Isolation of the Proto-oncogene c-myb from D. melanogaster

Alisa L. Katzen,* Thomas B. Kornberg,* and
J. Michael Bishop†
* Department of Biochemistry and Biophysics
† Department of Microbiology and Immunology and the
G. W. Hooper Foundation
University of California Medical Center
San Francisco, California 94143

Summary

We have isolated the proto-oncogene c-myb from Drosophila melanogaster. This gene is represented by a single locus at position 13E-F on the X chromosome, and is expressed in early embryos by transcription into two polyadenylated RNAs with lengths of approximately 3.0 and 3.8 kb. The gene may encode a protein with a molecular weight of at least 55,000 that shares a domain with c-myb(chicken) in which 91 of 125 (or 73%) of the amino acids are identical in the Drosophila and chicken genes. These findings represent the first rigorous identification of a Drosophila proto-oncogene that can encode what may be a nuclear protein, and they set the stage for a genetic analysis of how c-myb serves the normal organism.

Introduction

The oncogenes of retroviruses arose by transduction of cellular genes now known as proto-oncogenes or cellular oncogenes (reviewed in Bishop, 1983; Bishop and Varmus, 1982 and 1985). Representative proto-oncogenes have been found in every metazoan species examined to date, and the family of ras proto-oncogenes is also represented in Saccharomyces cerevisiae (Gallwitz et al., 1983; DeFeo-Jones et al., 1983; Powers et al., 1984). The remarkable conservation of individual proto-oncogenes across vast distances of evolutionary time has prompted the hypothesis that these genes provide vital functions for the organisms in which they are found. Moreover, several lines of evidence indicate that abnormalities of proto-oncogenes may contribute to the development of malignant tumors (reviewed in Varmus, 1984). It is therefore presumed that, in their normal state, proto-oncogenes are components of the genetic program that regulates the multiplication and differentiation of cells.

In order to explore the functions of proto-oncogenes, we have turned to the use of Drosophila melanogaster. Shilo and Weinberg first reported the presence of proto-oncogenes in the genome of Drosophila (1981), and six of these genes have now been isolated from Drosophila DNA—three representing the family of c-ras (Newman-Silberberg et al., 1984) and three related to c-src (Simon et al., 1983; Hoffman-Falk et al., 1983; Hoffman et al., 1983). All of these proto-oncogenes encode proteins that are likely to reside and act at the plasma membrane. The products of c-ras gene bind (Papageorge et al., 1982) and

hydrolyze (McGrath et al., 1984; Sweet et al., 1984) GTP while influencing the activity of adenyl cyclase (Kataoka et al., 1985; Toda et al., 1985), whereas c-src and its relatives encode tyrosine-specific protein kinases whose physiological functions are presently obscure (Hunter, 1984; Bishop and Varmus, 1982, 1985).

We now report the isolation and preliminary characterization of a Drosophila gene that corresponds to the proto-oncogene c-myb, described previously in birds and mammals (Klempnauer et al., 1982; Leprince et al., 1983; Franchini et al., 1983). Both c-myb and the retroviral oncogene v-myb, engendered by c-myb, encode proteins that are found in the nuclear matrix of the cell (Klempnauer et al., 1984; unpublished data of G. Ramsay and J. M. Bishop). The action of these proteins appears to be directed toward hemopoietic tissues, as these are the principal tissues in which expression of the cellular gene has been found (Gonda et al., 1982; Westin et al., 1982; Sheiness and Gardinier, 1984) and in which the viral gene induces tumors (Moscovici, 1975; Moscovici et al., 1981; Graf and Beug, 1978; Durban and Boettiger, 1981). Our findings enlarge the variety of proto-oncogenes found in D. melanogaster and set the stage for a genetic analysis of how c-myb serves the normal organism.

Results

Detection of c-myb in the DNA of D. melanogaster

The proto-oncogene c-myb was first encountered as a chicken gene that shared nucleotide sequences with the oncogene v-myb of avian myeloblastosis virus (Roussel et al., 1979; Klempnauer et al., 1982). The shared sequences represent six complete exons and part of a seventh exon in chicken c-myb (Figure 1A). In an effort to detect c-myb in Drosophila DNA, we used an isolated fragment from a molecular clone that encompasses the entirety of v-myb (diagrammed in Figure 1A) to prepare a radioactive probe. We then hybridized the probe, under conditions of low stringency, to restriction fragments of Drosophila DNA that had been fractionated by electrophoresis in agarose gels (Figure 2A). A single fragment was detected in DNA that had been digested with Eco RI, Hind III, or a combination of these two enzymes. We concluded that Drosophila might possess a single genetic locus with appreciable homology to v-myb.

Isolation of c-myb(Drosophila)

The v-myb probe (Figure 1A) was used to screen a library of Drosophila DNA, prepared in the lambda phage Charon 4A. A single genetic locus was isolated repeatedly as a series of overlapping clones. Each of these clones contained a 2.45 kb Eco RI fragment that hybridized with the radioactive probe for v-myb (see Figure 2D). By contrast, a 4.8 kb Eco RI fragment was detected in blots of Drosophila DNA (Figure 2A). This discrepancy can be attributed to the use of different strains of D. melanogaster in the two procedures. The c-myb loci of these strains are distinguished by
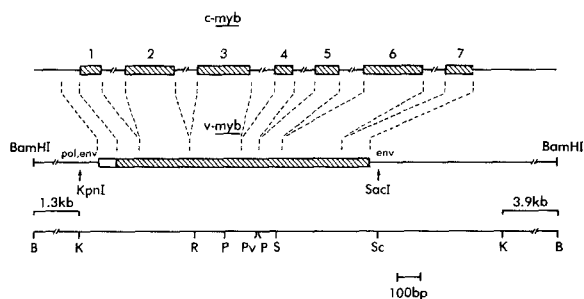
Figure 1. The Topographies of c-myb(chicken) and v-myb

Schematic representations of c-myb(chicken) and v-myb(AMV). Hatched boxes numbered 1 to 7 denote known exons of c-myb(chicken). The open box represents a small portion of v-myb that is derived from an intron of c-myb (Klempnauer et al., 1982; Klempnauer and Bishop, 1983a). The remainder of v-myb is derived from exons of c-myb, shown in the hatched boxes. Portions of viral structural genes in the AMV genome are denoted by pol and env. The designated Kpn I and Sac I sites define the 1.3 kb fragment that was used for preparation of v-myb probes. The restriction map of v-myb is shown below the schematic representation, with the following cleavage sites: B, Bam HI; K, Kpn I; P, Pst I; Pv, Pvu II; R, Eco RI; S, Sal I; and Sc, Sac I.

a restriction site polymorphism for Eco RI (unpublished observations of the authors).

