Volume 13 Number 7 1985 Nucleic Acids Research

Nucleotide sequence of the 3' exon of the human N-myc gene

Richard W.Michitsch and Peter W.Melera*

Memorial Sloan-Kettering Cancer Center, Graduate Program in Molecular Biology and Virology, 1275 York Avenue, New York, NY 10021, USA

Received 28 December 1984; Revised and Accepted 11 March 1985

ABSTRACT

We have analyzed a 3.8 kb Eco RI fragment of genomic DNA obtained from the amplified N-myc gene of human neuroblastoma cell line BE(2)-C. This fragment contains an exon with an open reading frame encoding approximately 170 amino acids of the carboxy-terminal end of the putative N-myc protein. Comparison of the inferred amino acid sequence of this peptide with that of the 3' domain of the human c-myc protein shows that locally conserved but dispersed regions of homology exist throughout the lengths of these peptides, while hydropathy plots indicate that the physical properties implied by their primary sequences are strikingly similar. Based upon these and other considerations, it is sugggested that the 3' domains of c-myc and N-myc may potentially share related functions.

INTRODUCTION

We have recently reported (1) the isolation of a cDNA clone, pBE(2)-C-59, from a partial cDNA library representing the poly A^+ RNA population of the human neuroblastoma cell line BE(2)-C. This clone was identified via a duplicate screening procedure (2) that made use of genomic DNA probes enriched for amplified sequences we had previously demonstrated to be present in these cells and located within two prominent HSRs (3). pBE(2)-C-59 was shown to hybridize with a 3.8 kb Eco RI genomic DNA fragment found to be amplified in all neuroblastoma cell lines and in one retinoblastoma cell line that displayed either HSRs or DMs. Nonneuroblastoma tumor cell types regardless of HSR/DM content did not amplify this sequence nor did a neuroblastoma cell line without HSRs or DMs. When used in Northern blot analysis, BE(2)-C-59 hybridized with two overproduced polyadenylated RNAs of 3.0 and 1.5 kb in BE(2)-C cells. Indeed, in each of the cell lines tested, the level of expression of these two mRNAs was, to a first approximation, commensurate with the level of amplification of the BE(2)-C-59 gene (1).

That the BE(2)-C-59 gene is most likely the so-called "N-myc gene," originally reported by Schwab et al. (4) to be amplified in human

neuroblastoma cells, is indicated by the following: 1) probes derived from the third exon of the human c-myc gene and hybridized under conditions of reduced stringency with Eco RI digested BE(2)-C DNA share homology with the same amplified 3.8 kb Eco RI fragment that hybridizes with BE(2)-C-59, and 2) probes derived from the second exon of the human c-myc gene hybridize under the same conditions of reduced stringency with an amplified 2.1 kb Eco RI fragment from BE(2)-C DNA that can be shown by partial digestion of genomic DNA to lie adjacent to the 3.8 kb fragment (1). This 2.1 kb fragment has hybridization characteristics similar to those of the amplified 2.0 kb Eco RI fragment reported by Schwab et al. (4) to be present in neuroblastoma cells and to share partial nucleotide sequence homology to the 5' domain of the v-myc gene. The 2.1 kb fragment also hybridizes with the same two overproduced 3.0 and 1.5 kb poly A^+ RNAs as the BE(2)-C-59 containing 3.8 kb fragment. These and other considerations have led us to suggest that BE(2)-C-59 represents the 3' end of an N-myc transcript (1). In order to validate that proposal, we present here the nucleotide sequence of both BE(2)-C-59 and that region of the amplified 3.8 kb Eco RI genomic fragment which encodes it and shares homology with the third exon of the human c-myc gene. Furthermore, we present evidence to suggest that the N-myc gene encodes a protein whose 3' domain may serve a function related to that of the analogous domain of the human c-myc protein.

MATERIALS AND METHODS

DNA and RNA Preparations. The neuroblastoma cell line, BE(2)-C (5), maintained in a 50:50 mixture of MEM and F-12 media, supplemented with 15% FBS, non-essential amino acids, penicillin and streptomycin, was used as the source of both DNA and RNA. High molecular weight DNA and poly$(A)^+$ RNA were prepared from nuclei and polysomal pellets, respectively, as previously described (6, 7).

Restriction Endonuclease Digestion of Genomic DNA. Restriction endonucleases Bgl II, Bst NI, Eco RI, Hinc II, Hinf I, Pst I, Sma I, and Taq I were purchased from Bethesda Research Laboratory and International Biotechnologies, Inc. Digestion of DNA with these enzymes was carried out in buffers and at temperatures suggested by the vendors.

