HOMOG program. Family B is of interest because the mother (III-2) is diagnosed with chronic SMA and the child (IV-1) with acute SMA. Presumably, a disease allele from generation I is transmitted to both carrier parents in generation III, then inherited in double copy by individual IV-1. Consistent with this model, the affected child is homozygous for markers D5S39 and D5S78. The affected mother presumably has one copy of the disease allele in common with the affected child, and one heterologous copy inherited from her father. Thus, this family is consistent with allelic variation among chronic and acute cases of SMA. Alternatively, the affected mother could have transmitted a sporadic dominant mutation for SMA, though this is unusual in cases of childhood-onset SMA[11,12].

The nine acute and seven chronic SMA families in this and our previous study[1] provide strong evidence in favour of genetic homogeneity between the two disorders. The combination of linkage data and genetic homogeneity make it very likely that the acute form of SMA maps to chromosomal region 5q. One chronic family (Fig. 2, A) and one acute family (Fig. 2, B) seem to be unlinked to 5q markers, although family A might conceivably provide data in support of linkage when more informative markers are identified. It must be noted that SMA can be confused with clinically overlapping neurological disorders[2-6], and that the subdivision between acute and chronic SMA is arbitrary, with overlap betwen the groups[13]. It is not yet possible to distinguish whether, or to what extent, the data in Fig. 2 support genetic heterogeneity as opposed to disease misclassification among the SMA family set. These issues should be resolved as additional families are typed with markers from chromosome 5q, and disease phenotype is carefully scrutinized relative to genotype. The results caution against early expectations for comprehensive prenatal diagnosis. □

Received 14 March; accepted 8 May 1990.

1. Brzustowicz, L. M. *et al. Nature* **344**, 540-541 (1990).
2. Pearn, J. H. in *Advances in Neurology* (ed. Rowland, L. P.) 36, 121-130 (Raven, New York, 1982).
3. Emery, A. E. H., Davie, A. M., Holloway, S. & Skinner, R. *J. neurol. Sci.* **30**, 375-394 (1976).
4. Munsat, T. L., Woods, R., Foeler, W. & Pearson, C. M. *Brain* **92**, 9-24 (1969).
5. Dubowitz, V. *Muscle Disorders in Childhood* 146-178 (Saunders, London and Philadelphia, 1978).
6. Brooke, M. H. in *A Clinicians View of Neuromuscular Diseases* 2nd edn 36-80 (Williams and Wilkins, London, 1985).
7. Pearn, J. H. *J. med. Genet.* **10**, 260-265 (1973).
8. Smith, C. A. B. *J.R. Stat. Soc. B* **15**, 153 (1953).
9. Lander, E. S. & Botstein, D. *Science* **236**, 1567-1570 (1987).
10. Ott, J. *Analysis on Human Genetic Linkage* (Johns Hopkins University Press, Baltimore 1985).
11. Pearn, J. *J. med. Genet.* **15**, 409-413 (1978).
12. Winsor, E. J., Murphy, E. G., Thompson, M. W. & Reed, T. E. *J. med. Genet.* **8**, 143-148 (1971).
13. Dubowitz, V. *Colour Atlas of Muscle Disorders in Childhood* 66-86 (Year Book Medical Publishers, Chicago, 1989).
14. Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3443-3446 (1984).
15. Keats, B., Ott, J., & Conneally, M. *Cytogenet. Cell Genet.* **51**, 459-502 (1989).
16. Leppert, M. *et al. Science* **238**, 1411-1413 (1987).
17. Gilliam, T. C. *et al. Genomics* **5**, 940-944 (1989).

ACKNOWLEDGEMENTS. We thank the SMA families for their cooperation and support throughout this project. We also thank Drs I. Hausmanowa-Petrusewicz, F. Ziter, M. Leppert and K. Zerres for meticulous diagnostic information and the coordination of blood collection; Dr. R. Regev, Dr. N. Thomas, L. Skerry and D. Hooker for help with the collection of families; and H. Klein, D. Bomze and Dr K. Das for help with tissue culture. We thank Drs J. Gusella and T. Carbone, and W. Hobbes and H. MacFarlane for their help in the early days of this project. This work was supported by the W. M. Keck Foundation, the Muscular Dystrophy Association of America, the Muscular Dystrophy Group of Great Britain and Northern Ireland and the MRC of Great Britain.