The DNA encompassed by the molecular clones was mapped with restriction endonucleases and by molecular hybridization with the radioactive probe for v-myb (Figure 3). These data and subsequent results with nucleotide sequencing (see below), indicated that the bulk of homology with v-myb is on a 0.35 kb domain of Drosophila DNA defined by Sau 3A and Eco RI restriction sites (Figure 3). Analysis of restriction digests with a probe prepared from this Sau 3A–Eco RI fragment revealed identical patterns of fragments from either Drosophila DNA or the clones of Drosophila DNA in phage lambda (Figures 2C and 2D). We concluded that the molecular clones had suffered no substantial damage during their isolation.

In an effort to detect Drosophila DNA related more distantly to v-myb, the probe for the Sau 3A–Eco RI fragment was also hybridized to restriction fragments of Drosophila DNA under conditions of low stringency. The resulting patterns were identical to those obtained by hybridization with either the same probe under conditions of high stringency (data not shown) or a probe for v-myb under conditions of low stringency (Figures 2A and 2B). These data indicate that Drosophila contains a single genetic locus with detectable homology to v-myb and that part or all of this locus has been isolated without substantial perturbation of its structure.

c-myb(Drosophila) Is Located on the X Chromosome of D. melanogaster

We determined the position of the c-myb(Drosophila) locus on the polytene chromosomes of D. melanogaster by hybridization in situ. The locus mapped to a single site near the border of bands 13E and 13F on the X chromosome. Similar results were obtained with either a ^{3}H-labeled probe prepared from a lambda clone of c-myb(Drosophila) (Figure 4) or a ^{35}S-labeled probe prepared from
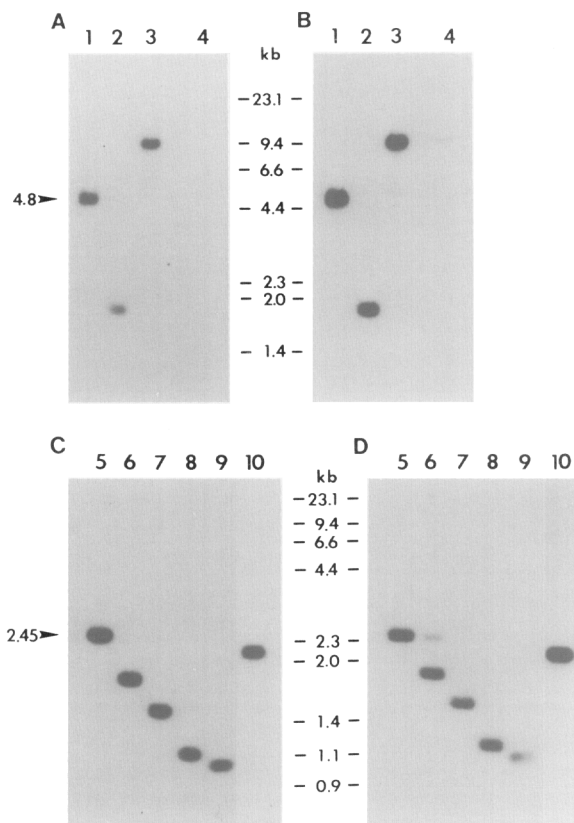

Figure 2. Identification of c-myb(Drosophila)

The autoradiograms were exposed for 2 days (A), 1 day (B), 2 days (C), and 1 hr (D). Molecular weight markers represented by Hind III fragments of λ phage DNA and Hae III fragments of φX-phage DNA are indicated in kilobases. Also indicated are the sizes of the hybridizing Eco RI fragments from each of the two strains.

(A) and (B) Hybridization with a probe for v-myb. D. melanogaster genomic DNA (strain: yellow, multiple wing hair, javelin; 8 μg) was digested with Eco RI (lane 1), Eco RI and Hind III (lane 2), Hind III (lane 3), and 2 μg of S. cerevisiae was digested with Eco RI (lane 4), fractionated on a 0.8% agarose gel, transferred to nitrocellulose, and hybridized under low stringency conditions with the ^{32}P-labeled Kpn I–Sac I fragment of v-myb (A) and the ^{32}P-labeled Eco RI–Sau 3A fragment of c-myb(Drosophila) (B).

(C) and (D) Hybridization with a probe for c-myb(Drosophila). D. melanogaster genomic DNA (strain: Canton S; 3 μg) (C), and 0.2 μg of phage DNA prepared from one of the Drosophila genomic clones, λM8 (the library was prepared with Canton S DNA) (D), were digested with Eco RI (lane 5), Eco RI and Hind III (lane 6), Eco RI and Cla I (lane 7), Eco RI and Xho I (lane 8), Eco RI and Sal I (lane 9), and Bam HI and Hind III (lane 10), fractionated on a 1.0% agarose gel, transferred to nitrocellulose, and hybridized under stringent conditions with the ^{32}P-labeled Eco RI–Sau 3A fragment of c-myb(Drosophila).

a plasmid subclone of the 2.45 kb Eco RI fragment that carries the homology with v-myb (data not shown).

Transcription of c-myb(Drosophila)

In order to detect and characterize transcription of c-myb (Drosophila), subclones in phage M13 were prepared that represented separately the two strands of the 0.35 kb Sau 3A–Eco RI fragment from c-myb(Drosophila) (see Figure 3). Radioactive probes prepared with these clones were used to analyze polyadenylated RNA isolated from a pool of Drosophila embryos that had been collected 0 to 6 hr
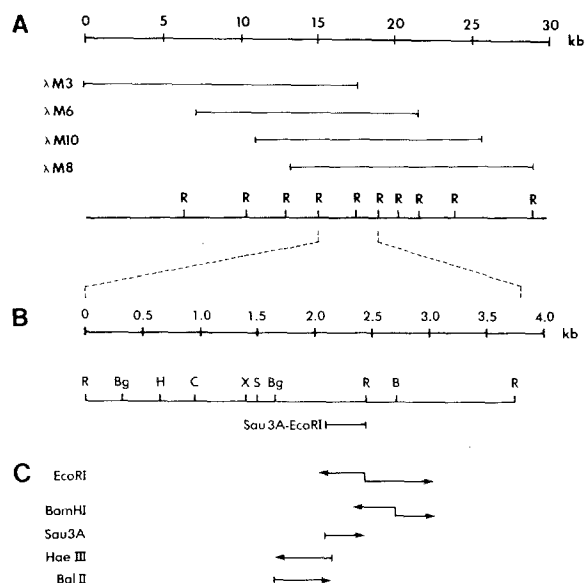
Figure 3. The Mapping and Sequencing of c-myb(Drosophila)

(A) Topography of phage clones. The relative positions of the inserts from four of the overlapping lambda phage clones isolated when a library of D. melanogaster DNA was screened with a v-myb probe are shown. Locations of Eco RI sites are indicated.

(B) Restriction map of the region containing homology to v-myb. The designated restriction fragment defined by Sau 3A and Eco RI sites contains the bulk of the homology and was used for preparation of c-myb(Drosophila) probes. Key to restriction enzyme cleavage sites: B, Bam HI; Bg, Bgl II; C, Cla I; H, Hind III; R, Eco RI; S, Sal I; X, Xho I.

