Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 3698–3702, June 1986
Biochemistry

Cellular *myc* (c-*myc*) in fish (rainbow trout): Its relationship to other vertebrate *myc* genes and to the transforming genes of the MC29 family of viruses

(homologues of viral oncogenes/DNA sequence analysis/teleost oncogenes)

R. J. VAN BENEDEN*, D. K. WATSON, T. T. CHEN†, J. A. LAUTENBERGER, AND T. S. PAPAS‡

Laboratory of Molecular Oncology, National Cancer Institute, Frederick, MD 21701

Communicated by William F. Harrington, January 29, 1986

ABSTRACT **We have isolated, cloned, and sequenced the rainbow trout (*Salmo gairdneri*) c-*myc* gene. The presumptive coding region of the trout c-*myc* gene shows extensive homology to the c-*myc* genes of chicken, mouse, and human. Comparison of nucleotide sequences reveals that human, mouse, chicken, and trout c-*myc* genes contain at least two coding exons, interrupted by introns of decreasing size of 1.38 kilobases (kb), 1.2 kb, 0.97 kb, and 0.33 kb, respectively. The exons are clearly delineated by donor-acceptor splice signals. The degree of nucleotide homology between trout, chicken, and human exon II is less than that observed for exon III. However, the greatest homology among these three genes is localized to two specific regions within exon II (*myc* boxes A and B). At the predicted amino acid level, fish c-*myc* shows considerable homology to vertebrate c-*myc* gene products. Trout c-*myc* is expressed in normal trout cells as a single 2.3-kb mRNA species, similar in size to other vertebrate transcripts.**

It is now widely accepted that retroviral oncogenes arose by transduction of portions of cellular genes known as proto-oncogenes. Typically, transforming retroviruses are defective in viral gene function, as a result of the substitution of viral gene(s) by portions of these proto-*onc* genes. In the presence of appropriate helper virus, they are able to initiate and maintain a variety of malignancies in a number of animal species after very short incubation periods and with a high degree of efficiency (1, 2).

The viral oncogene, v-*myc*, was first identified as a part of the transforming sequence of avian myelocytomatosis virus (MC29) (3, 4). The MC29 virus is a replication-defective retrovirus, lacking the entire polymerase (*pol*) gene and portions of the *gag* and envelope (*env*) genes, which have been substituted by the cell-derived *myc* sequence. MC29 induces a broad spectrum of malignant disease, including myelocytomas, renal and liver tumors, and, less typically, carcinomas, sarcomas, and erythroblastosis. This virus also induces morphological transformation of fibroblasts, epithelial cells, and macrophages in culture (1, 5–7). Cellular sequences homologous to v-*myc* have been identified by Southern blot analysis in a wide variety of species (8). However, prior to this report, nucleotide sequence analysis to prove that this hybridization is due to *myc* sequences had only been achieved in chicken (9), mouse (10), and human (11). In addition, it now appears that higher eukaryotes may possess multiple loci with limited *myc* homology, such as N-*myc* and L-*myc* (12, 13). N-*myc* has two distinct regions of homology to the second exon of c-*myc* (*myc* boxes A and B), which may represent a conserved functional domain or binding site region. In general, the cellular genes transduced by the retroviruses represent a family of genes that have been conserved throughout evolution, suggesting that they have an important role in cellular growth and/or differentiation (2). Therefore, comparison of *myc* gene sequences from very divergent species is an important step toward our understanding of the evolution and function of this gene. The most primitive class of vertebrates so far found by hybridization analysis to have *myc* gene sequences are the teleosts (bony fishes). We initiated our studies by cloning and sequencing the c-*myc* gene from rainbow trout (*Salmo gairdneri*). Rainbow trout were chosen because they are a commercially valuable food and sport fish, easily obtainable, and frequently used in studies of chemical-induced carcinogenesis (14). In addition, studies have reported the widespread occurrence of neoplasia in feral fish populations (15). Their phylogenetic position and high incidence of neoplasia make teleosts ideal model organisms for the study of both oncogene evolution and tumorigenesis.