Southern Blot Transfer and Analysis. DNA fragments were size-fractionated by electrophoresis on 0.8% agarose gels and transferred to nitrocellulose membrane filters as described by Southern (8) with modifications (6). Hybridization of the filters with ^{32}P-labeled nick-

translated probes (9) and washing of the filters after hybridization were carried out either according to the procedures of Southern (8), or under conditions of reduced stringency by hybridization in 40% de-ionized formamide, 10x Denhardt's solution, 5xSSC, and 50 µg/ml sheared salmon sperm DNA for 42 hr at 42°C, followed by two washes in 0.4xSSC, 0.2% SDS at 44°C for 45 min, and a third in the same buffer at 50°C (1).

Northern Blot Transfer and Analysis. BE(2)-C poly(A)$^+$ RNA was size-fractionated on 1.2% formamide/formaldehyde agarose gels, transferred to nitrocellulose membrane filters, hybridized and washed according to Thomas (10) as modified by Lewis et al. (6).

Cloning and Sequencing of DNA. Eco RI-digested genomic DNA from BE(2)-C cells was electrophoresed through 1% low melting agarose and the 3.8 kb region of the gel cut out and eluted (11). The resulting DNA fragments were cloned by direct ligation into the Eco RI site of pBR322, and the recombinants identified by screening (12) with pBE(2)-C-59 as probe (1). Subcloning of the 3.8 kb Eco RI fragment into pUC 8 (13) was carried out as described previously (14), while cloning of subfragments into bacteriophages M13mp18 and M13mp19 for dideoxy sequencing purposes (15) followed the procedures of Messing and Vieira (16). The chemical degradation method of Maxam and Gilbert (17) was used to determine the sequence of the cDNA clone pBE(2)-C-59.

Computer Analysis. The computer programs used for the analysis of the DNA sequence data are included in the IBI/Pustell Sequence Analysis System available from International Biotechnologies, Inc., P.O. Box 1565, New Haven, CT 06506, and included, among others, a protein-coding-region-locator program, homology matrix programs, and a peptide sequence hydropathy program.

RESULTS AND DISCUSSION

Restriction endonuclease mapping and subcloning of the genomic DNA surrounding the BE(2)-C-59 sequence. Figure 1A shows a partial restriction endonuclease map of the amplified BE(2)-C-59-containing 3.8 kb Eco RI fragment. After digestion with Bgl II, it was demonstrated by Southern procedures that homologies to both BE(2)-C-59 and to the 3' exon of the human c-myc gene resided within the 2.0 kb Bgl II-Eco RI fragment. As expected, homology with the 3' exon of c-myc was detected only when filter hybridization and washing were carried out using the reduced stringency protocol outlined in MATERIALS AND METHODS. Northern blot analysis (data not shown) of BE(2)-C poly(A)$^+$ RNA using each of the three fragments generated from a Bgl II digest of the 3.8 kb fragment indicated that only one, the 2.0 kb fragment,

