The EMBO Journal vol.6 no.11 pp.3359–3366, 1987

Structure and expression of the murine L-*myc* gene

Edith Legouy[1], Ronald DePinho, Kathryn Zimmerman, Robert Collum, George Yancopoulos, Lisa Mitsock[2], Ronald Kriz[2] and Frederick W.Alt

Departments of Biochemistry and Microbiology, College of Physicians and Surgeons of Columbia University, 630 West 168th Street, New York, NY 10032, and [2]Genetics Institute, 87 Cambridge Park Drive, Cambridge, MA 02131, USA

[1]Present address: Unité de Génétique Cellulaire de L'Institut Pasteur, 25 rue Dr Roux, Paris 75724, France

Communicated by E.Klein

We have isolated a 12 kb clone from the murine genome which we show by DNA transfection studies to contain an entire functional L-*myc* gene and the transcriptional promoter sequences necessary for its expression. We have also isolated a 3.1 kb cDNA sequence from a murine brain cDNA library which corresponds to most of the L-*myc* mRNA. We have identified the L-*myc* coding region within the genomic clone by a combination of S1 nuclease analyses, Northern blotting analyses and comparative nucleotide sequence analyses with the cDNA clone. The L-*myc* gene appears to be organized similarly to the other well-characterized *myc*-family genes, c-*myc* and N-*myc*. The predicted amino acid coding sequence of the L-*myc* gene indicates that the L-*myc* protein is significantly smaller than c- and N-*myc*, but is highly related. In particular, comparison of the N- and c-*myc* protein sequences reveals seven relatively conserved regions interspersed among non-conserved regions; the L-*myc* gene retains five of these conserved regions but lacks two others. In addition, a portion of one highly conserved region is encoded within a different region of the L-*myc* gene but, due to changes in the size of L-*myc* exons relative to those of N- and c-*myc*, maintains its overall position in the peptide backbone with respect to other conserved regions. We discuss these findings in the context of potential functional domains and the possibility of overlapping and distinct activities of *myc*-family proteins.

Key words: L-*myc* gene/structure/expression/mouse

Introduction

The *myc* family of cellular oncogenes is currently known to contain at least three members, the c-*myc*, N-*myc* and L-*myc* genes (reviewed by Alt *et al.*, 1986; DePinho *et al.*, 1987). The c-*myc* and N-*myc* genes have been demonstrated to have a very similar overall organization which consists of three exons with a major coding region occurring in exons two and three (Battey *et al.*, 1983; Kohl *et al.*, 1986; Stanton *et al.*, 1986; DePinho *et al.*, 1986). The coding regions of these genes exhibit extensive homology (Kohl *et al.*, 1986; Stanton *et al.*, 1986; DePinho *et al.*, 1986) and encode similar sized nuclear proteins (Slamon *et al.*, 1986; Ikegaki *et al.*, 1986; Ramsay *et al.*, 1986). Both genes also cooperate similarly with an activated *ras* oncogene to yield malignant transformation of primary rat embryo fibroblasts (Land *et al.*, 1983; Yancopoulos *et al.*, 1985; Schwab *et al.*, 1985). The L-*myc* gene was initially isolated on the basis of its high-level amplification in a subset of human small-cell lung carcinomas and its homology to a conserved region of exon two of the other *myc* genes (Nau *et al.*, 1985); a highly related sequence was isolated independently from the normal murine genome on the basis of homology with conserved regions of the second and third exons of the N-*myc* gene (Zimmerman *et al.*, 1986; see below). The overall relationship between the L-*myc* gene and the c- and N-*myc* genes was not previously characterized; however, limited nucleotide sequence analysis of the human L-*myc* clone confirmed the existence of two short homologous sequences shared among the genes and suggested that L-*myc* was a true member of the *myc* gene family (Nau *et al.*, 1985).

The normal function of the *myc* family genes is unknown, although it is believed that these genes may play some role in cellular differentiation and proliferation (reviewed by Kelly and Siebenlist, 1986; Marcu, 1987). The similar activities of N- and c-*myc* with regard to transformation suggest at least overlapping oncogenic activities. However, the conservation of the different *myc* genes as distinct sequences throughout vertebrate evolution (e.g. King *et al.*, 1986; Van Beneden *et al.*, 1986; DePinho *et al.*, 1987; Collum and Alt, unpublished data) and the distinct tissue- and stage-specific patterns of c-, N- and L-*myc* expression during murine development (Jacobovits *et al.*, 1985; Zimmerman *et al.*, 1986) suggest unique functions as well. The differential patterns of *myc* gene expression observed in normal cells and tissues is also reflected in the patterns of *myc* gene activation in different tumors. The c-*myc* gene is activated in a wide variety of different tumor types and by many different mechanisms in correspondence with its generalized expression patterns in developing murine tissues; in contrast, activation of the N- and L-*myc* genes has been observed thus far in a limited set of tumors and only by gene amplification (Schwab *et al.*, 1983; Kohl *et al.*, 1983; Nau *et al.*, 1985), again perhaps reflecting the much more restricted expression of these genes in normal cells (reviewed by Alt *et al.*, 1986). It has been suggested that differential or perhaps combinatorial expression of these genes may be related to the progression of cells through various differentiation pathways (Zimmerman *et al.*, 1986).

To define the relationship between the L-*myc* gene and other *myc*-family members, we have characterized in detail the structure of the murine L-*myc* gene. Our analyses clearly demonstrate that the L-*myc* gene has a similar overall organization to that of the other two family members; the putative coding regions are relatively conserved, but potential regulatory regions are totally divergent. Furthermore, the putative L-*myc* protein contains blocks of sequences highly conserved with those of N-*myc* and c-*myc* interspersed with totally divergent regions; the conserved regions are interspersed over the entire length of the L-*myc* protein and constitute a subset of the regions conserved between N- and c-*myc*. Taken together, our findings provide further support for an extended *myc* gene family with common and divergent functional activities.