Putative protein kinase product of the *Drosophila* segment-polarity gene *zeste-white3*

Esther Siegfried, Lizabeth A. Perkins*, Theresa M. Capaci* & Norbert Perrimon*

Department of Genetics and Howard Hughes Medical Institute*, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA

THE metameric pattern of the *Drosophila* embryo is regulated by a combination of maternal and zygotic genes. The segment-polarity class of genes are required for the correct patterning within each segmental unit. Mutations in any one of these genes results in deletions and duplications of parts of each segment[1]. The segment-polarity genes act coordinately by means of local cellular interactions to assign and maintain an identity for each cell in the segment, and to establish segment boundaries[2-5]. Here we describe the molecular characterization of a novel segment-polarity gene, *zeste-white3 (zw3)*. Embryos derived from germ lines that are homozygous for *zw3* mutations (*zw3* embryos) have phenotypes similar to embryos that are mutant for the segment-polarity gene *naked* (*nkd*)[6]. These embryos lack most of the ventral denticles, which are differentiated structures derived from the most anterior region of each segment. We have isolated the *zw3* gene and compared the structure of one maternal and one zygotic transcript encoded by the gene. The *zw3* gene is unique among the segment-polarity genes so far characterized, in that it encodes proteins that have homology to serine–threonine protein kinases. This indicates that *zw3* may play a part in a signal transduction pathway involved in the establishment of cell identity within each embryonic segment.

The *zw3* locus is located at chromosomal position 3B1, adjacent to the *period* (*per*) gene[7,8]. To identify the *zw3* gene at the molecular level we isolated overlapping phage encompassing 70 kilobases (kb) of genomic DNA distal to *per*. The breakpoints of two deficiencies, *Df(1)64j4* and *Df(1)64f1* (ref. 8), define the proximal and distal limits[6], respectively, of a 55-kb region of DNA within which the *zw3* gene must lie (Fig. 1*a* and *b*). Transcriptional analyses of this region revealed several related transcripts ranging in size from 2.5 to 6.3 kb (Fig. 1*d*). Although any one transcript may extend over as much as 40 kb of genomic DNA, all transcripts are detected by a single small genomic fragment (illustrated by filled bar in Fig. 1*a*) and are transcribed in the same direction (data not shown). The inversion $zw3^{b12}$, an allele of *zw3* (ref. 8), identifies this transcription unit as the *zw3* locus, as the inversion breakpoint disrupts the genomic DNA from which the transcripts are derived (Fig. 1*b* and *c*). Transcripts are detected at all developmental stages, including messenger RNA of maternal origin at 0-1 h after egg laying, and are developmentally regulated (Fig. 1*d*). Using a probe that is common to all the *zw3* transcripts we detected a uniform pattern of expression throughout the preblastoderm embryo and in all germ layers of the gastrulating embryo (Fig. 2). Expression is at its highest level during the period from fertilization through germ band elongation, and decreases during germ band retraction.

Using the genomic fragment that is common to all of the *zw3* transcripts we have isolated complementary DNAs from both ovarian[9] and embryonic 9-12 h[10] cDNA libraries. We have determined the nucleotide sequence of the largest ovarian cDNA (1,257 bp) and the largest embryonic cDNA (3,660 bp). The ovarian cDNA corresponds to the 2.5-kb maternal mRNA, and the embryonic cDNA is assumed to correspond to a transcript that is first detected at 6 h of embryonic development and that persists to the adult stage.

The embryonic cDNA contains a putative open reading frame of 2,199 nucleotides beginning at the ATG at position 378 (Fig. 3*b*). The context of this initiation codon is not a particularly good match to the *Drosophila* consensus for translation start sites ((C/A)AA(C/A)AUG) (ref. 11); there is another in-frame initiation codon at position 507, which is closer to the consensus. The open reading frame is followed by a 3′ untranslated region of 1,057 nucleotides. Several embryonic cDNAs have poly(A)

© 1990 Nature Publishing Group

tracts at their 3′ end, preceded by the sequence AAATTA which is a variant of the canonical polyadenylation sequence, AATAAA (ref. 12).

The ovarian cDNA contains an open reading frame of 867 nucleotides beginning with the ATG at position 391 and continuing to the end of the clone (Fig. 3*a*). This cDNA is not full-length and lacks sequence from the 3′ end of the transcript that it represents. The sequence flanking the initiation codon (TACG) is consistent with the *Drosophila* consensus at the highly conserved second position[12]. The ovarian and embryonic cDNAs are identical in a region of protein coding sequence starting at position 476 in the ovarian cDNA and position 1,162