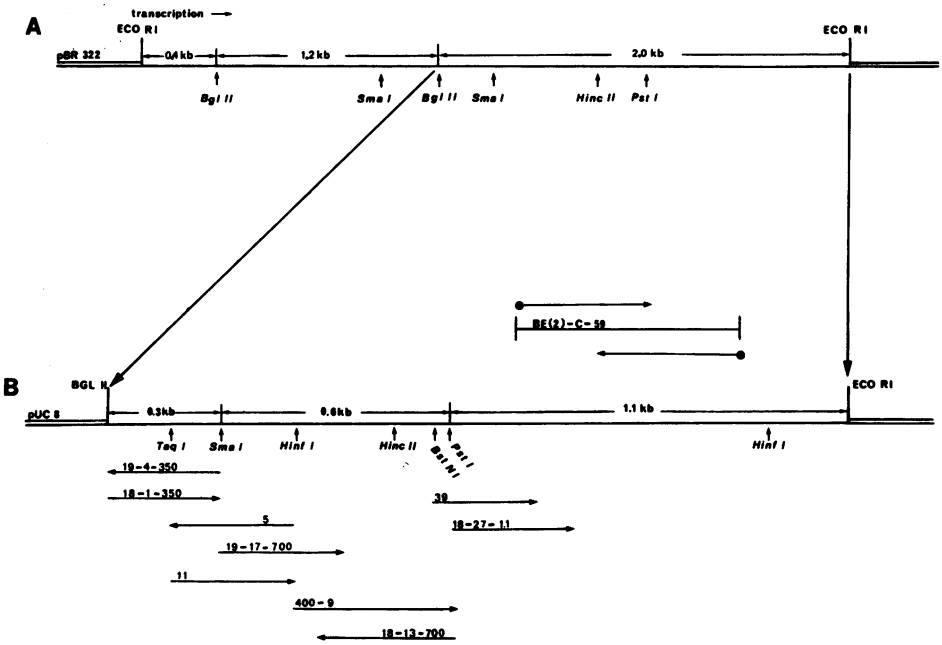

Fig. 1 Restriction maps and sequencing strategy for the analysis of the amplified 3.8 kb Eco RI genomic DNA fragment and the cDNA clone BE(2)-C-59.
Panel A. Partial restriction map of the cloned, amplified 3.8 kb Eco RI fragment from neuroblastoma cell line BE(2)-C, identified by its homology with the N-myc-specific cDNA clone pBE(2)-C-59 (1).
Panel B. Sequencing strategy applied to the 2.0 kb Bgl II-Eco RI fragment.
The numbers indicate the clone identification, and the arrows the direction in which the dideoxy sequencing was carried out. BE(2)-C-59 was sequenced by the chemical degradation method. The black dots indicate radiolabeled ends and the arrows the direction and extent of the sequence determined from those ends.

hybridized to the 3.0 kb and 1.5 kb RNAs overproduced by BE(2)-C cells and partially encoded by the amplified 3.8 kb Eco RI fragment (1). Hence the 0.4 and 1.2 kb Bgl II fragments were considered to be either intronic or flanking DNA. DNA sequence information presented in this report indicates that these fragments are intronic and that the direction of transcription is as indicated by the arrow. The 2.0 kb Bgl II-Eco RI fragment was subcloned into pUC 8 (Fig. 1B), and further Southern analysis indicated that the BE(2)-C-59 homology lay within the 1.1 kb Pst I-Eco RI fragment, whereas the partial c-myc homology was located within the 0.6 kb Sma I-Pst I fragment. Subcloning

of the 2.0 kb Bgl II-Eco RI fragment into bacteriophages M13mp18 and M13mp19 resulted in the clones represented by the horizontal arrows.

DNA sequence determinations and analysis. The nucleotide sequence of a major portion of the 2.0 kb Bgl II-Eco RI fragment shown in Figure 1B is presented in Figure 2. While we have not yet completed sequencing the region of the 2.0 kb Bgl II-Eco RI fragment containing the entire sequence of BE(2)-C-59, the identification of 181 bases of overlap between these two sequences clearly establishes this genomic fragment as that which encodes the BE(2)-C-59 transcript. Inspection of the BE(2)-C-59 sequence further shows that it terminates with 24 adenosine residues preceded 12 bases in the 5'-ward direction by the consensus polyadenylation signal sequence AATAAA. Since BE(2)-C-59 was initially cloned via reverse transcription of oligo(dT) primed poly A^+ RNA, it is reasonable to conclude that its poly(A) tail and polyadenylation signal sequence characterize it as the 3' end of an mRNA molecule. Hence the 2.0 kb Bgl II-Eco RI fragment contains a portion of the 3' end of the N-myc gene. Establishing the location of the exon encoding BE(2)-C-59 in genomic DNA by nucleotide sequence analysis allows us to establish the coding strand of the 2.0 kb fragment as being the complement of the sequence shown in Figure 2 and orients the direction of transcription within the 2.0 kb fragment; i.e., transcription proceeds from the Bgl II site toward the Eco RI site as indicated in Figure 1.

In order to determine whether BE(2)-C-59 or sequence 5'-ward of it contained an open reading frame, we subjected the entire sequence shown in Figure 2 to computer analysis using a program designed to locate the most probable amino acid coding region within an unknown DNA sequence. This program makes use of a codon bias table, in this case constructed from 23 randomly chosen human genes to establish the frequency with which the human genome uses a given codon to specify an amino acid. By applying this measure of codon bias to an unknown sequence, the program can indicate which frame (of the six possible since both strands are considered) is the most likely, versus simply open, reading frame. Additionally, all termination sequence locations are indicated. Although several open reading frames of varying length and bias were found on both strands of the sequence in Figure 2, the longest open reading frame and the one with the greatest bias toward that established by the human table we constructed, was found on the strand that we had established as the coding strand (see above), and began in the region of nucleotide 225, ending with the terminator TAG at position 744. Two other terminators in frame with the first were found at positions 789 and 801.