(C) Sequencing strategy. Fragments were cloned into M13 phage vectors and sequenced by the Sanger primer extension method. The restriction enzyme cleavage site from which each sequence was started is noted on the left.

Figure 4. Cytogenetic Mapping of c-myb(Drosophila)

The phage clone λM10 was nick-translated with tritiated nucleotides and hybridized to salivary gland polytene chromosomes prepared from third-instar Oregon R larvae. A single site of hybridization was detected in each nucleus and is designated by an arrow. The entire X chromosome is shown in (A), the pertinent region enlarged in (B).

after oviposition (Figure 5). The results indicated that c-myb(Drosophila) is transcribed in a single direction, from left to right as the gene is portrayed in Figure 3. The stable products of transcription include a predominant species with a size of approximately 3.8 kb and a much less abundant RNA of approximately 3.0 kb (Figure 5).

The Nucleotide Sequence of c-myb(Drosophila)

We focused our sequencing efforts on the region of c-myb (Drosophila) that displays detectable homology with v-myb. The strategy for sequencing is outlined in Figure 3. The nucleotide sequence is portrayed as the strand of DNA that presumably er codes the protein of c-myb(Drosophila) (Figure 6). The coding strand was identified by reference to both the transcription of the gene (Figure 5) and the sequence of c-myb(chicken) (Klempnauer et al., 1982).

Alignment of the c-myb(Drosophila) sequence with the available sequence of c-myb(chicken) revealed homology between the two genes that extends from the first nucleotide of the first identified codon of the chicken gene to residue 61 within the third known exon of the chicken gene. Across that domain, 253 of 375 nucleotides (or 67%) are identical in the Drosophila and chicken genes (Figure 6). This domain includes the entire region within v-myb that hybridized with a probe prepared from a phage clone of c-myb(Drosophila) (data not shown). However, the homology with vertebrate c-myb may be more extensive than
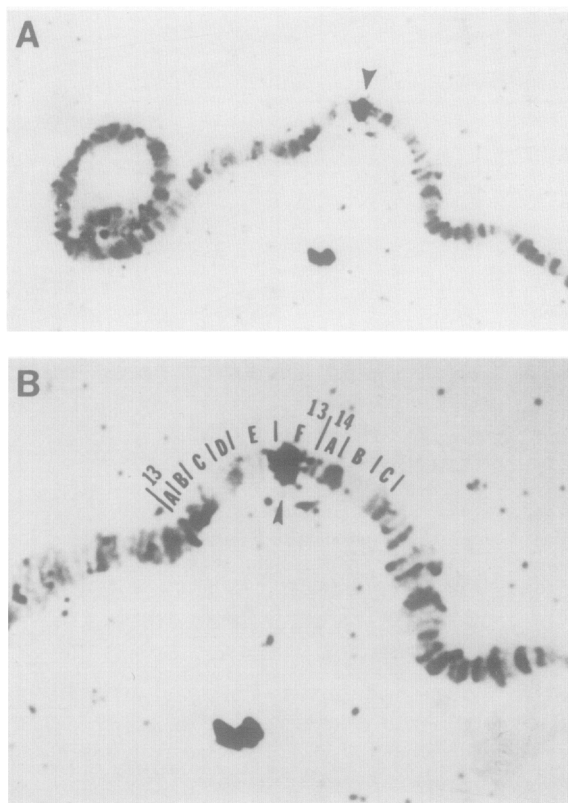
illustrated here because additional, presently unsequenced exons lie upstream in the chicken gene (Gonda and Bishop, 1983). In order to align the sequences of the chicken and Drosophila genes, it was necessary to exclude two previously identified introns from the chicken sequence. By contrast, the sequence of the Drosophila gene could be used without interruption to achieve the alignment. It therefore appears that c-myb(Drosophila) lacks at least two of the introns found in the homologous chicken gene.

The Amino Acid Sequence Encoded by c-myb(Drosophila)

The nucleotide sequence encodes an open reading frame that begins at residue 39 and extends rightward to the end of the sequence at residue 1430 (Figure 6). Additional provisional data (not shown) suggest that this same frame may continue for another 54 nucleotides or more, so the available sequence may encode a protein containing at least 482 amino acids (Figure 7). The protein may be even larger, since the reading frame may extend into regions of c-myb(Drosophila) that we have yet to identify and sequence (see Discussion). The five other possible reading frames are all interrupted repeatedly by termination codons, suggesting that none of these frames encodes a protein (analysis not shown).
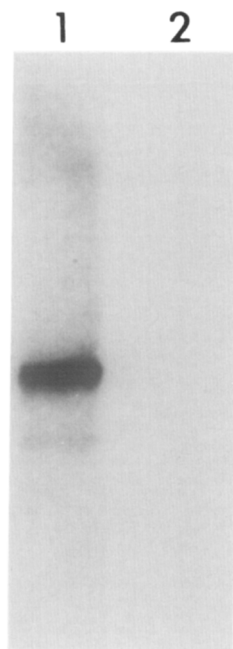
Figure 5. Transcription of c-myb(Drosophila)

Polyadenylated RNA from embryos collected between 0 and 6 hr after oviposition was fractionated on an agarose gel, transferred to a nitrocellulose filter, and hybridized with strand-specific ^{32}P-labeled DNA of the Sau 3A–Eco RI fragment of c-myb(Drosophila) (see Figure 3). Lane 2 was hybridized with the probe for the strand for which the nucleotide sequence is portrayed in Figure 6, and lane 1 with the probe for the other strand.

The homologous regions of amino acid sequences encoded by the chicken and Drosophila c-myb genes are aligned as shown in Figure 7 and the nonidentical residues in the sequence encoded by v-myb are as designated. Beginning with the first residue of the aligned sequences, the Drosophila and chicken proteins are identical at 91 of 125 amino acids (or 73%), including stretches of 13, 17, and 20 uninterrupted identities (residues 151–163, 191–207, and 240–259). Of the 34 nonidentical amino acids in this domain, 22% are either conservative or neutral substitutions according to the categories defined by McLachlan (1971). Downstream of this domain, however, the sequences of the chicken and Drosophila proteins show little or no resemblance to one another.

Can c-myb Be Detected in the Genome of S. cerevisiae

The discovery of a c-myb gene in Drosophila indicates that its origins predate the evolutionary divergence that gave rise to the arthropods. To investigate the possibility that c-myb might have even more ancient origins, we searched for myb-related sequences in the genome of S. cerevisiae. No hybridization was detected when restriction fragments of yeast DNA were fractionated by electrophoresis in agarose gels, transferred to nitrocellulose filters, and hybridized with either a v-myb or a c-myb(Drosophila) probe under conditions of low stringency (Figure 2A lane 4 and Figure 2B lane 4). Tenfold longer exposure than that shown in Figure 2 also failed to reveal any signal from the

lanes containing yeast DNA, despite the fact that fivefold more yeast genomes than Drosophila genomes were represented. In addition, no hybridization was detected when either of two libraries of S. cerevisiae DNA prepared in lambda phage vectors was screened with a v-myb probe under conditions of low stringency. Therefore, it appears that under these hybridization conditions, sequences homologous to v-myb cannot be detected in the genome of S. cerevisiae.