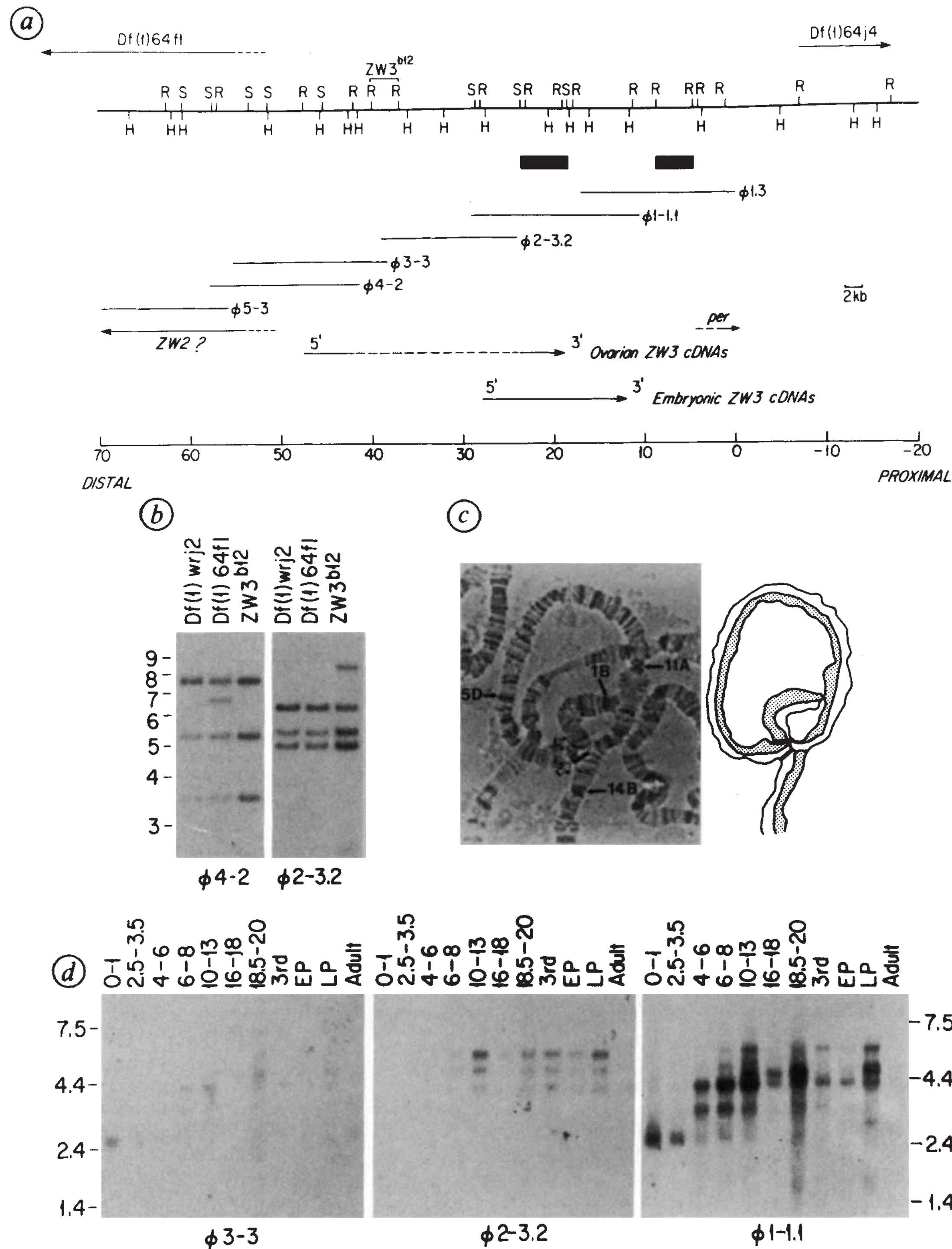


FIG. 1 The molecular organization of the *zw3* locus. *a*, Restriction map and the positions of the deficiencies *Df(1)64j4* (ref. 7) and *Df(1)64f1*, and the inversion $zw3^{b12}$ are shown. The lines above the molecular map indicate where the genomic DNA is present in the deficiencies, the stippled portions indicate uncertainty in the position of the breakpoint. The hatched bar represents the probe used to initiate the phage walk and the filled bar represents the probe used to isolate cDNAs. R, *Eco*RI; H, *Hin*dIII; S, *Sal*I. *b*, DNA from heterozygous flies (*Df(1)64f1/FM7* and $zw3^{b12}$*/FM7)* was digested with *Hin*dIII and probed with the recombinant phage 4-2 and 2-3.2. Alterations were detected with several restriction enzymes and this data was used to map the breakpoints (data not shown). *Df(1)wrj2* (ref. 8) is a large deficiency that removes DNA in the region of the genome represented by the phage walk. *c*, *In situ* hybridization to polytene chromosomes from 3rd instar larvae of the genotype, $zw3^{b12}$*/yf*, using the recombinant phage 2-3.2 as a probe. This phage hybridizes to genomic DNA on either side of the inversion, $zw3^{b12}$, which extends from 3B1 to 12F (ref. 8). Open arrows indicate the position of the hybridization and the thin black arrows indicate cytological positions on the chromosome. The stippled region in the panel on the right represents the *yf* chromosome. *d*, Northern-blot analysis of the *zw3* locus. A northern blot containing mRNA from different developmental stages was probed with three overlapping phage that cover the *zw3* locus. The embryonic stages represented are described in hours after egg laying (AEL), the other stages are third instar larvae (3rd), 0–24 h pupae (EP), 96–120 h pupae (LP) and adult.

