MOLECULAR AND CELLULAR BIOLOGY, Apr. 1989, p. 1545–1552
0270-7306/89/041545-08$02.00/0

Vol. 9, No. 4

Two N-*myc* Polypeptides with Distinct Amino Termini Encoded by the Second and Third Exons of the Gene

TOMI P. MÄKELÄ, KALLE SAKSELA, AND KARI ALITALO*

Departments of Virology and Pathology, University of Helsinki, Haartmaninkatu 3, 00290 Helsinki, Finland

Received 18 October 1988/Accepted 3 January 1989

The N-*myc* and c-*myc* genes encode closely related nuclear phosphoproteins. We found that the N-*myc* protein from human tumor cell lines appears as four closely migrating polypeptide bands (p58 to p64) in sodium dodecyl sulfate-polyacrylamide gels. This and the recent finding that the c-*myc* protein is synthesized from two translational initiation sites located in the first and second exons of the gene (S. R. Hann, M. W. King, D. L. Bentley, C. W. Anderson, and R. N. Eisenman, Cell 52:185–195, 1988) prompted us to study the molecular basis of the N-*myc* protein heterogeneity. Dephosphorylation by alkaline phosphatase reduced the four polypeptide bands to a doublet with an electrophoretic mobility corresponding to the two faster-migrating N-*myc* polypeptides (p58 and p60). When expressed transiently in COS cells, an N-*myc* deletion construct lacking the first exon produced polypeptides similar to the wild-type N-*myc* protein, indicating that the first exon of the N-*myc* gene is noncoding. Furthermore, mutants deleted of up to two thirds of C-terminal coding domains still retained the capacity to produce a doublet of polypeptides, suggesting distinct amino termini for the two N-*myc* polypeptides. The amino-terminal primary structure of the N-*myc* protein was studied by site-specific point mutagenesis of the 5′ end of the long open reading frame and by N-terminal radiosequencing of the two polypeptides. Our results show that the N-*myc* polypeptides are initiated from two alternative in-phase AUG codons located 24 base pairs apart at the 5′ end of the second exon. Both of these polypeptides are phosphorylated and localized to the nucleus even when expressed separately. Interestingly, DNA rearrangements activating the c-*myc* gene are often found in the 1.7-kilobase-pair region between the two c-*myc* translational initiation sites and correlate with the loss of the longer c-*myc* polypeptide. Thus the close spacing of the two N-*myc* initiation codons could explain the relative resistance of the N-*myc* gene to similar modes of oncogenic activation.

The N-*myc* gene was originally identified as an amplified DNA fragment from human neuroblastomas on the basis of its homology to the v-*myc* oncogene (43). The N-*myc* gene is closely related to the more extensively investigated c-*myc* gene: both contain three exons, with long 5′ and 3′ untranslated regions, and a predicted single long open reading frame beginning with an ATG codon in the second exon. Overall homology of the deduced amino acid sequences is 32%, and distinct regions throughout the proteins are nearly identical (6, 43, 50). Interestingly, even the nonhomologous coding regions are highly conserved between the human and mouse N-*myc* genes, which show an overall homology of over 85% (13, 23, 50, 51). This suggests that the two genes perform similar but not identical functions. In line with such deductions, overexpressed human c-*myc* and N-*myc* genes complement mutant c-*ras* oncogenes in the transformation of rat primary embryonic cells in culture (44, 57) and transform Rat-1A cells without the assistance of other oncogenes (47). During cell and tissue differentiation the level of expression of both genes is reduced (52, 53, 59).

Several differences also exist between the expression of the two *myc* genes and their oncogenic potential. The c-*myc* gene is expressed in a wide variety of cells and tissues, and its mRNA is subject to a rapid regulation by transcriptional and posttranscriptional mechanisms in response to a variety of growth factors (for a review, see reference 37). In contrast, the N-*myc* gene is expressed in various tissues during embryogenesis, but expression in adult tissues is highly restricted (20, 59). The expression pattern of the two genes appears to be also reflected in their oncogenic spectra. The c-*myc* gene has been transduced by several acute transforming retroviruses, and it has been found activated by retroviral insertions in avian bursal and rodent T-cell lymphomas. Chromosomal translocations deregulate c-*myc* expression in B-cell malignancies of mice, rats, and humans (for reviews, see references 7 and 22). In addition, amplification of the c-*myc* gene occurs in a wide variety of tumors (1). In contrast to the various mechanisms of c-*myc* activation, amplification is the only known mechanism of activation of the N-*myc* proto-oncogene. N-*myc* amplifications have been found predominantly in tumors with neuroendocrine characteristics, namely, neuroblastomas (43), retinoblastomas (30), and small-cell lung cancers (35), but sporadically also in other tumors (14, 41). Interestingly, amplification of the N-*myc* gene in neuroblastoma correlates with advanced disease stage (8).

The c-*myc* and N-*myc* genes encode DNA-binding nuclear phosphoproteins with short half-lives (2, 16, 19, 36, 38, 39, 45). The molecular functions of the proteins are not known. However, both N-*myc* and c-*myc* proteins induce DNA replication (10, 11, 18); in addition, the c-*myc* protein has been proposed to function in transcriptional regulation (21) and in RNA processing (48). Both the c-*myc* and N-*myc* proteins appear as two major bands in polyacrylamide gel electrophoresis after metabolic labeling and immunoprecipitation. The genomic structures of the c-*myc* and N-*myc* genes suggest that both proteins are translated starting from an AUG codon in the second exon. However, it was recently reported that the two c-*myc* bands represent two polypeptide chains with distinct N termini, with the translation of the longer of the two c-*myc* polypeptides initiating from a CUG codon in the first exon of c-*myc* (17). In this report we show

* Corresponding author.

that the N-*myc* protein is translated from two in-phase AUG initiation codons located only 24 bp apart in the second exon, resulting in two N-terminally distinct N-*myc* polypeptides.

MATERIALS AND METHODS

Preparation of antisera. A 345-base-pair (bp) *Pst*I-*Xho*I fragment from the second exon coding region of the N-*myc* gene was ligated between the *Pst*I and *Sal*I sites in the polylinker region of the procaryotic expression vector pEX-1 (49; a kind gift from K. K. Stanley). The resulting construct encodes a 120-kilodalton β-galactosidase–N-*myc* fusion protein containing 115 N-*myc*-derived amino acids. The N-*myc*–β-galactosidase protein was purified by preparative sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Gel slices containing the polypeptides were homogenized and used to immunize rabbits. The resulting antiserum was tested for specificity in immunoprecipitation and immunofluorescence experiments. The pan-*myc* antiserum (34) was a kind gift from Gerard Evan.

Construction of the N-*myc* expression vectors. For construction of plasmid SV-N-*myc* a 2.6-kilobase (kb) *Xba*I-*Bgl*II fragment from plasmid pNb-9 (44, 50; a kind gift from Manfred Schwab) spanning the second exon and a 2.2-kb *Bgl*II-*Eco*RI fragment from the same plasmid spanning the third exon were ligated into a 2.6-kb *Eco*RI-*Xba*I fragment containing plasmid sequences and the simian virus 40 (SV40) early promoter from the pSVc*myc*1 vector (ATCC 37149). The resulting 7.4-kb SV-N-*myc* plasmid thus has an internal 400-bp *Bgl*II-*Bgl*II deletion in the second intron. For deletion constructs, (i) SV-N-*myc* was digested with *Xho*I-*Mlu*I, and recessed ends were filled with the large fragment of DNA polymerase I and ligated (SV-N-ΔI); (ii) the 2.2-kb *Bgl*II-*Eco*RI fragment of SV-N-*myc* containing exon III was replaced with a 1.6-kb *Bgl*II-*Eco*RI fragment from the pSV2neo vector (ATCC 37149) containing the SV40 polyadenylation signal (SV-N-ΔII); and (iii) a 3.2-kb *Bam*HI-*Eco*RI fragment of SV-N-*myc* was replaced as in SV-N-ΔII (SV-N-ΔIII).

