The EMBO Journal vol.5 no.6 pp.1215–1219, 1986

Nucleotide sequence of the coding region of the mouse N-*myc* gene

Yoichi Taya, Saeko Mizusawa and Susumu Nishimura

Biology Division, National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo 104, Japan

Communicated by W. Fiers

A genomic clone for the mouse N-*myc* gene was isolated and the total nucleotide sequence (4807 bp) of the two coding exons and an intron located between them was determined. The amino acid sequence of the N-*myc* protein was deduced from the DNA sequence. This protein is composed of 462 amino acids, slightly larger than human and mouse c-*myc* proteins, and is rich in proline like the c-*myc* protein. Comparison of the amino acid sequences of the mouse N-*myc* and c-*myc* proteins showed that conserved sequences are located in eight regions: four regions are in the N-terminal half of the N-*myc* protein and are separated from each other by regions poorly homologous to those of the c-*myc* protein, and the four others are located in the C-terminal half, throughout which certain homology exists. A remarkable sequence containing 13 successive acidic amino acids is present in one of the conserved regions located in the middle of the N-*myc* protein.

Key words: DNA sequence/*myc* family gene/N-*myc* gene/oncogene

Introduction

The N-*myc* gene, which bears a DNA sequence homologous to that of c-*myc*, was originally detected as an amplified DNA sequence in human neuroblastomas (Schwab *et al.*, 1983; Kohl *et al.*, 1983; Montgomery *et al.*, 1983). Amplification of the N-*myc* gene was also observed in retinoblastomas (Lee *et al.*, 1984) and small cell lung cancers (Nau *et al.*, 1984; Nau *et al.*, 1985) A correlation of N-*myc* gene amplification with tumor progression was shown in neuroblastomas (Brodeur *et al.*, 1983; Seeger *et al.*, 1985).

The N-*myc* mRNA was found to be expressed in all neuroblastomas (Schwab *et al.*, 1984; Kohl *et al.*, 1984) and retinoblastomas (Lee *et al.*, 1984) examined, even in the absence of N-*myc* gene amplification. The amount of N-*myc* mRNA is greatly enhanced by this gene amplification. Its expression was not detectable in many other normal and tumor cells examined, except small cell lung cancers. Recently, however, it was found to be expressed at high levels in teratocarcinoma stem cells and embryos, and at lower levels in adult brain, testis and kidney (Jakobovits *et al.*, 1985).

The ability of the N-*myc* gene to contribute to transformation of cultured normal cells was shown by co-transfection with the *ras* gene (Schwab *et al.*, 1985; Yancopoulos *et al.*, 1985). This finding showed that N-*myc* has c-*myc*-like activity. Taken together, these findings suggest an important role of N-*myc* in the genesis and/or progression of specific cancers and in the differentiation of certain tissues.

A third *myc*-related gene, L-*myc*, was also found recently in some small cell lung cancers (Nau *et al.*, 1985). Therefore, it is intriguing to know how these *myc* family genes are selectively expressed in the development of cancers and the differentiation of tissues and what the function of these *myc* family gene proteins is. For studies on these problems, it is necessary to know the structure of the N-*myc* gene. However, only its partial sequences have so far been reported: Schwab *et al.* (1983) sequenced a 0.35-kbp fragment of the human N-*myc* gene, finding two blocks of 72 and 48 bp long that show homology to exon 2 of c-*myc*; Michitsch and Melera (1985) determined the sequence of a 1.67-kbp fragment that is homologous to exon 3 of c-*myc*. Here, we report the total nucleotide sequence of the coding regions of the mouse N-*myc* gene.

Results and discussion

Isolation of a genomic clone for the mouse N-myc gene

For isolation of a human N-*myc* clone, a human genomic library was constructed by ligation of DNA isolated from a human giant cell lung carcinoma propagated in a nude mouse into a λ Charon 4A phage (Taya *et al.*, 1984). The library was screened with the 2.0-kb *Eco*RI fragment of the human N-*myc* gene (Kohl *et al.*, 1983) as a probe. In this way, a weakly hybridized clone, λclone91, was obtained, in addition to a strongly hybridized clone that contains the human N-*myc* gene. The λclone91 seemed likely to contain the mouse N-*myc* gene, since human tumors maintained in nude mice are frequently contaminated with mouse cells.