MATERIALS AND METHODS

Isolation of Genomic Clones and Sequence Analysis. A genomic library was prepared from rainbow trout testes DNA, by partially digesting with *Eco*RI and cloning into Charon 4A. The library was propagated in *Escherichia coli* strain LE392 and screened by the method of Benton and Davis (16). Clones were identified initially by hybridization to v-*myc* under relaxed conditions (30% formamide/0.74 M NaCl/0.05 M NaH_2PO_4/0.005 M EDTA, pH 7.4, 37°C). Positive clones were further characterized by endonuclease restriction patterns and hybridization to specific probes from the chicken c-*myc* exon II [190-base-pair (bp) *Pst* I/*Hin*fI fragment) or exon III (849-bp *Cla* I/*Eco*RI fragment) (9). By these criteria, 20 positive clones were categorized into six different groups of overlapping clones. A 3.7-kb *Eco*RI DNA fragment (which contained sequences homologous to both chicken exon II and III) was isolated from λ clone RTC-1 and subcloned into pBR325. The purified insert from this subclone (named C181) was used to generate appropriate labeled restriction enzyme DNA fragments for sequence analysis of both strands by the method of Maxam and Gilbert (17) as described earlier (18, 19).

Sequence Comparison. Simultaneous comparisons of the deduced fish c-*myc* amino acid sequence with those of chicken and human were performed as described (20), using the matrix of pairwise amino acid comparisons (21). The gap penalty used was 8 (no special weighting was used for matched histidine, cysteine, or methionine residues).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: bp, base pair(s); kb, kilobase(s).

*Guest Researcher from the Chesapeake Bay Institute of the Johns Hopkins University, Shadyside, MD 20764. Current address: Program Resources, Inc., Frederick, MD 21701.
†On sabbatical from the Department of Biology, McMaster University, Hamilton, ON L8S 4K1, Canada.
‡To whom reprint requests should be addressed.

RESULTS

Molecular Cloning and Nucleotide Sequence Analysis of the Trout c-*myc* Gene. Twenty rainbow trout *myc*-related clones were isolated that hybridized to probes derived from chicken exon II or exon III (see *Materials and Methods*). Only three of these clones contained sequence homology to both chicken exons (22), and one (RTC-1) was subcloned (C181) and used for further analyses. Examination of the nucleotide sequence of the rainbow trout c-*myc* (Fig. 1) reveals two exons (II and III), as defined both by sequence homology to the second and third exons of the other c-*myc* genes and by consensus splice signals at positions I-1 and I-332. Trout exon II is shorter (664 bp) than exon II of chicken (702 bp), mouse (757 bp), and human (771 bp) (9–11). Alignment of the trout c-*myc* DNA sequence with that of chicken or human indicates that three codons are missing in the C181 sequence, assuming no insertions or deletions occur in the unsequenced 5′ region (see Fig. 3). If this assumption is true, trout exon II would have a coding capacity of 224 amino acids (672 bp). Trout exon III is longer (578 bp) than the third exon of chicken, mouse, and human, which all have 561 nucleotides.

Consistent with other c-*myc* genes, trout exons II and III are interrupted by an intron. The trout second intron (332 bp) is considerably shorter than those of chicken (971 bp), mouse [1.2 kilobases (kb)], or human (1.38 kb). This observation suggests a correlation of a smaller intron size in the c-*myc* gene in evolutionarily more primitive organisms.

The first c-*myc* exon has been localized for the chicken (23, 24), mouse (10), and human (24, 25) genes. Using DNA probes derived from either chicken or human exon I, we have been unable to detect homologous sequences in the fish, suggesting a more rapid divergence of these sequences relative to exons II and III.

