Thus the α-fibrinogen ATTAAC sequence is also within a region essential for optimum function of this promoter.

One copy of the ATTAAC sequence is present in the rat, mouse, and human albumin promoters in the opposite orientation at position −54, and deletion of this sequence results in a reduction in tissue-specific in vitro transcription (*5*). This region is protected from DNase I digestion by a protein in liver cells (*20*). These results suggest that HNF1 also interacts with the albumin promoter.

REFERENCES AND NOTES

1. M. Ott *et al.*, *EMBO J.* **3**, 2505 (1984).
2. R. H. Costa, E. Lai, J. E. Darnell, Jr., *Mol. Cell. Biol.* 6, 4697 (1986).
3. G. Ciliberto *et al.*, *Cell* **41**, 531 (1985).
4. R. W. Scott, T. F. Vogt, M. E. Croke, S. M. Tilghman, *Nature (London)* **310**, 562 (1984); R. E. Hammer, R. Krumlauf, S. A. Camper, R. L. Brinster, S. M. Tilghman, *Science* **235**, 53 (1987).
5. K. Gorski, M. Carneiro, U. Schibler, *Cell* **47**, 767 (1986).
6. L. M. Staudt *et al.*, *Nature (London)* **323**, 640 (1986); N. F. Landolfi, J. D. Capra, P. W. Tucker, *ibid.*, p. 548.
7. H. Ohlsson and T. Edlund, *Cell* **45**, 35 (1986).
8. J. R. de Wet *et al.*, *Mol. Cell. Biol.* 7, 725 (1987).
9. D. Galas and A. Schmitz, *Nucleic Acids Res.* **5**, 3157 (1978).
10. M. G. Fried and D. M. Crothers, *J. Mol. Biol.* **172**, 241 (1984); F. Strauss and A. Varshavsky, *Cell* **37**, 889 (1984).
11. H. Singh, R. Sen, D. Baltimore, P. A. Sharp, *Nature (London)* **319**, 154 (1986).
12. R. W. Carthew, L. A. Chodosh, P. A. Sharp, *Cell* **43**, 439 (1985); L. A. Chodosh, R. W. Carthew, P. A. Sharp, *Mol. Cell. Biol.* 6, 4723 (1986).
13. L. A. Chodosh *et al.*, *Science* **238**, 684 (1987); J. Morgan *et al.*, in preparation.
14. S. Baumhueter *et al.*, in preparation.
15. W. S. Dynan and R. Tjian, *Cell* **32**, 669 (1983); *ibid.* **35**, 79 (1983); M. R. Briggs, J. T. Kadonaga, S. P. Bell, R. Tjian, *Science* **234**, 47 (1986).
16. K. A. Jones, K. R. Yamamoto, R. Tjian, *Cell* **42**, 559 (1985); B. J. Graves, P. F. Johnson, J. L. McKnight, *ibid.* **44**, 565 (1986); R. B. Cohen, M. Sheffrey, C. G. Kim, *Mol. Cell. Biol.* 6, 821 (1986).
17. G. L. Long, T. Chandra, S. L. Woo, E. W. Davie, K. Kurachi, *Biochemistry* **23**, 4828 (1984).
18. G. R. Crabtree and J. A. Kant, *Cell* **31**, 159 (1982).
19. S. L. Woo *et al.*, *Regulation of Liver Gene Expression* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1987), p. 171.
20. S. Cereghini, M. Raymondjean, A. G. Carranca, P. Herbomel, M. Yaniv, *Cell* **50**, 627 (1987).
21. B. A. Spalholz, Y. C. Yang, P. M. Howley, *ibid.* **42**, 183 (1985).
22. T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982).
23. F. Graham and A. van der Eb, *Virology* **52**, 456 (1973); M. Wigler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 76, 1373 (1979).
24. C. M. Gorman, L. F. Moffat, B. H. Howard, *Mol. Cell. Biol.* **2**, 1044 (1982).
25. D. M. Fowlkes *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2313 (1984).