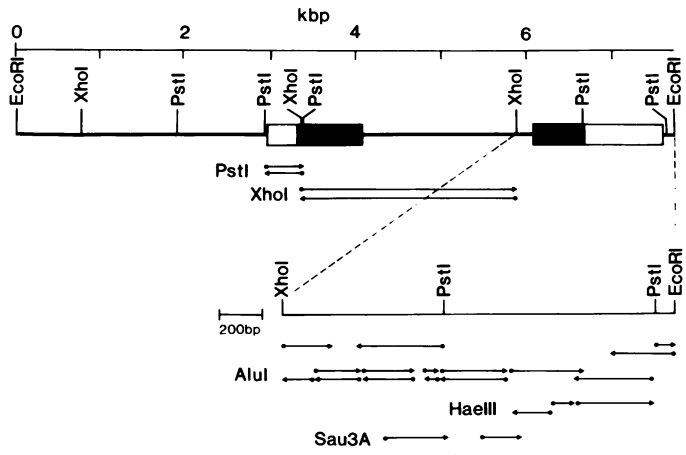
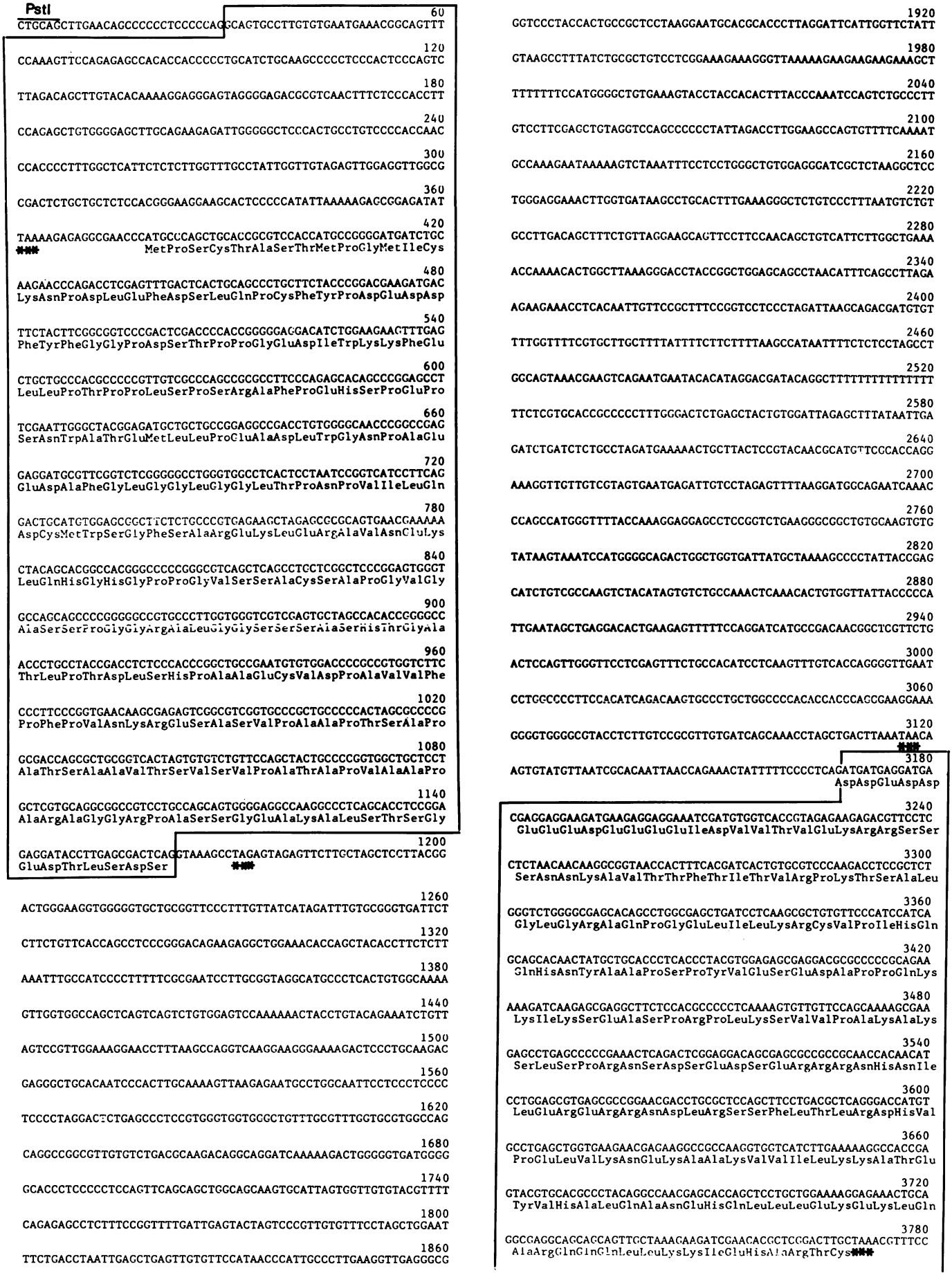


Fig. 1. Physical map of the DNA fragment containing the mouse N-*myc* gene. The boxed regions indicate coding exons. Solid boxes represent protein coding regions. The 5′ end of the gene is to the left. The strategy for determining the nucleotide sequence is shown by arrows, indicating the extents and directions of sequence analyses. A magnification of the *Xho*I–*Eco*RI fragment is shown. Fragments produced by the indicated restriction enzymes were subcloned in vectors M13mp18 and/or M13mp19, and sequenced by the dideoxy chain termination method using dc^{7}GTP in place of dGTP (Mizusawa *et al.*, 1986). The nucleotide sequence of the 0.45-kb *Pst*I fragment was determined by use of dGTP. The *Xho*I fragment of 2.5 kb was sequenced by the stepwise deletion method (Yanisch-Perron *et al.*, 1985)