© IRL Press Limited, Oxford, England

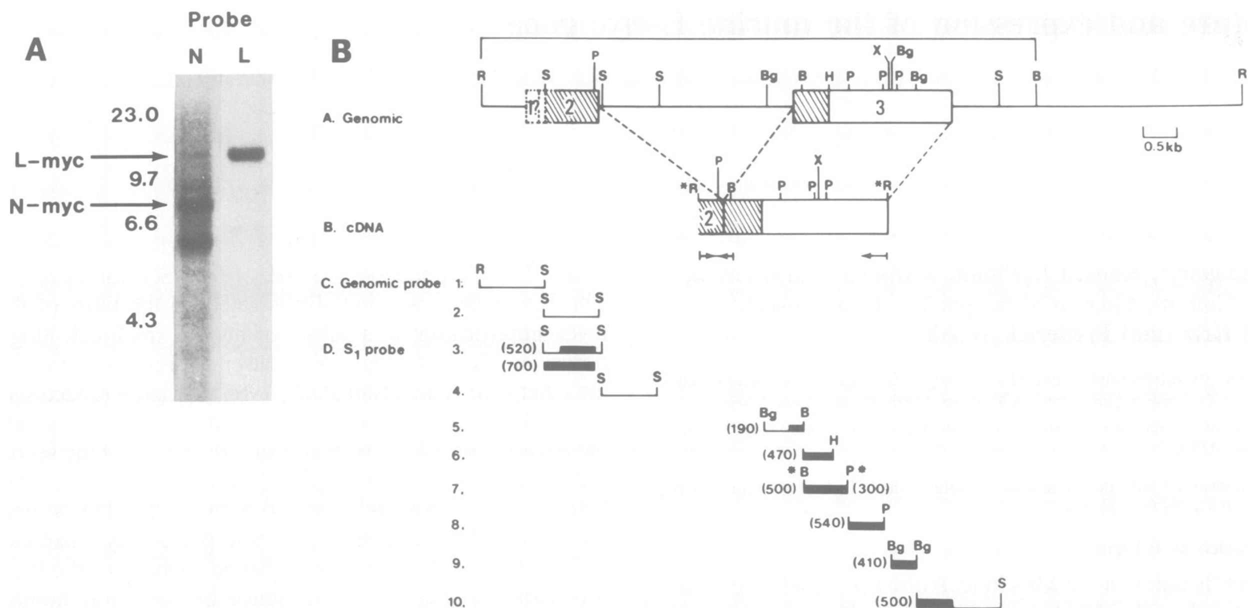

Fig. 1. Organization of the murine L-*myc* gene. **A.** Murine N- and L-*myc*-related sequences. Ten micrograms of BALB/c genomic DNA was digested with *Eco*RI, fractionated by agarose gel electrophoresis, transferred to nitrocellulose and assayed either for (**panel N**) hybridization to a human N-*myc* exon 2-specific probe (*Xho*I–*Bam*HI fragment; Kohl *et al.*, 1986) or for (**panel L**) hybridization to a mouse L-*myc* exon 2-specific probe (genomic probe 2, **panel B**). The human probe was hybridized under reduced stringency conditions (DePinho *et al.*, in preparation). **B.** Organization of the L-*myc* gene. Sections a and b show a partial restriction endonuclease map of the 12 kb murine L-*myc* genomic clone and a 3.2 kb murine L-*myc* cDNA clone. Restriction endonuclease sites are indicated as follows: B, *Bam*HI; Bg, *Bgl*II; H, *Hin*dIII; P, *Pst*I; R, *Eco*RI; S, *Sac*I; X, *Xho*I; the *Eco*RI sites created by linker addition during the cloning of the cDNA are indicated with an asterisk. The large bracket above the map of the genomic clone indicates the region that was sequenced. The boxes indicate the positions of the exons. The dashed box and the question mark indicate the position of a putative exon 1; see text for details. The shaded areas indicate putative translated sequences. The arrows below the cDNA map indicate the sequenced portions and the sites which were end-labeled. **Panel C** shows fragments of the genomic clone that were isolated for use as nick-translated probes. **Panel D** diagrams fragments of the genomic clone that were used for the S1 nuclease mapping; S1-resistant fragments after hybridization to total newborn brain poly(A) containing RNA are indicated by solid boxes; the size of the protected fragment is given in parentheses. An asterisk at the end of a fragment indicates that the probe was terminally labeled. The arrow indicates the presence of an S1-hypersensitive site in probe 7 which was not further characterized.

Results

Structure of the murine L-myc gene

A low stringency Southern blot assay of murine DNA for hybridization to an exon 2 probe derived from the human N-*myc* gene revealed a variety of cross-hybridizing fragments in addition to the 7.7 kb fragment containing the murine N-*myc* gene (Figure 1A, lane N); several of these *Eco*RI fragments also hybridized to a third exon N-*myc* probe (data not shown). On this basis a panel of murine *myc*-related clones was isolated from a genomic *Eco*RI library. One of these clones, a 12 kb *Eco*RI fragment, had strong homology to a probe specific for the human L-*myc* gene (Nau *et al.*, 1985), which is now known to correspond to an exon 2 specific probe (DePinho *et al.*, in preparation), and weak, but distinct, homology to probes derived from corresponding regions of the N-*myc* or c-*myc* genes (not shown). Under normal stringency, unique sequence probes prepared from this clone (e.g. probe 2, Figure 1B) hybridized strongly only to the homologous 12 kb fragment in murine DNA (Figure 1A; lane L) and to the polymorphic 10 kb or 6.6 kb fragment diagnostic of the human L-*myc* gene (Nau *et al.*, 1985) in human DNA (data not shown). Based on these types of analyses, we conclude that the isolated 12 kb fragment harbors sequences derived from the murine homologue of the L-*myc* gene; a conclusion confirmed by direct comparison of the nucleotide sequence of this fragment with that of the human L-*myc* gene (DePinho *et al.*, in preparation).

Restriction mapping and hybridization assays revealed two separated regions of the L-*myc* genomic clone which respectively had substantial sequence conservation with probes for conserved sequences within the second or third exons of the N-*myc* gene (data not shown), providing a tentative transcriptional orientation for the clone. To further define the structure of the L-*myc* gene, we employed a probe from the exon 2-homologous region to isolate several L-*myc* cDNA sequences from a murine newborn brain cDNA library (Figure 1B); newborn brain is the naturally occurring murine source with the highest known levels of L-*myc* expression (Zimmerman *et al.*, 1986). Comparison of the complete nucleotide sequence of an 8.8 kb subfragment of the genomic clone containing the N-*myc* homologous regions (Figure 2) with that of appropriate regions of the longest cDNA clone (Figures 1B and 2), coupled with extensive S1 nuclease mapping experiments, clearly defined the presence of two L-*myc* coding exons separated by an intron of approximately 2900 kb. These L-*myc* exons have considerable organizational homology to exons 2 and 3 of the N-*myc* and c-*myc* genes (Figures 1B and 2); we refer to them as exons 2 and 3 of the L-*myc* by analogy with those of other *myc* genes, although the putative exon 1 sequences of the gene have been less clearly defined (see below). The presence of a poly(A) tail defined the relative orientation of the cDNA clones and confirmed the orientation of the genomic clone predicted by the N-*myc* homologous regions (Figures 1 and 2). A consensus adenylation sequence, AATAAA (Proudfoot and Brownlee, 1976), occurs at position 7227 of the genomic clone (Figure 2), just upstream from the point at which the cDNA sequence diverges from the exon 3 genomic sequence due to the beginning of the poly(A) tract. The approximate exon 2–3 junction predicted from the S1 nuclease mapping studies (Figure 1D) was precisely mapped from the cDNA sequence to occur between positions 1786 and 4669 in the genomic clone; the sequences