Discussion

Identification and Isolation of c-myb(Drosophila)

We report here the identification and characterization of a Drosophila gene that appears to be a homolog of the vertebrate proto-oncogene c-myb. The gene was identified by the presence of a domain of 375 nucleotides in which 253 residues are identical to those found in a homologous region of c-myb(chicken). We suspect that the full extent of c-myb(Drosophila) has not been identified, since the size of the principal RNA transcribed from c-myb(Drosophila) is appreciably longer than the region of the gene we have characterized (approximately 3.8 kb vs. approximately 1.5 kb). It is therefore possible that the homology between the c-myb genes of chicken and Drosophila is more extensive than what we have demonstrated here.

Many of the identified proto-oncogenes belong to small families of genes. For example, six or more related proto-oncogenes that encode tyrosine-specific protein kinases have been identified, and there are at least three forms of c-ras in yeast, Drosophila, and mammals (reviewed in Bishop and Varmus, 1985). By contrast, the work reported here indicates that c-myb is a solitary genetic locus in D. melanogaster; we were unable to find even extensively divergent duplications of the gene. These findings conform to previous work showing that birds and mammals also appear to possess single versions of c-myb (Bergmann et al., 1981; Klempnauer et al., 1982; Leprince et al., 1983).

The Topography of c-myb(Drosophila)

The portion of c-myb(Drosophila) characterized here is occupied almost entirely by an uninterrupted open reading frame. The gene appears, therefore, to lack at least two of the introns found in c-myb(chicken). Differences among the topographies of homologous genes in Drosophila and vertebrates have been described previously for other proto-oncogenes (Hoffman et al., 1983; Neuman-Silberberg et al., 1984), and for highly conserved structural proteins such as actin and collagen (Fyrberg et al., 1981; Zakut et al., 1982; Monson et al., 1982). Removal of the introns from c-myb(chicken) leaves an open reading frame that aligns precisely with the open reading frame of the homologous Drosophila gene within the conserved domain. Either the introns entered c-myb after the evolutionary divergence that engendered the insect and vertebrate lineages, or the presence of introns antedated the divergence and they were removed with great precision prior to the emergence of Drosophila.
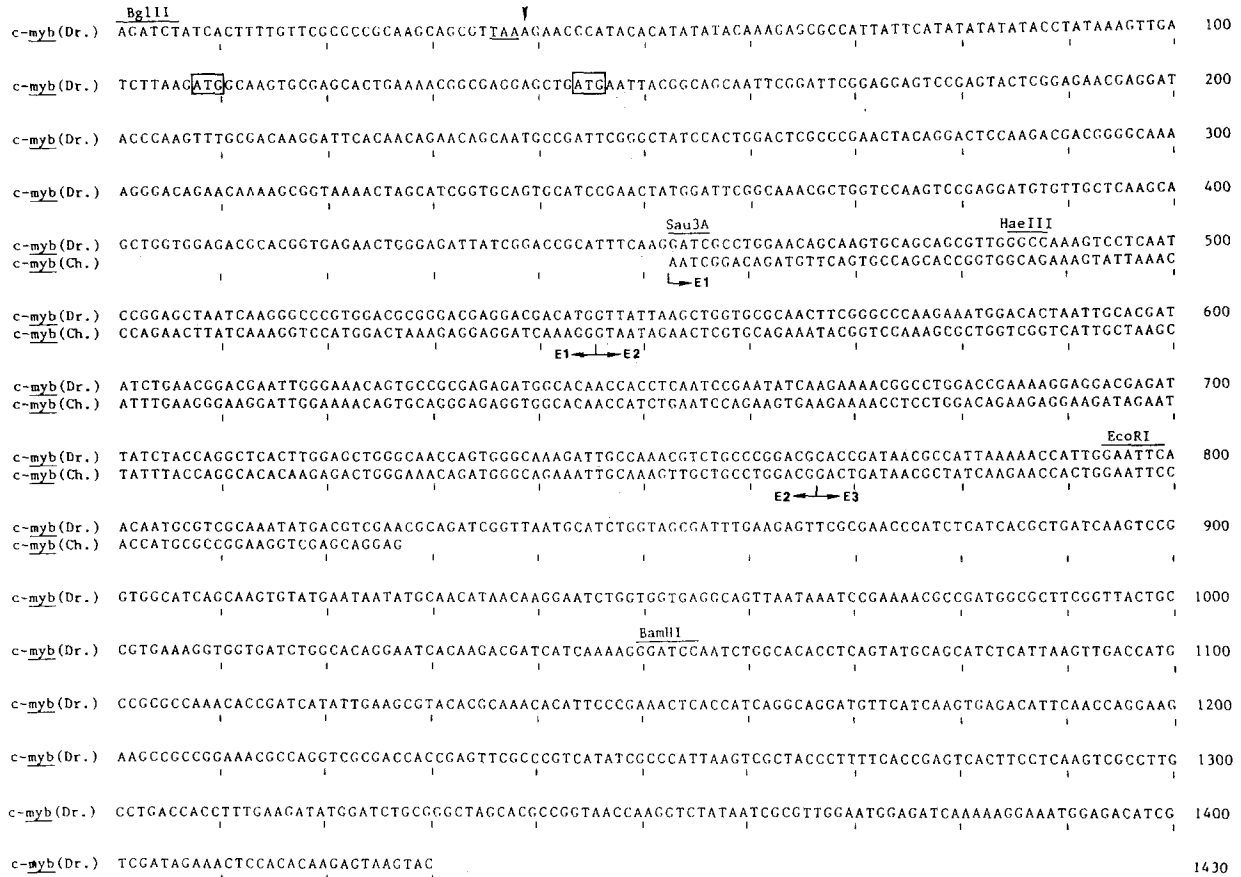
We cannot now define the boundaries of the coding do-

```
          BglII                                        ▼
c-myb(Dr.) AGATCTATCACTTTTGTTCGCCCCGCAAGCAGCGTTAAAGAACCCATACACATATATACAAAGAGCGCCATTATTCATATATATATACCTATAAAGTTGA   100

c-myb(Dr.) TCTTAAG[ATG]GCAAGTGCGAGCACTGAAAACGGCGAGGAGCTG[ATG]AATTACGGCAGCAATTCGGATTCGGAGCAGTCCGAGTACTCGGAGAACGAGGAT   200

c-myb(Dr.) ACCCAAGTTTGCGACAAGGATTCACAACAGAACAGCAATGCCGATTCGGGCTATCCACTGGACTCGCCCGAACTACAGGACTCCAAGACGACGGGCAAA   300

c-myb(Dr.) AGGGACAGAACAAAAGCGGTAAAACTAGCATCGGTGCAGTGCATCCGAACTATGGATTCGGCAAACGCTGGTCCAAGTCCGAGGATGTGTTGCTCAAGCA   400

                                                        Sau3A                              HaeIII
c-myb(Dr.) GCTGGTGGAGACGCACGGTGAGAACTGGGAGATTATCGGACCGCATTTCAAGGATCGCCTGGAACAGCAAGTGCAGCAGCGTTGGGCCAAAGTCCTCAAT   500
c-myb(Ch.)                                              AATCGGACAGATGTTCAGTGCCAGCACCGGTGGCAGAAAGTATTAAAC
                                                          L→E1

c-myb(Dr.) CCGGAGCTAATCAAGGGCCCGTGGACGCGGGACGAGGACGACATGGTTATTAAGCTGGTGCGCAACTTCGGGCCCAAGAAATGGACACTAATTGCACGAT   600
c-myb(Ch.) CCAGAACTTATCAAAGGTCCATGGACTAAAGAGGAGGATCAAACGGTAATAGAACTCGTGCAGAAATACGGTCCAAAGCGCTGGTCGGTCATTGCTAAGC
                                              E1→⊥→E2

c-myb(Dr.) ATCTGAACGGACGAATTGGGAAACAGTGCCCGCGAGAGATGGCACAACCACCTCAATCCGAATATCAAGAAAACGGCCTGGACCGAAAAGGAGCACGAGAT   700
c-myb(Ch.) ATTTGAAGGGAAGGATTGGAAAACAGTGCAGGGAGAGGTGGCACAACCATCTGAATCCAGAAGTGAAGAAAACCTCCTGGACAGAAGAGGAAGATAGAAT

                                                                                                EcoRI
c-myb(Dr.) TATCTACCAGGCTCACTTGGAGCTGGGCAACCAGTGGGCAAAGATTGCCAAACGTCTGCCCCGGACGGCACCGATAACGCCATTAAAAACCATTGGAATTCA   800
c-myb(Ch.) TATTTACCAGGCACACAAGAGACTGGGAAACAGATGGGCAGAAATTGCAAACTTGCTGCCTGGACGGACTGATAACGCTATCAAGAACCACTGGAATTCC
                                                          E2→⊥→E3

c-myb(Dr.) ACAATGCGTCGCAAATATGACGTCGAACGCAGATCGGTTAATGCATCTGGTAGCGATTTGAAGAGTTCGCCGAACCCATCTCATCACGCTGATCAAGTCCG   900
c-myb(Ch.) ACCATGCGCCGGAAGGTCCAGCAGGAG

c-myb(Dr.) GTGGCATCAGCAAGTGTATGAATAATATGCAACATAACAAGGAATCTGGTGGTGAGGCAGTTAATAAATCCGAAAACGCCGATGGCGCTTCGGTTACTGC   1000

                                                        BamHI
c-myb(Dr.) CGTGAAAGGTGGTGATCTGGCACAGGAATCACAAGACGATCATCAAAACGGATCCAATCTGGCACACCTCAGTATGCAGCATCTCATTAAGTTGACCATG   1100

c-myb(Dr.) CCGCGCCAAACACCGATCATATTGAAGCGTACACGCAAACACATTCCCGAAACTCACCATCAGGCAGGATGTTCATCAAGTGAGACATTCAACCAGGAAG   1200

c-myb(Dr.) AAGCCGCCGGAAACGCCAGCTCGCGACCACCGAGTTCGCCCGTCATATCGCCCATTAAGTCGCTACCCTTTTCACCGAGTCACTTCCTCAAGTCGCCTTG   1300

c-myb(Dr.) CCTGACCACCTTTGAAGATATGGATCTGCGGGCTAGCACGCCGGTAACCAAGCTCTATAATCGCCGTTGGAATGGAGATCAAAAAGGAAATGGAGACATCG   1400

c-myb(Dr.) TCGATAGAAACTCCACACAAGAGTAACTAC                                                                       1430
```

