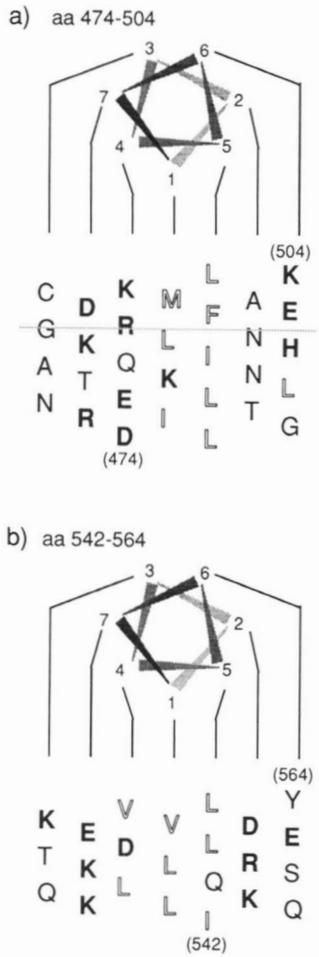

FIGURE 7.—Two regions of amphipathic α-helix. The α-helices are shown as viewed from the surface, split between amino acids 3 and 6 of the helical repeat. Hydrophobic amino acids are shown in open face, charged amino acids in bold. (a) Domain aa 474–504, the stippled line indicates the site of the predicted turn; (b) domain aa 542–564.

nonmesodermal tissues disappears, and by the end of germband retraction (early stage 13) RNA levels appear to be uniform throughout the cellular portions of the embryo (Figure 8e). Starting in stage 13, RNA levels gradually become elevated in the neural tissues. After stage 15, the level of RNA in non-neural tissue begins to decrease and by early stage 17, *unkempt* RNA is principally found only in the tissues of the central nervous system (Figures 8f). No RNA expression is detectable after mid-stage 17.

The preblastoderm and blastoderm RNA staining is present in all embryos from parents heterozygous for DNA deletions at the *unkempt* locus (*Df(3R)r94a* and unk^{r98b}), which remove sequences corresponding to the SM1 cDNA hybridization probe. This suggests there is a significant source of maternal *unkempt* RNA in the embryo. In a quarter of these embryos, those presumably homozygous for the *unk* deletions, RNA levels begin to decrease soon after gastrulation. In these embryos there is no apparent increase in RNA staining in invaginating mesodermal cells, although in some embryos staining decays more slowly in the mesoderm associated with the two midgut invaginations than in surrounding tissues (Figure 8g). This RNA appears to have disappeared by stage 15 in all the deficiency embryos (Figure 8h).

Lethal effects of *unk* mutations: Lethal alleles of *unkempt* were examined for their effects on embryonic development and larval growth. These studies were complicated by the *TM3* balancer chromosome originally used to maintain these stocks; this variant *TM3* chromosome (which we designated *TM3, daeh*) had a recessive embryonic lethality, in which embryos were blocked in head involution (MOHLER and VASILAKIS 1992).

Embryos homozygous for most alleles of *unkempt* (unk^{OR20}, unk^{R14}, unk^{R31}, unk^{r98b} and unk^{CHS20}) survive embryogenesis to hatch. Those *unk* mutants that are also deficient for *hedgehog* only show the expected *hedgehog*-associated phenotypes in homozygous embryos (PERRIMON and MAHOWALD 1987; MOHLER 1988; J. MOHLER and K. VANI, manuscript submitted) both in cuticle preparations and in stage 12 of embryogenesis. Only homozygotes of unk^{r93b} had a distinct embryonic lethal phenotype (hypertrophy of the ectoderm during stage 12), probably due to an additional mutation on the unk^{r93b} chromosome. In addition, examination of the nervous system of stage 16 and 17 homozygous unk^{R14} embryos following α-HRP staining (JAN and JAN 1982) revealed no gross abnormalities, consistent with their apparent embryonic viability. Because this set of embryonic viable *unk* alleles includes the unk^{r98b} mutation, which is deficient for the entire *unkempt* locus (as described above), there appears to be no absolute zygotic requirement for *unkempt* for embryonic viability.