Site-directed in vitro mutagenesis. For in vitro mutagenesis, four antisense oligonucleotides, 5′-GCAGCTCGGGAT CGGCTCG-3′, 5′-TGGACGTGGATCAGCTCGGCA-3′, 5′-CATGCCCGGGATGGTGGAC-3′, and 5′-TCTTGCAGAT CAGGCCCGGCA-3′, complementary to N-*myc* nucleotides 1619 to 1637, 1627 to 1647, 1643 to 1661, and 1651 to 1671, respectively, according to Stanton et al. (50), but containing one mismatch each (indicated by the bold letters), were synthesized with an Applied Biosystems (Foster City, Calif.) model 381A DNA synthesizer and further purified by polyacrylamide gel electrophoresis. A 1.2-kbp *Xba*I-*Bam*HI fragment from SV-N-*myc* was cloned into M13mp19, and single-stranded DNA was used as a template for the in vitro mutagenesis reaction. Mutagenesis was carried out using a kit from Amersham (Buckinghamshire, United Kingdom) as recommended by the supplier. In some experiments, after synthesis and ligation of the second strand, remaining single-stranded DNA was excluded with an anion-exchange resin (Qiagen; Diagen GmbH, Düsseldorf, Federal Republic of Germany) instead of a nitrocellulose filter. The recombinant phages were absorbed to nitrocellulose filters, and mutants were identified by differential hybridization with [γ-^{32}P]ATP-kinased mutant oligonucleotides. In some cases identification was by sequencing. Absence of other cloning artifacts was always confirmed by sequencing of the mutant clones followed by recloning of the mutant *Xba*I-*Bam*HI fragment back to the SV-N-*myc* expression vector.

Cell culture, transfections, and radioactive labeling. The colon carcinoma cell line COLO320HSR (ATCC) and the small cell lung cancer cell lines GLC8 (9; a kind gift from Charles Buys, State University of Groningen, Groningen, The Netherlands) and CORL88 (4; a kind gift from Pamela Rabbitts) were grown in tissue culture flasks in RPMI 1640 medium supplemented with 10% fetal calf serum. The neuroblastoma cell lines IMR-32 (ATCC) and Kelly (ATCC) and monkey COS-7 cells (15) were grown in Dulbecco modified Eagle medium supplemented with 10% fetal calf serum (GIBCO). For transfections, about 10^6 COS-7 cells on 10-cm plates were transfected with 3μg of the plasmids using DEAE dextran-mediated DNA transfer (33). Cells were analyzed 60 h after transfection. For labeling of proteins, cells were incubated for 1 h in methionine-free modified Eagle medium supplemented with [^{35}S]methionine (250 μCi/ml, Amersham SJ. 1215), or for 1 h in phosphate-free modified Eagle medium supplemented with $^{32}P_i$ (2.0 mCi/ml, Amersham PBS 43).

Immunoprecipitation and gel electrophoresis. Cell extracts were prepared by washing labeled cells twice with ice-cold phosphate-buffered saline (PBS), followed by lysis in 0.1% SDS–0.5% Nonidet P-40–0.5% sodium deoxycholate–50 mM NaCl–50 mM Tris hydrochloride (pH 7.5). All subsequent steps were carried out at 4°C. Sonicated cell lysates were cleared of insoluble material by centrifugation at 45,000 × *g* for 60 min, and the supernatants were incubated with 2 μl of the anti-N-*myc* antiserum for 4 h followed by the addition of 20 μl of a 50% suspension of protein A-Sepharose (Pharmacia) and further incubation with agitation for 1 h. The precipitates were washed six times with the lysis buffer, once with PBS, and once with water. Proteins were removed from the final protein A-Sepharose pellet by boiling for 3 min in Laemmli sample buffer (12) or for alkaline phosphatase (AP; Boehringer Mannheim) treatment as described by Barber and Verma (5). Control and AP-treated samples were incubated for 60 min at 37°C before addition of an equal volume of 2× Laemmli sample buffer (12). For competitive inhibition of antibody precipitation (blocking) experiments, 5 μg of the bacterial N-*myc* protein was incubated with 10 μl of antiserum overnight at 4°C before its addition to the lysates. Samples were analysed by 8% SDS-PAGE, followed by fixation in 10% acetic acid and impregnation with Amplify (Amersham) before drying and fluorography at −70°C.

Peptide mapping. Partial proteolytic maps of the immunoprecipitated N-*myc* proteins were prepared essentially as described by Cleveland et al. (12). Directly after gel electrophoresis, the gel was dried on a dialysis membrane, and bands of interest were identified by autoradiography, cut out from the gel, rehydrated with a 1:10-diluted Laemmli sample buffer, placed in sample wells of a 15% SDS-polyacrylamide gel together with 500 ng of trypsin, electrophoresed for 6 h, and fluorographed as described above.

Amino acid sequence analysis. The radiolabeled, immunoprecipitated N-*myc* polypeptides were separated in a preparative polyacrylamide gel and blotted electrophoretically in 25 mM Tris-glycine buffer (pH 8.3) containing 20% methanol onto a polyvinylidene difluoride membrane (Immobilon transfer membrane; Millipore Corp., Bedford, Mass.), which was then autoradiographed. Pieces of the filter containing the radioactive protein bands were cut out and used directly as a sample matrix for radiochemical protein sequence analysis using the Applied Biosystems model 421 automated protein sequencer. Radioactivity of collected fractions was determined by scintillation counting.

Immunofluorescence. For immunofluorescence studies the

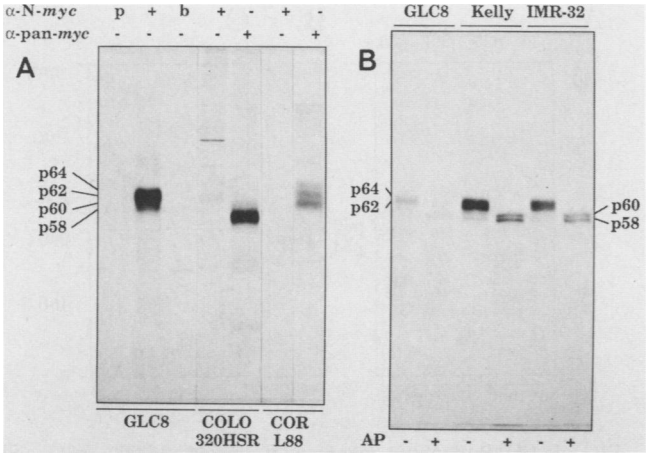

FIG. 1. (A) N-*myc* immunoprecipitation from cells expressing amplified *myc* genes. Lysates of [^{35}S]methionine-labeled GLC8 small-cell lung cancer cells expressing an amplified N-*myc* gene were immunoprecipitated with preimmune rabbit serum (p), anti-N-*myc* antiserum (α-N-*myc*), or immune serum preabsorbed with the N-*myc*–β-galactosidase fusion protein (b). Similar immunoprecipitations were carried out with lysates from COLO320HSR colon carcinoma cells expressing amplified c-*myc* and from CORL88 cells expressing amplified L-*myc*, using either anti-N-*myc* or pan-*myc* serum. The pan-*myc* antiserum recognizes both the c-*myc* and L-*myc* proteins (34). (B) Alkaline phosphatase treatment of the N-*myc* polypeptides. N-*myc* immunoprecipitates from GLC8 small cell lung carcinoma cells and Kelly and IMR-32 neuroblastoma cells were treated with AP (+) and analyzed together with untreated immunoprecipitates (−) as described for panel A.