26. We thank K. Shaw for editorial assistance, L. Chodosh and P. Sharp for the adenovirus major late promoter, J. Sklar for the human α-globin promoter, G. Long for the human α_1-antitrypsin promoter, and D. Clayton for review of the manuscript. Supported by grants HL33942 and CA39612 from the National Institutes of Health and a grant from the Eli Lilly Research Laboratories. G.C. was partially supported by the Ministère des Relations Extérieures (France).

28 May 1987; accepted 4 August 1987

Dorsal, an Embryonic Polarity Gene in *Drosophila*, Is Homologous to the Vertebrate Proto-oncogene, c-*rel*

RUTH STEWARD

The *Drosophila* gene, *dorsal*, is a maternal effect locus that is essential for the establishment of dorsal-ventral polarity in the developing embryo. The *dorsal* protein was predicted from the complementary DNA sequence; it is almost 50 percent identical, over an extensive region, to the protein encoded by the avian oncogene v-*rel*, its cellular homolog, c-*rel*, and a human c-*rel* fragment. The oncogene v-*rel* is highly oncogenic in avian lymphoid, spleen, and bone marrow cells.

THE FINAL IDENTITY OF EACH CELL along the dorsal-ventral axis in the *Drosophila* embryo is dependent on the interaction of maternal effect and zygotic gene products during oogenesis and early embryogenesis. The maternal effect gene products control the expression of zygotic genes that are essential for the elaboration of the asymmetric pattern along the dorsal-ventral axis. Exhaustive mutant screens have led to the identification of 12 such genes (*1*). Females homozygous for loss or reduction of function mutations in 11 of these genes (the "dorsal group") produce embryos missing the ventral pattern elements and body parts; all of the embryonic cells assume a fate normally assigned to cells at a more dorsal position. Genetic experiments suggest that the products of these 11 maternal effect genes participate in a complex morphogenetic pathway resulting in dorsal-ventral polarity in the embryo (*2*).

The *dorsal* locus was the first "dorsal group" gene to be identified (*3*). Females that are homozygous mutant for *dorsal* produce embryos that fail to establish normal dorsal-ventral polarity, irrespective of the genotype of the father. The *dorsal* phenotype is first observed during the formation of the cellular blastoderm, 2.5 to 3 hours after fertilization. While *dorsal* mutations cause severe perturbation along the dorsal-ventral axis of the developing embryo, anterior-posterior polarity appears to be normal (*4, 5*).

Although *dorsal* is a maternal effect locus and must be expressed during oogenesis, two observations suggest that the gene product is active or required early in embryogenesis. First, temperature-shift experiments indicate that the *dorsal* protein is active during a short period in early embryogenesis (between 1.25 and 2.5 hours post-fertilization, before cellular blastoderm formation) (*6*). Second, *dorsal* embryos can be partially rescued by injection of cytoplasm from wild-type, cleavage-stage embryos (*7*). Injection experiments also suggest that the distribution of the *dorsal* rescuing activity changes during early embryogenesis from a uniform distribution on both the ventral and dorsal sides of the embryo immediately after the egg is fertilized to an enrichment on the ventral side by the syncytial blastoderm stage (*8*). It seems likely that other genes in the dorsal group also participate in this asymmetric distribution of the *dorsal* rescuing activity.

The *dorsal* gene has been cloned and characterized (*9*). The *dorsal* transcription unit is about 14 kb in length and encodes a polyadenylated [poly(A)$^+$] RNA of about 2.8 kb (Fig. 1). We found that the gene is only transcribed in adult females and not in males, or at other stages during development (*9*). In females, *dorsal* expression is restricted to the nurse cells of the ovary, and the *dorsal* messenger RNA (mRNA) accumulates in a stable form in the maturing egg. The *dorsal* mRNA persists after fertilization, can be detected throughout the first 2 hours of embryogenesis, and then turns over rapidly and can no longer be detected by cellular blastoderm formation (2.5 hours). In situ hybridization to tissue sections shows that the *dorsal* mRNA is uniformly distributed throughout the ooplasm and cytoplasm of early embryos (*5, 10*).