METHODS. Overlapping phage were isolated from a phage EMBL3 *Drosophila* genomic library[22]. Sixteen cDNAs were isolated from a LamdaZap (Stratagene) cDNA library made to ovarian RNA[9] and the longest was sequenced. Embryonic cDNAs were isolated from a λgt11 library made from mRNA isolated from 9–12 h embryos[10]. Eight embryonic cDNAs were isolated and analysed and the longest was chosen for sequencing. Restriction maps have been determined for all of the cDNAs and the ends of all of them have been sequenced (data not shown). Southern transfer, northern transfer and all hybridizations were performed using standard techniques (Final washes were 0.2×SSC, 0.1% SDS at 65 °C)[23]. Chromosome squashes for the *in situ* hybridization to polytene chromosomes were as described[24], with minor modifications. The recombinant phage 2-3.2 was labelled by nick translation with biotinylated dUTP (ref. 25) and hybridization was detected with the Detek-1-HRp kit (Enzo Diagnostic Inc.).

© 1990 Nature Publishing Group

FIG. 2 Localization of *zw3* transcripts. *zw3* transcripts are detected uniformly throughout the embryo at (*a*) preblastoderm, stage 1–2 (~0–90 minutes AEL), (*b*) germ-band elongation, stage 11 (~6 hours AEL) and (*c*) germ-band retraction, stage 15 (~12 hours AEL). Embryos are oriented anterior to the left, dorsal side up. Phase contrast images are on the left, dark-field images on the right. Abbreviations: (am), anterior midgut; (pm) posterior midgut; (p) pharynx; (m) midgut; (vc) ventral cord.
METHODS. Wild-type embryos were collected and grown to the desired developmental stage. Fixation, embedding, sectioning and hybridizations were performed as described[26]. DNA probes generated from the fragment common to all the cDNAs (Fig. 1) were radioactively labelled by nick translation using [^{35}S]dCTP. Exposure times are 8 days (*a*), 12 days (*b*) and 30 days (*c*).

in the embryonic cDNA, and probably represent alternatively spliced mRNAs. The sequence at position 1,160 of the embryonic cDNA, where the ovarian and embryonic sequence diverge, is consistent with a splice junction[13] (Fig. 3*b*).

The sequence that is common to both ovarian and embryonic cDNAs encodes a protein that resemble the catalytic domain of a serine–threonine protein kinase[14] (Fig. 4). The predicted ovarian protein has 29 amino acids preceding the common domain, and the predicted embryonic protein has a unique amino terminus of 262 amino acids. This putative kinase catalytic

FIG. 3 Nucleotide and deduced amino acid (single-letter code) sequence of *zw3*. *a*, Sequence of the ovarian cDNA, cOV7. *b*, Sequence of the embryonic cDNA, cKZ5. The 53 nucleotides at the 3′ end represent sequence of a related embryonic cDNA which extends further in the 3′ direction and includes the poly(A) tract. Sequence common to cOV7 and cKZ5 is indicated by the bold letters. Numbers refer to the positions of the nucleotides. Double line indicates the sequence flanking the potential translation start site. Open reading frame of the embryonic cDNA extends 5′ of either one of the potential initiating codons; it is therefore possible that the cDNA is incomplete and the true initiation codon is not contained in this cDNA. Single line indicates the putative polyadenylation signal. Asterisks mark the sequence at the junction of unique and shared regions in the embryonic cDNA, which is consistent with a splice junction. Sequence in lower case is the *Eco*RI site into which the cDNAs are cloned. The ovarian cDNA is truncated and ends at an *Eco*RI site internal to the full-length cDNA.
METHODS. The cDNAs were sequenced as double-stranded DNA by the dideoxy chain termination method[27] with T7 polymerase (Pharmacia) or Sequenase (USB): cOV7 was sequenced on one strand using nested deletions and oligonucleotides to known sequence; cKZ5 was sequenced on both stands by a combination of subcloning, nested deletions and oligonucleotides.

