Volume 15 Number 16 1987 Nucleic Acids Research

Characterization of rat c-*myc* and adjacent regions

Kenshi Hayashi, Reiko Makino, Hideki Kawamura, Akira Arisawa and Kenji Yoneda

Biochemistry Division, National Cancer Center Research Institute 1-1, Tsukiji 5-Chome, Chuo-ku, Tokyo, Japan

Received May 13, 1987; Revised and Accepted July 5, 1987 Accession no. Y00396

ABSTRACT

Rat genomic regions covering c-myc were cloned from the DNA of both normal liver and two lines of Morris hepatomas, one of which had c-myc amplification. The three restriction maps showed perfect agreement within the overlapping regions. The 7 kb regions, which included the entire normal rat c-myc and the region 2.2 kb upstream, and one from the hepatomas, were sequenced and found to be identical. The coding regions of exons 2 and 3 were highly conserved between rat, mouse and man, but some differences in amino acids were noted. Exon 1 and the non-coding region of exon 3 showed limited homology between the three species. Rat exon 1 contained several nonsense codons in each frame and no ATG codon, indicating there to be no coding capacity in this exon. The 2.2 kb upstream regions and the introns compared showed unusual conservation between the rat and human genes. Some motifs, previously proposed as having a functional role in human c-myc, were also found in equivalent positions of the rat sequence.

Nucleas S1 protection mapping revealed the second promoter to be preferentially used in most tissues or in hepatoma cells, and the second poly A addition signal to be the only one functional in all the RNA sources examined.

INTRODUCTION

MYC was initially discovered in the form of a viral oncogene of an avian myelocytomatosis virus, MC 29 (1), and subsequently identified in various vertebrate genomes in the forms of its cellular counterpart, c-myc, and transducing v-myc in several oncogenic retroviruses (2,3). The structure of c-myc is highly conserved through evolutionary stages (4), and the product, c-myc protein, is known to be localized in the nucleus (5) and to bind to DNA (6). Several cellular functions, including growth competence or differentiation, seem to require the expression or suppression of c-myc, although the

precise role which the gene has in these functions is not yet clear. Various types of cancer cells overexpress c-myc mRNA (7), some of them as a result of rearrangement or amplification (8). Such gene overexpression or deregulation is suspected to be a cause of the unlimited growth of the cancer cells. Expression of c-myc seems to be regulated in a complicated manner involving transcription initiation (9), transcript elongation (10, 11) and mRNA degradation (12, 13). Such regulation requires sequences of functional significance to be located upstream, and within exons and introns. Recent studies using human c-myc have suggested both positive and negative regulatory sequences to be located upstream and within exon 1 and intron 1 (14, 15). Rearrangements resulting from translocations or retroviral insertions far upstream of the gene have been reported in some B-cell lymphomas (7), suggesting the importance of remote sequence elements in the regulation of the gene. One approach towards identifing such elements is the search for evolutionally conserved sequences in the gene and its neighboring regions. To date, an 8 kb sequence of the entire human c-myc and its flanking regions has been reported (16, 17, 18). Another mammalian c-myc sequence available is that of the mouse (19, 20, 21) although, in this case, information on the upstream region is limited and the sequence of intron 2 is unavailable.

In the present papar we report the cloning of normal rat c-myc and that of the Morris hepatomas, 5123D and 7794A. Overexpression of the c-myc mRNA in both hepatomas, and amplification of the gene in the latter, have been reported previously (22). Approximately 7 kb of both normal rat and Morris hepatoma 7794A, that including the gene and the 2.2 kb upstream region, were sequenced. The sequences from the two sources were exactly the same. Comparison between rat and human sequences revealed remarkably conserved stretches distributed throughout the 2.2 kb upstream region and both introns, as well as in the exons.

It is shown that the second of the two promoters is very much preferred as the mRNA initiation site in various tissues and cultured hepatoma cells, and also that, in all tissues and

cells examined, only the signal downstream of the dual termination (poly A addition) is used.

MATERIALS AND METHODS

Cloning

A transplanted Morris hepatoma 7794A genomic library was constructed by partial digestion of its DNA with Sau3A1, ligation with arms of EMBL3a (23) and in vitro packaging (24). The library was screened by plaque-hybridization (25) using the 3'-half of the v-myc sequence (26) as a probe, and three overlapping clones were obtained. A detailed restriction map was constructed for one of the clones, λM77-5.

c-myc clones from normal Buffalo rat liver (λB-14) and Morris hepatoma 5123D (λD-5) were obtained by complete digestion of their DNA with EcoRI, ligation and packaging into phage Charon 4A, and screening with a probe identical to that used for the isolation of λM77-5.

DNA sequencing of c-*myc* from Morris hepatoma 7794A

After appropriate subcloning of regions of λM77-5 into plasmids, the PstI-PstI 4 kb fragment (5'-flank, exon 1 and most of intron 1) or the SacI-HpaI 3 kb fragment (half of intron 1, exon 2, intron 2 and most of exon 3) were sonicated, cloned into M13mp10 using Deininger's method (27) and sequenced by dideoxy chain termination (28) using an M13 sequencing kit (Amersham). The HpaI-HpaI 0.6 kb fragment (3'-end of exon 3 and its flank) was separately cloned into M13mp10 and sequenced from both ends.