To extend the characterization of the *dorsal* gene and its product (or products) we isolated a series of *dorsal* complementary DNA (cDNA) recombinants from *Drosophi-*

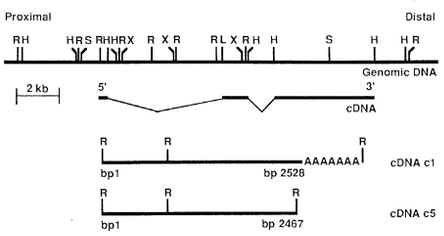

Fig. 1. Restriction map of the *dorsal* region, RNA coding region, and cDNAs. (R) Eco RI; (H) Hind III; (L) Sal I; (S) Sst I; (X) Xho I. The cDNAs were isolated from a cDNA library constructed from 0- to 2-hour poly(A)$^+$ embryonic RNA (*12*).

Department of Biology, Princeton University, Princeton, NJ 08544.

```
MFPNQNNGAAPGQGPAVDGQQSLNYNGLPAQQQQQLAQSTKNVRKKPYVKITEQPAGKAL   60

RFRYECEGRSAGSIPGVNSTPENKTYPTIEIVGYKGRAVVVVSCVTKDTPYRPHPHNLVG  120

KEGCKKGVCTLEINSETMRAVFSNLGIQCVKKKDIEAALKAREEIRVDPFKTGFSHRFQP  180

SSIDLNSVRLCFQVFMESEQKGRFTSPLPPVVSEPIFDKKAMSDLVICRLCSCSATVFGN  240

TQIILLCEKVAKEDISVRFFEEKNGQSVWEAFGDFQHTDVHKQTAITFKTPRYHTLDITE  300

PAKVFIQLRRPSDGVTSEALPFEYVPMDSDPAHLRRKRQKTGGDPMHLLLQQQQKQQLQN  360

DHQDGRQTNMNCWNTQNIPPIKTEPRDTSPQPLGFLIGRLLSSHPRRSRCRHRATTTTTA  420

RPRPTTWPPRSATNGQQQLMSPNHPQQQQQQQQYGATDLGSNYNPFAQQVLAQQQQHQQQ  480

QQQHQHQHQQQHQQQQQQQQQQQQQQQSLQFHANPFGNPGGNSWESKFSAAAVAAAAATAT  540

GAAPANGNSNNLSNLNNPFTMHNLLTSGGGPGNANNLQWNLTTNHLHNQHTLHQQQQLQQ  600

QQQQQYDNTAPTNNNANLNNNNNNNNNTAGNQADNNGPTLSNLLSFDSGQLVHINSEDQQI  660

LRLNSEDLQISNLSIST                                               667
```

Fig. 2. The predicted aa sequence encoded by cDNA c1. Bold underlining, region of homology to c-*rel* and v-*rel*. Normal underlining, regions of single aa residue repeats. The two Eco RI fragments were subcloned into a bluescript vector (*26*). Deletions were obtained as described (*27*). The DNA was sequenced by the method of Sanger (*28*). Single-letter abbreviations are shown (*29*).

la cDNA libraries constructed from poly(A)$^+$ RNA extracted from 0- to 2-hour embryos (*11, 12*). Two of these cDNAs (c1 and c5) (Fig. 1) are nearly full length (about 2.5 kb), and were chosen for DNA sequence analysis. The entire sequence of cDNA c1 was obtained on both strands, whereas the sequence of cDNA c5 was obtained only in part (all ends of the Eco RI fragments) to confirm the c1 sequence. At the 5′ end there are 274 bp of sequence that contains multiple termination sites, followed by a translation start codon. After a 2031-bp open reading frame, there is a stop codon followed by another stretch of sequence containing multiple termination codons. At +2224 bp there is a potential poly(A) addition site (AATAAA), and the poly(A) tail starts at +2252 bp, indicating that this cDNA extends to the 3′ end of the *dorsal* RNA. The open reading frame encodes a protein of 677 amino acids with a molecular weight of 75,600 and a *p*I of 10.2.

The predicted amino acid (aa) sequence of the *dorsal* protein (Fig. 2) was used in a computer search for homologous proteins (*13, 14*). The *Drosophila dorsal* gene encodes a protein closely related to the previously identified protein encoded by the turkey proto-oncogene c-*rel* and its corresponding viral oncogene v-*rel*, the transforming gene of the reticuloendotheliosis virus strain T (Rev-T) (*15*). This retrovirus is highly oncogenic in chicks and poults, and mediates transformation of spleen cells, but not of chicken embryo fibroblasts in tissue culture (CEF) (*16, 17*).