Figure 6. Nucleotide Sequence of c-myb(Drosophila)

The sequence of c-myb(Drosophila) is presented as the predicted coding strand, designated c-myb [Dr]. The nucleotides are numbered such that residue number one is the first 5' residue of available sequence. The restriction endonuclease cleavage sites that were used in the preparation of M13 subclones for sequencing (see Figure 3) are indicated. A single termination codon that is present in the predicted reading frame is underlined (residues 36–38), and the first and second codons for methionine (residues 108–110 and 144–146) are boxed. The arrowhead designates the beginning of the open reading frame, which continues throughout the remainder of the displayed sequence. c-myb(Drosophila) displays homology to c-myb(chicken) between residues 453 and 827. Beneath this region, the relevant portion of the c-myb(chicken) sequence is portrayed, designated c-myb[Ch.], beginning at the first nucleotide of the first identified codon in the chicken gene. Intron sequences are not shown, but the boundaries of the exons are designated by E1, E2, and E3 (Klempnauer et al., 1982).

main in c-myb(Drosophila). The open reading frame does not terminate within the sequence reported here, so the carboxy terminus of the encoded protein is not in view. By contrast, there are two AUG codons near the beginning of the open reading frame (nucleotide residues 108–110 and 144–146). The sequence surrounding the first of these AUGs displays the two most highly conserved features of eukaryotic initiation sites (Kozak, 1984a, 1984b), viz., an adenosine at position −3 and a guanosine at position +4. In contrast, the sequence surrounding the second AUG does not display any of the conserved features. For the moment, however, we cannot exclude the possibility that splicing bypasses the termination codon at nucleotide residues 36–38 and extends the open reading frame upstream of the domain we have characterized.

Expression of c-myb(Drosophila)

c-myb(Drosophila) is expressed in embryonic tissues. The gene is transcribed in a single direction, and the RNA generated by transcription has the polarity required to carry the coding domain suggested by the nucleotide sequence of the gene. The data presented here show expression only in early embryos (0–6 hr following oviposition). However, preliminary work indicates that the gene is expressed in roughly equal abundance at all major stages of Drosophila development (data not shown).

The c-myb gene of vertebrates encodes a 75 kilodalton (kd) protein found in the nucleus of cells (Klempnauer et al., 1983b, 1984). Since the complete coding domain of c-myb(Drosophila) has not yet been characterized, we cannot offer a full description of the encoded protein. However, the available data suggest that the Drosophila protein has a mass of at least approximately 55 kd. We have no indication as to the biochemical function of either the Drosophila or vertebrate gene product.