Rainbow Trout c-*myc* Is Transcriptionally Active. Sequences homologous to *myc* are expressed in avian and mammalian tissues (26–28). Illustrated in Fig. 2 is an RNA blot of total trout liver RNA hybridized to a trout c-*myc* probe. This probe (1.5-kb *Eco*RI/*Pst* I fragment of C181) contains nearly the entire second and all of the third exon (see Fig. 4). A single transcript of ≈2.3 kb was detected, slightly smaller than the 2.5-kb species reported for avian and mammalian *myc*.

Comparison of the Nucleotide and Predicted Protein Sequences of Cellular and Viral *myc* Genes. The sequences of human (11) and chicken (9) cellular *myc* genes as well as those of the avian acute transforming retroviruses MC29 (30), MH2 (31), and OK10 (32) have been reported. We sought to make pairwise comparisons of the trout c-*myc* sequences with those of the chicken and human. To obtain a consistent alignment between the trout, chicken, and human c-*myc* and the v-*myc* gene products, the predicted amino acid sequences were compared (19). The result of such an alignment is shown in Fig. 3. The high degree of nucleotide and amino acid homology between the three cellular sequences is evident from the data presented in Table 1. For all three sequences, the region homologous to the MC29 v-*myc* gene can be organized into two exons. Overall, the third exon is more highly conserved than the second exon at both the nucleotide and amino acid levels, as indicated by both the greater homology and the lower number of gaps required for optimal alignment. However, several segments of the second exon are very highly conserved among the three organisms, including the two *myc* box regions (A and B) homologous to