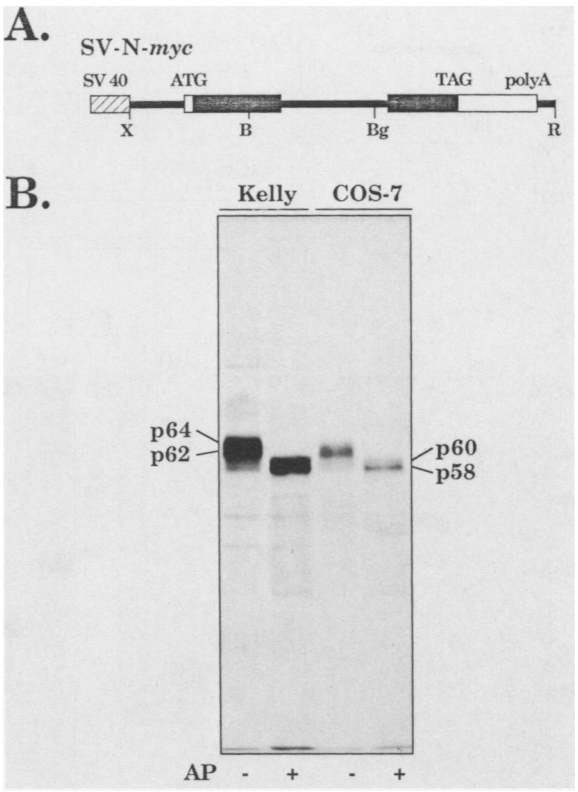

FIG. 2. Structure of the SV-N-*myc* expression vector and analysis of its protein products. (A) Structure of the SV-N-*myc* expression vector with exons II and III shown as bars: black box, long open reading frame; thin line, intron sequences; stippled box, SV40 promoter sequences. A 400-bp *Bgl*II fragment from the N-*myc* second intron was deleted in construction (for details, see Materials and Methods). Transcription of the vector is initiated from the SV40 promoter and terminates after the N-*myc* polyadenylation signal in the noncoding part of the third exon. Vector sequences are not shown. Abbreviations: X, *Xba*I; B, *Bam*HI; Bg, *Bgl*II; R, *Eco*RI; polyA, polyadenylation signal. (B) COS-7 cells were transfected with SV-N-*myc* and analyzed as described for Fig. 1. Kelly neuroblastoma cells were used as a control.

cells were trypsinized and transferred onto cover slips 1 day after transfection. The next day the transfectants were fixed at room temperature after the cover slips were rinsed twice with PBS by incubating them in 4% (wt/vol) paraformaldehyde in PBS for 30 min; the transfectants were then permeabilized with 0.2% Nonidet P-40 in PBS for 2 min. Immunostaining and analysis were performed as described (2).

RESULTS

The two N-*myc* polypeptides are encoded by exons II and III. Immunoprecipitation of the N-*myc* protein from [^{35}S] methionine-labeled neuroblastoma and lung carcinoma cells expressing amplified N-*myc* genes yielded a closely migrating doublet of polypeptides of M_r 64,000 (p64) and 62,000 (p62) in polyacrylamide gel electrophoresis. A fainter doublet of polypeptides of M_r 60,000 (p60) and 58,000 (p58) was seen just beneath the higher-molecular-weight doublet (Fig. 1). The specificity of the immunoprecipitations was controlled with preimmune serum (Fig. 1A, lane p) and with anti-N-*myc* antiserum blocked with gel-purified N-*myc*–β-galactosidase fusion protein used for the immunizations (Fig. 1A, lane b). In addition, our N-*myc* antiserum did not recognize specific bands from COLO320HSR colon carcinoma or CORL88 lung carcinoma cells expressing amplified c-*myc* or L-*myc* genes, respectively, while the pan-*myc* antibodies (34) precipitated a c-*myc* polypeptide doublet from the COLO320HSR cells and three L-*myc* polypeptides from CORL88 cells (Fig. 1A).

After treatment of [^{35}S]methionine-labeled N-*myc* immunoprecipitates with AP, only the two faster-migrating N-*myc* bands (p58 and p60) were seen, suggesting that p62 and p64 are phosphorylated forms of p58 and p60 (Fig. 1B). Similar treatment of N-*myc* immunoprecipitates from $^{32}P_i$-labeled cells removed all radioactivity from the polypeptides (data not shown). The ratio of [^{35}S]methionine-labeled, AP-treated p58 and p60 bands was constant in several experiments and was maintained during a pulse-chase with unlabeled methionine (data not shown), suggesting that the two polypeptides do not have a precursor-product relationship.

To further study the nature of the N-*myc* polypeptide doublet we constructed an expression vector containing the second and third exons of the N-*myc* gene under the SV40 early promoter and the SV40 origin of replication (Fig. 2A). The SV-N-*myc* construct was transfected into COS-7 cells, which are able to replicate SV40 origin-containing plasmids (15). Three days after transfection, the cells were labeled with [^{35}S]methionine and lysed and the lysates were immunoprecipitated with the anti-N-*myc* antiserum. The SV-N-*myc* construct retained the capacity to encode polypeptides comigrating with each of the corresponding untreated and AP-treated polypeptides from Kelly cells (Fig. 2B). These results indicated that both N-*myc* polypeptides are encoded solely by exons II and III.

Deletion mutagenesis suggests that the N-*myc* polypeptides differ in their N termini. The region causing the heterogene-