The homologous sequence begins at aa 46 of the putative *dorsal* protein and extends over a 295-aa stretch (Fig. 2). The putative *dorsal* protein has an approximately 47% identity with either v-*rel* or c-*rel* in the region of overlap, while the DNA homology is less than 45% in the same region (Fig. 3). The homology between *dorsal* and v-*rel*/c-*rel* is even more striking if one takes into account that many of the differences between the *Drosophila* and avian proteins represent conservative amino acid changes. In this case, the *dorsal* and v-*rel*/c-*rel* homology would be about 80%. In the *dorsal* protein the homology starts at aa 42 and extends to aa 341. In the v-*rel* protein (503 aa) the homology begins at aa 16. The first 12 aa of the v-*rel* protein represent the NH_2-terminal of the retroviral *envelope* gene followed by 472 aa derived from several exons of the turkey c-*rel* gene. The 19 COOH-terminal aa are again derived from the viral envelope sequence, but are spliced to the v-*rel* sequence from a different frame (*15*).

The homology starts at aa 5 of c-*rel*. However, unlike v-*rel*, the actual NH_2- and COOH-termini of c-*rel* are not known (cDNA or genomic sequences upstream and downstream from the terminal aa indicated in Fig. 3 have not been isolated). Hence, it is possible that the homology between the *dorsal* and the c-*rel* protein extends further toward the NH_2-terminus than indicated in Fig. 3. At the COOH-terminus the *dorsal* protein has about 300 aa that are not homologous to the turkey gene. It is most unlikely that there is another stretch of homology in the yet unknown c-*rel* COOH-terminal protein (*15*). The human c-*rel* sequences identified so far are homologous to exon 4 and 5 of the turkey c-*rel* gene, and the sequence of these two proteins differ only in 13 aa (*18*). The identity between the *dorsal* protein and the human *rel* fragment is 49%, and the homology is over 80% if the conservatively changed aa are counted (Fig. 3).

In the region of homology, one possible serine phosphorylation site (Arg-Arg-X-Ser) is identical in *dorsal*, c-*rel*, v-*rel*, and human c-*rel* (at aa 311 to 315 of the *dorsal* protein; Fig. 3). A possible glycosylation site (Asn-X-Thr/Ser) is conserved in *dorsal*, c-*rel*, and v-*rel* (at aa 78 of the *dorsal* protein; Fig. 3). The *dorsal* protein contains 12 cysteine residues, 10 of which are found in the region of homology. Of the ten cysteine residues in the c-*rel*/v-*rel* proteins, seven are in the region of homology and six are conserved in the *dorsal* protein, while one cysteine is displaced by one aa position in the *dorsal* protein.

The COOH-terminal region of the putative *dorsal* protein shows features that are typical of a number of other *Drosophila* proteins. There are several stretches of single amino acid residues. At +1417 bp, there are 34 glutamine residues [M or OPA repeat (*19, 20*)], followed at +1585 bp by 11 alanine residues, another 12 glutamine residues at +1777 bp, and finally, 14 aspara-

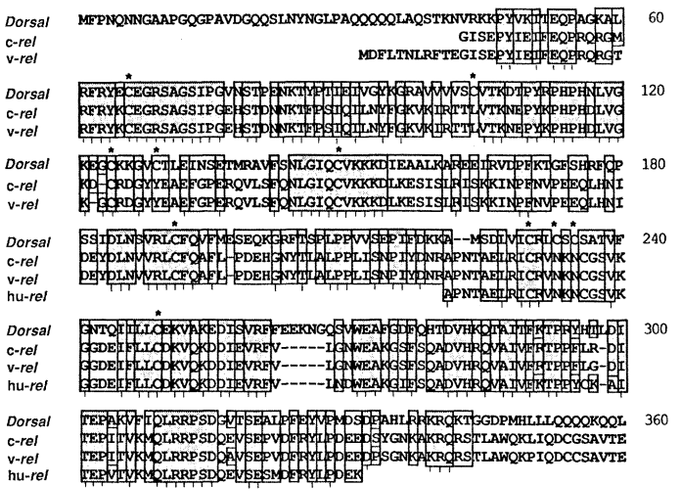

Fig. 3. Comparison of the predicted *dorsal* NH_2-terminal aa sequence with that of c-*rel*, v-*rel*, and hu-*rel*. Vertical lines under the boxes, identical aa; open boxes, conservatively changed aa. Cysteines are marked by *. The aa sequence homology search was done with the FastP program (*13*) and the Bionet aa sequence library (*14*).