The Evolutionary Conservation of c-myb

At least four proto-oncogenes and/or the retroviral oncogenes that they engender encode proteins found in the nucleus of the cell; namely, myb, myc, fos, and ski (re-

```
                                          ▼               ▼
c-myb(Dr.)   RTHTHIYKERHYSYIYTYKVDLKMASASTENGEELMNYGS   40

c-myb(Dr.)   NSDSEESEYSENEDTQVCDKDSQQNSNADSGYPLDSPELQ   80

c-myb(Dr.)   DSKTTGQKGQNKSGKTSIGAVHPNYGFCKRWSKSEDVLLK  120

c-myb(Dr.)   QLVETHGENWEIIGPHFKDRLEQQVQQRWAKVLNPELIKG  160
c-myb(Ch.)              NRTDVQCQHRWQKVLNPELIKG
                                          N

c-myb(Dr.)   PWTRDEDDMVIKLVRNFGPKKWTLIARYLNGRIGKQCRER  200
c-myb(Ch.)   PWTKEEDQRVIELVQKYGPKRWSVIAKHLKGRIGKQCRER
                  H                  D

c-myb(Dr.)   WHNHLNPNIKKTAWTEKEDEIIYQAHLELGNQWAKIAKRL  240
c-myb(Ch.)   WHNHLNPEVKKTSWTEEEDRIIYQAHKRLGNRWAEIAKLL

c-myb(Dr.)   PGRTDNAIKNHWNSTMRRKYDVERRSVNASGSDLKSSRTH  280
c-myb(Ch.)   PGRTDNAIKNHWNSTMRRKVEQE
                   V

c-myb(Dr.)   LITLIKSGGISKCMNNMQHNKESGGEAVNKSENADGASVT  320

c-myb(Dr.)   AVKGGDLAQESQDDHQKGSNLAHLSMQHLIKLTMPRQTPI  360

c-myb(Dr.)   ILKRTRKHIPETHHQAGCSSSETFNQEEAAGNARSRPPSS  400

c-myb(Dr.)   PVISPIKSLPFSPSHFLKSPCLTTFEDMDLRASTPVTKVY  440

c-myb(Dr.)   NRVGMEIKKEMETSSIETPHKSKY                  464
```

Figure 7. Deduced Amino Acid Sequence of c-myb(Drosophila)
The deduced open reading frame was translated beginning at nucleotide residue 39 (after the UAA termination codon; see Figure 6) and continuing to the end of the available nucleotide sequence. The positions of the methionines encoded by the first and second AUG codons are indicated by arrowheads. c-myb(Drosophila) displays homology to c-myb(chicken) between amino acid residues 139 and 263. The corresponding region of the protein encoded by the chicken gene (Klempnauer et al., 1982) is portrayed beneath the Drosophila sequence. Residues that are identical in the two sequences are boxed. The four amino acid substitutions between v-myb and this portion of c-myb (chicken) are indicated below the chicken sequence. Amino acids are designated by the conventional single-letter code: A, alanine; R, arginine; N, asparagine; D, aspartic acid; C, cysteine; E, glutamic acid; Q, glutamine; G, glycine; H, histidine; I, isoleucine; L, leucine; K, lysine; M, methionine; F, phenylalanine; P, proline; S, serine; T, threonine; W, tryptophan; Y, tyrosine; V, valine.

viewed in Bishop and Varmus, 1985). There are several lines of evidence suggesting that the functions of myb and myc may be related: the two genes display distant but perceptible similarities between their structures (Ralston and Bishop, 1983); they both encode proteins that reside in the matrix of the nucleus (Eisenman et al., 1985; unpublished work of G. Ramsay and J. M. Bishop); and they both cooperate with a mutant version of the proto-oncogene c-ras in the transformation of rat embryo cells (Land et al., 1983; H. Land and L. F. Parada, personal communication). This description of c-myb(Drosophila) represents the first rigorous demonstration that a member of this class of proto-oncogenes can be found in invertebrates.

Only a portion of chicken and Drosophila c-myb has been conserved during evolution. The conserved domain comprises a minimum of 125 amino acids, 91 of which are identical in the chicken and Drosophila proteins and 28 of which are conservative or neutral substitutions. The same domain represents one of two principal regions of conservation when the chicken and human c-myb genes are compared (L. Souza, personal communication), and also corresponds to a tandem duplication of sequence identified previously in v-myb and c-myb(chicken) (Ralston and

Bishop, 1983). The conservation of c-myb can be dramatized further by comparison to v-myb. Within the conserved domain of c-myb, the amino acid sequences of the chicken gene and v-myb differ by substitutions at four amino acid residues (Figure 7). At three of these positions, the chicken and Drosophila v-myb genes are identical; the fourth position has sustained a conservative substitution (Leu in Drosophila, Val in chicken). By contrast, three of the four substitutions in v-myb are nonconservative. The divergence of v-myb at these positions presumably arises from the propensity of retroviral genomes to mutate during replication (Varmus and Swanstrom, 1982) and from the absence of the selective pressures that have conserved the Drosophila and chicken genes. The domain of c-myb conserved in Drosophila, birds, and mammals presumably represents a portion of the myb protein that is essential for its function. However, the possibility must be considered that a functional domain might be conserved between proteins with different physiological roles. An example of this phenomenon is a domain thought to facilitate aggregation that is conserved between the small heat-shock proteins of Drosophila and the α-crystallins of mammals (Ingolia and Craig, 1982). However, the extent and degree of homology between these proteins are more limited than what we have found with c-myb.

How ancient is c-myb? Identification of the gene in Drosophila implies that the gene was extant at the time of the evolutionary divergence that gave rise to insects, that is, approximately 800 million years ago. However, we have failed to find c-myb in S. cerevisiae, and it is therefore possible that, in contrast to c-ras, c-myb arose with the emergence of metazoan organisms.

The Physiological Role of c-myb

Vertebrate c-myb is expressed principally in hemopoietic cells (Gonda et al., 1982; Westin et al., 1982; Sheiness and Gardinier, 1984), a fact that is thought to reflect the specialized function of the gene. Can c-myb have an analogous function in Drosophila, whose hemopoietic system is far less complex than that of vertebrates (Rizki, 1978)? We hope to pursue this issue by isolating and analyzing mutants of c-myb(Drosophila). Two features of the gene may be helpful: it is located at 13E–F on the X chromosome in the vicinity of previously identified genetic markers; and it is represented by a single genetic locus, a fact that should preclude complementation of mutations in c-myb by other genes. The exploration of c-myb in Drosophila may help to illuminate principles that govern the genesis of developmental pathways during the course of evolution.

Experimental Procedures

Molecular Cloning

translation to a specific activity of about 1×10^8 cpm/μg (Rigby et al., 1977). Filters were preincubated at 42°C for 4–12 hr in hybridization solution: 35% formamide (v/v), 3 × SSC (1 × SSC is 0.15 M sodium chloride, 0.015 M sodium citrate, pH 7.0), 20 mM Hepes pH 7.0, 2.5 × Denhardt's solution (1 × Denhardt's solution is 0.2 mg/ml Ficoll, 0.2 mg/ml polyvinylpyrrolidone, 0.2 mg/ml BSA), and 0.2 mg/ml salmon sperm DNA. Filters were hybridized in fresh hybridization solution containing 2–6×10^5 cpm/ml of radioactively labeled probe at 42°C for 36–48 hr. The filters were rinsed in 4 × SSC, 0.1% SDS, and 0.1% sodium pyrophosphate for 2–4 hr at 50°C.

Characterization of isolated clones and subcloning of fragments into the plasmid vector PUC8 were performed by standard procedures (Maniatis et al., 1982).