DNA sequencing of c-*myc* from normal rat

A 2.1 kb HindIII-HindIII fragment (5'-flank to 5'-end of exon 1) and a HindIII-XbaI 4.3 kb fragment (5'-end of exon 1 to end of coding region of exon 3) were cloned into M13mp18 or mp19, progressively deleted by the method of Yanisch-Perron et al. (29) and se quenced. When necessary, 7-deaza-2'-dGTP was used in place of dGTP in the sequencing reaction to alleviate the compaction of bands in the GC-rich regions (30).

Computer analyses

Sequences from gel readings were assembled using the GENIAS (Mitsui Knowledge Inc.) or GENETYX programs (Japan Soft-

ware Development Company). All other computer analyses were done by UWGCG (31) or IDEAS (32) programs run on the VAX11/750 of the Institute of Medical Science, Tokyo University.

Nuclease S1 protection assay(33)

Fragments spanning either promoter sites or poly A addition sites were cloned into M13mp18. Strand specific probes (10^8 cpm/μg) were synthesized using a sequencing primer (Amersham), and isolated by alkaline agarose gels after digestion with appropriate enzymes (34). Probes (1-1.5 x 10^5 cpm) were hybridized with either total or poly A^+ RNA in 0.3 M NaCl, 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA (3'-end) or in 80% formamide, 0.4 M NaCl, 20 mM PIPES, pH 6.8, and 10 mM EDTA (5'-end) at 45 °C overnight. The annealed mixture was digested with nuclease S1 (Takara) at 37 °C for 30 min. The protected fragments were run in 8 M urea, 4% acrylamide gels.

RESULTS

Structure of rat c-*myc*

Fig 1 shows the restriction map of the rat c-myc locus cloned from three independent sources, normal Buffalo rat liver and Morris hepatomas 5123D and 7794A. The three clones had exactly the same restriction sites in the overlapping regions. The approximate location and orientation of c-myc were determined by Southern hybridizations of appropriately restricted fragments to the 5'- or 3'-half of v-myc. There are some discrepancies between the restriction map shown in Fig. 1 and that of the Lou/Wsl rat reported by Pear et al. (35), especially at the exon 1 region. The difference in rat strain may possibly explain this discrepancy. We examined the HindIII and BamHI sites of the exon 1 region in both Sprague-Dawley and Wistar rat strains, and the results agreed with the map presented here.

The regions spanning approximately 7 kb of λM77-5 (from Morris hepatoma 7794A with amplified c-myc) and λB-13 (from normal Buffalo rat liver) were sequenced independently using the strategy outlined in Fig. 1. The two sequences turned out to be exactly the same, suggesting no structural change to have occurred during c-myc amplification in the Morris hepatoma,