PstI
60
CTGCAGCTTGAACAGCCCCCCTCCCCCAGGCAGTGCCTTGTGTGAATGAAACGGCAGTTT
120
CCAAAGTTCCAGAGAGCCACACCACCCCCTGCATCTGCAAGCCCCCTCCCACTCCCAGTC
180
TTAGACAGCTTGTACACAAAAGGAGGGAGTAGGGGAGACGCGTCAACTTTCTCCCACCTT
240
CCAGAGCTGTGGGGAGCTTGCAGAAGAGATTGGGGGCTCCCACTGCCTGTCCCCACCAAC
300
CCACCCCTTTGGCTCATTCTCTCTTGGTTTGCCTATTGGTTGTAGAGTTGGAGGTTGGCG
360
CGACTCTGCTGCTCTCCACGGGAAGGAAGCACTCCCCCATATTAAAAAGAGCGGAGATAT
420
TAAAAGAGAGGCGAACCCATGCCCAGCTGCACCGCGTCCACCATGCCGGGCATGATCTGC
*** MetProSerCysThrAlaSerThrMetProGlyMetIleCys
480
AAGAACCCAGACCTCGAGTTTGACTCACTGCAGCCCTGCTTCTACCCGGACGAAGATGAC
LysAsnProAspLeuGluPheAspSerLeuGlnProCysPheTyrProAspGluAspAsp
540
TTCTACTTCGGCGGTCCCGACTCGACCCCACCGGGGGAGGACATCTGGAAGAACTTTGAG
PheTyrPheGlyGlyProAspSerThrProProGlyGluAspIleTrpLysLysPheGlu
600
CTGCTGCCCACGCCCCCGTTGTCGCCCAGCCGCGCCTTCCCAGAGCACAGCCCGGAGCCT
LeuLeuProThrProProLeuSerProSerArgAlaPheProGluHisSerProGluPro
660
TCGAATTGGGCTACGGAGATGCTGCTGCCGGAGGCCGACCTGTGGGGCAACCCGGCCGAG
SerAsnTrpAlaThrGluMetLeuLeuProGluAlaAspLeuTrpGlyAsnProAlaGlu
720
GAGGATGCGTTCGGTCTCGGGGGCCTGGGTGGCCTCACTCCTAATCCGGTCATCCTTCAG
GluAspAlaPheGlyLeuGlyGlyLeuGlyGlyLeuThrProAsnProValIleLeuGln
780
GACTGCATGTGGAGCGGCTTCTCTGCCCGTGAGAAGCTAGAGCGCGCAGTGAACGAAAAA
AspCysMetTrpSerGlyPheSerAlaArgGluLysLeuGluArgAlaValAsnGluLys
840
CTACAGCACGGCCACGGGCCCCCGGGCGTCAGCTCAGCCTCCTCGGCTCCCGGAGTGGGT
LeuGlnHisGlyHisGlyProProGlyValSerSerAlaCysSerAlaProGlyValGly
900
GCCAGCAGCCCCGGGGGCCGTGCCCTTGGTGGGTCGTCGAGTGCTAGCCACACCGGGGCC
AlaSerSerProGlyGlyArgAlaLeuGlyGlySerSerSerAlaSerHisThrGlyAla
960
ACCCTGCCTACCGACCTCTCCCACCCGGCTGCCGAATGTGTGGACCCCGCCGTGGTCTTC
ThrLeuProThrAspLeuSerHisProAlaAlaGluCysValAspProAlaValValPhe
1020
CCCTTCCCGGTGAACAAGCGAGAGTCGGCGTCGGTGCCCGCTGCCCCCACTAGCGCCCCG
ProPheProValAsnLysArgGluSerAlaSerValProAlaAlaProThrSerAlaPro
1080
GCGACCAGCGCTGCGGTCACTAGTGTGTCTGTTCCAGCTACTGCCCCGGTGGCTGCTCCT
AlaThrSerAlaAlaValThrSerValSerValProAlaThrAlaProValAlaAlaPro
1140
GCTCGTGCAGGCGGCCGTCCTGCCAGCAGTGGGGAGGCCAAGGCCCTCAGCACCTCCGGA
AlaArgAlaGlyGlyArgProAlaSerSerGlyGluAlaLysAlaLeuSerThrSerGly
1200
GAGGATACCTTGAGCGACTCAGGTAAAGCCTAGAGTAGAGTTCTTGCTAGCTCCTTACGG
GluAspThrLeuSerAspSer ***
1260
ACTGGGAAGGTGGGGGTGCTGCGGTTCCCTTTGTTATCATAGATTTGTGCGGGTGATTCT
1320
CTTCTGTTCACCAGCCTCCCGGGACAGAAGAGGCTGGAAACACCAGCTACACCTTCTCTT
1380
AAATTTGCCATCCCCTTTTTCGCGAATCCTTGCGGTAGGCATGCCCTCACTGTGGCAAAA
1440
GTTGGTGGCCAGCTCAGTCAGTCTGTGGAGTCCAAAAAACTACCTGTACAGAAATCTGTT
1500
AGTCCGTTGGAAAGGAACCTTTAAGCCAGGTCAAGGAAGGGAAAAGACTCCCTGCAAGAC
1560
GAGGGCTGCACAATCCCACTTGCAAAAGTTAAGAGAATGCCTGGCAATTCCTCCCTCCCC
1620
TCCCCTAGGACTCTGAGCCCTCCGTGGGTGGTGGGCTGTTTGCGTTTGGTGCGTGGCCAG
1680
CAGGCCGGCGTTGTGTCTGACGCAAGACAGGCAGGATCAAAAAGACTGGGGGTGATGGGG
1740
GCACCCTCCCCCTCCAGTTCAGCAGCTGGCAGCAAGTGCATTAGTGGTTGTGTACGTTTT
1800
CAGAGAGCCTCTTTCCGGTTTTGATTGAGTACTAGTCCCGTTGTGTTTCCTAGCTGGAAT
1860