Larvae *trans*-heterozygous for any two *unk* lethal alleles (or homozygous for unk^{r98b}, unk^{R14}, or unk^{OR20}) develop into mature, wandering third-instar larva and usually pupariate. Examination of imaginal structures in *unk* mutant larvae revealed no gross morphological abnormalities; in particular, imaginal discs were of normal size and structure. Because growth of the imaginal discs is usually extremely sensitive to lethal mutations in genes essential for general mitotic functions (GATTI and BAKER 1989), *unkempt* function is probably not required for normal cell growth and division. However, development of homozygous and *trans*-heterozygous *unk* mutant larva is slower than their heterozygous wild-type sibs. At 25°, homozygous unk^{r98b} (which is completely deficient for *unk*), homozygous unk^{R14} and *trans*-heterozygous unk^{r98b}/unk^{R14} mutant larvae reach the third larval instar approximately one day later than their heterozygous sibs, and the first of these *unk* larvae pupariate three days after their heterozyous sibs (approximately 8 days post-egglay). The developmental delay to pupariation

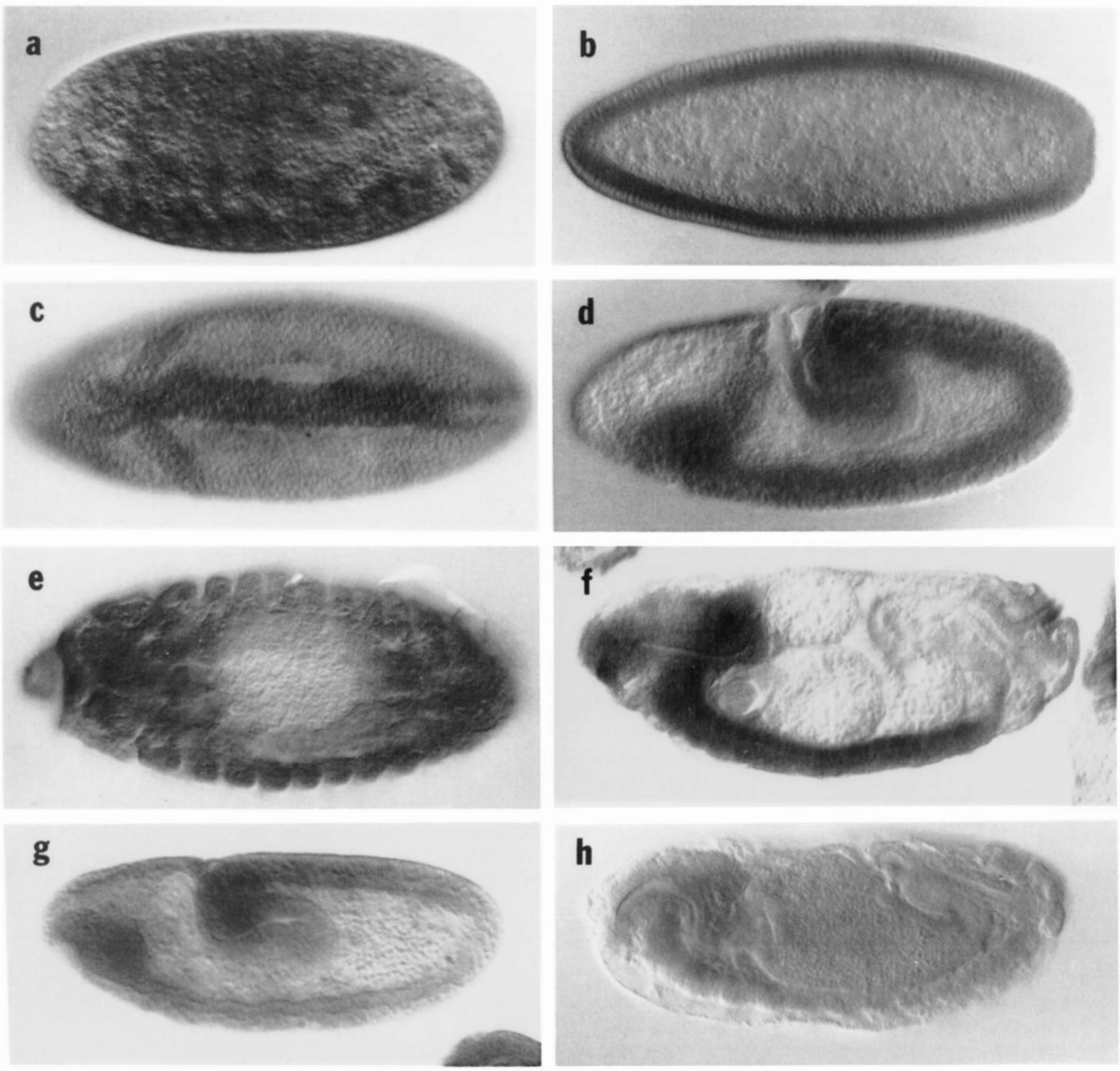

FIGURE 8.—*In situ* hybridization of *unk* cDNA to embryos. (a-f) Wild-type embryos. (a) Late stage 2 embryo; (b) early stage 5 blastoderm (cycle 14); (c) early gastrulation (stage 8), ventral view; (d) stage 9 extended germband, lateral view; (e) late stage 13, ventral view; (f) late stage 16, lateral view. Pharyngeal and posterior spiracle staining disappears before stage 17. (g and h) Homozygous unk^{98b} embryos. (g) Stage 9 lateral view; (h) stage 15, lateral view.

is shorter in the other mutant combinations; *trans*-heterozygous unk^{CHS20}/unk^{r93b} mutant larvae begin puparation at approximately the same time as their heterozygous sibs, and other heteroallelic combinations of *unk* alleles show intermediate delays in pupariation. In addition, most (>95%) developmentally delayed unk^{R14} and unk^{R14}/unk^{r98b} mutant larvae possessed one or more amorphous melanotic tumors at various locations in the abdomen; however, such melanotic tumors were not observed in larvae mutant for most other *unk* alleles.

Homozygous unk^{r98b}, homozygous unk^{R14} and *trans*-heterozygous unk^{r98b}/unk^{R14} individuals stop development during the mid-pupal phase. Because unk^{r98b} is a deficiency for this locus, this mid-pupal block represents the effect of a null mutation in this locus. *trans*-Heterozygous individuals for unk^{r93b} or unk^{CHS20} and other lethal *unk* alleles continue development to become mature pharate adults whose cuticle tans within the pupal case. Like the leaky unk^{R16} viable allele, these *unk* pharate adults possess an "unkempt" phenotype (small rough eyes, held-out wings and crossed scutellar bristles).