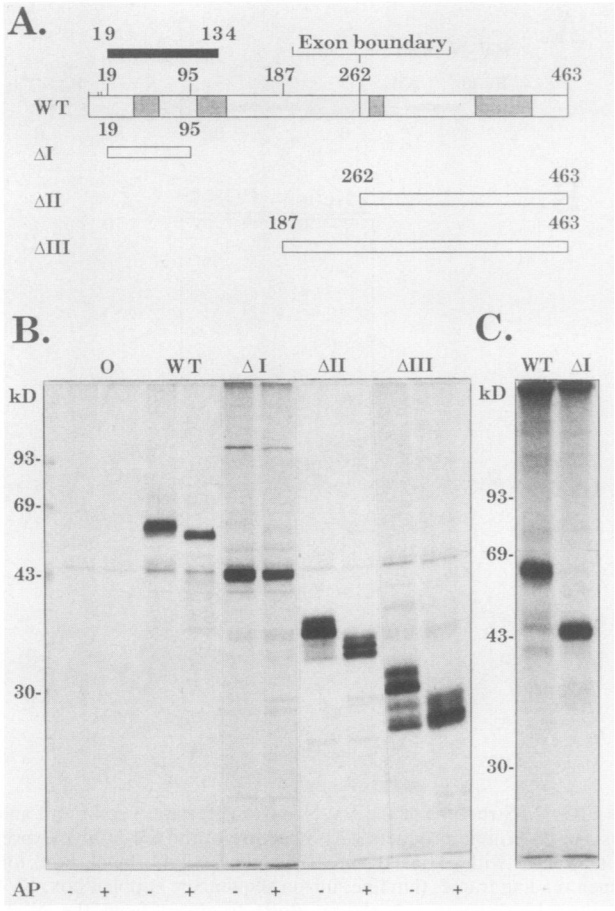

FIG. 3. Deletion mutagenesis of the SV-N-*myc* vector and analysis of the resulting protein products. (A) The N-*myc* protein (WT) is shown schematically, with regions homologous between different *myc* proteins shaded. The black bar on top shows the fragment expressed in the pEX-1 vector for the production of N-*myc* antibodies. Regions of the wild-type protein deleted in the different constructs are shown by open bars (ΔI, ΔII, ΔIII). Numbering denotes the corresponding amino acids. (B) Immunoprecipitation of [^{35}S]methionine-labeled COS-7 cells transfected with the indicated deletion constructs as described for Fig. 1. Lanes: 0, mock transfection; WT, wild-type construct; ΔI, SV-N-ΔI; ΔII, SV-N-ΔII; ΔIII, SV-N-ΔIII. The mobilities of molecular weight markers are indicated on the left. (C) Immunoprecipitation of $^{32}P_i$-labeled COS-7 cells transfected with either the wild type (WT) or the SV-N-ΔI construct (ΔI).

ity of the N-*myc* protein was narrowed down by introducing C-terminal and internal deletions into the SV-N-*myc* expression vector. The vector was modified by deleting (i) an internal fragment of the amino-terminal region of the second exon (SV-N-ΔI), (ii) the third exon (SV-N-ΔII), or (iii) C-terminal portions of the second exon in addition to the third exon (SV-N-ΔIII) (Fig. 3A). The black bar in the Fig. 3A also shows the N-*myc* sequences expressed in bacteria and used for immunization (amino acids 19 through 134). All numbering of amino acids starts from the proline codon at nucleotide 1629 according to the numbering of Stanton et al. (50), thus assuming the cleavage of the initiator methionine (see below). Immunoprecipitates of the deleted proteins expressed in COS-7 cells are shown in Fig. 3B. Several N-*myc* polypeptides were expressed from each one of the deleted constructs, and in each case a characteristic polypeptide doublet was seen after treatment with AP. The

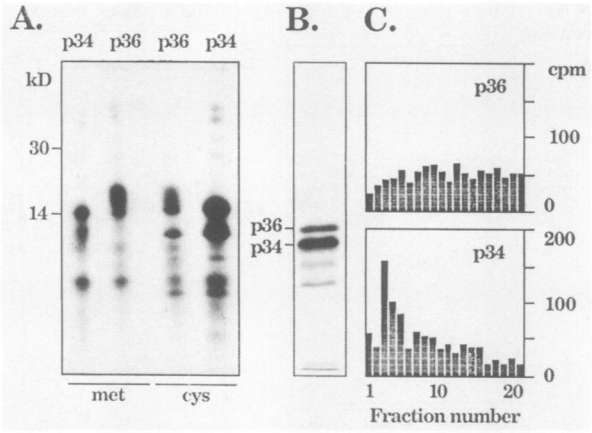

FIG. 4. (A) Proteolytic mapping of the M_r 36,000 (p36) and 34,000 (p34) polypeptides. The p36 and p34 bands of SV-N-ΔII polypeptides labeled with [^{35}S]methionine or [^{35}S]cysteine were separated in a preparative polyacrylamide gel, rehydrated, and subjected to partial proteolysis by 15% SDS-PAGE by the method of Cleveland et al. (12). (B and C) Amino-terminal sequence analysis of p36 and p34 polypeptides synthesized from the SV-N-ΔII expression vector. A portion of the sample used for the sequence analysis is shown in a 10% polyacrylamide gel (B). The bands indicated were transferred onto a polyvinylidene difluoride membrane, cut out, and subjected to automated Edman degradation. The radioactive yields (counts per minute) from the amino acid derivatives are plotted against the degradation cycle number (Fraction number).

SV-N-ΔI construct lacked N-*myc* amino acids 19 through 95 (Fig. 3A), resulting in a predicted polypeptide of 389 amino acids instead of the wild-type polypeptide of 463 amino acids. Two polypeptides of M_r 45,000 and M_r 43,000 were precipitated from cells expressing this construct. Interestingly, hydrolysis of the phosphoester groups with alkaline phosphatase did not visibly alter the migration of the ΔI polypeptide doublet, although the amount of its posttranslational phosphorylation was similar to that of the full-length N-*myc* protein as judged from $^{32}P_i$-labeled immunoprecipitates (Fig. 3B and C).

Deletions II and III (Fig. 3A) resulted in predicted polypeptides of 364 and 210 amino acids, terminating in 102 amino acids encoded by the second intron of the N-*myc* gene (SV-N-ΔII) or in 23 amino acids encoded by plasmid sequences (SV-N-ΔIII), respectively. SV-N-ΔII yielded major polypeptides of M_r 36,000 and 34,000 and minor ones of M_r 33,000 and 32,000, corresponding to wild-type bands of M_r 64,000, 62,000, 60,000, and 58,000, respectively (Fig. 3B), while SV-N-ΔIII yielded the corresponding four bands of M_r 30,000, 28,000, 26,000, and 24,000. In summary, the data obtained from the deletion mutants and from the SV-N-*myc* expression vector indicated that the region responsible for the two bands seen in AP-treated N-*myc* immunoprecipitates is located close to the 5′ end of the second exon.

N-terminal amino acid sequence analysis. The improved electrophoretic separation and the increased yield of the SV-N-ΔII polypeptide products (Fig. 4B) permitted N-terminal amino acid sequence analysis of the M_r 36,000 (p36) and 34,000 (p34) polypeptides. To confirm the common origin of the p36 and p34 polypeptides, the separated bands were subjected to partial proteolysis in polyacrylamide gels. Most of the proteolytic products of the p34 and p36 polypeptides were similar when labeled with [^{35}S]cysteine, and the [^{35}S]methionine-labeled proteolytic products furthermore shared common bands (Fig. 4A). For sequencing,