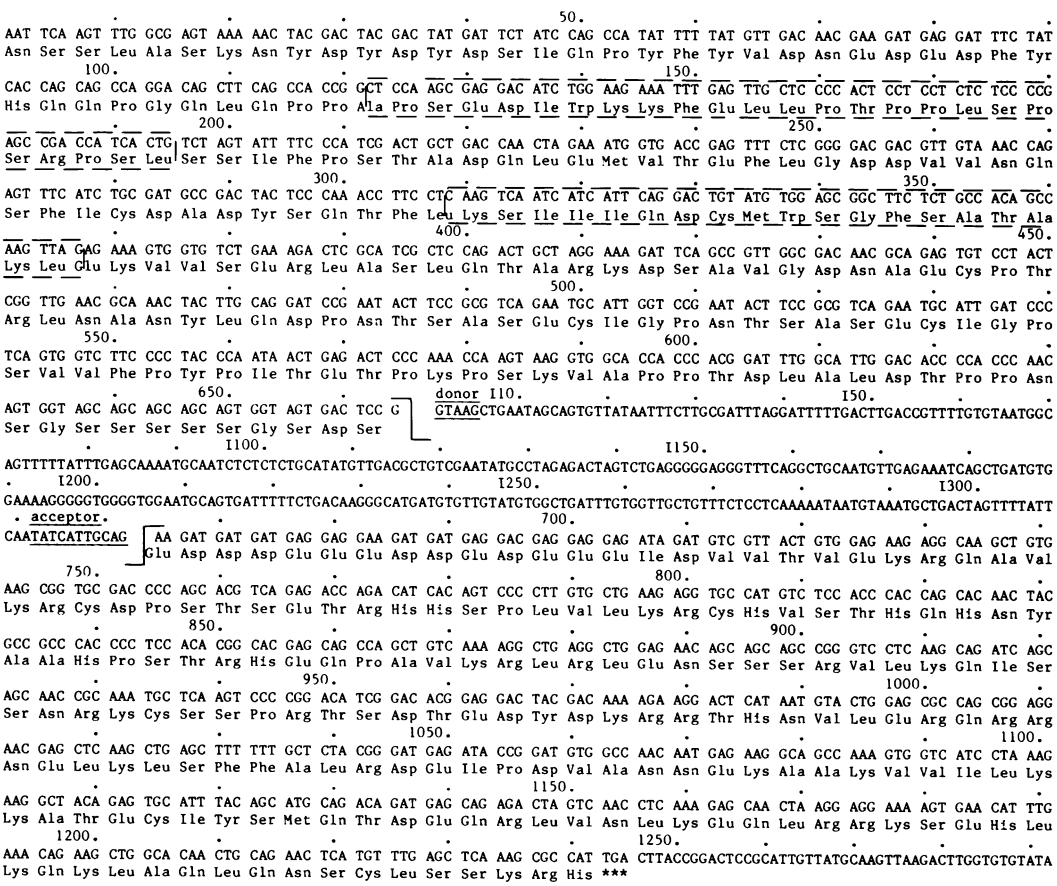

FIG. 1. Nucleotide sequence of rainbow trout c-*myc* locus. The positions of exons with *myc* homology (1–663, 664–1242), intron sequence (I-1 to I-332), and 3′ flanking cellular sequences are presented. The predicted amino acid sequence is given below the nucleotide sequence. Donor (⌐) and acceptor (¬) splice signals and the translational termination site are indicated. *myc* boxes are highlighted by dashed boxes.

FIG. 2. RNA blot of rainbow trout c-*myc*. Total RNA was isolated from rainbow trout liver, resolved in 1.2% formaldehyde/agarose gels (29), and hybridized to the 1.5-kb *Eco*RI/*Pst* I fragment of C181 under stringent conditions (50% formamide/0.75 M NaCl/0.075 M Na citrate, 42°C). Mobility of the 28S and 18S ribosomal RNAs are as indicated.

Table 1. Degree of *myc* homology between three cellular genes

	Triple matches	Trout vs. chicken	Trout vs. human	Chicken vs. human
		Amino acid		
Total gene	55	64	62	68
Exon II	52	60	58	64
Exon III	58	69	66	73
Box A	93	94	94	96
Box B	89	95	89	95
		Nucleic acid		
Total gene	53	62	60	70
Exon II	49	58	55	67
Exon III	58	68	66	74
Box A	70	74	77	81
Box B	81	85	83	92

The percentage homology was determined from alignment of the three sequences (see Fig. 3). The ratio used was the number of matches over the average number of non-gap residues in the region examined. If the constraint of having a consistent alignment between the three sequences is removed, it is likely that somewhat higher percentage matches could be obtained from the pairwise comparisons.

N-*myc* (12). It should be noted, however, that several other regions within the second exon as well as the central portion of the third exon are as highly conserved as these *myc* boxes, suggesting other significant functional domains in the *myc* gene. It is of interest to note that the three sequences of fish, chicken, and human each have inserted regions in the second exon unique to themselves (cf. Fig. 3). There are two such unique regions in the human sequences (HU1 and HU2),