DISCUSSION

Sequence analysis of the *unkempt* gene of Drosophila reveals it is a divergent member of the Cys_3His class of zinc-finger proteins. Comparison of the Cys_3His fingers of *unk* with those of the Cys_3His serum response proteins suggests a generalized consensus for the Cys_3His finger: (**YK** $^{T}/_{S}$ xx**C** x_{4-12}**G**x**C**xx**G**xx**C**x $^{F}/_{Y}$ x**H**xxx $^{E}/_{D}$ x_{7-9}). *unkempt* encodes an abundant ubiquitous embryonic RNA, which is present in most embryonic tissues during the initial half of embryogenesis and predominantly expressed in the central nervous system in later stages. There appears to be no obvious zygotic requirement for *unkempt* in those tissues during embryogenesis, embryos deficient for *unkempt* survive through hatching, although *unkempt* gene function is required post-embryonically for late larval/early pupal development.

Assuming that the presence of abundant *unkempt* RNAs indicates the *unkempt* gene product acts in some processes of embryogenesis, our failure to detect a requirement for the *unk* gene during embryogenesis probably reflects in part the maternal contribution of *unkempt* RNA levels. Because *unkempt* RNAs are detected most predominantly at early stages of embryogenesis, at high levels prior to blastoderm formation in both wild-type embryos and those homozygous for *unkempt* deficiencies, much of the early embryonic RNA is of maternal origin. Although much of this RNA disappears during the extended-germband stage, replaced by zygotic RNAs first in the invaginating mesoderm and later in the central nervous system, the maternal levels of *unkempt* RNA are likely to be sufficient for at least early zygotic requirements for *unk* activity. Perdurance of the *unkempt* protein product may account for the lack of a phenotype associated with the loss of later embryonic expression. Alternatively, there may be another gene expressed zygotically during late embryogenesis that is functionally redundant for *unk* function. Although we could not detect a similar gene to *unkempt* by low stringency hybridization (J. MOHLER and S. MOHAMMADI, unpublished results), it is not clear how similar such a gene would have to be to provide such a redundant function.