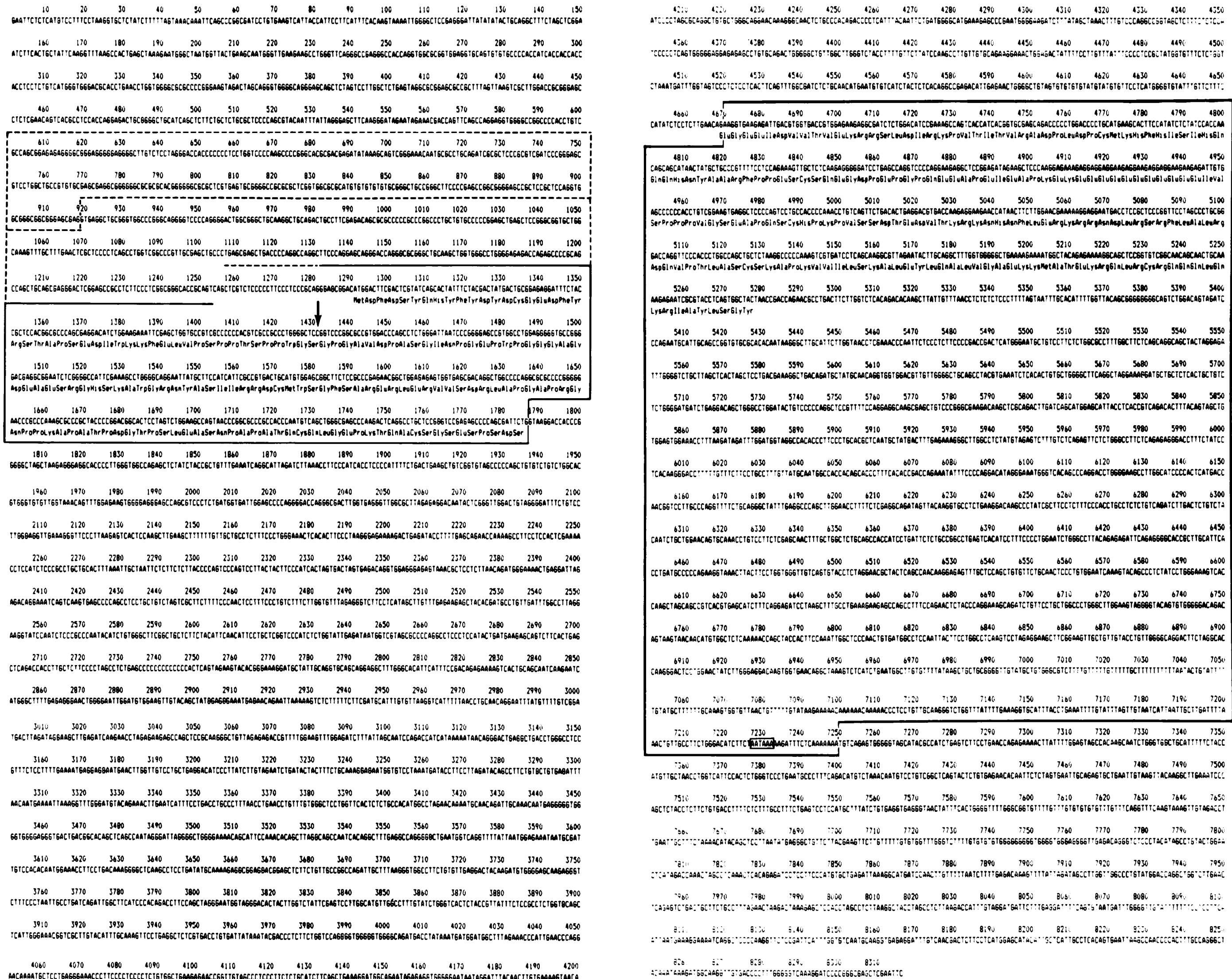

Fig. 2. Complete nucleotide sequence of mouse L-*myc* gene. Restriction fragments from the genomic clone were cloned into the M13 phage, and the sequence was performed by the dideoxynucleotide method (Sanger *et al.*, 1977). Solid or dashed boxes around the sequence indicate the position of the determined or putative exon sequences; see text for details. The dashed line at nucleotide 1260 indicates an imprecision in the localization of the 5′ end of exon 2 by the S1 nuclease method; see Figure 1. The dashed box indicates the possible location of a putative exon 1 based on evidence described in detail in the text. The arrow at nucleotide 1431 indicates the position of the 5′ end of the cDNA clone as determined by comparison of the genomic and cDNA nucleotide sequences; a corresponding polyadenylation signal present at nucleotide 7227 is boxed.

CTG/GTAAGG spanning the 3′ boundary of exon 2 and TCC-TCTTGAACAG/AA spanning the 5′ boundary of exon 3 are acceptable donor and acceptor splice junctions (Mount, 1982). Restriction endonuclease analyses confirmed the co-linearity of the cDNA and genomic clone within the exon 3 sequence (Figure 1B); thus exon 3 extends for approximately 2600 bp.

Our longest cDNA sequence (approximately 3.2 kb) extended through most of exon 2 to position 1431 (arrow, Figure 2) but ended just before the 5′ border of this exon, approximately located by S1 nuclease protection experiments (Figure 1B). Thus the cDNA did not allow accurate positioning of the upstream exon 2 boundary. In this regard utilization of genomic probe 3 in the S1 experiments revealed two protected bands, suggesting one exon 2 border that terminates at approximately position 1250–1300 and another extending further 5′ (Figure 1D). Notably a human L-*myc* pseudogene has a spliced exon 2 at a region conserved within the murine gene at a point corresponding to position 1281 (DePinho *et al.*, in preparation). Together, these data support the notion that at least one form of the murine L-*myc* mRNA derives from three exons with the 5′ boundary of exon 2 probably occurring at position 1281; an acceptable splice acceptor sequence spans this position.