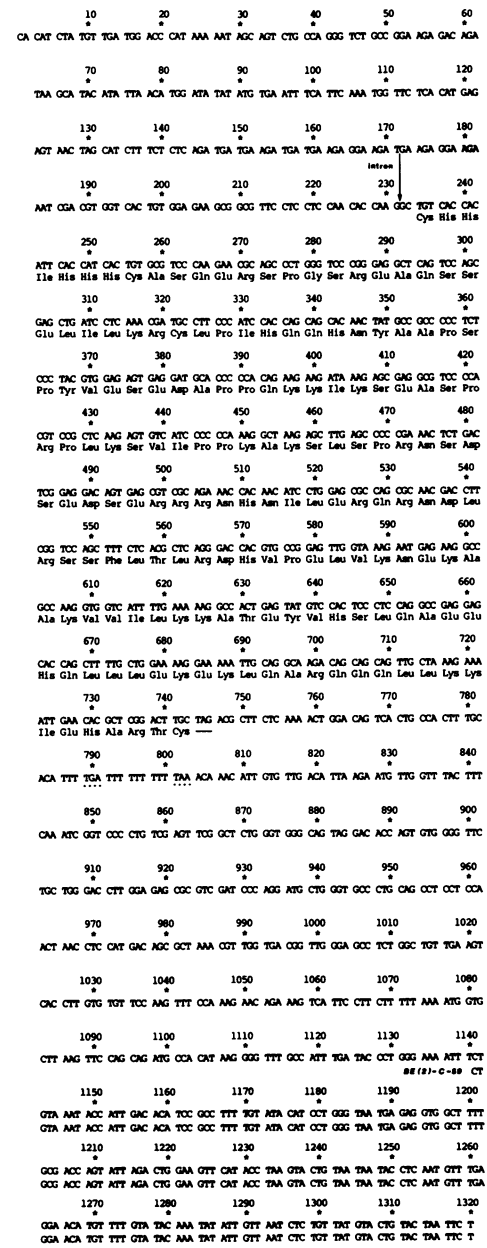

Fig. 2 Nucleotide sequence of the 3' exon of the human N-myc gene. The sequence begins approximately 10 bases 3'-ward of the Bgl II site shown in Figure 1B and proceeds 1320 nucleotides toward the Eco RI site. Overlap with BE(2)-C-59 begins at position 1141. The consensus polyadenylation signal sequence AATAAA is underlined and is located 12 bases upstream from the 24-residue poly(A) tail. The vertical arrow at position 231-232 indicates the position of the arbitrarily positioned intron/exon junction.

According to the computer analysis, the beginning of this reading frame, i.e., the intron/exon junction, is located somewhere in the vicinity of nucleotide 225. However, in the absence of S_1 nuclease data or a cDNA sequence bridging this junction we cannot, even by applying the consensus sequences compiled for such junctions (18, 19), accurately assign its position. Therefore, for the purpose of this report, we have done the following: visual inspection of the sequence shows that the 5' boundary of the reading frame cannot lie 5'-ward of the triplet at position 174 since the adjacent triplet in that direction is the terminator TGA. Hence, the reading frame and, therefore, the exon begins 3'-ward of the triplet at position 174. Based upon consensus (18,19), there are at least two possible intron/exon locations 3'-ward of that position and they are at positions 231-232 and 338-339. Since the protein-coding-region-locator program indicated that position 338-339 is within a region where the bias toward human coding sequence is high and also indicates that there is no bias toward human coding sequence 5'-ward of position 225, we have arbitrarily placed the position of the junction at position 231-232.

Comparison of the human c-myc and N-myc 3' exons. Since we had shown (Fig. 1) that the region of the 2.0 kb Bgl III-Eco RI fragment containing the 3' exon of N-myc also contained partial nucleotide sequence homology with the 3' exon of the human c-myc gene, we analyzed the extent of the homology between these two exons. The actual sequences compared by computer included bases 1-954 of the N-myc sequence (Fig. 2) and the entire protein encoding region of the human c-myc gene (20). The results are presented in Figure 3 and show only those portions of the respective sequences within which all of the significant homology was found. When aligned according to the highest percentage of match over the greatest distance, i.e., nucleotides 576 through 631 of N-myc, where the homology with the c-myc sequence, i.e., nucleotides 3534 through 3589, exceeds 80%, it is clear that the protein encoding regions of these two exons are in frame with one another, and that frame is the one shown in Figure 2 to be the most probable N-myc reading frame.

When the inferred amino acid sequence of the peptide encoded by the 3' exon of N-myc is compared to the amino acid sequence of its c-myc counterpart, a region of considerable homology can be identified between the amino acids encoded by nucleotides 430 and 663 of N-myc and 3489-3620 of c-myc (Fig. 3). The overall amino acid sequence homology in this region is 68% with the most conservation occurring between the amino acids encoded by nucleotides 591-632 of N-myc and 3549-3590 of c-myc where the homology is 93%. Hence, these two peptides have maintained a substantial degree of amino acid sequence homology