© 1990 Nature Publishing Group

```
                          ** *  * *               * * *
Draf1          EEILIGPRIGSGSFGTVYRA...HWHGP...VAVKTLNVKTPSPAQLQAF
Cdc28     MSGELANYKRLEKVGEGTYGVVYKALDLRPGQGQRVVALKKIRLESEDEGVPSTA
Cdc2          MENYQKVEKIGEGTYGVVYKA...RHKLSGRIVAMKKIRLEDESEGVPSTA
                |     ||:|::|||: |    :    :| :||:||:   :
zw3           EVSYTDTKVIGNGSFGVVFQA...KLCDTGELVAIKKVLQDRRFKN.....
                          ======

            *               *
Draf1     KNEVAMLK....KTRHCNILLFMGCVSKPS.......LAIVTQWCE...GSSLY
Cdc28     IREISLLKELKDD....NIVRLYDIVHSDAHK.....LYLVFEFLDLDLKRYME
Cdc2      IREISLLKEVNDENNRSNCVRLLDILHAES.K.....LYLVFEFLDMDLKKYMD
           ||: ::: ::     : |:|:||    :: : |      | || |::   : |
zw3       .RELQIMRKLE....HCNIVKLLYFFYSSGEKRDEVFLNLVLEYIPETVYKVAR

                                *  **        * **   **
Draf1     KHVHVSETKFKLNTLIDIGRQVAQGMDYLHAKNIIHRDLKSNNIFLHEDLS.VK
Cdc28     GIPKDQP..LGADIVKKFMMQLCKGIAYCHSHRILHRDLKPQNLLINKDGN.LK
Cdc2      RISETGATSLDPRLVQKFTYQLVNGVNFCHSRRIIHRDLKPQNLLIDKEGN.LK
            : |  | :    :  : |||  :: ::||  |:|||:||||||:| |:   ||
zw3       QYAKTKQT.IPINFIRLYMYQLFRSLAYIHSLGICHRDIKPQNLLLDPETAVLK
                                                    -----

          * ***                    *  * * *                * ** *
Draf1     IGDFGLATAKTRWSGEKQANQPTGSILWMAPEVIRMQELNPYSFQSDVYAFGIV
Cdc28     LGDFGLARAFG..VPLRAYTHEIVTLWYRAPEV..LLGGKQYSTGVDTWSIGCI
Cdc2      LADFGLARSFG..VPLRNYTHEIVTLWYRAPEV..LLGSRHYSTGVDIWSVGCI
          | ||| |:       :   | ::  |:: :|||||:  : |:  |:| :|:|| ||:
zw3       LCDFGSAKQLL..HGEPNVSY.ICSRYYRAPEL..IFGAINYTTKIDVWSAGCV
                                  ---------
Draf1     MYELLAECLPYGHISNKDQIL......FMVGRGLLRPDMSQVRSDARRHSKRLA
Cdc28     FAEMCNRKPIFSGDSEIDQIFKIFRVLGTP.NEAIWPDIVYLPDFKPSFPQWRR
Cdc2      FAEMIRRSPLFPGDSEIDEIFKIFQVLGTP.NEEVWPGVTLLQDYKSTFPRWKR
           ||::    |:||||| :|::   :   |||||  |::      :    ::|  ||   |
zw3       LAELLLGQPIFPGDSGVDQLVEVIKVLGTPTREQIREMNPNYTEFK..FPQIKS

                               *          *
Draf1     EDCIKYTP.KDRPLFRPLLNML...ENMLRTLPKIHRSASEPNLT
Cdc28     KDLSQVVP.SLDPRGIDLLDKLLAYDPINRISARRAAIHPYFQES
Cdc2      MDLHKVVP.NGEEDAIELLSAMLVYDPAHRISAKRALQQNYLRDF
              ||     :  :||:|:|  :|  |  |:  ||:: :|:    :   :
zw3       HPWQKVFRIRTPTEAINLVSLLLEYTPSARITPLKACAHPFFDEL
```

FIG. 4 Comparison of *zw3* amino-acid sequence (single-letter code) and other serine-threonine protein kinases. cdc2, *S. pombe cdc2*[15]; cdc28, *S. cerevisiae CDC28*[17]; Draf1, the *Drosophila* homologue of the mammalian *c-raf* (refs 28, 29). Comparison of zw3 and cdc2 proteins reveals 34% identity over 295 amino acids. Straight vertical line indicates identity between residues. Colon indicates a conservative change, this is based on structural similarity as described[14]. Conserved residues in all protein kinases are indicated by asterisks, which mark residues that are conserved in 65 different kinases examined and described[14]. Asterisk and a bold letter indicate residues that are identical in 62 of the 65 sequences. Asterisk alone indicates a position where residues of similar structure are conserved in 63 of 65 kinases. Double line indicates the region containing the nucleotide-binding site (GXGXXG) found in many nucleotide-binding proteins, including protein kinases. Single lines indicate sequence that is indicative of a serine-threonine kinase rather than a tyrosine kinase[18]. The complete protein sequences for cdc2 and cdc28 proteins are shown. Amino acids 265-623 of Draf1 are shown, as are residues 397-688 for the predicted protein of cKZ5.