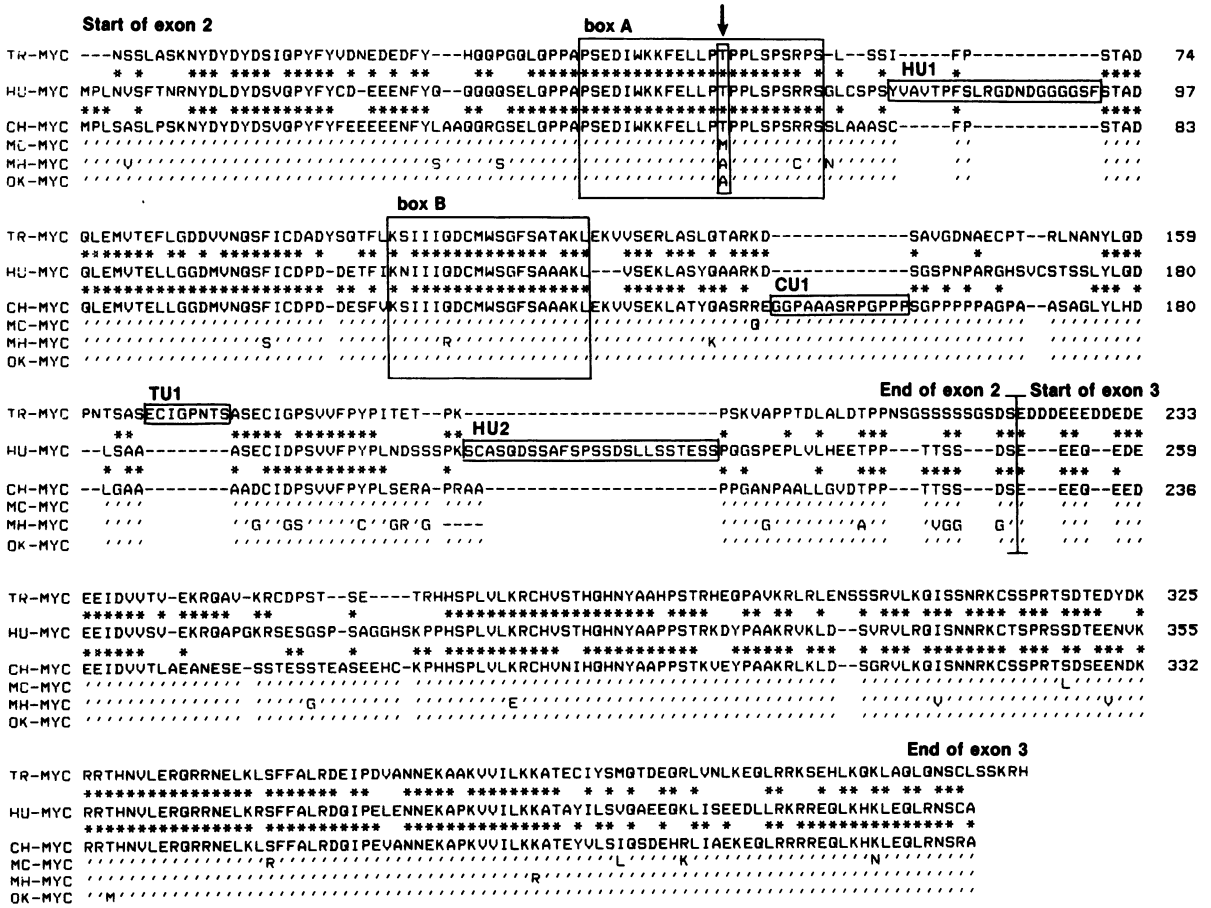

FIG. 3. Comparison of the deduced protein sequences of the rainbow trout (TR-MYC), human (HU-MYC), and chicken (CH-MYC), cellular *myc* genes with the v-*myc* genes of avian retroviruses MC29 (MC-MYC), MH2 (MH-MYC), and OK10 (OK-MYC). The amino acids are abbreviated by the standard one-letter code. Position 1 is assigned to the presumptive initiation codon in the second exon. Dashes represent gaps inserted by the three-protein alignment program (20), using the parameters described in *Materials and Methods*. Asterisks represent identity between human, trout, or chicken predicted amino acid sequences. Apostrophes represent identity of a viral residue with the corresponding chicken residue. The position (codon 61 in chicken) where each of the viral sequences differs from the chicken progenitor is highlighted by an arrow. The division between the second and third exons and the regions corresponding to *myc* boxes A and B are indicated.