TTCTGACCTAATTGAGCTGAGTTGTGTTCCATAACCCATTGCCCTTGAAGGTTGAGGGCG
1920
GGTCCCTACCACTGCCGCTCCTAAGGAATGCACGCACCCTTAGGATTCATTGGTTCTATT
1980
GTAAGCCTTTATCTGCGCTGTCCTCGGAAAGAAAGGGTTAAAAAGAAGAAGAAGAAAGCT
2040
TTTTTTTCCATGGGGCTGTGAAAGTACCTACCACACTTTACCCAAATCCAGTCTGCCCTT
2100
GTCCTTCGAGCTGTAGGTCCAGCCCCCCTATTAGACCTTGGAAGCCAGTGTTTTCAAAAT
2160
GCCAAAGAATAAAAAGTCTAAATTTCCTCCTGGGCTGTGGAGGGATCGCTCTAAGGCTCC
2220
TGGGAGGAAACTTGGTGATAAGCCTGCACTTTGAAAGGGCTCTGTCCCTTTAATGTCTGT
2280
GCCTTGACAGCTTTCTGTTAGGAAGCAGTTCCTTCCAACAGCTGTCATTCTTGGCTGAAA
2340
ACCAAAACACTGGCTTAAAGGGACCTACCGGCTGGAGCAGCCTAACATTTCAGCCTTAGA
2400
AGAAGAAACCTCACAATTGTTCCGCTTTCCGGTCCTCCCTAGATTAAGCAGACGATGTGT
2460
TTTGGTTTTCGTGCTTGCTTTTATTTTCTTCTTTTAAGCCATAATTTTCTCTCCTAGCCT
2520
GGCAGTAAACGAAGTCAGAATGAATACACATAGGACGATACAGGCTTTTTTTTTTTTTTT
2580
TTCTCGTGCACCGCCCCCTTTGGGACTCTGAGCTACTGTGGATTAGAGCTTTATAATTGA
2640
GATCTGATCTCTGCCTAGATGAAAAACTGCTTACTCCGTACAACGCATGTTCGCACCAGG
2700
AAAGGTTGTTGTCGTAGTGAATGAGATTGTCCTAGAGTTTTAAGGATGGCAGAATCAAAC
2760
CCAGCCATGGGTTTTACCAAAGGAGGAGCCTCCGGTCTGAAGGGCGGCTGTGCAAGTGTG
2820
TATAAGTAAATCCATGGGGCAGACTGGCTGGTGATTATGCTAAAAGCCCCTATTACCGAG
2880
CATCTGTCGCCAAGTCTACATAGTGTCTGCCAAACTCAAACACTGTGGTTATTACCCCCA
2940
TTGAATAGCTGAGGACACTGAAGAGTTTTTCCAGGATCATGCCGACAACGGCTCGTTCTG
3000
ACTCCAGTTGGGTTCCTCGAGTTTCTGCCACATCCTCAAGTTTGTCACCAGGGGTTGAAT
3060
CCTGGCCCCTTCCACATCAGACAAGTGCCCTGCTGGCCCCACACCACCCAGCGAAGGAAA
3120
GGGGTGGGGCGTACCTCTTGTCCGCGTTGTGATCAGCAAACCTAGCTGACTTAAATAACA
***
3180
AGTGTATGTTAATCGCACAATTAACCAGAAACTATTTTTCCCCTCAGATGATGAGGATGA
AspAspGluAspAsp
3240
CGAGGAGGAAGATGAAGAGGAGGAAATCGATGTGGTCACCGTAGAGAAGAGACGTTCCTC
GluGluGluAspGluGluGluGluIleAspValValThrValGluLysArgArgSerSer
3300
CTCTAACAACAAGGCGGTAACCACTTTCACGATCACTGTGCGTCCCAAGACCTCCGCTCT
SerAsnAsnLysAlaValThrThrPheThrIleThrValArgProLysThrSerAlaLeu
3360
GGGTCTGGGGCGAGCACAGCCTGGCGAGCTGATCCTCAAGCGCTGTGTTCCCATCCATCA
GlyLeuGlyArgAlaGlnProGlyGluLeuIleLeuLysArgCysValProIleHisGln
3420
GCAGCACAACTATGCTGCACCCTCACCCTACGTGGAGAGCGAGGACGCGCCCCCCGCAGAA
GlnHisAsnTyrAlaAlaProSerProTyrValGluSerGluAspAlaProProGlnLys
3480
AAAGATCAAGAGCGAGGCTTCTCCACGCCCCCTCAAAAGTGTTGTTCCAGCAAAAGCGAA
LysIleLysSerGluAlaSerProArgProLeuLysSerValValProAlaLysAlaLys
3540
GAGCCTGAGCCCCCGAAACTCAGACTCGGAGGACAGCGAGCGCCGCCGCAACCACAACAT
SerLeuSerProArgAsnSerAspSerGluAspSerGluArgArgArgAsnHisAsnIle
3600
CCTGGAGCGTGAGCGCCGGAACGACCTGCGCTCCAGCTTCCTGACGCTCAGGGACCATGT
LeuGluArgGluArgArgAsnAspLeuArgSerSerPheLeuThrLeuArgAspHisVal
3660
GCCTGAGCTGGTGAAGAACGAGAAGGCCGCCAAGGTGGTCATCTTGAAAAAGGCCACCGA
ProGluLeuValLysAsnGluLysAlaAlaLysValValIleLeuLysLysAlaThrGlu
3720
GTACGTGCACGCCCTACAGGCCAACGAGCACCAGCTCCTGCTGGAAAAGGAGAAACTGCA
TyrValHisAlaLeuGlnAlaAsnGluHisGlnLeuLeuLeuGluLysGluLysLeuGln
3780
GGCCAGGCAGCAGCAGTTCCTAAAGAAGATCGAACACGCTCCGACTTGCTAAACGTTTCC
AlaArgGlnGlnGlnLeuLeuLysLysIleGluHisAlaArgThrCys***