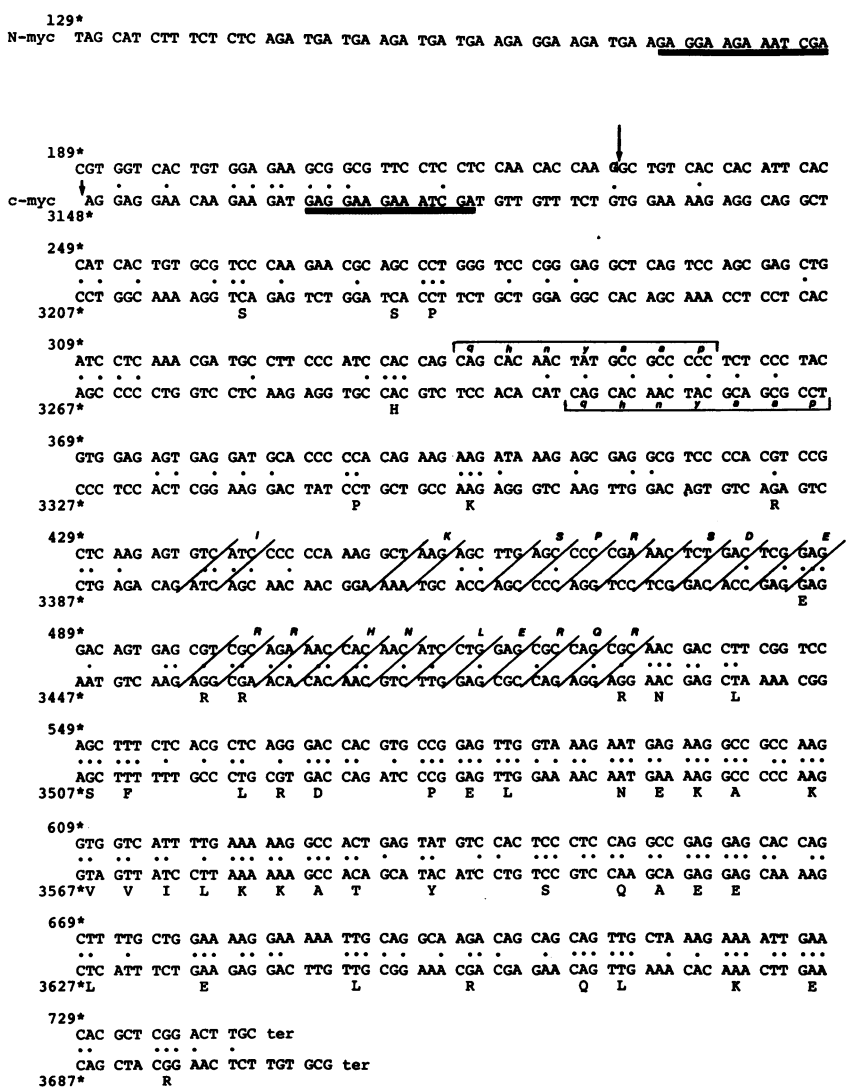

Fig. 3 Nucleotide sequence comparison of the 3' exons of the human N-myc and c-myc genes. The nucleotide sequences of both exons are displayed beginning with nucleotide No. 129 of N-myc and nucleotide No. 3143 of c-myc, which is the first nucleotide 3'-ward of the c-myc 3' exon splice site (20) and ending in both cases at the respective termination codons. The vertical arrow at position 3147-3148 of c-myc indicates the known position of its 3' intron/exon splice junction (20) while the vertical arrow at position 231-232 of the N-myc sequence is shown for reference purposes only. The alignment of the exons has been dictated by their nucleotide sequence homology, and each homologous nucleotide is indicated by a black dot. No gaps or other alterations have been introduced to optimize the alignment. The single letter amino acid code

has been used and where upper case letters are shown immediately underneath a line of c-myc sequence, the indication is that homologous amino acids exist at that point between N-myc and c-myc. Two other regions of nucleotide and amino acid sequence homology are indicated by diagonal lines and horizontal brackets. The diagonal lines indicate homologies that are one codon out of phase and the brackets a region of homology three codons out of phase. In both cases the actual reading frame is that established by the major nucleotide sequence homology. Upper case letters above the N-myc sequence and between the diagonals indicate amino acid sequence homologies between the two sequences, as do the lower case letters within the brackets.