METHODS. Genbank and EMBL databases were searched using the program TFASTA of the University of Wisconsin Genetics Computer Group. Alignments were generated using BESTFIT[30].

domain has the most similarity with the *cdc2* family of genes which includes *cdc2* from *Schizosaccharomyces pombe*[15], its human homologue[16] and *CDC28* from *Saccharomyces cerevisae*[17]. The amino acid comparison of the putative *zw3*-encoded kinase catalytic domain with *cdc2* reveals 34% identity over the entire length of the *cdc2*-encoded protein. The predicted *zw3*-encoded protein contains greater similarity in limited regions of the catalytic domain that seem to be functionally significant; the region of the nucleotide binding site, and two other regions within the catalytic domain that are indicative of a serine-threonine kinase (Fig. 4)[14,18]. In addition to these specific regions, there are residues that are conserved in the catalytic domain of most protein kinases and that are also conserved in the putative catalytic domain of *zw3* (Fig. 4)[14]. Searches of available databases with the unique 5′ coding regions of the ovarian and embryonic cDNAs do not reveal significant homology to any other known sequences. We believe that the similarity between *zw3* and *cdc2* reflects the fact that *zw3* encodes a putative serine-threonine kinase rather than its having analogous functions. The two functionally equivalent kinases, *CDC28* from *S. cerevisiae* and *cdc2* from *S. pombe* share >60% identity[19,20].

The *zw3* gene encodes several transcripts that are expressed uniformly throughout the embryo. We have sequenced cDNAs representing two developmental transcripts and these encode at least two proteins that resemble serine-threonine kinases. The *zw3* gene may be involved in the signal transduction of information between cells that is required to establish cell identity within each segment[2-5]. The nature of the *zw3* gene product and the lack of spatially restricted expression suggests that *zw3* may interact with a gene product that is spatially localized to the anterior portion of each segment, the region affected in *zw3* mutant embryos. The zygotic gene *nkd* is a good candidate for such a gene because mutations in *zw3* and *nkd* result in similar embryonic phenotypes. It is possible that the *nkd* gene product is a substrate, or alternatively an activator, of the putative *zw3*-encoded kinase. This possible interaction between *nkd* and *zw3* remains to be tested, as *nkd* has yet to be cloned and its expression pattern is unknown. In addition to its role in embryonic segmental patterning, *zw3* is required for normal epidermal cell differentiation[6,21]. In the absence of *zw3* gene product, cells undergo a fate change and assume a neural fate. Zygotic *zw3* gene product may be involved in mediating the choice between neural and epidermal cell fate. The *zw3* gene may be of importance in mediating intercellular interactions that determine cell fate. (The *zw3* locus has recently been renamed *shaggy*[21]. We have used the original nomenclature.) □

Received 22 December 1989; accepted 5 April 1990.

1. Nusslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795-801 (1980).
2. Martinez-Arias, A., Baker, N. E. & Ingham, P. W. *Development* **103**, 157-170 (1988).
3. DiNardo, S., Sher, E., Heemskerk-Jongens, J., Kassis, J. A. & O'Farrell, P. H. *Nature* **332**, 604-609 (1988).
4. Hooper, J. E. & Scott, M. P. *Cell* **59**, 751-765 (1989).
5. Nakano, Y., Guerrero, I., Hidalgo, A., Taylor, A., Whittle, J. R. S. & Ingham, P. W. *Nature* **341**, 508-573 (1989).
6. Perrimon, N. & Smouse, D. *Devl Biol* **135**, 287-305 (1989).
7. Reddy, P. *et al. Cell* **38**, 701-710 (1984).
8. Judd, B. H., Shen, M. W. & Kaufman, T. C. *Genetics* **71**, 139-156 (1972).
9. Caudy, M., Vassin, H., Brand, M., Tuma, R., Jan, L. Y. & Jan, Y. N. *Cell* **55**, 1061-1067 (1988).
10. Zinn, K., McAllister, L. & Goodman, C. S. *Cell* **53**, 577-587 (1988).
11. Cavener, D. R. *Nucleic Acids Res.* **15**, 1353-1361 (1987).
12. Birnstiel, M. L., Busslinger, M. & Strub, K. *Cell* **41**, 349-359 (1985).
13. Shapiro, M. B. & Senapathy, P. *Nucleic Acids Res.* **15**, 7155-7174 (1987).
14. Hanks, S. K., Quinn, A. M. & Hunter, T. *Science* **241**, 42-52 (1988).
15. Hindeley, J. & Phear, G. A. *Gene* **31**, 129-134 (1984).
16. Lee, M. G. & Nurse, P. *Nature* **327**, 31-35 (1987).
17. Lorincz, A. T. & Reed, S. I. *Nature* **307**, 183-185 (1984).
18. Hanks, S. *Proc. natn. Acad. Sci. U.S.A.* **84**, 388-392 (1987).
19. Beach, D., Durkacz, B. & Nurse, P. *Nature* **300**, 706-709 (1981).
20. Booher, R. & Beach, D. *Molec. cell. Biol.* **6**, 3523-3530 (1986).
21. Simpson, P., El Messal, M., Moscoso del Prado, J. & Ripoll, P. *Development* **103**, 391-401 (1988).
22. Blackman, R., Grimaila, R., Koehler, M. & Gelbart, W. *Cell* **49**, 497-505 (1987).
23. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
24. Gall, J. G. & Pardue, M. L. *Meth. Enzym.* **21**, 470-480 (1971).
25. Langer-Sofer, P. R., Levine, M. & Ward, D. C. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4381-4385 (1982).
26. Hafen, E., Kuroiwa, A. & Gehring, W. J. *Cell* **37**, 833-841 (1984).
27. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).