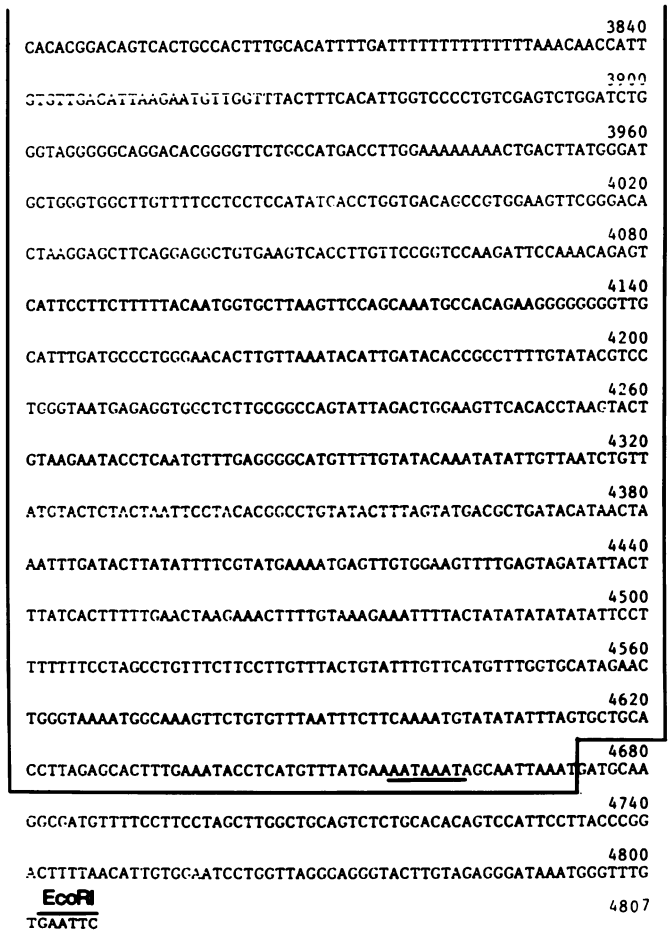


Fig. 2. Nucleotide sequence and predicted amino acid sequence of the mouse N-*myc* gene. The exons are enclosed in continuous lines. The end of the 3′-exon was deduced by consideration of the nucleotide sequence homology with that of the human N-*myc* gene (Michitsch and Melera, 1985). A polyadenylation signal is underlined. Asterisks indicate nonsense codons that limit the extent of the open reading frames for the protein coding regions.