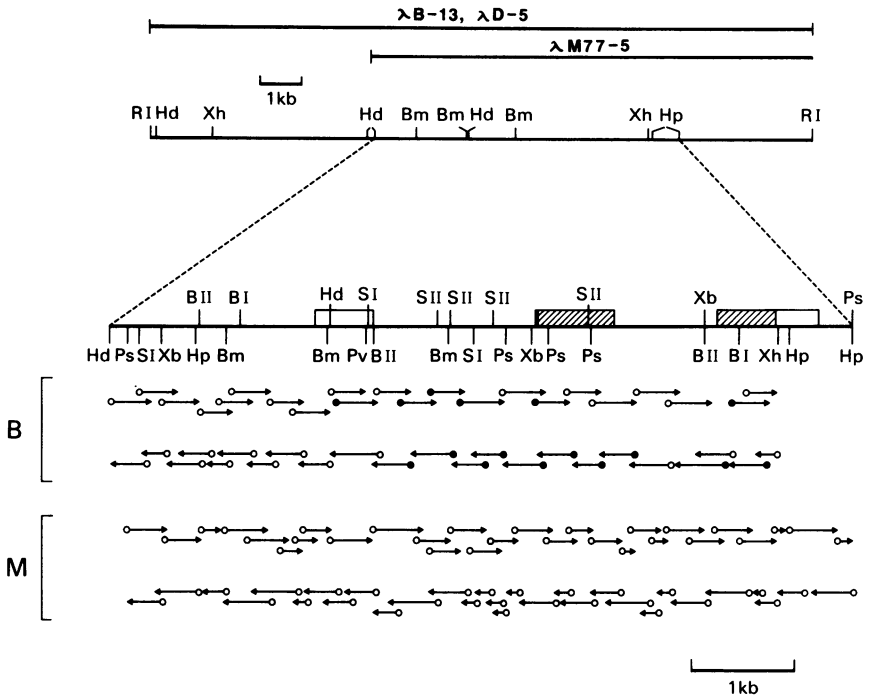


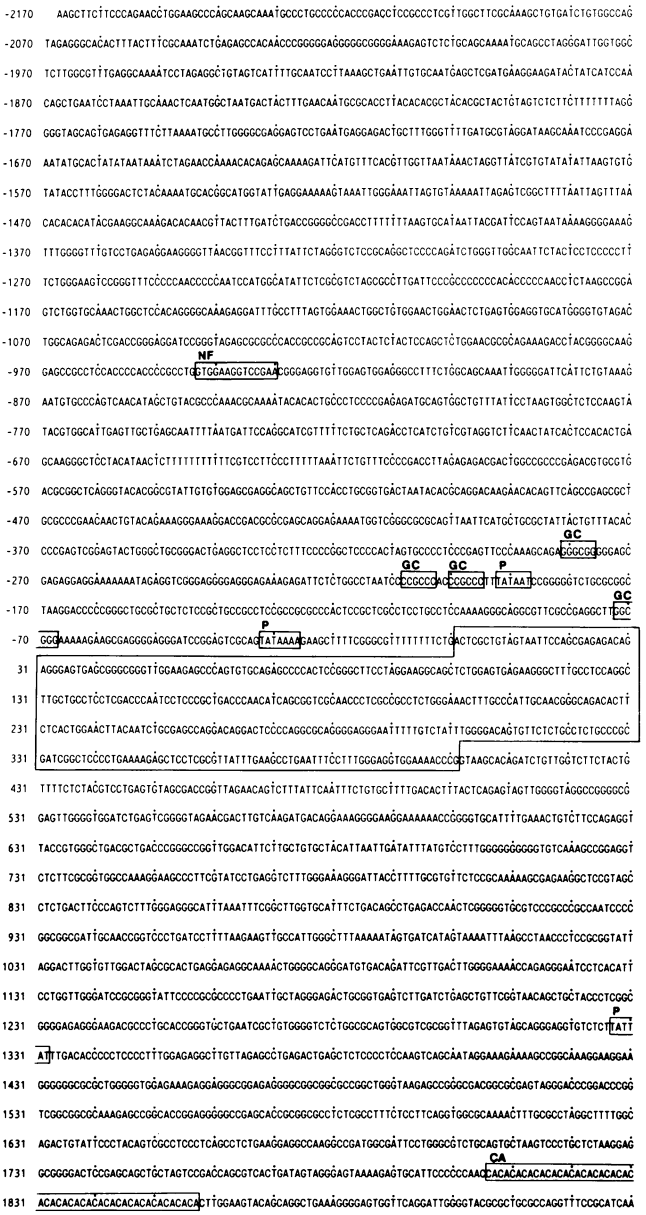
Fig. 1 Restriction map and sequencing strategies of rat c-myc. The regions covered by clones λB-13 (from normal rat), λD-5 (from Morris hepatoma 5123D) and λM77-5 (from Morris hepatoma 7794A) are indicated at the top. Restriction enzymes are abbreviated as follows: RI, Eco RI; Hd, Hind III; Xh, Xho I; Bm, BamH I; Hp, Hpa I; Ps, Pst I; SI, Sac I; Xb, Xba I; BII, Bgl II; BI, Bgl I; Pv, Pvu I; SII, Sac II. Coding and non-coding exons are indicated with hatched and open boxes on the map. Arrows starting from open and filled circles show dideoxy sequencing reactions using normal and 7-deaza GTP, respectively. B and M to left of arrows indicate sequencing of λB-13 and λM77-5 regions, respectively.

7794A. The agreed sequence is presented in Fig. 2. The positions of the exons were identified by homology with human (18) and mouse (20) sequences and also from results of nuclease S1 mapping as described below.

Sequence motifs in the upstream region

Expression of eukaryotic genes are understood to be regulated by proteins which recognize and bind to specific sequence motifs usually located at sites upstream of the genes.

Some such regulatory proteins have been characterized and their recognition sequences determined. Siebenlist et al. (36) has identified a sequence, GTGGAAGGNANCCAA, at -1504 to -1491 and