through a region located within approximately 30 to 60 amino acids of their respective carboxy termini. However, upon further inspection we have been able to identify three other regions of homology within these peptides. Two of these occur at both the nucleotide and amino acid sequence level and, while existing within the same reading frame as the major homology discussed above, are out of phase with it. The first begins at nucleotide 441 in N-myc and continues through nucleotide 533. Its homologous c-myc sequence begins at nucleotide 3396, one triplet out of phase in the 5'-ward direction with its homologous triplet in N-myc, and extends through base 3488 (see Fig. 3). The most significant amino acid sequence homology in this region, however, begins at N-myc nucleotide 465 and contains two areas of c-myc homology, one (465-480) where 6 of 8 amino acids (75%) are homologous to c-myc and another (501-533) where 9 of 11 (82%) are homologous. While the overall homology throughout this region is 65%, these local regions of high homology have clearly been conserved within these two peptides and maintained at equal distances from their respective carboxy-termini. The second of these regions begins at N-myc nucleotide 339 and ends at nucleotide 359. This sequence of 21 nucleotides encodes 7 amino acids that share 100% amino acid sequence homology (81% nucleotide sequence homology) with a sequence in c-myc beginning at nucleotide 3306 and ending at 3326. Although three triplets out of phase in the 3'-ward direction with N-myc, this region has also been highly conserved between these two peptides and, like the other regions of homology shared between them, has been maintained equidistant from the carboxy terminus. The third region of homology is 14 nucleotides in length and is found between nucleotides 175 and 188 of N-myc, 57 nucleotides 5'-ward of the site we have arbitrarily shown as the N-myc splice site at position 231-232. The 100% homologous c-myc sequence is located between nucleotides 3165 and 3178 and lies 17 nucleotides 3'-ward of the 3' exon splice site (20), within the protein encoding region of the third exon. Since we have not precisely located the 3' intron/exon N-myc splice junction we cannot unequivocally state

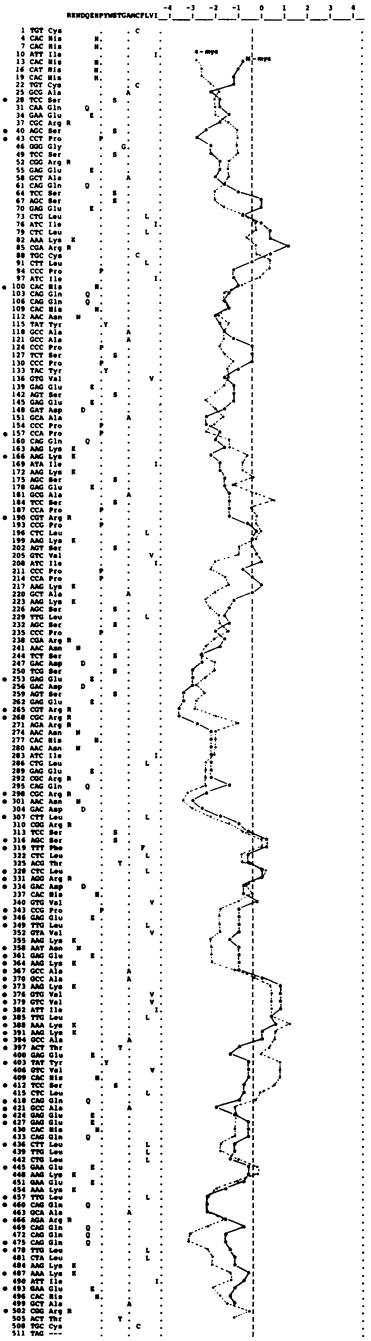

Fig. 4 Comparative hydropathy plots of the peptides encoded by the 3' exons of N-myc and c-myc. The figure shows the actual plot for N-myc (solid line), over which the plot of c-myc (dotted line) has been superimposed. The nucleotide sequence alignment shown in Figure 3 was used and no attempt was made to optimize the homology. The nucleotide sequence shown on the left margin is that of N-myc. Nucleotide No. 1 corresponds to nucleotide position 234 in Figure 2. The black dots along the left margin indicate the homologous amino acids shared between N-myc and c-myc as indicated by the nucleotide sequence alignment shown in Figure 3. The brackets indicate those regions where the out-of-phase homologies exist. See Figure 3 for details. The plot itself (23) displays a relative measure of hydrophobicity over a 9-amino acid window, with the degree of hydrophobicity increasing to the right. The amino acids that comprise the sequence are also shown and plotted according to their assigned hydropathic values (23), the most hydrophobic being isoleucine and the least, arginine. The brackets indicate those regions where the amino acid homology between N-myc and c-myc, although out of phase, has been maintained.

that the 14 bp sequence resides within the N-myc intron. However, we believe that this is likely to be the case since 1) the reading frame of the 14 bp sequence in N-myc differs from that in c-myc and therefore would not result in the maintenance of any amino acid sequence homology between the two peptides and 2) the 14 bp sequence lies very close (2 bases 3'-ward) to the feasible limit, based upon reading frame, of the the 3' N-myc exon and a considerable distance 5'-ward from any of the most likely intron/exon junctions or from that region where the codon usage begins to approach that of the bias established by our computer program. It is unusual to find exon sequence in one gene that apparently has been highly conserved as intron sequence in another. Since the 14 bp sequences lie in close proximity to splice junctions, and assuming that c-myc and N-myc evolved from the same ancestoral gene, it may be that both 14 bp sequences were once exonic. Some genetic event, i.e., insertion, inversion, etc., could then have occurred to place the 14 bp sequence in N-myc within an intron, while at the same time, modifying the domain encoded by the 3' N-myc exon. In any case, the 14 bp sequence, in addition to its amino acid encoding role in c-myc, may serve some basic, as yet unknown, function in both of these genes; hence its retention in N-myc.