This possibility was tested by Southern blot analysis of *Eco*RI digests of mouse and human DNAs using as probes the human N-*myc* 2.0-kb *Eco*RI fragment and the 7.7-kb *Eco*RI fragment derived from the λclone91. With the former probe, a weak 7.7-kb band and a strong 2.0-kb band were detected in mouse and human DNAs, respectively. On the other hand, with the latter probe a strong 7.7-kb band and a weak 2.0-kb band were detected in mouse and human DNAs, respectively (data not shown). Thus the λclone91 was concluded to be a mouse N-*myc* clone.

Subsequently, the 7.7-kb *Eco*RI fragment of the mouse N-*myc* gene was subcloned into a plasmid, pUC19, and its nucleotide sequence was analysed.

Nucleotide sequence of the mouse N-myc gene

A restriction map of the 7.7-kb *Eco*RI fragment was constructed as shown in Figure 1. The region that hybridized with the 2.0-kb *Eco*RI fragment of the human N-*myc* was localized on the 0.45-kb *Pst*I fragment (position 2.90−3.35 kb) and the 2.5-kb *Xho*I fragment (position 3.34−5.84 kb). In addition, the 1.9-kb *Xho*I−*Eco*RI fragment (position 5.84−7.7 kb) was found to hybridize with the 3.8-kb *Eco*RI fragment of human N-*myc*, which contains the 3′ coding region (Michitsch *et al.*, 1984; Michitsch and Melera, 1985) (data not shown). Thus, the total DNA sequence of these regions (4807 bp) was determined by the dideoxy nucleotide method (Figure 2).

The putative 5′ and 3′ coding exons (boxed) can be identified from this sequence because the former bears homology to exon 2 of c-*myc* (Colby *et al.*, 1983, Battey *et al.*, 1983; Bernard *et al.*, 1983) as well as the corresponding human N-*myc* region (Schwab *et al.*, 1983) and the latter shows homology to exon 3 of c-*myc* and to the 3′-coding region of human N-*myc* (Michitsch and Melera, 1985). The intron between the two coding regions was concluded to be as indicated in Figure 2 for the following reasons: Positions 1162 and 3167 should be the donor and acceptor sites, respectively, for splicing since the sequences around these positions are consistent with the consensus sequence for the splice junction (Mount, 1982). In addition, the amino acid sequence adjacent to the donor site (Ser Asp Ser, residues 259−261) is identical to that of c-*myc* (Figure 3). The Glu, Asp-rich sequence (residues 262−274), near the acceptor site also resembles that of c-*myc* (Figure 3). Furthermore, nonsense codes (positions 1171−1173 and 3116−3118) in the same reading frames as these coding regions are located near this junction. Michitsch and Melera (1985) proposed a splicing acceptor site of the 3′-exon of the human N-*myc* protein, which corresponds to the nucleotide 3254 of mouse N-*myc* (Figure 2). If splicing takes place at position 3254, two unusual features have to be explained: first, part of the coding sequence of c-*myc* is conserved in an intron of N-*myc*, as they noticed, and second, there is no homology in the amino acid sequence between positions 3254 and 3313 although the nucleotide sequences of mouse and human N-*myc* are highly homologous. Assuming that one nucleotide is

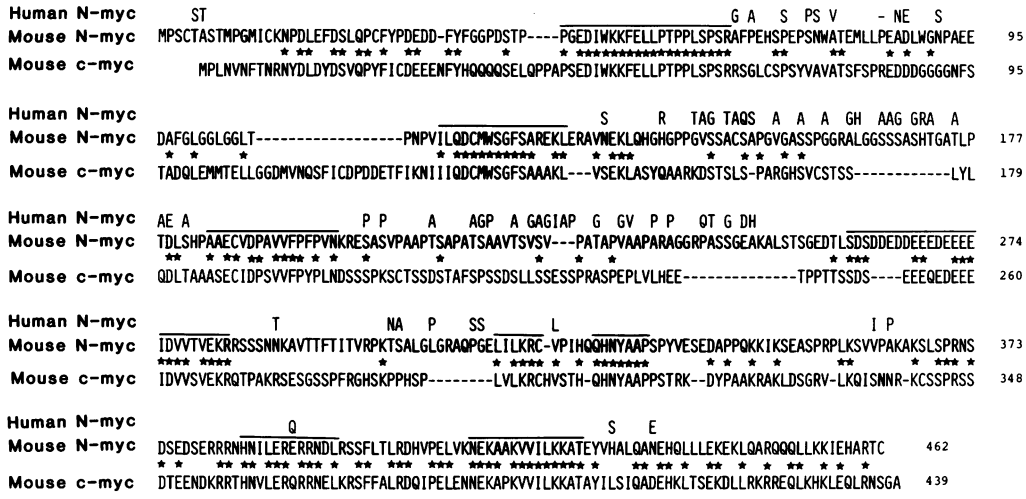


Fig. 3. Comparison of the amino acid sequences of mouse and human N-*myc* with that of c-*myc*. Each human residue differing from the murine is indicated above it. The human N-*myc* sequence and the mouse c-*myc* sequence are from Kohl *et al.* (1986) and Bernard *et al.* (1983), respectively. Solid stars indicate identical amino acids in mouse N-*myc* and c-*myc*. Strongly conserved sequences in N-*myc* and c-*myc* are overlined.