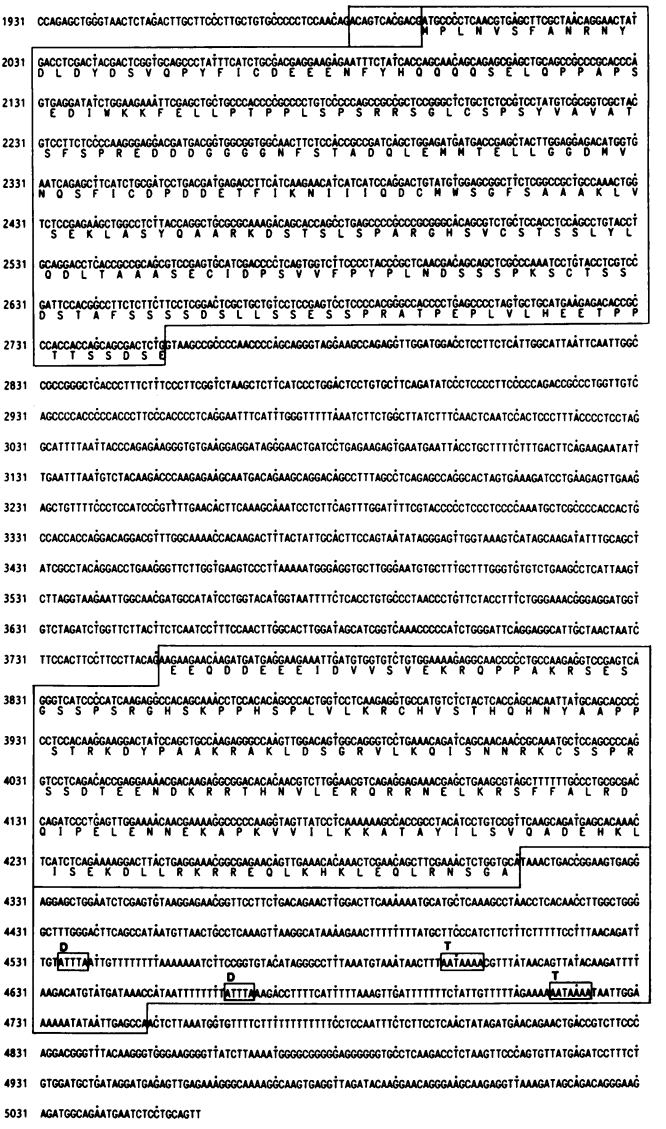


Fig. 2 Nucleotide sequence of rat c-myc and adjacent regions. Nucleotides are numbered from 1 at the cap site of the second promoter. Exons are enclosed by large boxes. Several sequence motifs are boxed with notations as follows. NF, sequence resembling nuclear factor I binding site; GC, putative SP1 binding site; P, TATA box; CA, (CA)n repeat; D, ATTTA sequence which is part of the putative mRNA destabilizing sequence; T, poly A addition signal. The fourth putative SP1 binding site deviates from the canonical sequence.

at -870 to -857 (the nucleotide adjacent to the 5'-side of the second transcription initiation site is designated -1) of human c-myc, both of which have been suggested as the binding sites of nuclear factor 1 (37). The exactly matched sequence does not occur in the 2 kb upstream region of the rat sequence; instead, a related sequence, GTGGAAGGTCCGAA, was found at -944 to -931. This, as well as the human sequence, bears some resemblance to the enhancer core sequences of viral and cellular genes.

Several putative Sp1 (38) binding sites, GGGCGG or its complement, were found in proximity to the rat c-myc promoter, and most of these were also found in their corresponding positions in the human gene.

No apparent CCAAT box was found within the region 0.5 kb upstream of the gene. Three TATA boxes, two at the beginning of exon 1 and one in intron 1, were found in the rat c-myc. These three promoters are also seen in the human and mouse genes. In human and mouse c-myc, the first two promoters are known to function in various tissues and cells, the third usually being cryptic and activated when the upstream promoters are removed by translocation in some B-cell lymphomas (19, 20). We found the second promoter to be predominant in rats, as shown in the following section.

Dual termination (poly A addition) signals, each associated with TG motifs, are located at the end of exon 3 (39). Human and mouse c-myc also have dual terminators at these sites (18, 20), suggesting this to have some, as yet unknown, significance. So far, the second terminator is the only one functional in the various rat tissues and cells examined (data shown below). It is interesting to note that each AATAAAA sequence was preceded by ATTTA and T-clusters approximately 60 bases upstream. Recently, a similar sequence was shown to be responsible for the instability of granulocyte-macrophage colony stimulating factor (GM-CSF) mRNA (48).

Another conspicuous feature of rat c-myc is the presence of $(CA)_{26}$ in intron 1. A similar repeat, $(CA)_{20}$, was found at the corresponding position in the mouse (20), but not the

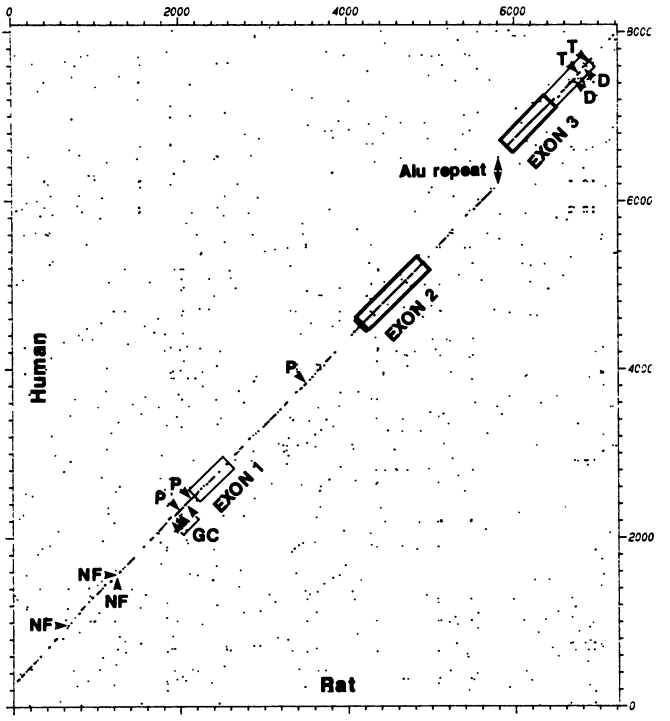


Fig. 3 Dot matrix comparison of rat and human c-myc sequences. Each dot represents 11 matches out of 13 consecutive nucleotides. Coding and non-coding exons are indicated with bold and light boxes respectively. Arrows perpendicular to the homology line indicate motifs conserved in rat and human sequences at identical positions. Vertical and horizontal arrows indicate motifs found only in rat and only in human sequences, respectively, at the indicated positions. All arrows are marked as described in the legend to Fig. 2.