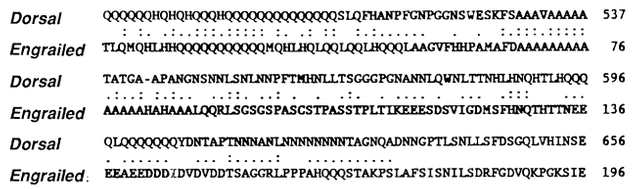

Fig. 4. Comparison of the predicted *dorsal* COOH-terminal aa sequence with that of the NH_2-terminal aa sequence of *engrailed*.

gine residues at +1837 bp. Within this region of the protein, the *dorsal* protein is homologous to the glutamine repeat of the *Notch* protein (*20*), and to several wheat proteins (*21*). In a 144-aa overlap, 26% of the COOH-terminal *dorsal* aa are identical to the *engrailed* protein (*12*), including one of the glutamine repeats and the alanine repeat. In both proteins the spacing between the two repeats is identical (*dorsal* protein, aa 479 to 632; *engrailed* protein aa 19 to 172; Fig. 4). The homology is 73%, if conservatively changed amino acids are taken into account. The functional significance of the single aa stretches in *Drosophila* proteins is not yet known. They are found in proteins with different apparent functions and in different positions; for example, they occur near the NH_2-terminus in *engrailed* and the COOH-terminus in *dorsal*. The presumptive *dorsal* protein seems to consist of at least two large domains: the NH_2-terminal half showing homology to the *rel* proteins, and the COOH-terminal half exhibiting "species specific" characteristics.

The extent of homology between the *dorsal* and *rel* proteins indicates that they share a common function and possibly have closely related structures. Both proteins contain a comparable number of cysteines (*dorsal* has 13; *rel* has 10), and most are found at identical sites in the region of homology, whereas the COOH-terminal half of each protein contains very few and the last third no cysteine residues at all. The spacing of the cysteine residues does not follow any known pattern.

Very little is known about the c-*rel* gene. A 4-kb poly(A)$^+$ RNA hybridizing to c-*rel* sequences has been observed in spleen, muscle, and hematopoietic tissue (*22*). The v-*rel* protein is phosphorylated (*17*), and it is possible that the serine (aa 314 in the *dorsal* protein) that is conserved in all four proteins is one of the aa that might be phosphorylated.

A variety of studies suggest that the *dorsal* gene product functions at one of the last, if not the last, step in the dorsal-ventral pathway controlled by the maternal effect genes. It may represent the link between this pathway and the differential activation of the zygotic genome (*1, 2*). The *dorsal* protein may function in the nucleus; a positively charged aa region at position 337 (five out of six aa are positively charged) corresponds to the position of four positively charged aa in c-*rel* and v-*rel* that have been shown, by site-directed mutagenesis, to be responsible for nuclear localization of the v-*rel* protein (*23*). The sequence of another "dorsal group" gene, *snake*, reveals strong homology to a serine protease that in all probability must be cleaved to be activated (*24*). Thus, the dorsal-ventral pattern genes may represent a morphogenetic pathway that causes the asymmetric distribution of a morphogen, possibly the *dorsal* protein, by an activation cascade. This morphogen directly or indirectly controls activation of zygotically active genes expressed differentially in cells arrayed along the dorsal-ventral axis.

Several genes important for invertebrate development have been shown to be homologous to vertebrate growth control protein genes (*20, 25*). The study of these and related genes may enhance understanding of the regulatory pathways of growth control in normal development in vertebrates and the changes associated with oncogenesis.