missing between positions 282 and 290 of the human N-*myc* DNA sequence (corresponding to positions 3304−3312 in Figure 2) because of technical difficulty in sequencing the GC-rich region, the splicing site is probably at position 145 of human N-*myc* (corresponding to position 3167 in Figure 2) for the following reasons: (i) The nucleotide sequences around this splicing site in the mouse and human N-*myc* genes are more consistent than those around the site proposed by Michitsch and Melera with consensus for an acceptor site for splicing. (ii) The amino acid sequence from residues 253−307 (Figure 3) of the mouse N-*myc* protein, which is highly homologous to the c-*myc* protein, is also conserved in the amino acid sequence of the human N-*myc* protein deduced from the reading frame. Another putative splice acceptor site is located at position 30. The sequence adjacent to this site is completely identical to the consensus sequence of the acceptor site (Mount, 1982), which suggests the presence of a noncoding exon in the upstream area like exon 1 of c-*myc*. This assumption must be confirmed by cloning and sequencing N-*myc* cDNA.

The initiation methionine codon was concluded to be as shown in Figure 2, from the presence of a nonsence codon (position 361−363) closely upstream to it and the fact that the nucleotide sequence downstream but not upstream from this methionine codon is very similar in the mouse and human N-*myc*, which we partially sequenced (data not shown). Possibly, however, the second ATG codon (position 403−405) is the initiation codon, since the nucleotide sequence around it is more preferable for translation initiation (Kozak, 1983) than that of the first ATG.

Comparison of the amino acid sequences of the N-myc and c-myc proteins

As shown in Figure 3, the N-*myc* protein is rich in proline, like the c-*myc* protein. The homologous regions in N-*myc* and c-*myc* are found in at least eight clusters. By comparison of nucleotide sequences, Schwab *et al.*, (1983) found two homologous regions in exon 2 of human c-*myc* and the 0.35-kbp fragment of human N-*myc*. These two regions correspond to amino acids 45−65 and 111−127 of mouse N-*myc* in Figure 3. Here, we noticed rather weak homologies in a region close to the N-terminus (residues 15−36).

A unique sequence containing 13 Asp and Glu residues is present in the middle of the N-*myc* protein (residues 261−274). The c-*myc* protein has a similar acidic region although it is shorter than that of N-*myc*. This region could have a role in interacting with basic proteins such as the histones of chromatin because the c-*myc* protein is suggested to be a DNA binding protein localized in the nucleus (Persson and Leder, 1984; Hann and Eisenman, 1984) and the N-*myc* protein is expected to have the same property. As shown in Figure 3, the N-terminal half consists of conserved sequences and poorly homologous sequences, while the C-terminal half (residues 316−462) shows some homology, including those in four well-conserved regions, throughout its length, as was observed between human N-*myc* and c-*myc* by Michitsch and Melera (1985). Rabbitts *et al.* (1983) and Papas and Lautenberger (1985), found that threonine at position 58 of c-*myc* is replaced by other amino acids in v-*myc* and translocated c-*myc* of Raji Burkitt's lymphoma cells, although the surrounding sequence is entirely conserved. The sequence of mouse as well as human N-*myc* is also conserved in this region, but serine is substituted for threonine in L-*myc* (Nau *et al.*, 1985).

While this paper was in preparation, Kohl *et al.* (1986) reported the nucleotide sequence of the human N-*myc* gene. They showed the existence of a 5′-noncoding exon by cloning and sequencing of a cDNA. The splice sites between exon 2 and exon 3 and the polyadenylation site indicated in their structure are consistent with ours. The deduced amino-acid sequence of their human N-*myc* protein is also very similar to that of our mouse N-*myc* protein (Figure 3). However, the 5′-boundary of exon 2 is different from the site we indicated. Our sequence CCCTCCCCCAG/G is more consistent than their sequence GTCGGTTGCAG/T with the consensus sequence (Mount, 1982) although they determined the boundary by sequencing the cDNA.