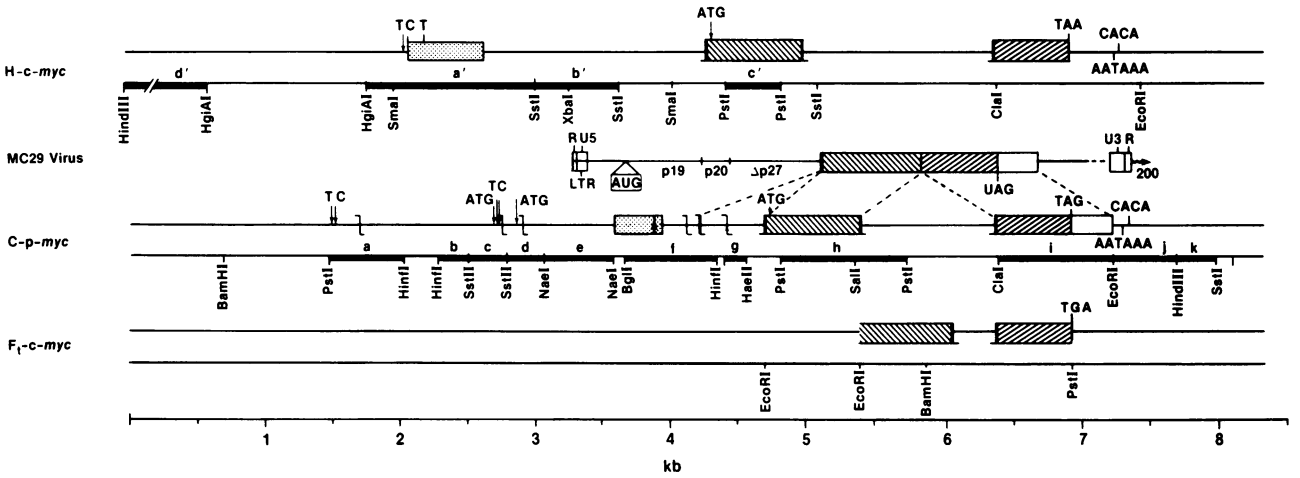

FIG. 4. Summary of the major features of the structure of the human (H-c-*myc*), chicken (C-p-*myc*), trout (F$_t$-c-*myc*), and MC29 *myc* genes. The hatched regions represent the second (▭) and third (▭) exons; stippled regions represent the first exon. The potential regulatory sites shown are as follows: T, TTATA box; C, capping site; AATAAA, polyadenylylation signal; CACA, poly(A) addition site; (⌊) and (⌋) represent donor and acceptor signals.

whereas the chicken and trout each have a single unique region (CU1 and TU1, respectively). Other than these noted insertions, discounting other smaller insertion and deletion events, the second and third exons seem to have gradually evolved by the accumulative acceptance of point mutations. The human and chicken sequences are somewhat more similar to each other than they are to fish. This observation is consistent with a phylogenetic model in which birds and mammals share a common lineage that diverged from the line that led to modern fishes (33).

DISCUSSION

Our results with rainbow trout demonstrate that a c-*myc* gene is present in a lower vertebrate and that it is transcriptionally active. The highlights of the *myc* genes are diagrammatically represented in Fig. 4 and can be summarized as follows: (*i*) Human, chicken, and trout c-*myc* genes contain at least two coding exons interrupted by introns of 1.38 kb, 0.97 kb, and 0.27 kb, respectively. This decrease in size, from human to fish, suggests an evolutionary correlation of smaller-sized introns in primitive organisms. Similar findings are apparent in comparisons of other proto-oncogene homologues. Reports of c-*myb* from *Drosophila* demonstrate that this gene lacks at least two of the introns found in the homologous chicken gene (34). Introns from chicken and mouse c-*ets* also follow this pattern (D.K.W., unpublished observation). (*ii*) In MC29, an open reading frame is observed extending from the ATG initiation codon located at the 5′ end of the *gag* gene into the *myc* sequence. The two exons in the fish, chicken, and human *myc* genes as defined by consensus donor–acceptor splice signals and by their alignment with the v-*myc* sequence share a common reading frame. The coding regions of the chicken and human genes terminate at the same position; however, the fish c-*myc* termination codon is found farther downstream. (*iii*) Several lines of evidence imply that mRNA from normal cells may be generated from proto-*myc* sequences. Direct examination of the nucleotide sequence of cloned *myc* gene from chicken, mouse, and human reveals consensus transcriptional signals. Hybridization studies using RNA extracted from appropriate cells show major *myc*-related species of 2.3–2.5 kb. (*iv*) Comparison of the predicted amino acid sequences of three avian retroviral v-*myc* genes (MC29, MH2, and OK10) to that of the chicken proto-oncogene from which they were derived and the other known c-*myc* oncogenes reveals a threonine at position 61 (numbered from the first ATG in chicken exon II) that is invariant in all c-*myc* genes but is substituted in v-*myc* genes by methionine (MC29) or alanine (MH2, OK10) (see Fig. 3 and ref. 35). It is tempting to draw a direct analogy of the base substitution in v-*myc* to that known for the *ras* gene codon 12 (36), which is associated with the transforming ability of *ras*. This idea appears to be in conflict with a recent report demonstrating that constructs containing either v-*myc* (containing *gag–myc* sequences) or c-*myc* (mouse plasmacytoma) are able to elicit the same transformation of rat embryo fibroblasts *in vitro* (37). However, the effect of this point mutation at position 61 on the *in vivo* spectrum of pathogenicity has not been ascertained.