REFERENCES AND NOTES

1. K. V. Anderson, *Trends Genet.* **3**, 91 (1987); ——— and C. Nusslein-Volhard, in *Primers in Developmental Biology* (Macmillan, New York, 1984), p. 269.
2. K. V. Anderson, G. Jurgens, C. Nusslein-Volhard, *Cell* **42**, 779 (1985).
3. C. Nusslein-Volhard, in *Determinants of Spatial Organization*, S. Subteiny and I. R. Konigsberg, Eds. (Academic Press, New York, 1979), p. 185.
4. P. W. Ingham, K. R. Howard, D. Ish-Horowicz, *Nature (London)* **318**, 439 (1985).
5. R. Steward, L. Ambrosio, T. Schupbach, in *Molecular Approaches to Developmental Biology*, R. E. Firtel and E. H. Davidson, Eds. (Liss, New York, 1987), p. 39.
6. K. V. Anderson and C. Nusslein-Volhard, in *Gametogenesis and the Early Embryo*, J. G. Gall, Ed. (Liss, New York, 1986), p. 177.
7. ———, *Nature (London)* **311**, 223 (1984).
8. P. Santamaria and C. Nusslein-Volhard, *EMBO J.* **2**, 1695 (1983).
9. R. Steward, F. McNally, P. Schedl, *Nature (London)* **311**, 262 (1984).
10. R. Steward, L. Ambrosio, P. Schedl, *Cold Spring Harbor Symp. Quant. Biol.* **50**, 223 (1985).
11. J. S. Poole, L. M. Kauvar, B. Drees, T. Kornberg, *Cell* **40**, 37 (1985).
12. E. Frei *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **50**, 127 (1985.
13. D. J. Lipman and W. R. Pearson, *Science* **227**, 1435 (1985).
14. Bionet: National Computer Resource for Molecular Biology, Intelligenetics, Palo Alto, CA.
15. R. M. Stephens, N. R. Rice, R. R. Hiebsch, H. R. Bose, R. V. Gilden, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6225 (1983); K. C. Wilhelmsen, K. Eggleton, H. M. Temin, *J. Virol.* **52**, 172 (1984).
16. H. Beug *et al.*, *Virology* **115**, 295 (1981).
17. T. D. Gilmore and H. M. Temin, *Cell* **44**, 791 (1986).
18. E. Brownell, S. J. O'Brien, W. G. Nash, N. R. Rice, *Mol. Cell. Biol.* **5**, 2826 (1985).
19. W. McGinnis, M. S. Levine, E. Hafen, A. Kuriowa, W. J. Gehring, *Nature (London)* **308**, 428 (1984); K. A. Wharton, B. Yedvobnick, V. G. Finnerty, S. Artavanis-Tsakonas, *Cell* **40**, 55 (1985).
20. K. A. Wharton, K. M. Johansen, T. Xu, S. Artavanis-Tsakonas, *Cell* **43**, 567 (1985).
21. M. Sumner-Smith, J. A. Rafalski, T. Sugiyama, M. Stoll, D. Soell, *Nucleic Acids Res.* **13**, 3905 (1985).
22. I. S. Y. Chen, K. C. Wilhelmsen, H. M. Temin, *J. Virol.* **45**, 104 (1983).
23. T. D. Gilmore, personal communication.
24. R. DeLotto, P. Spierer, *Nature (London)* **323**, 688 (1986).
25. I. Greenwald, *Cell* **43**, 583 (1985); R. W. Padgett, D. St. Johnston, W. M. Gelbart, *Nature (London)* **325**, 81 (1987); E. Hafen, K. Basler, J.-E. Edstroem, G. M. Rubin, *Science* **236**, 55 (1987); F. Rijsewik *et al.*, *Cell* **50**, 649 (1987).
26. Stratagene Cloning Systems, San Diego, CA.
27. S. Henikoff, *Gene* **28**, 351 (1984).
28. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
29. Single-letter abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

30. I thank P. Schedl for his interest and enthusiasm, D. Mullen for help with sequencing and computer analysis, L. Bell, L. Brennan, I. Greenwald, and M. Resh for discussions and comments, and the Horace W. Goldsmith Foundation for assistance. Supported by a grant from the NSF and grants from the March of Dimes Birth Defect Foundation and the American Cancer Institute to P. Schedl.

20 July 1987; accepted 16 September 1987