In considering the nature of the function provided by *unkempt* protein, it is tempting to speculate that the Cys_3His fingers may indicate a possible role in transcriptional regulation, based on their limited similarity to Cys_2His_2 and Cys_4 zinc-finger proteins (see review by EVANS and HOLLENBERG 1988). DUBOIS and coworkers (1990) have investigated the TIS11/Nup475/TTP Cys_3His protein for its similarity to other zinc-finger proteins and have shown that the TIS11/Nup475/TTP protein is a nuclear antigen and binds zinc. However, there is no published evidence that either the TIS11/Nup475/TTP protein or the closely related cMG1 protein is capable of binding DNA, so a role of the Cys_3His motif in transcription regulation is purely speculative as yet.

Although the TIS11/Nup475/TTP protein has been shown to be a nuclear antigen, it is not obvious that the function of this Cys_3His repeat is restricted only to the nuclear compartment of the cell. Thus, it is relevant to consider other roles for this motif that might occur in both the nuclear and cytoplasmic compartment. Based on the two obvious structural features of the *unkempt* protein, the Cys_3His fingers and the long α-helical domain, we can hypothesize two speculative cytoplasmic roles for the *unkempt* protein based on each of these features, although numerous other possibilities abound. One hypothetical cytoplasmic role for the *unkempt* protein might be as an RNA-binding protein mediated by the Cys_3His fingers. The zinc finger motif has been implicated in the function of three RNA-binding proteins: TFIIIA, p43, and eIF2β. TFIIIA is a transcription factor with nine Cys_2His_2 fingers that is required for the synthesis of 5SRNA in Xenopus oocytes, and its activity is blocked by its binding to the 5S product sequestered in the cytoplasm (HONDA and ROEDER 1980; PELHAM and BROWN 1980; BROWN, SANDER and ARGOS 1985); p43 is an accessory product that also binds 5SRNA in Xenopus oocytes (JAHO *et al.* 1990). eIF2β is a eucaryotic translation initiation factor with a single Cys_4 finger that may have a role in binding to either mRNA or rRNA (PATHAK *et al.*, 1988). The structural analogies between these fingers and the Cys_3His fingers of *unk* could equally indicate a possible RNA-binding function or a role involving DNA binding.

An example of another speculative cytoplasmic function based instead on the long α-helical domain of the *unkempt* protein is that this protein might function as a modulator of cytoskeletal fiber assembly. Several cytoskeletal elements [notably myosin (MCLACHLAN and KARN 1983) and intermediate filaments (STEINERT 1978)] assemble via a long amphipathic α-helical domain. The long amphipathic α-helical domain of *unk* could function as a competitive inhibitor for multimerization of the cytoskeletal fiber, and, in such a case, the Cys_3His domains might function to provide metal ion-mediated modulation of the proposed association between *unk* protein and the fiber-monomer. Such an activity of the *unk* protein is in keeping with its expression in tissues with cytoplasmic myosin activity (gastrulating tissues and differentiating axonal fibers of the central nervous system; YOUNG, PESACRETA and KIEHART 1991).

At the moment, there is no reason to presuppose any defined physiological function for the *unkempt* gene. Clearly, we need to know more about the intracellular localization of the *unkempt* protein, the nature of the biochemical activities of the protein, and

whether it can bind other proteins or nucleic acid, before we can determine a possible role of *unkempt* in the morphogenesis of the Drosophila embryos.

The authors wish to thank the 1989 class of Laboratory in Molecular Biology and 1990 class of the Project Laboratory in Molecular Biology (especially ROMY ANTOINE and ELIZABETH HALO) for their assistance in mapping the *unk* region and screening for RFLP deletions. We also wish to thank M. NOLL, N. BROWN, V. PIRROTTA and P. SCHEDL for DNA libraries and thank N. CHU, D. KALDERON and S. MOUNT for reading drafts of the manuscript. This work was funded by U.S. Public Health Service grant R29HD22751.

LITERATURE CITED

AUSUBEL, F. M., R. BRENT, R. E. KINGSTON, D. D. MOORE, J. G. SEIDMAN, J. A. SMITH and K. STRUHL, 1987 *Current Protocols in Molecular Biology*. Wiley, New York.