© 1990 Nature Publishing Group

28. Mark, G. E., MacIntyre, R. J., Digan, M. E., Ambrosio, L. & Perrimon, N. *Molec. cell. Biol.* **7,** 2134–2140 (1987).
29. Nishida, Y. *et al. EMBO J.* **7,** 775–781 (1988).
30. Devereux, J., Haeberli, M. & Smithies, O. *Nucleic Acids Res.* **12,** 387–395 (1984).

ACKNOWLEDGEMENTS. We thank M. Rosbash for *per* DNA, K. Zinn and B. Hay for cDNA libraries, J. Hall for stocks and B. Noll and K. McCall for technical assistance. We also thank B. Finkelstein, J. Klingensmith, B. Rutledge and G. Thomas for comments on the manuscript. This work was supported by the Howard Hughes Medical Institute, and the NIH (N.P. and E.S.).

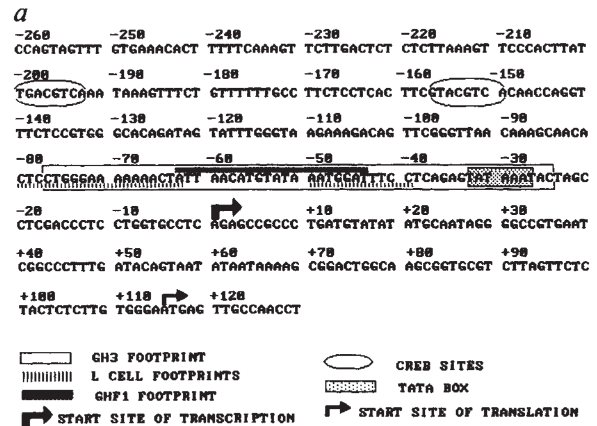
Regulation of the pituitary-specific homeobox gene *GHF1* by cell-autonomous and environmental cues

Alison McCormick, Helen Brady, Lars E. Theill & Michael Karin

Department of Pharmacology, M036, School of Medicine, Center for Molecular Genetics, University of California, San Diego, La Jolla, California 92093, USA

HOMEODOMAIN proteins[1,2] function in determination of mating type in yeast[3], segmentation in fruit flies[4] and cell-type specific gene expression in mammals[5]. In *Drosophila*, expression of homeobox genes is controlled by cell-autonomous interactions between regulatory proteins and environmental clues[6–8]. Similar controls may operate during mammalian limb development[9,10] and frog embryogenesis[11,12]. But, the exact way in which expression of homeodomain proteins is regulated in these systems is not clear and requires biochemical analysis of homeobox gene transcription. We now describe such an analysis of the *GHF1* gene, which encodes a mammalian homeodomain protein specifying expression of the growth hormone (GH) gene in anterior pituitary somatotrophs[13–15]. *GHF1* is transcribed in a highly restricted manner and the presence of GHF1 protein is correlated both temporally and spatially with activation of the *GH* gene during pituitary development[16]. Analysis of the *GHF1* promoter indicates that transcription is also controlled by cell-autonomous interactions involving positive autoregulation by GHF1, and environmental cues that modulate the intracellular level of cyclic AMP and thereby the activity of cAMP response element binding protein (CREB)[17], a ubiquitous transactivator that binds to the *GHF1* promoter.

We examined the regulation of *GHF1* promoter activity by protein factors in cell lines derived from a GH-expressing anterior pituitary tumour (GH3 or GC cells) compared with L cells which do not express GH and can eliminate GH and GHF1 expression on fusion to GH3 cells[18]. *GHF1* was isolated from a rat genomic library[19] and its start site of transcription was