human, gene. The alternating purine-pyrimidine stretch is known to be capable of forming Z-DNA (41), possibly serving to release tortional stress and facilitate chromatin structure loosening.

Gross conservation

Fig. 3 shows the dot matrix homology analysis of the overall c-myc locus between rat and man. The homologies, determined by maximum matching with a gap penalty of 2 (42), of exons 1,2 and 3 were 70, 88 and 79%, respectively. The coding region alone showed homology of 88%. Interestingly, a striking

```
H          T                    Y          Q
R  MPLNVSFANR NYDLDYDSVQ PYFICDEEEN FYHQQQQSEL QPPAPSEDIW BKFELLPTPP   60
M        N T

H                        - P   L G N      S           V
R  LSPSRRSGLC SPSYVAVATS FSPREDDDGG GGNFSTADQL EMMTELLGGD MVNQSFICDP  120
M                                                    Q

H                                                   G PN
R  DDETFIKNII IQDCMWSGFS AAAKLVSEKL ASYQAARKDS TSLSPARGHS VCSTSSLYLQ  180
M

H    S                          A  Q  S   P          T      QGS
R  DLTAAASECI DPSVVFPYPL NDSSSPKSCT SSDSTAFSSS SDSLLSSESS PRATPEPLVL  240
M                                        P                   S

H                  E                   A G          PSAG
R  HEETPPTTSS DSEEEQDDEE EIDVVSVEKR QPPAKRSESG SSPSRGHSKP PHSPLVLKRC  300
M                  E                   T            F

H                           V    V    R         T           V
R  HVSTHQHNYA APPSTRKDYP AAKRAKLDSG RVLKQISNNR KCSSPRSSDT EENDKRRTHN  360
M

H                                                        E  Q     E
R  VLERQRRNEL KRSFFALRDQ IPELENNEKA PKVVILKKAT AYILSVQADE HKLISEKDLL  420
M                                                        I     T

H                  C
R  RKRREQLKHK LEQLRNSGA  439
M
```

Fig. 4 Amino acid sequence of rat c-myc. Residues different in human (above) and mouse (below) c-myc are also indicated. Underlined regions indicate myc boxes.

conservation outside exons can be seen clearly as a diagonal line extending across the whole region. The overall homologies of the 5'-flank (2.2 kb), intron 1 and intron 2, were 68, 65 and 67%, respectively.

The interruption of the diagonal line by the insertion of an Alu family sequence at the 3'-end of intron 2 should also be noted (16, 18). The lack of such a middle repetitive sequence in the rat sequence implies that this insertion has no essential role in c-myc function.

Coded protein

The putative amino acid sequence of rat c-myc is presented in Fig. 4, with those of mouse and human for comparison. The discrepancies between rat and human c-myc numbered 38 out of 439 while those between rat and mouse c-myc numbered 10 out of 439. The non-conservative changes among sequence from the

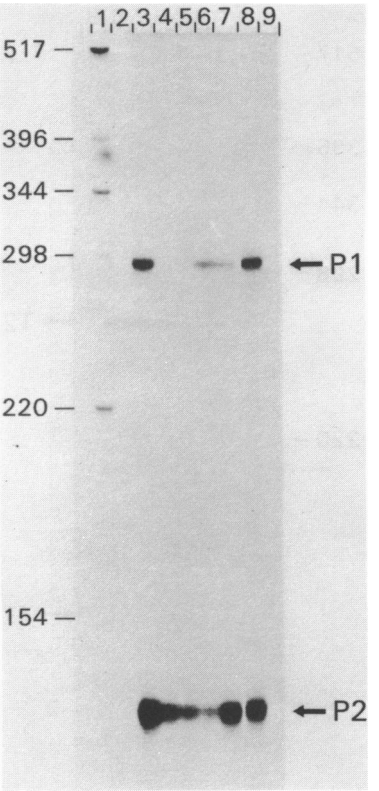


Fig. 5 Nuclease S1 protection assay of transcription initiation sites. A BstNI - BstNI fragment (from -734 to +125, taking the cap site of the mRNA from P2 as +1) was subcloned into M13mp19, and a continuously labeled single stranded probe prepared. The probe was annealed with various RNAs and analyzed. Lane 1, size marker prepared by end-filling pBR322 Hinf I-digest with α-^{32}P-dATP. Lane 2, probe DNA diluted and directly applied to the gel. Lane 3, 2µg liver total RNA of rat receiving cycloheximide 3 hr before killing. Lane 4, 20µg normal liver poly A$^+$ RNA. Lane 5, same as lane 4 but 14µg. Lane 6, 5µg normal spleen poly A$^+$ RNA. Lane 7, 5µg liver poly A$^+$ RNA 2 hrs after partial hepatectomy. Lane 8, 8 µg total RNA of Morris hepatoma 7794 A cultured cells. Lane 9, 20µg Eschericlia coli tRNA.