We have not located the putative first exon of the murine gene, but again comparison with the human L-*myc* pseudogene reveals that in the pseudogene a region homologous to positions 700–919 of the murine genomic clone is directly linked to the exon 2-homologous sequences (DePinho *et al.*, in preparation). Again an acceptable donor splice sequence spans position 919 in the murine gene. Additional Northern blotting data support the conclusion that sequences from this region are found in the L-*myc* mRNA (see below). Whether the longer probe 3 protected fragment (Figure 1B) derives from an alternative form of the L-*myc* mRNA or from precursor sequences remains to be determined. It is notable that several different forms of the c-*myc* and N-*myc* RNAs derive from variations in 5′ initiation and splicing patterns (Battey *et al.*, 1983; Bernard *et al.*, 1983; Kohl *et al.*, 1986; Zim-

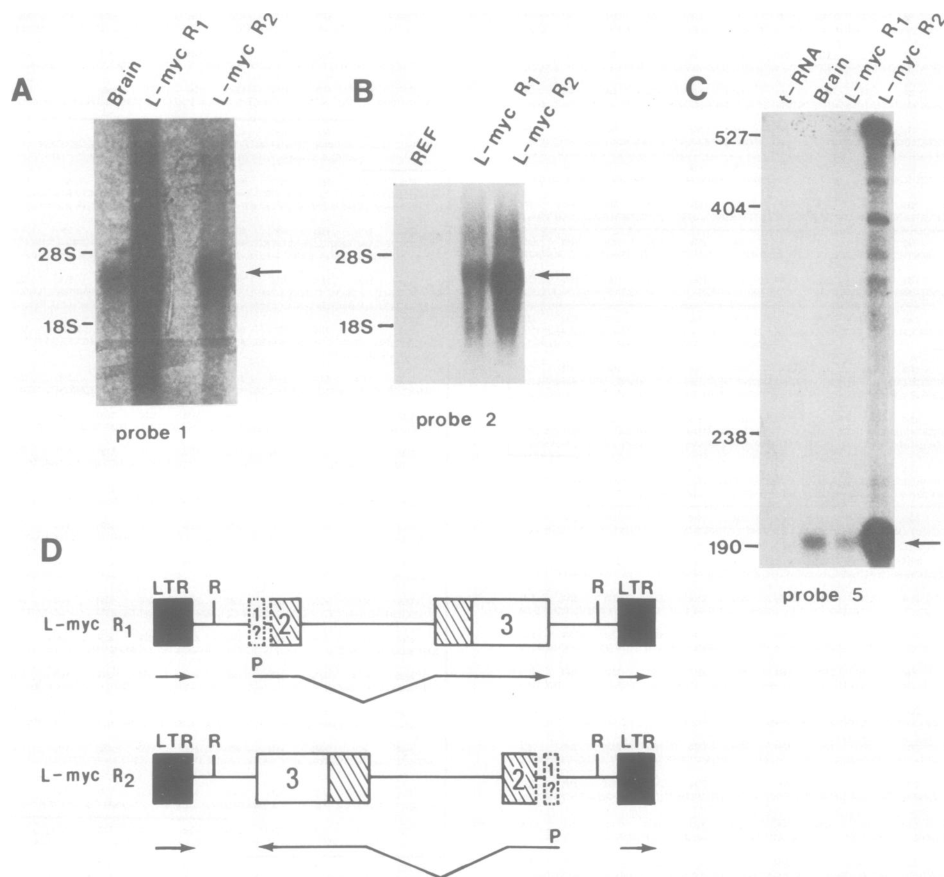

Fig. 3. Expression of an introduced L-*myc* gene in transformed rat embryo fibroblasts. **Panels A** and **B**: Ten micrograms of total RNA prepared from the indicated sources was fractionated by formaldehyde–agarose gel electrophoresis, transferred to nitrocellulose and assayed for hybridization to either L-*myc* genomic probe 1 (**panel A**) or probe 2 (**panel B**); see Figure 1 for description of probes. **Panel C**: Genomic probe 5 (see Figure 1) was uniformly labeled with ^{32}P and assays for hybridization with 20 μg of total RNA of the indicated samples were performed by the S1 nuclease digestion method (Kohl *et al.*, 1986). In this figure, 'REF' refers to normal, untransformed rat embryo fibroblasts; 'BRAIN' refers to newborn mouse forebrain. **Panel D** shows a map of the L-*myc* retroviral constructs which were prepared as described in the text and in Materials and methods. The arrows indicate the direction of the transcription from promoters located in the viral LTR or from the putative L-*myc* promoter (P).

merman and Alt, unpublished data); additional evidence from the analysis of human L-*myc* cDNA sequences supports the possibility of several forms of the L-*myc* mRNA which differ in their 5′ sequences (DePinho *et al.*, in preparation). Thus one form of the L-*myc* gene could consist of a first exon of at least several hundred bp, a second exon of approximately 550 bp and a third exon of 2600 bp. The combined size of these sequences plus a poly(A) tract could account for the size of the murine L-*myc* mRNA of approximately 3.6–4.0 kb (Zimmerman *et al.*, 1986).

Expression of the L-myc genomic clone

Due to the lack of a full-length cDNA clone, and no high level expression source to allow ready analysis of the L-*myc* RNA sequences, we resorted to transfection assays to further define coding and functional aspects of the cloned L-*myc* gene. Because the major defined coding regions of N-*myc* and c-*myc* are found within the second and third exons, it seemed possible that the 12 kb L-*myc* genomic clone could contain a complete copy of the L-*myc* gene. To test this possibility the 12 kb clone was inserted between two directly repeated Moloney murine leukemia virus long terminal repeats (LTRs) in the retroviral vector pVcos7. The L-*myc* genomic clone was inserted in both transcriptional orientations relative to the promoter in the LTR. We have found that direct transfection of such *myc* expression vectors into permanent cell lines often does not yield transfectants which express the introduced genes at very high levels (Zimmerman and Alt, unpublished data); however primary rat embryo fibroblasts malignantly transformed by LTR-N-*myc* plus Ha-*ras* expression vectors almost always have substantial expression levels of the introduced N-*myc* gene (Yancopoulos *et al.*, 1985). Therefore to readily assay expression and also to obtain a preliminary assessment of the functionality of the L-*myc* genomic clone, we co-transfected the two types of vectors (L-*myc* R1 and R2; Figure 3D), independently with an activated Ha-*ras* gene into primary rat embryo fibroblasts (REFs), and searched for the appearance of transformed foci which overgrew the normal monolayers. LTR-vectors with L-*myc* clones in either orientation in conjunction with the activated Ha-*ras* gene led to the appearance of transformed foci which also grew in soft agar and formed tumors in syngeneic rats (data not shown). These preliminary data suggest that, like the c-*myc* and N-*myc* genes (Land *et al.*, 1983; Yancopoulos *et al.*, 1985; Schwab *et al.*, 1985), the L-*myc* gene can cooperate with activated *ras* genes to cause tumorigenic transformation. However we did not perform a large enough set of co-transformation assays to strictly confirm this point or to determine the relative activity of L-*myc* in the assay.