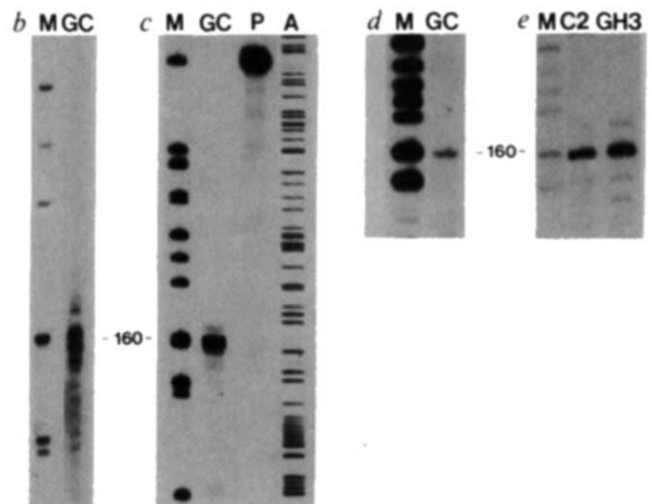


FIG. 1 Identification of the rat *GHF1* promoter. *a*, Nucleotide sequence of the *GHF1* promoter region. The locations of the TATA box, CREB binding sites, a GHF1 binding site and the GH3 and L cell-specific footprints are indicated. *b* and *c*, RNAse mapping. Total GC cell RNA was hybridized to 780- or 310-nucleotide (nt) *GHF1* antisense riboprobes complementary to positions −560 to +160 (*b*) or −90 to +160 (*c*), digested with RNase and the protected products were separated on a sequencing gel and visualized by autoradiography. The predominant 160-nt protected fragment corresponds to transcripts initiated 27 bp downstream of the TATA box indicated in *a*. M, relative molecular mass (M_r) size markers; P, undigested probe; A-A, specific sequence ladder. *d*, Primer extension analysis of total GC cell RNA with a *GHF1* specific primer. Primer extended products were seperated on a sequencing gel and visualized by autoradiography. *e*, Primer extension analysis of *GHF1* RNA transcribed *in vitro*. The p2.0GHF1-BS template was incubated with either C2 or GH3 whole cell extracts to generate *in vitro* synthesized *GHF1* RNA. The same *GHF1* primer used in *d* was used to analyse the RNA synthesized *in vitro*.

METHODS. Plaque-forming units (1×10^6) of a rat genomic library[19] were screened with a *GHF1* cDNA probe[13]. Duplicate nitrocellulose filters were washed stringently (50 °C, in $0.1\times$SSC, 50% formamide) and positive phages were rescreened with a *GHF1* cDNA 5′-specific probe. Of 16 positive clones, three hybridized to the 5′ probe and the largest was subsequently subcloned into pBluescript (BS, Stratagene) as *Eco*RI–*Eco*RV, *Eco*RV–*Eco*RV and *Eco*RV–*Sal*I fragments. Only the 2.0 kb *Eco*RI–*Eco*RV subclone (p2.0GHF1-BS) hybridized to northern blots of GC cell RNA (data not shown). The complete nucleotide sequence of this fragment was determined on both strands as previously described[32]. For RNase protection 20 μg total RNA extracted from GC cells was resuspended in 20 μl hybridization buffer in the presence of 1×10^6 c.p.m. of the different riboprobes[32]. Probes were synthesized from p2.0GHF1-BS using the T7 RNA polymerase as described by Stratagene. The 780-nt probe was generated by linearizing with *Sac*I(−560), whereas the 310-nt probe was generated by linearizing with *Hpa*I(−90). Both riboprobes end at the *Eco*RI site at position +160 of *GHF1* followed by 60 nt of Bluescript sequences. Hybridization was for 5 min at 80 °C and overnight at 45 °C. Hybrids were digested as described[32] and protected RNA fragments were separated on a 6% polyacrylamide, 7M urea sequencing gel, and visualized by autoradiography. For primer extension analysis, 20 μg total GC cell RNA was hybridized with 20 fmol ^{32}P-labelled *GHF1*-specific primer (5′-GAATTCAGAGGTATAAAGGTATCTGCCGACCTGAAA-GGTTGGAACT-3′) in 20 μl of 250 mM KCl, 10 mM Tris buffer pH 7.5, 1 mM EDTA at 55 °C for 1 h. Primer extension was performed as described[18]. For *in vitro* analysis whole cell extracts were prepared as described[18]. Subclone p2.0GHF1-BS (100 ng) was incubated with 75 μg of C2 or GH3 whole cell extracts for 15 min on ice in 30 μl of 50 mM HEPES buffer, pH 7.9, 100 mM KCl, 12 mM $MgCl_2$, 1 mM DTT, 1 mM EDTA, 20% glycerol and 10 U of Promega RNasin. Transcription was initiated by addition of 20 μl of 1 mM (each) rNTP and 5% polyvinylalcohol. After 1 h at 30 °C, reactions were stopped by addition of 150 μl transfer RNA stop mix (200 mM NaCl, 20 mM EDTA, 1% SDS, 5 μg tRNA) and analysed by primer extension as described above with the *GHF1*-specific primer.

© 1990 Nature Publishing Group