Comparison of the peptides encoded by the 3' exons of c-myc and N-myc. It is the protein encoded by the N-myc gene that is of ultimate interest and since the c-myc protein is known to be located in the nucleus (21, 22) and its 3' domain, implicated in DNA binding (21), we sought to compare the putative 3' exon encoded peptide of N-myc with its c-myc counterpart. We had already noted (see above) that local regions of high amino acid sequence homology were dispersed at fixed distances from the carboxy termini of both peptides. In order to estimate what effect these local regions of homology would have on the overall physical nature of these peptides, we constructed hydropathy plots (23) of each and then compared them. The results are shown in Figure 4. The regions of amino acid sequence homology are indicated by the black dots and brackets along the left-hand margin of the Figure and are easily distinguished in the plots as those areas where the hydropathy profiles are most similar. It should be emphasized that we have superimposed the two plots according to the major nucleotide sequence homology shown in Figure 3. No gaps or alterations have been introduced into either sequence. Although this does not appreciably affect the profiles in the region of nucleotides 232-298 where the homology is one codon out of phase (positions 465-533 and 3420-3488 of the N-myc and c-myc sequences, respectively, Fig. 3), it does obscure the fact that the region in the vicinity of nucleotides 106-124, where the homology is three

codons out of phase (positions 339-359 and 3306-3326 of the sequences in Fig. 3), is also very similar. Hence the individual stretches of amino acid sequence homology, as expected, lend similar hydropathic character to their respective regions. Perhaps of greater significance, however, is the striking similarity in spatial organization displayed by the two peptides even outside those areas of amino acid sequence homology. The distribution of hydrophobic and hydrophilic pockets along the length of the peptides is remarkably similar, although local heterogeneity is obviously present. For example, although clearly variant in the area between nucleotides 160 and 262, the general trend through this region in both peptides reflects a gradual increase then a decrease in hydrophobicity. Hence, even though dissimilarity does exist between these peptides, their regions of highly conserved amino acid sequence homology, size, hydropathic characteristics, and overall spatial organization, clearly indicate their relatedness.

C-myc protein has been shown to reside in the mammalian cell nucleus (21, 22), be involved with establishment/immortalization functions in cellular transformation (24), to be associated with the nuclear matrix (25) and to display a DNA binding activity (21, 25). Interestingly, the first three of these four properties are shared with the E1a protein of adenovirus (26, 27), while the third is apparently not. When we compared the hydropathic profile of the 3' domain of E1a protein (28) with those of c-myc and N-myc, we found that while probably all related (29), the two myc peptides are clearly more similar to one another than either is to E1a (data not shown). Moreover, the amino acid sequence homologies shown to exist between the 3' domains of E1a and c-myc (29) are almost entirely different from, and are much less extensive than, those that exist between the 3' domains of c-myc and N-myc (Fig. 3). Hence, while N-myc protein may also be shown to be a nuclear protein like E1a and c-myc, the functional role of its 3' domain may prove to more closely resemble that of c-myc.

While the data presented here clearly establish that transcripts of the human N-myc gene contain substantial amounts of c-myc-related information in their 3' exons, and therefore indicate that these two cellular genes are related, we may only speculate as to the functional role of N-myc. Since the amplification and overexpression of this gene and possibly its activation as well have been shown to be associated solely with tumors derived from cells of the neural crest (1, 3, 4, 30-33), the function of N-myc may be restricted to specific cell lineages, and in that sense differs from that predicted for c-myc, whose amplification and overexpression have been demonstrated to occur in

a variety of different tumor types (34-37) and whose expression is associated with cellular proliferation (37). On the other hand, in a manner similar to c-myc (24), the role of N-myc in tumorigenesis probably does not involve transformation per se since its amplification-mediated overexpression occurs predominantly in advanced stages of neuroblastoma (38), although a more fundamental role for the N-myc gene during the formation of retinoblastomas, has been suggested (33). Clarification of N-myc function and the possibility that it may play a complementary or supplementary role to that of c-myc in the growth regulation of highly specialized cells, such as those from the neural crest, or that it may be specifically involved with differentiative functions in these cells (39), remains to be determined.

ACKNOWLEDGEMENTS

We thank Joseph P. Davide for the sequence of pBE(2)-C-59. This work was supported in part by grants from the National Cancer Institute to the Sloan-Kettering Institute for Cancer Research and from the Special Projects Committee of the Society of the Memorial Sloan-Kettering Cancer Center, the Milton Schamuck Foundation, and the Jeanette Warren Davison Fund for Cancer Research to P.W.M.