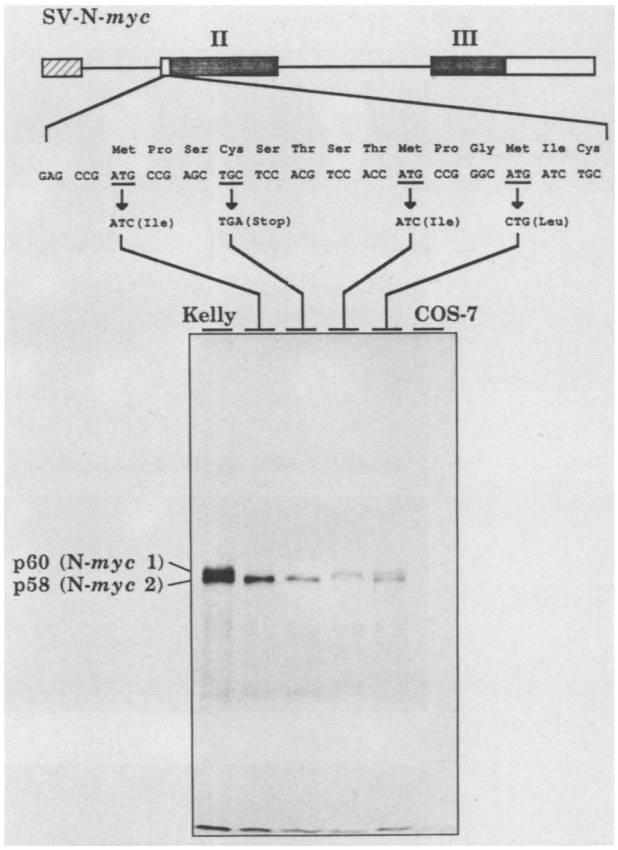

FIG. 5. Site-specific mutagenesis of the N-*myc* amino terminus: analysis by transient transfections. (Top) Structure of the SV-N-*myc* expression vector with sequences surrounding the three possible ATG initiation codons. Arrows point to the mutant codons and corresponding amino acids produced by oligonucleotide-directed point mutagenesis. (Bottom) Mutant SV-N-*myc* polypeptides expressed in COS-7 cells, analyzed as described for Fig. 1, and compared with polypeptides immunoprecipitated from labeled Kelly neuroblastoma cells. Shown is also a control precipitation from nontransfected COS-7 cells. All immunoprecipitates were treated with AP.

immunoprecipitated p34 and p36 were transferred to polyvinylidene difluoride membranes, and the isolated polypeptides were then directly subjected to automated Edman degradation as detailed in Materials and Methods. A graphic presentation of the radioactivity recovered from successive cycles of degradation is shown in Fig. 4C. Analysis of the p36 polypeptide failed to give values of radioactivity over background, but the p34 polypeptide showed a methionine residue as amino acid 3, suggesting that this polypeptide may initiate at codon 1650 (the second of the three AUGs; see below) if the N-terminal methionine residue of p34 polypeptide is cleaved cotranslationally. Moreover, these data strongly suggested that the two sequenced polypeptides differ in their N termini. To define these differences, we introduced point mutations into this region.

In vitro point mutagenesis of the N-terminal coding region. The top part of Fig. 5 shows the 5′ end of the long open reading frame of the N-*myc* gene. Three ATG codons are found in the beginning of the second exon in the same open reading frame at nucleotides 1626, 1650, and 1659 according to the sequence of Stanton et al. (50), but only the second ATG is in a sequence context (TCC ACC **ATG** C) which even partially fits the consensus sequence for translational initiation (GCC [A/G]CC **ATG** G; 27). To determine the translational initiation codons, we used oligonucleotide-directed site-specific mutagenesis to construct mutant SV-N-*myc* plasmids for the transfection assay. The mutagenesis strategy is shown in Fig. 5; the first and the second ATGs were changed into ATC (SV-N-*myc*-2 and SV-N-*myc*-1, respectively) coding for isoleucine, thus preventing translational initiation. In a similar way the third ATG was changed into a CTG encoding leucine. In addition, a TGA stop codon was introduced between the first and second ATGs at position 1635, thus preventing the elongation of any polypeptides initiated upstream of it.

Results obtained from expression of the mutant constructs in COS-7 cells are presented in Fig. 5. Substitution of the upstream ATG with ATC or introduction of a TGA stop codon between the first and second ATGs resulted in complete abolition of the slower-migrating band (p64). This indicated that the first ATG serves as the initiator codon for the p64 N-*myc* polypeptide, referred to as N-*myc*-1. Similarly, the mutation of the second ATG into an ATC prevented synthesis of the faster-migrating N-*myc* band (p62), showing that the second ATG serves as the initiation codon for the p62 polypeptide, referred to as N-*myc*-2. In contrast, a CTG triplet encoding leucine in place of the third ATG had no effect on the polypeptide pattern, confirming that the third ATG does not act as an initiation codon for the N-*myc* protein.

The mutant and wild-type N-*myc* polypeptides are nuclear phosphoproteins. Both wild-type and separately expressed mutant polypeptides (N-*myc*-1 and N-*myc*-2) had similar mobilities in SDS-PAGE and showed a mobility shift upon treatment with phosphatase (Fig. 6), suggesting that the presence of both N-*myc* polypeptides is not necessary for posttranslational phosphorylation. The diffuse appearance of the non-AP-treated bands of the N-*myc*-1 and N-*myc*-2 polypeptides may indicate some further heterogeneity in their phosphorylation. Immunofluorescence analysis of COS-7 cells transfected with SV-N-*myc*-1, SV-N-*myc*-2, or the wild-type vectors is shown in Fig. 7. Cells staining for N-*myc* were seen in all cultures transfected with the above constructs (Fig. 7, left-hand panels). In the positive cells, N-*myc* fluorescence was observed only in the nucleus as shown by the concomitant staining with Hoechst DNA fluorochrome (H 33258; right-hand panels of Fig. 7). The distribution of N-*myc* protein within the nuclei was very similar both for the wild-type SV-N-*myc* construct and for the SV-N-*myc*-1 and SV-N-*myc*-2 constructs expressing only one polypeptide.

DISCUSSION

Immunoprecipitation analysis using a polyclonal antiserum produced against a bacterially expressed fragment of the human N-*myc* protein revealed considerable heterogeneity of the N-*myc* polypeptides. This polymorphism could be in part attributed to posttranslational modification by protein phosphorylation, since only two polypeptide bands remained after AP treatment. The difference between the two polypeptides was further analyzed using an N-*myc* expression vector in COS-7 cells. This vector was used for deletion analysis, amino acid sequencing, and site-specific mutagenesis of the N-*myc* gene. Our results show that the N-*myc* protein is translated from two alternative AUG codons which are located 24 nucleotides apart in the beginning of the second exon of the gene. Use of two initiation codons results in two N-*myc* polypeptides, termed N-*myc*-1 (465 amino

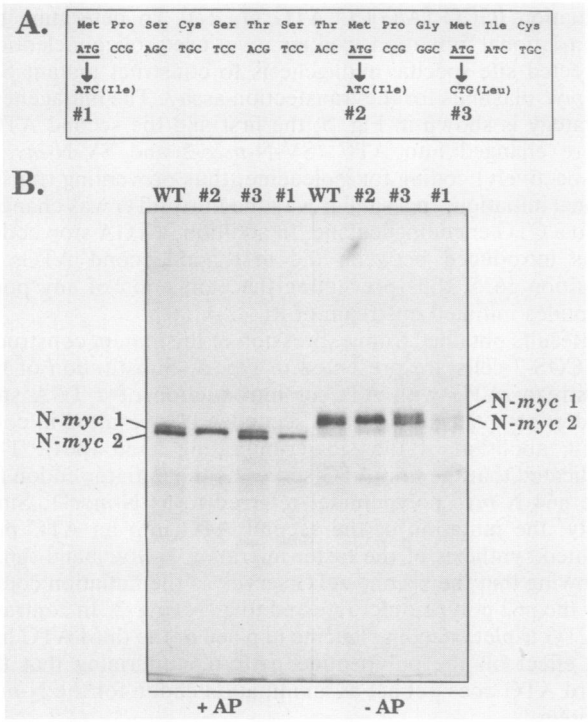

FIG. 6. Posttranslational modification of the separately expressed N-*myc* polypeptides. (A) Mutants analyzed in panel B are indicated by corresponding numbers. Note the 8-amino-acid difference between the N termini of the N-*myc* polypeptides; the translational initiation of the longer polypeptide (N-*myc*-1) is the first ATG codon, and the shorter polypeptide (N-*myc*-2) is initiated from the second ATG. (B) The wild-type SV-N-*myc* expression vector (WT) and vectors with one of three ATGs mutated as shown in panel A were analyzed as described for Fig. 1. Shown on the left side of the autoradiogram are immunoprecipitates treated with AP (+AP), and on the right side are the same samples without AP treatment (−AP). Note the relative lack of resolution of the native samples compared to the corresponding AP-treated ones and the mobility shift seen in all cases after AP treatment.