Materials and methods

Cloning of the mouse N-myc gene

A human genomic library was constructed by ligation of nude mouse tumor DNA of a human giant cell lung carcinoma to λ Charon 4A phage arms (Taya *et al.*, 1984). The library was screened by plaque hybridization at 42°C in the presence of 50% formamide as described previously (Taya *et al.*, 1984) with a ^{32}P-labelled probe of the 2.0-kb *Eco*RI fragment of the human N-*myc* clone pN-*myc* (Kohl *et al.*, 1983). A weakly hybridized clone was obtained by screening 200 000 independent plaques. This clone was identified as a mouse N-*myc* clone as described in the text.

DNA sequencing

The various fragments shown in Figure 1 were subcloned in vectors M13mp18 and/or M13mp19 (Yanisch-Perron *et al.*, 1985) and sequenced by the dideoxy nucleotide method (Sanger *et al.*, 1977) with [α-^{32}P]dATP (Amersham, 410 μCi/mmole). The nucleotide sequences of most parts were determined by use of 7-deazadeoxyguanosine triphosphate (dc^{7}GTP) in place of dGTP to facilitate determination of sequences of GC rich regions (Mizusawa *et al.*, 1986).

Acknowledgements

We thank Drs F.Alt and N.Kohl for providing pN-*myc*, and Mr N.Arai for help in the isoaltion of λ clones. This work was supported by a Grant-in-Aid from the Ministry of Health and Welfare for a Comprehensive 10-Year Strategy for Cancer Control, Japan, by a Grant-in-Aid for Cancer Research from the Ministry of Health and Welfare and by a Naito Foundation Natural Science Scholarship.

References

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.M. (1983) *EMBO J.*, **2**, 2375–2383.

Brodeur,G.M., Seeger,R.C., Schwab,M., Varmus,H.E. and Bishop,J.M. (1983) *Science*, **224**, 1121–1124.

Colby,W.W., Chen,E.Y., Smith,D.H. and Levinson,A.D. (1983) *Nature*, **301**, 722–725.

Hann,S. and Eisenman,R. (1984) *Mol. Cell. Biol.*, **3**, 829–838.

Jackobovits,A., Schwab,M., Bishop,J.M. and Martin,G.R. (1985) *Nature*, **318**, 188–191.

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.E., Gee,C.E. and Alt,F. (1984) *Science*, **226**, 1335–1337.

Kohl,N.E., Legouy,E., DePinho,R.A., Nisen,P.D., Smith,R.K., Gee,C.E. and Alt,F.W. (1986) *Nature*, **319**, 73–77.

Kozak,M. (1983) *Microbiol. Rev.*, **47**, 1–45.

Lee,W.-H., Murphree,A.L. and Benedict,W.F. (1984) *Nature*, **309**, 458–460.

Michitsch,R.W., Montgomery,K.T. and Melera,P.W. (1984) *Mol. Cell. Biol.*, **4**, 2370–2380.

Michitsch,R.W. and Melera,P.W. (1985) *Nucleic Acids Res.*, **13**, 2545-2558.

Mizusawa,S., Nishimura,S. and Seela,F. (1986) *Nucleic Acids Res.*, **14**, 1319–1324.

Montgomery,K.T., Biedler,J.L., Spengler,B.A. and Melera,P.W. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5724–5728.

Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.

Nau,M.M., Carney,D.N., Battey,J., Johnson,B., Little,C., Gazdar,A. and Minna,J.D. (1984) *Curr. Top. Microbiol. Immunol.*, **113**, 172–177.

Nau,M.M., Brooks,B.J., Battey,J., Sausville,E., Gazdar,A.F., Kirsch,I.R., McBride,O.W., Bertness,V., Hollis,G.F. and Minna,J.D. (1985) *Nature*, **318**, 69–73.

Papas,T.S. and Lautenberger,J.A. (1985) *Nature*, **318**, 237.

Persson,H. and Leder,P. (1984) *Science*, **225**, 718–721.

Rabbitts,T.H., Hamlyn,P.H. and Baer,R. (1983) *Nature*, **306**, 760-765.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Schwab,M., Alitalo,K., Klempnauer,K.-H., Varmus,H.E., Bishop,J.M., Gilbert,F., Brodeur,G., Goldstein,M. and Trent,J. (1983) *Nature*, **305**, 245–248.

Schwab,M., Ellison,J., Bush,M., Rosenau,W., Varmus,H.E. and Bishop,J.M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4940–4944.

Schwab,M., Varmus,H.E. and Bishop,J.M. (1985) *Nature*, **316**, 160–162.

Seeger,R.C., Brodeur,G.M., Sather,H., Dalton,A., Siegel,S.E., Wong,K.Y. and Hammond,D. (1985) *New Engl. J. Med.*, **313**, 1111–1116.

Taya,Y., Hosogai,K., Hirohashi,S., Shimosato,Y., Tsuchiya,R., Tsuchida,N., Fushimi,M., Sekiya,T. and Nishimura,S. (1984) *EMBO J.*, **3**, 2943–2946.

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.

Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.

Received on 3 February 1986; revised on 1 April 1986.