Analysis of Cloned and Genomic DNAs
DNA was prepared from phage and plasmid clones by standard procedures (Maniatis et al., 1982). D. melanogaster cellular DNA was prepared from adult flies by a procedure that involves homogenization of frozen flies, removal of large debris by filtration through cheesecloth, pelleting of nuclei through sucrose, and extraction of DNA from the isolated nuclei. Yeast cellular DNA was provided by C. Hammond. DNA was digested with restriction endonucleases purchased from New England Biolabs or Boehringer Mannheim and used according to the suppliers' instructions. Cleaved DNA was fractionated by agarose gel electrophoresis and transferred to nitrocellulose filters by the method of Southern (1975). Conditions for low stringency hybridizations were the same as those used for screening of phage libraries. Hybridization under high stringency conditions differed from that under low stringency conditions as follows: the hybridization solution contained 50% formamide and the filters were washed in 0.1 × SSC, 0.5% SDS, 0.1% sodium pyrophosphate for 2–4 hr at 68°C.

In Situ Hybridization
The method of Pardue and Gall (1975), as modified by Bonner and Pardue (1976), was used for the preparation of, and in situ hybridization to, salivary gland polytene chromosomes from third-instar Oregon R larvae. Two radioactive probes for the c-*myb*(Drosophila) locus were prepared. ^{3}H-labeled DNA was prepared from one of the lambda phage clones, λM10, by nick translation to a specific activity of approximately 10^7 cpm/μg DNA, and ^{35}S-labeled DNA was prepared from a plasmid subclone containing the 2.4 kb Eco RI fragment (see Figures 1C, 1D, and 3) by nick translation to a specific activity of approximately 10^6 cpm/μg DNA.

Analysis of c-*myb*(Drosophila) RNA
Polyadenylated RNA was isolated from Drosophila embryos as described by Poole et al. (1985). The RNA was electrophoretically fractionated on an agarose–formaldehyde gel and transferred to nitrocellulose (Maniatis et al., 1982). Strand-specific ^{32}P-labeled c-*myb* (Drosophila) probes were prepared by the method of B. Drees (personal communication). The 0.35 kb Sau 3A–Eco RI fragment designated in Figure 3 was cloned into the M13 phage vectors mp8 and mp9 (Messing and Vieira, 1982). A 17 base single-stranded oligonucleotide purchased from P. L. Biochemicals was annealed to the single-stranded template (Messing and Vieira, 1982), and the primer extension method was used to extend the strand in the presence of [^{32}P]dCTP (Sanger et al., 1977). The double-stranded DNA was cleaved at the far end of the insert with a restriction enzyme, and the newly synthesized, labeled DNA was isolated from a low-melting-temperature agarose gel after denaturation and electrophoretic fractionation. The conditions used for hybridization and washing were the same as those described for Southern blot analysis at high stringency.

Sequence Analysis
All DNA sequencing was obtained by the chain terminator technique (Sanger et al., 1977) after fragments were subcloned into the M13 phage vectors mp8 or mp9 (Messing and Vieira, 1982). The reported sequence was obtained from at least two independent, overlapping clones and/or from both strands of DNA, as depicted in Figure 3.

Acknowledgments

References

Benton, W. D., and Davis, R. W. (1977). Screening lambda-gt recombinant clones by hybridization to single plaques *in situ*. Science *196*, 180–182.

Bergmann, D. G., Souza, L. M., and Baluda, M. A. (1981). Vertebrate DNAs contain nucleotide sequences related to the transforming gene of avian myeloblastosis virus. J. Virol. *40*, 450–455.

Bishop, J. M. (1983). Cellular oncogenes and retroviruses. Ann. Rev. Biochem. *52*, 301–354.

Bishop, J. M., and Varmus, H. E. (1982). Functions and origins of retroviral transforming genes, in Molecular Biology of Tumor Viruses, Part III RNA Tumor Viruses, R. Weiss, N. Teich, H. Varmus, and J. Coffin, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Press), pp. 999–1108.

Bishop, J. M., and Varmus, H. E. (1985). Functions and origins of retroviral transforming genes, in Molecular Biology of Tumor Viruses, Part III RNA Tumor Viruses, R. Weiss, N. Teich, H. Varmus, and J. Coffin, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Press), in press.

Bonner, J. J., and Pardue, M. L. (1976). Ecdyson-stimulated RNA synthesis in imaginal discs of Drosophila melanogaster. Chromosoma *58*, 87–99.

DeFeo-Jones, D., Scolnick, E. M., Koller, R., and Dhar, R. (1983). ras-related gene sequences identified and isolated from *Saccharomyces cerevisiae*. Nature *306*, 707–709.

Durban, E. M., and Boettiger, D. (1981). Differential effects of transforming avian RNA tumor viruses on avian macrophages. Proc. Natl. Acad. Sci. USA *78*, 3600–3604.

Eisenman, R. N., Tachibana, C. Y., Abrams, H. D., and Hann, S. R. (1985). v-*myc* and c-*myc*-encoded proteins are associated with the nuclear matrix. Mol. Cell. Biol. *5*, 114–126.

Franchini, G., Wong-Staal, F., Beluda, M. A., Lengel, C., and Tronick, S. R. (1983). Structural organization and expression of human DNA sequences related to the transforming gene of avian myeloblastosis virus. Proc. Natl. Acad. Sci. USA *80*, 7385–7389.

Fyrberg, E. A., Bond, B. J., Hershey, N. D., Mixter, K. S., and Davidson, N. (1981). The actin genes of Drosophila: protein coding regions are highly conserved but intron positions are not. Cell *24*, 107–116.

Gallwitz, D., Donath, C., and Sander, C. (1983). A yeast gene encoding a protein homologous to the human c-*has*/*bas* proto-oncogene product. Nature *306*, 704–707.

Gonda, T. J., and Bishop, J. M. (1983). Structure and transcription of the cellular homolog (c-*myb*) of the avian myeloblastosis virus transforming gene (v-*myb*). J. Virol. *46*, 212–220.

Gonda, T. J., Sheiness, D. K., and Bishop, J. M. (1982). Transcripts from the cellular homologs of retroviral oncogenes: distribution among chicken tissues. Mol. Cell. Biol. *2*, 617–624.

Graf, T., and Beug, H. (1978). Avian leukemia viruses. Interaction with their target cells *in vivo* and *in vitro*. Biochim. Biophys. Acta *516*, 269–299.

Hoffman, F. M., Fresco, L. D., Hoffman-Falk, H., and Shilo, B.-Z. (1983). Nucleotide sequences of the Drosophila src and abl homologs: con-

servation and variability in the src family oncogenes. Cell 35, 393–401.

Hoffman-Falk, H., Einat, P., Shilo, B.-Z., and Hoffman, F. M. (1983). Drosophila melanogaster DNA clones homologous to vertebrate oncogenes: evidence for a common ancestor to src and abl cellular genes. Cell 32, 589–598.

Hunter, T. (1984). The Proteins of Oncogenes. Sci. Amer. 251, 70–79.