To examine the nature of the L-*myc* transcripts produced from these vectors, we isolated a permanent REF line derived with each vector, grew these to large numbers and assayed their total RNA for hybridization to various L-*myc* probes. Hybridization of RNA prepared from various sources to an exon 2-specific probe revealed that although normal REFs do not produce L-

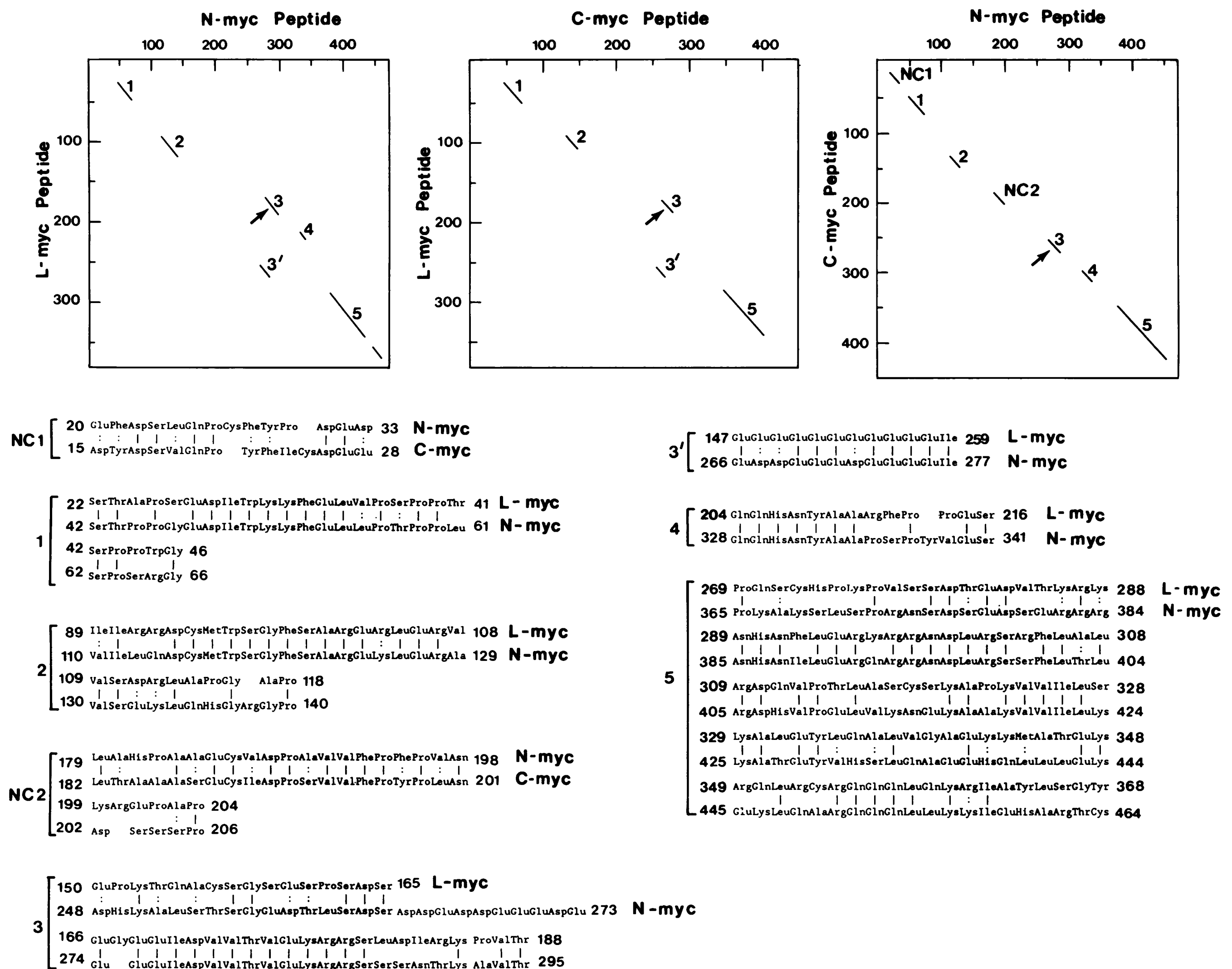

Fig. 4. Comparison of the c-, N- and L-*myc* proteins. **Upper**: The three boxes illustrate the results of a dot-matrix computer analysis of protein sequence homology among members of the murine *myc* gene family. The comparison was performed using the Beckman MicroGenie Program with a window of 8 and a stringency of 60%. From left to right the boxes show a comparison between the mouse L-*myc* and human N-*myc* proteins; the mouse L-*myc* and mouse c-*myc* proteins; and mouse c-*myc* and human N-*myc* proteins. Regions of homology shared by all three proteins are numbered 1, 2, 3, 4 (less conserved in c-*myc*) and 5. Region 3′ in the L-*myc* protein is a possible transposition of a portion of region 3 in N-*myc* and c-*myc*. Regions NC1 and NC2 are shared by N-*myc* and c-*myc*, but are not found in L-*myc*. The arrow in the middle of the homology box 3 indicates the regions encoded by exon 2 or 3; the nucleotides encoding this region span this splice junction. **Lower**: The amino acid alignment of the different homology boxes is indicated; not all c-*myc*/N-*myc* alignments are shown. A solid line between amino acids indicates homology; a double dot indicates a conservative substitution.

myc transcripts identical in size to authentic 4 kb L-*myc* mRNA (Figure 3B, arrow; in other blots the major transcripts were shown to co-migrate with murine brain L-*myc* mRNA; see also panel A). In addition S1 nuclease assays with probe 5 (Figure 1B) confirmed that these transcripts protected a 190 bp band diagnostic of a normal L-*myc* exon 3 splice junction (Figure 3C). Strikingly, the steady state levels of L-*myc* RNA produced from the L-*myc* R2 vector (with the gene in the opposite transcriptional orientation from the LTR promoter) were even higher than those produced from the L-*myc* R1 vector.

Together our results indicate that all of the sequences necessary for generation of full length L-*myc* transcripts, including a transcriptional promoter, are contained within the 12 kb L-*myc* genomic clone, although the adjacent LTR may provide enhancer function to obviate potential tissue-specific expression constraints on the introduced L-*myc* genes. Assay of RNA from the transformants, or from newborn brain, for hybridization to genomic probe 1, which lies upstream of the exon 2 sequences, demonstrated hybridization to normal size L-*myc* mRNA in all samples; however an extensive lane background was present in the L-*myc* R1 sample, presumably due to a heterogeneous series of transcripts initiated from the viral LTR (Figure 3A). This finding confirms that L-*myc* transcripts initiate upstream of the exon 2 region and supports the existence of the putative exon 1 region. The c-*myc* gene appears to contain two independent promoter regions that each contain CAAT and TATA motifs upstream from their respective initiation sites (Battey *et al.*, 1983), while transcription of the N-*myc* gene appears to occur from multiple sites in the absence of a TATA motif (Kohl *et al.*, 1985; DePinho *et al.*, 1986). Although it was not possible to define the normal L-*myc* promoter due to the very low abundance of L-*myc* mRNA in known naturally occurring sources, a region with similarities to the N-*myc* and/or c-*myc* promoter regions occurs just upstream of the putative exon 1 sequences, starting with a potential TATA motif at position 691 and extending approximately 100 bp in the 5′ direction (Figure 2).