Comparison of c-*myc* from fish, chicken, and human with the known v-*myc* oncogenes has facilitated our understanding of the evolution of this gene. Such comparison suggests that additional functional domains may exist. In higher vertebrates, amplification, rearrangement, and/or increased expression of the *myc* gene has been correlated with certain malignant conditions (38). Specific chromosomal translocations in mouse and human B-cell tumors have been reported to involve the c-*myc* locus (10, 39, 40). Isolation of the *myc* proto-oncogene from rainbow trout provides a probe to screen teleost tumors for analogous translocations and/or rearrangements of the *myc* oncogene in these systems.

We thank J. Maizel and J. Owens for assistance with the three-protein sequence alignment program. This work was supported in part by grants from Natural Sciences and Engineering Research Council (Canada) (A6527) and Department of Fisheries and Oceans to T.T.C. and Grant DEB 82-07006 from the National Institutes of Health and Maryland Sea Grant RF44 to D. A. Powers of the Chesapeake Bay Institute and the Department of Biology of the Johns Hopkins University.

1. Weiss, R. A., Teich, N. M., Varmus, H. E. & Coffin, J. M., eds. (1982) *RNA Tumor Viruses, Molecular Biology of Tumor Viruses*, 2nd Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
2. Vande Woude, G. F. & Gilden, R. V. (1985) in *Cancer: Principles and Practice of Oncology*, eds. De Vita, V. T., Jr., Hellman, S. & Rosenberg, S. A. (Lippincott, Philadelphia), Vol. 1, pp. 33–47.
3. Duesberg, P. H., Bister, K. & Vogt, P. K. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4320–4324.
4. Mellon, P., Pawson, A., Bister, K., Martin, G. S. & Duesberg, P. H. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5874–5878.
5. Graf, R. & Beug, H. (1978) *Biochim. Biophys. Acta* **516**, 269–299.
6. Ramsay, G., Graf, T. & Hayman, M. J. (1980) *Nature (London)* **288**, 170–172.