three species appear to define several variable regions: positions 6 to 33, 78 to 103, 161 to 164, 210 to 232, 272 to 285 and 411 to 417. Two long regions were conserved among the three species: 104 to 160 and 355 to 410, the first including the so called "myc box" sequence (43) which is conserved among

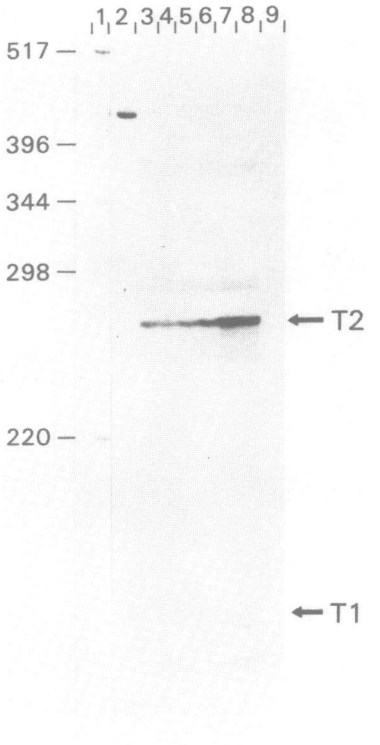


Fig. 6 Nuclease S1 protection assay of poly A addition sites. A HpaI - BstNI fragment (from 280 bp upstream to 70 bp downstream of the second poly A addition site) was subcloned into M13mp19 and a continuously labeled single stranded probe prepared. The probe was annealed with various RNAs and analyzed. The lane samples are the same as those in Fig. 5.

myc gene families. In human c-myc, the corresponding regions were recently shown to be essential for transformation of cells and nuclear localization of the produced protein (44).

Promoter usage

Mammalian c-myc has a conserved dual promoter at the 5'-end of the first exon, and a differential usage of each, depending on tissue, is suggested (45). We examined the sites at which the c-myc promoters, P1 and P2, were used in various normal rat tissues and rat cancer cells using nuclease S1 protection assay (33) (Fig. 5). The results indicated, in most cases, the second promoter, P2, to be used predominantly for

c-myc transcription initiation. In the spleen, however, the c-myc mRNA starting from P1 was also significant. Previous reports have shown, in mouse spleen, P2 to be very much the preferred initiation site (45). There may be some species difference, therefore, in promoter usage.

In the rat liver, the mRNA dominance starting from P2 did not change upon induction of regeneration or after cycloheximide administration. In either case, it is known that there is a remarkable increase of c-myc mRNA (46).

It has been shown previously that inhibition of protein synthesis by cylcloheximide stops c-myc mRNA turnover (12, 13). The fact that the proportion with a P2 preference did not change before and after cycloheximide administration indicates the sequence between P1 and P2 not to contribute to the mRNA stability.

Poly A addition site

The dual poly A addition signals are also conserved among rat, mouse and man. S1 mapping of the 3'-terminus of c-myc mRNA from various sources indicated the downstream signal to be used exclusively in all cases examined (Fig. 6). The upstream signal, however, may be functional under some circumstances, as it has been suggested previously that poly A addition could occur at the upstream signal of human c-myc, when it was expressed in Xenopus oocytes (47).

The exact position of poly A addition in the mRNA seems, however, to be heterogenous, as evidenced by several faint bands observed near the major band in Fig. 6. The relative amount of mRNA having these heterogeneous termini differed neither between normal and cancer cells nor between resting and proliferating cells.

DISCUSSION

We cloned the c-myc of a normal rat and two lines of transplanted hepatomas, one of which contained the gene amplified 10 fold (22). Restriction maps and the sequences determined from the independent sources agreed, suggesting no mutations nor rearrangements to have occurred in the amplified gene and its flanking regions. This is in agreement with

earlier notions of no significant structural changes being found in the amplified c-myc genes of some human malignant cells (48).

Apart from the exon sequences, homologous stretches between rat and human c-myc were found throughout the compared region (Fig. 3). The evolutionary distance between rat and man is estimated to be approximately 100 million years. Considering the changing rate of sequences with no selective pressure to be 0.66% per million years (49), the "meaningless" sequences should diverge to give almost no homology between the two species. The remarkable conservation in the surrounding regions of c-myc, therefore, suggests: (1) there to be many sites, in the region several kb upstream and in the introns of c-myc, where transacting factors bind and contribute to gene regulation, (2) such regions to have some biological functions not directly related to c-myc. In this regard, it is of interest that recent reports have suggested a new transcriptional activity that seems to start from heterogeneous sites in c-myc and extend upstream of the gene (10, 11).