The L-myc protein

Major forms of the c-*myc* and N-*myc* proteins are initiated from ATG codons that are located near the beginning of exon 2. In both of these genes exon 1 appears to be primarily an untranslated leader sequence, although an additional form of the c-*myc* protein has recently been demonstrated to be initiated from a variant (non-ATG) initiation codon in exon 1 (King *et al.*, 1987). An ATG codon also occurs at a similar position within the 5′ region of L-*myc* exon 2 (position 1291). From this translation start site an open reading frame extends for 1122 bp (allowing for splicing) and terminates with a TAA stop codon after position 5277. Multiple termination codons interrupt this sequence in the other two translational reading frames. Thus the putative murine L-*myc* protein would contain 368 amino acids with a predicted size of approximately 40 kd, a size significantly smaller than the murine N-*myc* (462 amino acids; DePinho *et al.*, 1986) or c-*myc* (450 amino acids; Stanton *et al.*, 1984) proteins. However the substantial homology of the putative L-*myc* protein to N-*myc* and c-*myc* strongly argues that this is the correct reading frame (Figure 4). A dot matrix computer comparison of the N-*myc* and c-*myc* proteins reveals seven clusters of amino acid sequences that are relatively conserved with respect to composition and position on the peptide backbone; the amino acid sequence linking these regions is quite divergent (Figure 4; third panel). Comparison of the L-*myc* peptide with the other two demonstrates that five of these regions including those most conserved between N-*myc* and c-*myc* (regions 1, 2, 3 and 5) are also highly conserved in the L-*myc* protein. In some cases the L-*myc* homology to N-*myc* is even greater than the N-*myc*/c-*myc* homology (e.g. regions 2 and 5). Two regions that are relatively conserved between the N-*myc* and c-*myc* (NC1 and NC2) are absent from the L-*myc* protein; they are encoded in regions of exon 2 where L-*myc* appears to have major deletions relative to the N-*myc* and c-*myc* proteins. The major size difference between L-*myc* and the other two *myc* proteins appears to result from the relatively short coding region contributed by exon 2 of L-*myc*; the exon 3 coding regions are similar in size for the three proteins (Figure 4).

A portion of homology region 3 between L-*myc* and the other two proteins appears to be interrupted compared with that of the N-*myc* (or c-*myc*) proteins in that it lacks a major internal stretch of acidic amino acids (polyglutamic acid stretch in c-*myc*, not shown, or a polyglutamic and aspartic acid stretch in N-*myc*, Figure 4, bottom). Homology region 3 is encoded by the 3′ portion of exon 2 (ending in Ser Asp Ser; Figures 2 and 4) and the 5′ portion of exon 3 (Figure 4, arrow); the splice junctions of these exons appear to be conserved between the two genes, suggesting that the truncated acidic amino acid stretch is missing from just within the beginning of the exon 3 coding sequences of L-*myc*. Notably, a polyglutamic acid stretch is encoded further into the exon 3 sequence of L-*myc* and appears on the L-*myc* peptide backbone in a position identical to the corresponding sequences in the N-*myc* and c-*myc* proteins (conserved region 3′; Figure 4 upper and lower).

Discussion

The myc gene family

Our characterization of the murine L-*myc* gene clearly demonstrates that it represents an additional member of the *myc* family of cellular oncogenes (Figure 4). The full extent of this family is unknown; other potential members of the human and murine *myc* family have been isolated by cross-hybridization, but the only well-characterized of these is an L-*myc* pseudogene (DePinho *et al.*, in preparation). Our preliminary mapping data confirm that the *myc* family is a dispersed family in both the human and the mouse, with members lying on many different chromosomes (R.DePinho *et al.*, in preparation; D'Eustachio and Alt, unpublished data). The c-, N- and L-*myc* genes are highly conserved in vertebrate evolution, indicating fundamental but distinct roles in normal cellular processes (DePinho *et al.*, 1986; DePinho *et al.*, in preparation; Collum and Alt, unpublished data); yet despite intensive investigation, the normal role of *myc* gene products has not been elucidated. Various studies have suggested that c-*myc* has a role in cellular differentiation and proliferation (reviewed by Marcu, 1987). A role in differentiation for N-*myc* was suggested on the basis of its decreased expression in retinoic acid differentiated human neuroblastomas (Thiele *et al.*, 1985); the strikingly restricted tissue- and developmental stage-specific expression patterns of both N- and L-*myc* in early development and in tumor cells are also consistent for some stage-specific role for these genes.

Deregulated c-*myc* expression has been implicated in the genesis of various classes of tumors and by a number of different mechanisms including promoter (enhancer) insertion, viral transduction and amplification (reviewed by Varmus, 1984; Cole, 1987). In contrast, deregulated N-*myc* expression has been implicated in a very restricted set of tumors including neuroblastomas (Schwab *et al.*, 1983; Kohl *et al.*, 1983; Seeger *et al.*, 1985), small-cell lung carcinoma (Nau *et al.*, 1986) and retinoblastoma (Kohl *et al.*, 1983; Lee *et al.*, 1984); deregulated L-*myc* expression has thus far only been implicated in the genesis of a subset of small-cell lung carcinomas (Nau *et al.*, 1985). Likewise, in contrast to the many mechanisms of c-*myc* activation, amplification has thus far been the only mechanism observed in the context of deregulated N- or L-*myc* expression in naturally occurring tumors. The reasons for the restricted patterns of N-*myc* and L-*myc* activation, both with respect to tumor type and mechanism (amplification) is not clearly understood, but may in part reflect the normal restricted expression patterns of these genes (see Alt *et al.*, 1986 for further discussion). In any case, despite the striking similarities in the structure of the genes and their products, these differences again suggest distinct roles in normal physiology as well.