*To whom correspondence should be sent

REFERENCES

1. Michitsch,R.W., Montgomery,K.T. and Melera,P.W. (1984) Mol. Cell. Biol. 4, 2370-2380.
2. Brison,O., Ardeshir,F. and Stark,G.W. (1982) Mol. Cell. Biol. 2, 578-587.
3. Montgomery,K.R., Biedler,J.L., Spengler,B.A. and Melera,P.W. (1983) Proc. Natl. Acad. Sci. USA 80, 5724-5728.
4. Schwab,M., Alitalo,K., Hempnauer,K.H., Varmus,H.E., Bishop,J.M., Gilbert,F., Brodeur,G., Goldstein,M. and Trent,J. (1983) Nature 305, 245-248.
5. Biedler,J.L., Meyers,M.B. and Spengler,B.A. (1983) Adv. Cell. Neurobiol. 4, 267-307.
6. Lewis,S.A., Davide,J.P. and Melera,P.W. (1982) Proc. Natl. Acad. Sci. USA 79, 6961-6965.
7. Melera,P.W., Hession,C.A., Davide,J.P., Scotto,K.W., Biedler,J.L., Meyers,M.B. and Shanske,J. (1982) J. Biol. Chem. 257, 12939-12949.
8. Southern,E.M. (1975) J. Mol. Biol. 98, 503-517.
9. Rigby,P.W., Dieckman,M., Rhodes,C. and Berg,P. (1977) J. Mol. Biol. 113, 237-251.
10. Thomas,P.S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5205.
11. Parker,R.C. and Seed,B. (1980) Methods Enzymol. 65, 358-361.
12. Grunstein,M. and Hogness,D. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965.
13. Vieira,J. and Messing,J. (1982) Gene 19, 259-268.
14. Melera,P.W., Davide,J.P., Hession,C.A. and Scotto,K.W. (1984) Mol. Cell.

Biol. 4, 38-48.
15. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
16. Messing,J. and Vieira,J. (1982) Gene 19, 269-276.
17. Maxam,A.M. and Gilbert,W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
18. Breathnach,R. and Chambon,P. (1981) Annu. Rev. Biochem. 50, 349-383.
19. Mount,S.M. (1982) Nucleic Acids Res. 10, 459-472.
20. Colby,W.W., Smith,D.H. and Levinsen,A.D. (1983) Nature 301, 722-725.
21. Persson,H. and Leder,P. (1984) Science 225, 718-721.
22. Hann,T. and Eisenmann,R. (1984) Mol. Cell. Biol. 4, 2486-2497.
23. Kyte,J. and Doolittle,R.F. (1982) J. Mol. Biol. 157, 105-132.
24. Land,H., Parada,L.F. and Weinberg,R.A. (1983) Nature 304, 596-602.
25. Eisenman,R.N., Tachibana,C.Y., Abrams, H.D. and Hann,S.R. (1985) Mol. Cell. Biol. 5, 114-126.
26. Feldman,L.T. and Nevins,J.R. (1983) Mol. Cell. Biol. 3, 829-838.
27. Flint,S.J. (1980) in DNA Tumor Viruses, Tooze,J. Ed., pp. 547-576, Cold Spring Harbor Laboratory, New York.
28. Perricaudet,M., leMoullec,J-M. and Tiollais,P. (1980) Nature 288, 174-176.
29. Ralston,R. and Bishop,J.M. (1983) Nature 306, 803-806.
30. Kohl,N.E., Gee,C.E. and Alt,F.W. (1984) Science 226, 1335-1337.
31. Kohl,N.E., Kanda,N., Schreck,R.R., Bruns,G., Latt,S.A., Gilbert, F. and Alt,F.W. (1983) Cell 35, 359-367.
32. Schwab,M., Ellison,J., Busch,M., Rosonar,W., Varmus,H.E. and Bishop,J.M. (1984) Proc. Natl. Acad. Sci. USA 81, 4940-4944.
33. Lee,W-H., Morphree,A.L. and Benedict,W.F. (1984) Nature 309, 458-460.
34. Collins,S. and Groudine,M. (1982) Nature 298, 679-681.
35. Alitalo,K., Schwab,M., Lim,C.C., Varmus,H.E. and Bishop,J.M. (1983) Proc. Natl. Acad. Sci. USA 80, 1707-1711.
36. Little,C.D., Nau,M.M., Larney,D.H., Gazdar,A.F. and Minna,J.D. (1983) Nature 306, 194-196.
37. Persson,H. and Leder,P. (1984) Science 225, 687-693.
38. Brodeur,G.M., Seeger,R.C., Schwab,M., Varmus,H.E. and Bishop,J.M. (1984) Science 224, 1121-1124.
39. Thiele,C-J., Reynolds,C.P. and Isreal,M.A. (1985) Nature 313, 404-406.