BIER, E., H. VAESSIN, S. SHEPHARD, K. LEE, K. MCCALL, S. BARBEL, L. ACKERMAN, R. CARRETTO, T. UEMURA, E. GREE, L. Y. JAN and Y. N. JAN, 1989 Searching for pattern and mutation in the Drosophila genome with a P-lacZ vector. Genes. Dev. **3:** 1273–1287.

BROWN, N. H., and F. C. KAFATOS, 1988 Functional cDNA libraries from Drosophila embryos. J. Mol. Biol. **203:** 425–437.

BROWN, R. S., C. SANDER and P. ARGOS, 1985 The primary structure of transcription factor TFIIIA has 12 consecutive repeats. FEBS Lett. **186:** 271–274.

CAMPOS-ORTEGA, J. A., and V. HARTENSTEIN, 1985 *The Embryonic Development of Drosophila melanogaster*. Spring-Verlag, Berlin.

CHOU, P. Y., and G. D. FASMAN, 1978 Empirical predictions of protein conformation. Annu. Rev. Biochem. **47:** 251- 276.

DUBOIS, R. N., M. W. MCLANE, K. RYDER, L. F. LAU and D. NATHANS, 1990 A growth factor-inducible nuclear protein with a novel cysteine-histidine repetitive sequence. J. Biol. Chem. **265:** 19185–19191.

EVANS, R. M., 1988 The steroid and thyroid hormone receptor superfamily. Science **240:** 889–894.

EVANS, R. M., and S. M. HOLLENBERG, 1988 Zinc fingers: Gilt by association. Cell **52:** 1–3.

FRIGERIO, G., M. BURRI, D. BOPP, S. BAUMGARTNER and M. NOLL, 1988 Structure of segmentation gene *paired* and *Drosophila* PRD gene set as part of a gene network. Cell **47:** 735–746.

GATTI, M., and B. S. BAKER, 1989 Genes controlling essential cell-cycle functions in *Drosophila melanogaster*. Genes Dev. **3:** 438–453.

GERACE, L., and B. BURKE, 1988 Functional organization of the nuclear envelope. Annu. Rev. Cell Biol. **4:** 335–374.

GOMPERTS, M., J. C. PASCALL and K. D. BROWN, 1990 The nucleotide sequence of a cDNA encoding an EGF-inducible gene indicates the existence of a new family of mitogen-induced genes. Oncogene **5:** 1081–1083.

HARRISON, S. C., 1991 A structural taxonomy of DNA-binding domains. Nature **353:** 715–720.

HARTSHORNE, T. A., H. BLUMBERG and E. T. YOUNG, 1986 Sequence homology of the yeast regulatory protein ADR1 with Xenopus transcription factor TFIIIA. Nature **320:** 283–286.

HAZELRIGG, T., R. LEVIS and G. RUBIN, 1984 Transformation of *white* locus DNA in Drosophila: dosage compensation, *zeste* interaction and position effects. Cell **36:** 469–481.

HENIKOFF, S., 1984 Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28:** 351–359.

HERSCHMAN, H. R., 1991 Primary response genes induced by growth factors and tumor promotors. Annu. Rev. Biochem. **60:** 281–320.

HONDA, B. M., and R. G. ROEDER, 1980 Association of a 5S gene transcription factor with 5S RNA and altered levels of the factor during cell differentiation. Cell **22:** 119–126.

JAN, L. Y., and Y. N. JAN, 1982 Antibodies to horseradish peroxidase as specific neural markers in Drosophila and grasshopper embryos. Proc. Natl. Acad. Sci. USA **79:** 2700–2704.

JOHO, K. D., M. K. DARBY, E. T. CRAWFORD and D. D. BROWN, 1990 A finger protein structurally similar to TFIIIA that binds exclusively to 5S RNA in Xenopus. Cell **61:** 293–300.

KIDWELL, M., 1986 P-element mutagenesis, in *Drosophila, a Partical Approach*, edited by D. B. ROBERTS. IRL Press, Oxford.

LAI, W. S., D. J. STUMPO and P. J. BLACKSHEAR, 1990 Rapid insulin-stimulated accumulation of an mRNA encoding a proline-rich protein. J. Biol. Chem. **265:** 16556–16563.

LA ROCHELLE, W. J., M. MAY-SIROFF, K. C. ROBBINS and S. A. AARONSON, 1991 A novel mechanism regulating growth factor association with the cell surface: identification of a PDGF retention domain. Genes Dev. **5:** 1191–1199.