Structure and expression of myc genes

The c- and N-*myc* genes consist of three exons in humans and mice (Battey *et al.*, 1983; Stanton *et al.*, 1984, 1986; Bernard *et al.*, 1983; Kohl *et al.*, 1986; DePinho *et al.*, 1986). At least one form of the L-*myc* gene also appears to be organized into this characteristic three exon−two intron pattern (Figure 1). We have not unequivocally defined the murine exon 1 L-*myc* sequences, but the gene has an even larger (2000 bp) 3′ untranslated region than the N- or c-*myc* genes (Figure 2), accounting for the larger size of the L-*myc* mRNA. Notably the L-*myc* 3′ untranslated region (Figure 2) has no obvious homology to that of N- or c-*myc*, yet like the untranslated regions of c- and N-*myc*, it also appears to be conserved with the corresponding regions of L-*myc* genes in other species (DePinho *et al.*, in preparation). It has been suggested that the evolutionary conservation of the untranslated regions of these genes suggests an important regulatory role (DePinho *et al.*, 1986), as described for similar regions of the c-*fos* gene (reviewed by Verma, 1986). If the untranslated regions of these genes are indeed important for regulation of expression, the divergence of the sequences between different family members would be consistent with the differen-

tial expression patterns of the different *myc*-family genes in normal tissues, in tumors, and in permanent cell lines (Zimmerman *et al.*, 1986). However thus far DNA transfection experiments in cell lines have failed to define the elements which restrict the expression of *myc* genes in various cell types (Legouy *et al.*, 1987).

Activities of myc proteins

The c- and N-*myc* gene products are nuclear-associated phosphoproteins that have a strong affinity *in vitro* for nucleic acids (Donner *et al.*, 1982; Watts *et al.*, 1985; Slamon *et al.*, 1986; Ikegaki *et al.*, 1986; Ramsay *et al.*, 1986). Whether or not *myc* proteins interact with specific DNA or RNA sequences to modulate specific gene expression remains open to speculation (e.g. Nisen *et al.*, 1986; Legouy *et al.*, 1987). Recently an extensive analysis of the transforming activity and cellular localization of the products of mutagenized c-*myc* genes has provided a preliminary dissection of potential functional regions within the protein (Stone *et al.*, 1987); it is of interest to compare the regions noted by these studies with the conserved homology sequences among the three *myc* proteins (Figure 4) that have been elucidated by our analysis. Two domains found to be essential for the REF co-transformation assay, in which N-*myc* and apparently L-*myc* also function, included a small portion of the second exon which contains homology region 2 (highly conserved region among all *myc* proteins; Figure 4) and a major portion of the C terminus of the protein which included homology region 5 (the largest highly conserved region among the three proteins; Figure 4). Equally notable is the fact that a region whose complete integrity was not considered essential contained the highly conserved homology sequence 1 (Figure 4), and another region which was considered dispensible in the co-transformation assay contained several highly conserved sequences including homology sequence 3 as well as NC2 and 4 (Figure 4). Clearly there appear to be major functional regions based on sequence conservation which are not scored by the co-transformation assay. In this regard some of the mutagenized c-*myc* genes were also tested for their ability to transform the established Rat-1 line which is known to be susceptible to transformation by introduced high-level c-*myc* expression (Stone *et al.*, 1987). Again mutants which had defects in the region of conserved homology sequence 5 were totally inactive, underscoring the importance of this highly conserved region; but strikingly mutants in the region 3 area that were active in the cooperation assay were inactive in the Rat-1 assay, suggesting the possibility that regions of c-*myc* important for co-transformation may not be identical to those necessary for transformation of established cell lines. It will be of interest to test mutagenized N- and L-*myc* genes in the Rat-1 assay and to assess the potential role of the highly conserved region 3 sequences in that system.

Hybrid gene studies indicated several sequences within the c-*myc* protein that could direct the movement of the muscle pyruvate kinase gene to the nucleus (Stone *et al.*, 1987). One of these sequences, which occurs at the beginning of homology region 5 (Figure 4), is highly conserved among all three genes, suggesting that L-*myc*, like c-*myc* and N-*myc*, will also be found in the nucleus. Furthermore, the studies of Stone *et al.* (1987) also implicated domains containing conserved homology regions 2 and 5 as necessary for interaction with nuclear components, suggesting that these conserved sequences may play similar general roles in the N-*myc* and L-*myc* proteins. The seven conserved regions between c-*myc* and N-*myc* and the six between L-*myc* and others, as well as the divergent regions between the proteins, should provide more restricted targets for future mutational analyses of overlapping and potentially divergent activities.

Materials and methods

Genomic and cDNA cloning

Genomic DNA from the placenta of BALB/c mice was digested to completion with *Eco*RI and cloned into the *Eco*RI site of lambda phage Charon 30 essentially as described previously (DePinho *et al.*, 1984). This library was screened for clones which hybridized to exon 2- and 3-specific probes derived from human N-*myc* cDNA clones (Kohl *et al.*, 1986). A 12 kb clone which hybridized to both probes was subcloned into plasmid vector pUC18 for further analysis.

A cDNA library was prepared from poly(A) containing total RNA isolated from the brain of newborn BALB/c mice as described previously (Gubler and Hoffman, 1983; Kohl *et al.*, 1986). This library was screened for clones which hybridized to an exon 2 L-*myc*-specific probe as described in the text. Several clones were isolated, the longest of which was subcloned into pUC18 for further analysis.

DNA sequencing

The nucleotide sequence of the L-*myc* genomic clone was determined by the dideoxynucleotide method (Sanger *et al.*, 1977), following shotgun subcloning into bacteriophage M13. The nucleotide sequence of indicated regions of the cDNA clone was determined by the method of Maxam and Gilbert (1977).

Analysis of RNA and DNA

DNA and RNA preparation, restriction endonuclease digestions, probe preparations by the nick-translation method, blotting procedures and hybridization conditions were as described previously (Kohl *et al.*, 1983).

S1 nuclease assays and probes

The S1 probes were derived from the 12 kb murine genomic clone as indicated in Figure 1 and either uniformly labeled or end-labeled as described previously (Kohl *et al.*, 1986). Hybridization, digestion with S1 nuclease and analysis by electrophoresis through polyacrylamide gels was performed as previously described (Kohl *et al.*, 1986).

Computer analysis of DNA and protein sequence

Computer analyses of the L-*myc* sequence were performed on an IBM-AT personal computer using the Microgenie Sequence Analysis Program (copyright 1983 by SciSoft, Inc.).

Rat embryo fibroblast co-transformation

The 12 kb L-*myc* genomic clone was subcloned in both transcriptional orientations into retroviral vector pVcos7 and co-transfected with the activated *ras* plasmid pT24-*ras* as described previously (Yancopoulos *et al.*, 1985). Transformed foci were isolated and derived cell lines grown to large numbers for RNA preparation.