7. Bister, K. & Duesberg, P. H. (1982) *Adv. Viral Oncol.* **1**, 3–43.
8. Shilo, B.-Z. & Weinberg, R. A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6789–6792.
9. Watson, D. K., Reddy, E. P., Duesberg, P. H. & Papas, T. S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2146–2150.
10. Stanton, L. W., Fahrlander, P. D., Tesser, P. M. & Marcu, K. B. (1984) *Nature (London)* **310**, 423–425.
11. Watson, D. K., Psallidopoulos, M. C., Samuel, K. P., Dalla-Favera, R. & Papas, T. S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3642–3645.
12. Schwab, M., Alitalo, K., Klempnauer, K.-H., Varmus, H. E., Bishop, J. M., Gilbert, F., Brodeur, G., Goldstein, M. & Trent, J. (1983) *Nature (London)* **305**, 245–248.
13. Nau, M. M., Brooks, B. J., Battey, J., Sausville, E., Gazdar, A. F., Kirsch, I. R., McBride, O. W., Bertness, V., Hollis, G. F. & Minna, J. D. (1985) *Nature (London)* **318**, 69–73.
14. Hendricks, J. D., Meyers, T. R., Casteel, J. L., Nixon, J. E., Loveland, P. M. & Bailey, G. S. (1984) in *Use of Small Fish in Carcinogenicity Testing*, ed. Hoover, K. L. (National Cancer Institute, Frederick, MD), Monograph 65, pp. 129–137.
15. Sindermann, C. J. (1979) *Fish. Bull.* **76**, 717–749.
16. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
17. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
18. Rushlow, K. E., Lautenberger, J. A., Papas, T. S., Baluda, M. A., Perbal, B., Chirikjian, J. G. & Reddy, E. P. (1982) *Science* **216**, 1421–1423.
19. Reddy, E. P., Reynolds, K., Watson, D. K., Schulz, R., Lautenberger, J. & Papas, T. S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2500–2504.
20. Murata, M., Richardson, J. S. & Sussman, J. L. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3073–3077.
21. MacLachlan, A. D. (1971) *J. Mol. Biol.* **61**, 409–424.
22. Van Beneden, R. J., Watson, D. K., Chen, T. T. & Papas, T. S. (1986) in *Advances in Gene Technology: Molecular Biology of the Endocrine System* (Cambridge Univ. Press, Cambridge, U.K.), pp. 142–143.
23. Linial, M. & Groudine, M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 53–57.
24. Papas, T. S., Kan, N. K., Watson, D. K., Flordellis, C. S., Psallidopoulos, M. D., Lautenberger, J., Samuel, K. P. & Duesberg, P. (1984) in *Cancer Cells 2: Oncogenes and Viral Genes*, eds. Vande Woude, G. F., Levine, A. J., Topp, W. C. & Watson, J. D. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) pp. 153–163.
25. Hamlyn, P. H. & Rabbitts, T. H. (1983) *Nature (London)* **304**, 135–139.
26. Hayward, W. S., Neel, B. G. & Astrin, S. M. (1981) *Nature (London)* **290**, 475–479.
27. Slamon, D. J. & Cline, M. J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7141–7145.
28. Eva, A., Robbins, K. C., Anderson, P. R., Srinivasen, A., Tronick, S. R., Reddy, E. P., Ellmomre, N. W., Galen, A. T., Lautenberger, J. A., Papas, T. S., Westin, E. H. & Wong-Staal, F. (1982) *Nature (London)* **295**, 116–119.
29. Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977) *Biochemistry* **16**, 4743–4751.
30. Reddy, E. P., Reynolds, K., Watson, D. K., Schultz, R., Lautenberger, J. A. & Papas, T. S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2500–2504.
31. Kan, N. C., Flordellis, C. S., Garon, C. F., Duesberg, P. H. & Papas, T. S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6566–6570.
32. Hayflick, J., Seeburg, P. H., Ohlsson, R., Pfeifer-Ohlsson, S., Watson, D., Papas, T. & Duesberg, P. H. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2718–2722.
33. Romer, A. S. (1966) *Vertebrate Paleontology* (Univ. of Chicago Press, Chicago) p. 13.
34. Katzen, A. L., Kornberg, T. B. & Bishop, J. M. (1985) *Cell* **41**, 449–456.
35. Papas, T. S. & Lautenberger, J. A. (1985) *Nature (London)* **318**, 237.
36. Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. (1982) *Nature (London)* **300**, 149–152.
37. Mougneau, E., Lemieux, L., Rassoulzadegan, M. & Cuzin, F. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5758–5762.
38. Yunis, J. J. (1983) *Science* **221**, 227–236.
39. Erikson, J., ar-Rushdi, A., Drwinga, H. L., Nowell, P. C. & Croce, C. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 820–824.
40. Gelmann, E. P., Psallidopoulos, M. C., Papas, T. S. & Dalla-Favera, R. (1983) *Nature (London)* **306**, 799–803.