LEVIS, R., T. HAZELRIGG and G. M. RUBIN, 1985 Effects of genomic position on the expression of transduced copies of the *white* gene of *Drosophila*. Science **229:** 558–561.

MCLACHLAN, A. D., and J. KARN, 1983 Periodic features in the amino acid sequence of nematode myosin rod. J. Mol. Biol. **164:** 605.

MOHLER, J., 1988 Requirements for *hedgehog*, a segmental polarity gene, in patterning larval and adult cuticle of Drosophila. Genetics **120:** 1061–1072.

MOHLER, J., and G. VASILAKIS, 1992 A block in early head involution in *deah* mutant embryos. Drosophila Inform. Serv. **71:** (in press).

MOHLER, J., K. VANI, S. LEUNG and A. EPSTEIN, 1991 Segmentally restricted, cephalic expression of a leucine zipper gene during Drosophila embryogenesis. Mech. Dev. **34:** 3–10.

MURRE, C., P. S. MCCAW, H. VAESSIN, M. CAUDY, L. Y. JAN, Y. N. JAN, C. V. CABRERA, J. N. BUSKIN, S. D. HAUSCHKA, A. B. LASSAR, H. WEINTRAUB and D. BALTIMORE, 1989 Interaction between heterologous helix-loop-helix proteins generates complexes that bind specifically to a common DNA sequence. Cell **58:** 537–544.

O'SHEA, E. K., R. RUTKOWSKI, W. F. STAFFORD III and R. S. KIM, 1989 Preferential heterodimer formation by isolated leucine zippers from fos and jun. Science **245:** 646–648.

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.

PATHAK, V. K., P. J. NIELSEN, H. TRACHSEL and J. W. B. HERSHEY, 1988 Structure of the β subunit of translational initiation factor eIF-2. Cell **54:** 633–639.

PAVLETICH, N. P., and C. O. PABO, 1991 Zinc finger–DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. Science **252:** 809–817.

PELHAM, H. R. B., and D. D. BROWN, 1980 A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. Proc. Natl. Acad. Sci. USA **77:** 4170–4174.

PERKINS, K. K., A. ADMON, N. PATEL and R. TJIAN, 1990 The Drosophila Fos-related AP-1 protein is a developmentally regulated transcription factor. Genes Dev. **4:** 822–834.

PERRIMON, N. and A. P. MAHOWALD, 1987 Multiple functions of segment polarity genes in Drosophila. Dev. Biol. **119:** 587–600.

ROSENBERG, U. B., C. SCHRÖDER, A. PREISS, A. KIENLIN, S. COTE, I. RIEDE and JÄCKLE, H., 1986 Structural homology of the product of the Drosophila *Krüppel* gene with Xenopus transcription factor IIIA. Nature **319:** 336–339.

ROBERTSON, H. M., C. R. PRESTON, R. W. PHILLIS, D. JOHNSON-SCHLITZ, W. K. BENZ and W. R ENGELS, 1988 A stable

genomic source of P element transposase in *Drosophila melanogaster*. Genetics **118:** 461–470.

ROSE, G. D., 1978 Prediction of chain turns in globular proteins on a hydrophobic basis. Nature **272:** 586–590.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

STEINERT, P. M., 1978 Structure of the three-chain unit of the bovine brain epidermal keratin filament. J. Mol. Biol. **123:** 49–70.

TAUTZ, D., and C. PFEIFLE, 1989 A non-radioactive *in situ* hybridization method for the localization of specific RNAs. Chromosoma **98:** 81–85.

VARNUM, B. C., R. W. LIM, V. P. SUKHATME and H. R. HERSCHMAN, 1989 Nucleotide sequence of cDNA encoding TIS11, a message induced in Swiss 3T3 cells by the tumor promoter tetradecanoyl phorbol acetate. Oncogene **4:** 119–120.

VINSON, C. R., P. B. SIGLER and S. L. MCKNIGHT, 1989 Scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science **246:** 911–916.

YOUNG, P. E., T. C. PESACRETA and D. P. KIEHART, 1991 Dynamic changes in the distribution of cytoplasmic myosin during Drosophila embryogenesis. Development **111:** 1–14.

Communicating editor: T. SCHÜPBACH