Gazin et al. (50) have suggested that human c-myc exon 1 codes for a protein which reacts with antibodies raised to synthetic peptides having amino acid sequences deduced from regions of the nucleotide sequence of the exon. They assumed a minor mRNA species starting upstream of the two major c-myc promoters, and the initiation codon provided by the upstream sequence. In the corresponding region, we could find no amino acid sequences possibly related to those used in raising the antibodies. Such a protein cannot, therefore, be coded in the rat c-myc region.

It is known that c-myc mRNA turns over rapidly (half life 15 - 30 min) in various cells in vivo and in vitro (12, authors' unpublished data). The short life time is one of the characteristics of some mRNAs involved in cell cycling, e. g. those of c-fos (51, 52), c-myb (53) and some growth factors. These mRNAs have a region in common containing ATTTA and T clusters. Shaw et al. (40) have shown one such regions from GM-CSF mRNA to be responsible for the instability of the mRNA. An RNase, possibly, specifically recognizing this sequence

motif, could be involved in the metabolism of the short lived mRNAs. One candidate for this enzyme is RNase L, which is known to degrade U-rich sequences preferentially (54). Accelerated degradation of c-myc mRNA by interferon (55) also supports the idea of the enzyme's involvement, since interferon causes a rapid synthesis of 2'-5' oligo A, an activator of this RNase (56). We found the non-coding region in exon 3 to be responsible for the rapid turnover of rat c-myc mRNA (manuscript in preparation), and this region to contain two sets of ATTTA and T-clusters. Mouse and human c-myc also have two motifs at identical positions in the non-coding regions of exon 3. In all cases, these motifs were located approximately 60 nucleotides upstream of the poly A addition signals.

Inhibition of protein synthesis by cycloheximide is known to stop c-myc mRNA degradation (12), and result in mRNA accumulation in various types of cell. One interpretation of this effect of cycloheximide is that the inhibitor "freezes" ribosomes on the mRNA, thus sterically protecting the mRNA from RNase attack. The destabilizing sequence, which is likely to be the site recognised by the involved RNase is, however, located downstream of the termination codon and expected to be exposed whether or not the ribosomes were frozen on the upstream sequence of the mRNA. Perhaps the effect of cycloheximide should be otherwise explained, such as the drug inactivating the RNase.

Interestingly, poly A addition at the upstream signal, if such occurs, removes one of the destabilizer sequences from the mRNA. Such shortened mRNA may have a longer life and, therefore, may accumulate even in the absence of transcriptional activation. Although we have not yet found such a shortened mRNA version, the conservation of the poly A addition signal's duality hints at such a regulatory mechanism being required at some stage of the life cycle.

Finally, we are tempted to emphasize that the clones and the rat c-myc sequence data could be useful in studying the possible involvement of this gene in cancer development since the rat is the most widely used aminal in experimental carcinogenesis.

ACKNOWLEDGEMENTS

We thank Dr. J. M. Bishop for supplying pv-myc, Dr. M. Yamamoto for allowing us to use Vax11/750 and Dr. S. Nishimura for suggesting the use of 7-deaza GTP in the sequencing reactions. This work was supported by a Grant-in-Aid from the Ministry of Health and Welfare of Japan for a Comprehensive 10-Year Strategy for Cancer Control and Grants-in-Aid for Cancer Research from the Ministry of Education, Science and Culture of Japan.

REFERENCES

1. Duesberg, P.H., Bister, K. and Vogt. P. K. (1977) Proc. Nat. Acad. Sci. USA 74, 4320-4324.
2. Kan, N. C., Flordellis, C. S., Garon, C. F., Duesberg, P. H. and Papas, T. S. (1983) Proc.. Nat. Acad. Sci. USA 80, 6566-6570.
3. Hayflick, J., Seeburg, P. H., Ohlsson, R., Pfeifer-Ohlsson, S., Watson, D., Papas, T. and Duesberg, P. H. (1985) Proc. Nat. Acad. Sci. USA 82, 2718-2722.
4. Shilo, B.-Z. and Weinberg. R. A. (1981) Proc. Nat. Acad. Sci. USA 78, 6789-6792.
5. Abrams, H. D., Rohrschneider, L. R. and Eisenman, R. N. (1982) Cell 29, 427-439.
6. Watt, R., Shatzman, A. R. and Rosenberg, M. (1985) Mol. Cel. Biol. 5, 448-456.
7. Klein, G. and Klein, E. (1985) Nature 315, 190-195.
8. Bishop, J. M. (1983) Annu. Rev. Biochem. 52, 301-354.
9. Greenberg, M. E. and Ziff. E. (1984) Nature 311, 433-438.
10. Nepveu, A. and Marcu, K. (1986) EMBO J. 5, 2859-2865.
11. Bentley, D. L. and Groudine, M. (1986) Nature 321, 702-706.
12. Blanchard, J.-M., Piechaczyk, M., Dani, C., Chambard, J-C., Franchi, A., Pouyssegur, J. and Jeanteur, P. (1985) Nature 317, 443-445.
13. Linial, M., Gunderson, N. and Groudine, M. (1985) Science 230, 1126-1132.
14. Chung, J., Sinn, E., Reed, R. R. and Leder, P. (1986) Proc. Nat. Acad. Sci. USA 83, 7918-7922.
15. Remmers, E., Yang, J.-Q. and Marcu, K. B. (1986) EMBO J. 5, 899-904.
16. Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G. and Leder, P. (1983) Cell 34, 779-787.
17. Colby, W. W., Chen, E. Y., Smith, D. H. and Levinson, A. D. (1983) Nature 301, 722-725.
18. Gazin, C., Dupont De Dinechin, S., Hampe, A., Masson, J-M., Martin, P., Stehelin, D. and Galibert, F. (1984) EMBO J. 3, 383-387.
19. Stanton, L. W., Watt, R. and Marcu, K. B. (1983) Nature 303, 401-406.
20. Bernard, O., Cory, S., Gerondakis, S., Webb, E. and Adams, J. M. (1983) EMBO J. 2, 2375-2383.