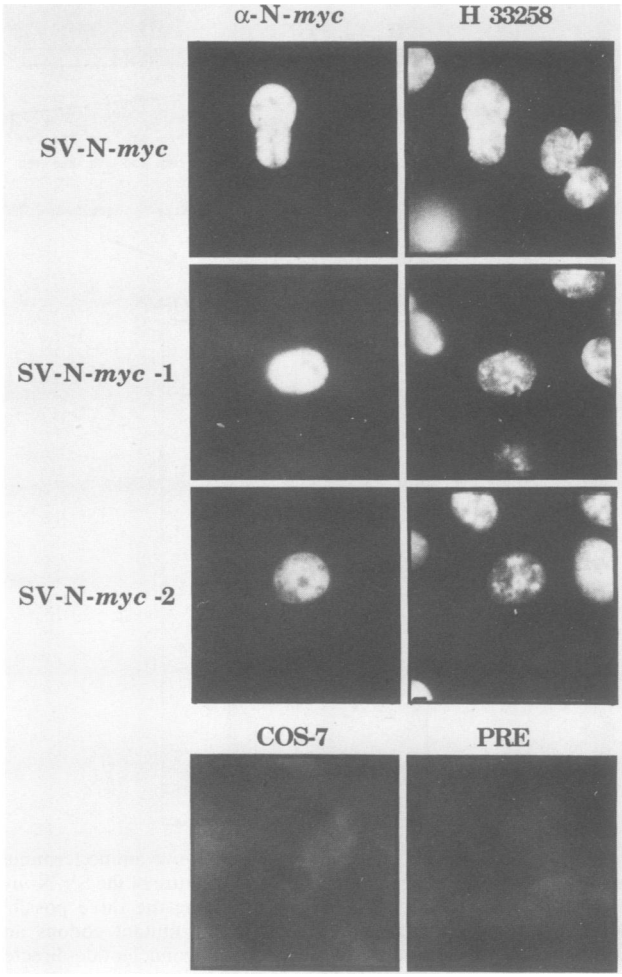

FIG. 7. Indirect immunofluorescence and DNA staining of transfected COS-7 cells. COS-7 cells were fixed and processed for immunofluorescence 60 h after their transfection with the SV-N-*myc*, SV-N-*myc*-1, or SV-N-*myc*-2 constructs (see Fig. 5 and 6 for the polypeptides immunoprecipitated in parallel experiments). Left panels, Fluorescence with the anti-N-*myc* antiserum; right panels, corresponding microscopic fields examined for fluorescence using the DNA fluorochrome Hoechst 33258. Mock-transfected COS-7 cells (COS-7) and SV-N-*myc*-transfected COS-7 cells stained with preimmune serum (PRE) served as controls.

acids) and N-*myc*-2 (458 amino acids). Even when expressed independently, both polypeptide products are phosphorylated and localize to the nucleus.

The probable coding sequence of the N-*myc* gene was first deduced from the genomic sequences of both the human (23, 50) and mouse (13, 51) N-*myc* genes. Three ATG codons were identified at the 5′ end of the single long open reading frame of N-*myc*. The first of these ATGs was suggested to function as a translational initiation codon, since this codon and the downstream coding sequence were conserved in human and murine N-*myc* (23). Alternatively, the second ATG was regarded as the most likely initiation codon (50) because it was in the best sequence context conforming to established sites of translational initiation (27). Our data indicate that at least for the human N-*myc* protein both of these predictions hold true.

Eucaryotic bifunctional mRNAs were first suggested to exist by virtue of the presence of ATG codons upstream from the initiation sites of long open reading frames and by analogy to similar viral sequences known to direct the synthesis of two polypeptides (25). Interestingly, although upstream ATGs are relatively rare in mammalian genes, two thirds of proto-oncogene sequences contain these elements (26). A functional role for short upstream open reading frames has been suggested both for the yeast transcription factor *GCN4* (54) and for the *lck* proto-oncogene (32). In the *lck* proto-oncogene the upstream region contains three out-of-frame ATG codons which are eliminated in the two known *lck* gene rearrangements, resulting in oncogenic activation and an overexpression of the p56 *lck* gene product (32). A similar translational control might exist for N-*myc* and c-*myc*, in which the use of the upstream initiation codon could regulate the accessibility of the downstream initiation codon to the ribosomal apparatus. Additional translational control may be exerted by the several upstream open reading frames of the N-*myc* gene. The potential open reading frames and corresponding translational initiation sites in the L-*myc* gene have not yet been tested, but immunoprecipitated and dephosphorylated L-*myc* protein appears as a single band in SDS-PAGE (K. Saksela, T. P. Mäkelä, and K. Alitalo, EMBO J., in press), suggesting the absence of a similar polymorphism.

The evolutionary conservation of two translational initiation codons in the N-*myc* and c-*myc* genes (17; unpublished

results) suggests a functional significance for this phenomenon. Comparison of the unique 8-amino-acid N-terminal stretch of the N-*myc*-1 polypeptide reveals no homology with c-*myc* or other GenBank protein sequences. However, several roles have been attributed to short N-terminal regions of proteins (for a review, see reference 40). An N-terminal domain is responsible for the short half-life of the adenovirus E1A protein (46), which possibly involves proteolytic cleavage of an N-terminal peptide (31). On the other hand, four N-terminal amino acids of the β-tubulin subunit autoregulate the stability of β-tubulin mRNA (58). Interestingly, inhibition of translation is also known to stabilize the mRNAs of several oncogenes, including N-*myc* and c-*myc*, resulting in elevated levels of the corresponding mRNAs (29; unpublished results), and recently it has been postulated that ongoing translation of the *fos* mRNA itself is required for its degradation (55). In light of these data it is tempting to speculate that the level of N-*myc* mRNA would be regulated by selective translation of only one of the N-*myc* polypeptides.

The deduced amino acid sequence of N-*myc* contains a C-terminal stretch of leucines present in several enhancer binding proteins and termed the leucine zipper (28). This region is known to stabilize the heterodimer of the *fos* and *jun* proteins and to enhance the DNA binding of this complex (24, 42). Although experimental evidence of a dimeric complex of *myc* polypeptides has only been reported for the v-*myc* protein (3), the two dissimilar *myc* polypeptides might have specific roles in such a dimeric complex.

The biological significance of two N-*myc* polypeptides is unclear. During differentiation of mouse erythroleukemia cells, a relative increase of the longer c-*myc* polypeptide (c-*myc*-1), apparently owing to a stabilization of c-*myc*-1, has been reported (56). A similar change in the relative amounts of the c-*myc* polypeptides occurs in fibroblastic cells when they reach confluency in cell culture (S. Hann, personal communication). Therefore, it is interesting that synthesis of c-*myc*-1 was found to be inhibited in 11 of 17 examined Burkitt's lymphoma cell lines, owing either to a chromosomal translocation with a breakpoint in the first intron or to point mutations disrupting the translational initiation sequences (17). Thus, the absence of c-*myc*-1 protein may be one of several mechanisms deregulating c-*myc* in Burkitt's lymphomas. We are currently studying the possibility of a similar differential regulation of the two N-*myc* polypeptides. However, no evidence of a disruption of synthesis of one of the N-*myc* polypeptides in naturally occurring tumors has yet been reported.