Acknowledgements

This work was supported by NIH grants CA23767-06 and CA42335 and American Cancer Society grant CD-269. F.A. is an Irma T.Hirschl Career Scientist and a Malinckrodt Scholar. E.L. was a fellow of the European Molecular Biology Organization, R.D. a recipient of NIH physician scientist award AI-00602 and K.Z. a trainee of NIH grant CA 09503.

References

Alt,F.W., Kellems,R.E., Bertine,J.R. and Schimke,R.T. (1978) *J. Biol. Chem.*, **253**, 1357–1371.

Alt,F.W., DePinho,R.A., Zimmerman,K., Legouy,E., Hatton,K., Ferrier,P., Tesfaye,A., Yancopoulos,G.D. and Nisen,P. (1986) *Cold Spring Harbor Symp. Quant. Biol.*, Vol. LI, 931–941.

Battey,J., Moulding,C., Taub,R., Murphy,W., Stewart,T., Potter,H., Lenoir,G. and Leder,P. (1983) *Cell*, **34**, 779–787.

Bernard,O., Cory,S., Gerondakis,S., Webb,E. and Adams,J. (1983) *EMBO J.*, **2**, 2375–2383.

Brodeur,G.M., Seeger,R.C., Schwab,M., Varmus,H.E. and Bishop,J.M. (1984) *Science*, **224**, 1121–1124.

Cole,M.D. (1986) *Annu. Rev. Genet.*, **20**, 361–384.

DePinho,R., Kruger,K., Andrews,W., Lutzger,S., Baltimore,D. and Alt,F.W. (1984) *Mol. Cell. Biol.*, **4**, 2905–2910.

DePinho,R.A., Legouy,E., Feldman,L.B., Kohl,N.E., Yancopoulos,G.D. and Alt,F.W. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1827–1831.

DePinho,R., Hatton,K., Ferier,P., Zimmerman,K., Legouy,E., Tesfaye,A., Collum,R., Yancopoulos,G., Nisen,P. and Alt,F.W. (1987) *Ann. Clin. Res.*, **18**, 284–289.

Gubler,U. and Hoffman,B.J. (1983) *Gene*, **25**, 262–269.

Ikegaki,N., Bukovsky,J. and Kennett,R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5929–5933.
Jakobovits,A., Schwab,M., Bishop,J.M. and Martin,G.R. (1985) *Nature*, **318**, 188–191.
Kelly,K. and Siebenlist,U. (1986) *Annu. Rev. Immunol.*, **4**, 317–338.
Kelly,K., Cochran,B., Stiles,C. and Leder,P. (1983) *Cell*, **35**, 603–610.
King,M.W., Roberts,J.M. and Eisenman,R.N. (1987) *Mol. Cell. Biol.*, **6**, 4499–4508.
Klein,G. and Klein,E. (1985) *Immunol. Today*, **6**, 208–215.
Kohl,N.E., Kanda,N., Schreck,R.R., Bruns,G., Latt,S.A., Gilbert,F. and Alt, F.W. (1983) *Cell*, **35**, 359–367.
Kohl,N., LeGouy,E., DePinho,R., Smith,R., Gee,C. and Alt,F.W. (1986) *Nature*, **319**, 73–77.
Land,H., Parada,L. and Weinberg,R. (1983) *Nature*, **304**, 596–599.
Lee,W.H., Murphee,A.L. and Benedict,W.F. (1984) *Nature*, **309**, 458–460.
LeGouy,E., DePinho,R., Zimmerman,K., Ferrier,P., Collum,R. and Alt,F.W. (1987) In Alt,F.W., Harlow,E. and Ziff,E. (eds), *Nuclear Oncogenes*, in press.
Marcu,K.B. (1987) *Bioessays*, **6**, 28–31.
Maxam,A. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 1800–1804.
Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.
Nau,M.M., Brooks,B.J., Carney,D.N., Gazdar,A.F., Battey,J.F., Sausville,E.A. and Minna,J.D. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1092–1096.
Nau,M., Brooks,B., Battey,J., Sausville,E., Gasdar,A., Kirsh,I., McBride,O., Bertness,V., Hollis,G. and Minna,J. (1985) *Nature*, **318**, 69–73.
Nisen,P.D., Zimmerman,K., Cotter,S., Gilbert,F. and Alt,F.W. (1986) *Cancer Res.*, **46**, 6217–6222.
Persson,H. and Leder,P. (1984) *Science*, **225**, 718–721.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211–214.
Ramsay,G., Stanton,L., Schwab,M. and Bishop,J.M. (1986) *Mol. Cell. Biol.*, **6**, 4450–4457.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schwab,M., Alitalo,K., Klempnauer,K., Varmus,H., Bishop,J.M., Gilbert,F., Brodeur,G., Goldstein,M. and Trent,J. (1983) *Nature*, **305**, 245–248.
Schwab,M., Varmus,H.E. and Bishop,J.M. (1985) *Nature*, **316**, 160–162.
Seeger,R., Brodeur,G., Sather,H., Dalton,A., Siegel,S., Wong,K. and Hammond, O. (1985) *New Engl. J. Med.*, **313**, 1111–1119.
Sheiness,D. and Bishop,J.M. (1979) *J. Virol.*, **31**, 514–521.
Slamon,D., Boon,T.C., Seeger,R., Keith,D., Chazin,V., Lee,W. and Souza,L.M. (1986) *Science* **232**, 768–772.
Stanton,L.W., Farlander,P.D., Tesser,P.M. and Marcu,K.B. (1984) *Nature*, **310**, 423–425.
Stanton,L.W., Schwab,M. and Bishop,J.M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1772–1776.
Stone,J., de Lange,T., Ramsay,G., Jakobivits,E., Bishop,J.M., Varmus,H.E. and Lee,W. (1987) *Mol. Cell. Biol.*, in press.
Thiele,C.J., Reynolds,C.P. and Israel,M.A. (1985) *Nature*, **313**, 404–406.
Van Beneden,R.J., Watson,D.K., Chen,T.T., Lantenberger,J.A. and Papas,T.S. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3698–3702.
Varmus,H. (1984) *Annu. Rev. Genet.*, **18**, 553–612.
Verma,I.M. (1986) *Trends Genet.*, **2**, 93–96.
Yancopoulos,G.D., Nisen,P.D., Tesfaye,A., Kohl,N.E., Goldfarb,M.P. and Alt, F.W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 5455–5459.
Zimmerman,K.A., Yancopoulos,G.D., Collum,R.G., Smith,R.K., Kohl,N.E., Denis,K.A., Nau,M.M., Witte,O.N., Toran-Allerand,D., Gee,C.E., Minna,J.D.S. and Alt,F.W. (1986) *Nature*, **319**, 780–783.

Received on July 24, 1987