21. Keath, E. J., Kelekar, A. and Cole, M. D. (1984) Cell 37, 521-528.
22. Hayashi, K., Makino, R. and Sugimura, T. (1984) Gann 75, 475-478.
23. Frischauf, A.-M., Lehrach, H., Poustka, A. and Murray, N. (1983) J. Mol. Biol. 170, 827-842.
24. Enquist, L. and Sternberg, N. (1979) Methods in Enzymol. 68, 281-298.
25. Benton, W. D. and Davis, R. W. (1977) Science 196, 180-182.
26. Vennström, B., Moscovici, C., Goodman, H. M. and Bishop, J. M. (1981) J. Virol. 39, 625-631.
27. Deininger, P. L. (1983) Analyt. Biochem. 129, 216-223.
28. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Nat. Acad. Sci. USA 74, 5463-5467.
29. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119.
30. Mizusawa, S., Nishimura, S. and Seela, F. (1986) Nuc. Acids Res. 14, 1319-1324.
31. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nuc. Acids Res. 12, 387-395.
32. Kanehisa, M. (1984) Nuc. Acids Res. 12, 417-428.
33. Berk, A. J. and Sharp. P. (1977) Cell 12, 721-
34. Davis, L. G., Dibner, M. D. and Battey, J. F. in Basic Methods in Molecular Biology, Elsevier N.Y.
35. Pear, W. S., Ingvarsson, S., Steffen, D., Münke, M., Francke, U., Bazin, H., Klein, G. and Sümegi, J. (1986) Proc. Nat. Acad. Sci. USA 83, 7376-7380.
36. Siebenlist, U., Hennighausen, L., Battey, J. and Leder, P. (1984) Cell 37, 381-391.
37. Briggs, M. R., Kadonaga, J. T., Bell, S. P. and Tjian, R. (1986) Science 234, 47-52.
38. Nagata, K., Guggenheimer, R. A. and Hurwitz, J. (1983) Proc. Nat. Acad. Sci. USA 80, 6177-6181.
39. Birnstiel, M. L., Busslinger, M. and Strub, K. (1985) Cell 41, 349-359.
40. Shaw, G. and Kamen, R. (1986) Cell 46, 659-667.
41. Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. and Rich, A. (1979) Nature 282, 680-686.
42. Needleman, S. B. and Wunsch, C. D. (1970) J. Mol. Biol. 48, 443-453.
43. 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.
44. Stone, J., De Lange, T., Ramsay, G., Jabocits, E., Bishop, J. M., Varmus, H. and Lee, W. (1987) Mol. Cel. Biol. 7, 1697-1709.
45. Stewart, T. a., Bellvé, A. R. and Leder, P. (1984) Science 226, 793-710.
46. Makino, R., Hayashi, K. and Sugimura, T. (1984) Nature 310, 697-698.
47. Nishikura, K., Goldflam, S. and Vuocolo, G. A. (1985) Mol. Cel. Biol. 5, 1434-1441.
48. Shibuya, M. and Yamaguchi, S. (1986) in 5th International Workshop on Immune Deficient Animals (Karger), in press.
49. Britten, J. (1986) Science 231, 1393-1398.

50. Gazin, C., Rigolet, M., Briand, J. P., Van Regenmortel, M.H.V. and Galibert, F. (1986) EMBO J. 5, 2241-2250.
51. Kruijer, W., Cooper, J. A., Hunter, T. and Verma, I. M. (1984) Nature 312, 711-716.
52. Meijlink, F., Curran, T., Miller, A. D. and Verma, I. (1985) Proc. Nat. Acad, Sci. USA 82, 4987-4991.
53. Thompson, C. B., Challoner, P. B., Neiman, P. E. and Groudine, M. (1986) Nature 319, 374-380.
54. Floyd-Smith, G., Slattery, E. and Lengyel, P. (1981) Science 212, 1030-1032.
55. Einat, M., Resnitzky, D. and Kimchi, A. (1985) Nature 313, 597-600.
56. Williams, B. R. G., Golgher, R. R., Brown, R. E., Gilbert, C. S. and Kerr, I. M. (1979) Nature 282, 582-586.