Our results show that the second and third exons of the N-*myc* gene contain the coding capacity for both N-*myc* polypeptides. Thus the genomic organization of the N-*myc* gene differs from that of the c-*myc* gene, where the first exon is also needed to produce the two c-*myc* polypeptides (17). The short distance between the two N-*myc* polypeptide initiation sites may make the N-*myc* protein less vulnerable to the mutations or translocations found in c-*myc* of tumor cells.

ACKNOWLEDGMENTS

We thank Saija Pihkanen and Elina Roimaa for technical assistance and Marja-Liisa Huhtala for help in radiosequencing.

The study was carried out under a research contract with the Finnish Life and Pension Insurance Companies and was supported also by grants from the Finnish Medical Foundation Duodecim, Emil Aaltonen Foundation, Ida Montin Foundation, and the Finnish Cancer Organizations.

LITERATURE CITED

1. **Alitalo, K., P. Koskinen, T. P. Mäkelä, K. Saksela, L. Sistonen, and R. Winqvist.** 1987. *myc* oncogenes: activation and amplification. Biochim. Biophys. Acta **907**:1–32.
2. **Alitalo, K., G. Ramsay, J. M. Bishop, S. Ohlsson-Pfeifer, W. W. Colby, and A. D. Levinson.** 1983. Identification of nuclear proteins encoded by viral and cellular *myc* oncogenes. Nature (London) **306**:274–277.
3. **Bader, J. P., and D. A. Ray.** 1985. MC29 virus-coded protein occurs as monomers and dimers in transformed cells. J. Virol. **53**:509–514.
4. **Baillie-Johnson, H., P. R. Twentyman, N. E. Fox, G. A. Walls, P. Workman, J. V. Watson, N. Johnson, J. G. Reeve, and N. M. Bleehen.** 1985. Establishment and characterisation of cell lines from patients with lung cancer (predominantly small cell carcinoma). Br. J. Cancer **52**:495–504.
5. **Barber, J. B., and I. M. Verma.** 1987. Modification of *fos* proteins: phosphorylation of c-*fos*, but not v-*fos*, is stimulated by 12-tetradecanoylphorbol-13-acetate and serum. Mol. Cell. Biol. **7**:2201–2211.
6. **Battey, J., C. Moulding, R. Taub, W. Murphy, T. Stewart, H. Potter, G. Lenoir, and P. Leder.** 1983. The human c-*myc* oncogene: structural consequences of translocation into the IgH locus in Burkitt lymphoma. Cell **34**:779–787.
7. **Bishop, J. M.** 1987. The molecular genetics of cancer. Science **235**:305–311.
8. **Brodeur, G. M., R. C. Seeger, M. Schwab, H. E. Varmus, and J. M. Bishop.** 1984. Amplification of N-*myc* in untreated human neuroblastomas correlates with advanced disease stage. Science **224**:1121–1124.
9. **Broers, J. L. V., D. N. Carney, L. de Ley, G. P. Vooijs, and F. C. S. Ramaekers.** 1985. Differential expression of intermediate filament proteins distinguishes classic from variant small-cell lung cancer cell lines. Proc. Natl. Acad. Sci. USA **82**:4409–4413.
10. **Cavalieri, F., and M. Goldfarb.** 1988. N-*myc* proto-oncogene expression can induce DNA replication in Balb/c 3T3 fibroblasts. Oncogene **2**:289–291.
11. **Classon, M., M. Henriksson, J. Sümegi, G. Klein, and M.-L. Hammaskjöld.** 1987. Elevated c-*myc* expression facilitates the replication of SV40 DNA in human lymhoma cells. Nature (London) **330**:272–274.
12. **Cleveland, D. W., S. G. Fisher, M. W. Kirschner, and U. K. Laemmli.** 1977. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. J. Biol. Chem. **252**:1102–1106.
13. **DePinho, R. A., E. Legouy, L. B. Feldman, N. E. Kohl, G. D. Yancopoulos, and F. W. Alt.** 1986. Structure and expression of the murine N-*myc* gene. Proc. Natl. Acad. Sci. USA **83**: 1827–1831.
14. **Garson, J. A., J. Clayton, P. McIntyre, and J. T. Kemshead.** 1986. N-*myc* oncogene amplification in rhabdomyosarcoma at release. Lancet **i**:1496.
15. **Gluzman, Y.** 1981. SV40-transformed simian cells support the replication of early SV40 mutants. Cell **23**:175–182.
16. **Hann, S. R., H. D. Abrams, L. R. Rohrschneider, and R. N. Eisenman.** 1983. Proteins encoded by v-*myc* and c-*myc* oncogenes: identification and localization in acute leukemia virus transformants and bursal lymphoma cell lines. Cell **34**:789–798.
17. **Hann, S. R., M. W. King, D. L. Bentley, C. W. Anderson, and R. N. Eisenman.** 1988. A non-AUG translational initiation in c-*myc* exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. Cell **52**:185–195.
18. **Iguchi-Ariga, S. M. M., T. Itani, Y. Kiji, and H. Ariga.** 1987. Possible function of the c-*myc* product: promotion of cellular DNA replication. EMBO J. **6**:2365–2371.
19. **Ikegaki, N., J. Bukowsky, and R. H. Kennet.** 1986. Identification and characterization of the N-*myc* gene product in human neuroblastoma cells by monoclonal antibodies with defined specificities. Proc. Natl. Acad. Sci. USA **83**:5929–5933.
20. **Jakobovits, A., M. Schwab, J. M. Bishop, and G. R. Martin.** 1985. Expression of N-*myc* in teratocarcinoma stem cells and mouse embryos. Nature (London) **318**:188–191.