Ingolia, T. D., and Craig, E. A. (1982). Four small Drosophila heat shock proteins are related to each other and to mammalian alpha-crystallin. Proc. Natl. Acad. Sci. USA 79, 2360–2364.

Kataoka, T., Powers, S., Cameron, S., Fasano, O., Goldfarb, M., Broach, J., and Wigler, M. (1985). Functional homology of mammalian and yeast RAS genes. Cell 40, 19–26.

Klempnauer, K.-H., and Bishop, J. M. (1983a). Transduction of c-myb into avian myeloblastosis virus: locating points of recombination within the cellular gene. J. Virol. 48, 565–572.

Klempnauer, K.-H., Gonda, T. J., and Bishop, J. M. (1982). Nucleotide sequence of the retrovirus leukemia gene v-myb and its cellular progenitor c-myb: the architecture of a transduced oncogene. Cell 31, 453–463.

Klempnauer, K.-H., Ramsay, G., Bishop, J. M., Moscovici, M. G., Moscovici, C., McGrath, J. P., and Levinson, A. D. (1983b). The product of the retroviral transforming gene v-myb is a truncated version of the protein encoded by the cellular oncogene c-myb. Cell 33, 345–355.

Klempnauer, K.-H., Symonds, G., Evan, G. I., and Bishop, J. M. (1984). Subcellular localization of proteins encoded by oncogenes of avian myeloblastosis virus and avian leukemia virus E26 and by the chicken c-myb gene. Cell 37, 537–547.

Kozak, M. (1984a). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNA's. Nucl. Acids Res. 12, 857–872.

Kozak, M. (1984b). Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. Nature 308, 241–246.

Land, H., Parada, L. F., and Weinberg, R. A. (1983). Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Nature 304, 596–601.

Leprince, D., Saule, S., deTaisne, C., Gegonne, A., Begue, A., Righi, M., and Stehelin, D. (1983). The human DNA locus related to the oncogene myb of avian myeloblastosis virus (AMV): molecular cloning and structural characterization. EMBO. J. 2, 1073–1078.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., and Efstratiadis, A. (1978). The isolation of structural genes from libraries of eucaryotic DNA. Cell 15, 687–701.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, New York: Cold Spring Harbor Press).

McGrath, J. P., Capon, D. J., Goeddel, D. V., and Levinson, A. D. (1984). Comparative biochemical properties of normal and activated human ras p21 protein. Nature 310, 644–649.

McLachlan, A. D. (1971). Tests for Comparing Related Amino-acid Sequences. Cytochrome c and Cytochrome c_{551}. J. Mol. Biol. 61, 409–424.

Messing, J., and Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19, 269–276.

Monson, J. M., Natzle, J., Friedman, J., and McCarthy, B. J. (1982). Expression and novel structure of a collagen gene in Drosophila. Proc. Natl. Acad. Sci. USA 79, 1761–1765.

Moscovici, C. (1975). Leukemic transformation with avian myeloblastosis virus: present status. Curr. Topics Microbiol. Immunol. 71, 79–101.

Moscovici, C., Samarut, J., Gazzolo, L., and Moscovici, M. G. (1981). Myeloid and erythroid neoplastic responses to avian defective leukemia viruses in chickens and in quail. Virology 113, 765–768.

Neuman-Silberberg, F. S., Schejter, E., Hoffmann, F. M., and Shilo, B.-Z. (1984). The Drosophila ras oncogenes: structure and nucleotide sequence. Cell 37, 1027–1033.

Papageorge, A., Lowy, D., and Scolnick, E. M. (1982). Comparative bio-

chemical properties of p21 ras molecules coded for by viral and cellular ras genes. J. Virol. 44, 509–519.

Pardue, M. L., and Gall, J. G. (1975). Nucleic acid hybridization to the DNA of cytological preparations. Meth. Cell. Biol. 10, 1–16.

Poole, S. J., Kauvar, L. M., Drees, B., and Kornberg, T. (1985). The engrailed locus of Drosophila: structural analysis of an embryonic transcript. Cell 40, 37–43.

Powers, S., Kataoka, T., Fasano, O., Goldfarb, M., Strathern, J., Broach, J., and Wigler, M. (1984). Genes in S. cerevisiae encoding proteins with domains homologous to the mammalian ras proteins. Cell 36, 607–612.

Ralston, R., and Bishop, J. M. (1983). The protein products of the oncogenes myc, myb and adenovirus E1a are structurally related. Nature 306, 803–806.

Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977). Labeling of DNA to high specific activity by nick-translation. J. Mol. Biol. 113, 237–258.

Rizki, T. M. (1978). The circulatory system and associated cells and tissues. In The Genetics and Biology of Drosophila, Volume 2b, M. Ashburner and T. R. F. Wright, eds. (New York: Academic Press), pp. 397–452.

Roussel, M., Saule, S., Lagrou, C., Rommens, C., Beug, H., Graf, T., and Stehelin, D. (1979). Three new types of viral oncogene of cellular origin specific for haematopoietic cell transformation. Nature 281, 452–455.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.

Sheiness, D., and Gardinier, M. (1984). Expression of a proto-oncogene (proto-myb) in hemopoietic tissues of mice. Mol. Cell. Biol. 4, 1206–1212.

Shilo, B.-Z., and Weinberg, R. A. (1981). Genes homologous to vertebrate oncogenes are conserved in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA 78, 6789–6792.

Simon, M. A., Kornberg, T. B., and Bishop, J. M. (1983). Three loci related to the src oncogene and tyrosine-specific protein kinase activity in Drosophila. Nature 302, 837–839.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503–517.

Sweet, R. W., Yokoyama, S., Kamata, T., Feramisco, J. R., Rosenberg, M., and Gross, M. (1984). The product of ras is a GTPase and the T24 oncogenic mutant is deficient in this activity. Nature 311, 273–275.

Toda, T., Uno, I., Ishikawa, T., Powers, S., Kataoka, T., Broek, D., Cameron, S., Broach, J., Matsumoto, K., and Wigler, M. (1985). In yeast, RAS proteins are controlling elements of adenylate cyclase. Cell 40, 27–36.

Varmus, H. E. (1984). The molecular genetics of cellular oncogenes. Ann. Rev. Genet. 18, 553–612.

Varmus, H. E., and Swanstrom, R. (1982). Replication of retroviruses. In The Molecular Biology of Tumor Viruses. Part III. RNA Tumor Viruses, R. Weiss, N. Teich, H. Varmus, and J. Coffin, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 369–512.

Westin, E. H., Gallo, R. C., Arya, S. K., Eva, A., Souza, L. M., Baluda, M. A., Aaronson, S. A., and Wong-Staal, F. (1982). Differential expression of the AMV gene in human hematopoietic cells. Proc. Natl. Acad. Sci. USA 79, 2194–2198.

Zakut, R., Shani, M., Givol, D., Neuman, S., Yaffe, D., and Nudel, U. (1982). Nucleotide sequence of the rat skeletal muscle actin gene. Nature 292, 857–859.