21. **Kingston, R. E., A. S. Baldwin, Jr., and P. A. Sharp.** 1984. Regulation of heat shock protein 70 gene expression by c-*myc*. Nature (London) **312**:280–282.
22. **Klein, G., and E. Klein.** 1985. Evolution of tumors and the impact of molecular oncology. Nature (London) **315**:190–195.
23. **Kohl, N. E., E. Legouy, R. A. dePinho, P. D. Nisen, R. K. Smith, C. E. Gee, and F. W. Alt.** 1986. Human N-*myc* is closely related in organization and nucleotide sequence to c-*myc*. Nature (London) **319**:73–77.
24. **Kouzarides, T., and E. Ziff.** 1988. The role of the leucine zipper in the fos-jun interaction. Nature (London) **336**:646–651.
25. **Kozak, M.** 1986. Bifunctional messenger RNAs in eukaryotes. Cell **47**:481–483.
26. **Kozak, M.** 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15**:8125–8148.
27. **Kozak, M.** 1987. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. J. Mol. Biol. **196**:947–950.
28. **Landschultz, W. H., P. F. Johnson, and S. L. McKnight.** 1988. The leucine zipper: a hypothetic structure common to a new class of DNA binding proteins. Science **240**:1759–1764.
29. **Lau, L. F., and D. Nathans.** 1987. Expression of a set of growth-related immediate early genes in BALB/c 3T3 cells: coordinate regulation with c-*fos* or c-*myc*. Proc. Natl. Acad. Sci. USA **84**:1182–1186.
30. **Lee, W.-H., A. L. Murphree, and W. F. Benedict.** 1984. Expression and amplification of the N-*myc* gene in primary retinoblastoma. Nature (London) **309**:458–460.
31. **Lucher, L. A., K. H. Brackmann, J. S. Symington, and M. Green.** 1986. Posttranslational modification at the N terminus of the human adenovirus type 12 E1A 235R tumor antigen. J. Virol. **58**:592–599.
32. **Marth, J. D., R. W. Overell, K. E. Meier, E. G. Krebs, and R. M. Perlmutter.** 1988. Translational activation of the *lck* proto-oncogene. Nature (London) **332**:171–173.
33. **McClutchan, J. H., and J. S. Pagano.** 1968. Enhancement of the infectivity of simian virus 40 deoxyribonucleic acid with diethylaminoethyl-dextran. J. Natl. Cancer Inst. **41**:351–357.
34. **Moore, J. P., D. C. Hancock, T. D. Littlewood, and G. I. Evan.** 1987. A sensitive and quantitative enzyme-linked immunosorbence assay for the c-*myc* and N-*myc* proteins. Oncogen. Res. **2**:65–80.
35. **Nau, M. M., B. J. Brooks, Jr., D. N. Carney, A. F. Gazdar, J. G. Battey, E. A. Sausville, and J. D. Minna.** 1986. Human small-cell lung cancers show amplification and expression of the N-*myc* gene. Proc. Natl. Acad. Sci. USA **83**:1092–1096.
36. **Persson, H., and P. Leder.** 1984. Nuclear localization and DNA binding properties of a protein expressed by human c-*myc* oncogene. Science **225**:718–721.
37. **Piechaczyk, M., J.-M. Blanchard, and P. Jeanteur.** 1987. c-*myc* regulation still holds its secret. Trends Genet. **3**:47–51.
38. **Ramsay, G., G. I. Evan, and J. M. Bishop.** 1984. The protein encoded by the human proto-oncogene c-*myc*. Proc. Natl. Acad. Sci. USA **81**:7742–7746.
39. **Ramsay, G., L. Stanton, M. Schwab, and J. M. Bishop.** 1986. Human proto-oncogene N-*myc* encodes nuclear proteins that bind DNA. Mol. Cell. Biol. **6**:4450–4475.
40. **Rechsteiner, M., S. Rogers, and K. Rote.** 1987. Protein structure and intracellular stability. Trends Biochem. Sci. **12**:390–394.
41. **Saksela, K., J. Bergh, and K. Nilsson.** 1986. Amplification of the N-*myc* oncogene in an adenocarcinoma of the lung. J. Cell. Biochem. **31**:297–304.
42. **Sassone-Corsi, P., L. J. Ransone, W. W. Lamph, and I. M. Verma.** 1988. Direct interaction between fos and jun nuclear oncoproteins: role of the "leucine zipper" domain. Nature (London) **336**:692–695.
43. **Schwab, M., K. Alitalo, K.-H. Klempnauer, H. E. Varmus, J. M. Bishop, F. Gilbert, G. Brodeur, M. Goldstein, and J. Trent.** 1983. Amplified DNA with limited homology to the *myc* cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumor. Nature (London) **305**:245–248.
44. **Schwab, M., H. E. Varmus, and J. M. Bishop.** 1985. Human N-*myc* gene contributes to neoplastic transformation of mammalian cells in culture. Nature (London) **316**:160–162.
45. **Slamon, D. J., T. C. Boone, R. C. Seeger, D. E. Keith, V. Chazin, H. C. Lee, and L. M. Souza.** 1986. Identification and characterization of the protein encoded by the human N-*myc* oncogene. Science **232**:768–772.
46. **Slavicek, J. M., N. C. Jones, and J. D. Richter.** 1988. Rapid turnover of adenovirus E1A is determined through a co-translational mechanism that requires an amino-terminal domain. EMBO J. **7**:3171–3180.
47. **Small, M., N. Hay, M. Schwab, and J. M. Bishop.** 1987. Neoplastic transformation by the human gene N-*myc*. Mol. Cell. Biol. **7**:1638–1645.
48. **Spector, D. L., R. A. Watt, and N. F. Sullivan.** 1987. The v- and c-*myc* oncogene proteins colocalize in situ with small nuclear ribonucleoprotein particles. Oncogene **1**:5–12.
49. **Stanley, K. K., and J. P. Luzio.** 1984. Construction of a new family of high efficiency bacterial expression vectors: identification of cDNA clones coding for human liver proteins. EMBO J. **3**:1429–1434.
50. **Stanton, L. W., M. Schwab, and J. M. Bishop.** 1986. Nucleotide sequence of the human N-*myc* gene. Proc. Natl. Acad. Sci. USA **83**:1772–1776.
51. **Taya, Y., S. Mizusawa, and S. Nishimura.** 1986. Nucleotide sequence of the coding region of the mouse N-*myc* gene. EMBO J. **5**:1215–1219.
52. **Thiele, C. J., C. P. Reynolds, and M. A. Israel.** 1985. Decreased expression of N-*myc* prcedes retinoic acid-induced morphological differentiation of human neuroblastoma. Nature (London) **313**:404–406.
53. **Westin, E. H., F. Wong-Staal, E. P. Gelmann, R. D. Favera, T. S. Papas, J. A. Lautenberger, A. Eva, E. P. Reddy, S. R. Tronick, S. A. Aaronson, and R. C. Gallo.** 1982. Expression of cellular homologues of retroviral *onc* genes in human hematopoietic cells. Proc. Natl. Acad. Sci. USA **79**:2490–2494.
54. **Williams, N. P., P. P. Mueller, and A. G. Hinnebusch.** 1988. The positive regulatory function of the 5′-proximal open reading frame in *GCN4* mRNA can be mimicked by heterologous, short coding sequences. Mol. Cell. Biol. **8**:3827–3836.
55. **Wilson, T., and R. Treisman.** 1988. Removal of poly(A) and consequent degradation of c-*fos* mRNA facilitated by 3′ AU-rich sequences. Nature (London) **336**:396–399.
56. **Wingrove, T. G., R. Watt, P. Keng, and I. G. Macara.** 1988. Stabilization of the *myc* proto-oncogene proteins during Friend murine erythroleukemia cell differentiation. J. Biol. Chem. **263**:8918–8924.
57. **Yancopoulos, G. D., P. D. Nisen, A. Tesfaye, N. E. Kohl, M. P. Goldfarb, and F. W. Alt.** 1985. N-*myc* can cooperate with *ras* to transform normal cells in culture. Proc. Natl. Acad. Sci. USA **82**:5455–5459.
58. **Yen, T. J., P. S. Machlin, and D. W. Cleveland.** 1988. Autoregulated instability of β-tubulin mRNAs by recognition of the nascent amino terminus of β-tubulin. Nature (London) **334**: 580–585.
59. **Zimmerman, K. A., G. D. Yancopoulos, R. G. Collum, R. K. Smith, N. E. Kohl, K. A. Denis, M. M. Nau, O. N. Witte, D. Toran-Allerand, C. E. Gee, J. D. Minna, and F. W. Alt.** 1986. Differential expression of the *myc* family genes during murine development. Nature (London) **319**:780–783.