Isolation and Characterization of Complementary DNA for N-cym, a Gene Encoded by the DNA Strand Opposite to N-myc

Barbara C. Armstrong and Geoffrey W. Krystal

Department of Medicine, Division of Hematology/Oncology, and Department of Microbiology/Immunology, Medical College of Virginia, McGuire Veterans Administration Medical Center, Richmond, Virginia 23249

Abstract
The N-myc oncogene has been implicated in the pathogenesis of a number of human tumors, including childhood neuroblastoma and adult small cell lung cancer. We have isolated and characterized complementary DNA clones derived from a transcription unit, N-cym, located on the opposite DNA strand to N-myc, with extensive overlap existing between the 5' ends of the two transcription units. The N-cym gene, which can encode a 109-amino acid protein, is expressed during fetal development, as well as in tumor cell lines containing amplified N-myc loci, where it is expressed at very high levels. Although other examples of overlapping, opposite-strand eukaryotic genes exist, N-myc and N-cym are unique in that they appear to be coregulated in tumor cell lines under basal growth conditions and in response to the differentiating agent retinoic acid. This coregulation suggests that their protein products may be functionally interrelated during normal development and oncogenesis.

Introduction
DNA segments encompassing eukaryotic genes have traditionally been viewed as possessing a single transcription unit with the ability to encode a single peptide or family of peptides, where diversity can be generated by alternative RNA processing or alternative translational reading frames. Recently, however, there has been an accumulation of instances in which the DNA strand opposite or antisense to a characterized gene has been demonstrated to give rise to stable RNA species, complementary to regions of the primary transcript or processed RNA of the characterized gene. The region of overlap between the opposite-strand transcription units has been shown to be both at the 5' (1-4) and 3' (5-11) ends of the transcription units. Alternatively, entire antisense transcription units have been localized to introns of expressed genes (12-15). In general, it appears that transcription from complementary strands is regulated independently, with the exception of the dhir locus (1-3), where only a minority of divergent transcripts actually overlap.

Nuclear run-on transcriptional analysis of the myc gene family revealed that there was nearly as much antisense as sense transcription across sequences encompassing exon 1 of the N-myc gene (16). We subsequently demonstrated that the antisense primary transcription resulted in both stable polyadenylated and nonpolyadenylated RNAs (17). The latter class of unprocessed RNAs was demonstrated to form in vivo RNA duplexes with a subpopulation of N-myc sense transcripts. The physiological role for these RNA duplexes remains to be proved.

RNA protection experiments using RNA fractionated by oligo(dT)-cellulose chromatography suggested that the polyadenylated antisense RNA was processed and did not form a duplex with sense transcripts. Northern blots of polyadenylated RNA probed with an antisense-specific single-stranded probe derived from N-myc intron 1 detected a predominant 0.9-kb and a lesser 1.8-kb species.

The N-myc gene was originally described as an amplified gene common to many human childhood neuroblastomas and homologous to the v-myc oncogene (18). When cotransfected with an activated ras oncogene, it can transform rat embryo fibroblasts in a standard assay for transformation potential (19, 20). The precise function of the N-myc protein is not known, but it is thought to be a transcription factor. It is a nuclear protein that possesses regions homologous to functional domains of known transcriptional regulatory proteins, and it forms a complex with the max protein, which also forms a heteromeric complex with c-myc, which binds DNA in a sequence-specific fashion (21). Amplification of the N-myc gene correlates well with both neuroblastoma tumor stage and patient prognosis (22, 23). In addition to neuroblastomas, the gene is commonly amplified and expressed in adult SCLC tumors and tumor-derived cell lines (24, 25). The amplification unit of the N-myc locus in neuroblastomas is at least 140 kb in length (26), leaving open the possibility that other expressed genes surrounding the 7-kb N-myc locus may influence the transformed phenotype of tumors displaying N-myc amplification. We therefore undertook the cDNA cloning of the polyadenylated antisense transcripts not only to characterize the relatively unique genomic organization of the two transcription units, but also to eventually determine the relationship between expression of the antisense gene, N-myc expression, and the transformed phenotype.

Received 1/9/92.
1 This work was supported by a Merit Review Award from the Department of Veterans Affairs.
2 To whom requests for reprints should be addressed, at McGuire VA Medical Center, Room 5A-128, 1201 Broad Rock Road, Richmond, VA 23249.

3 The abbreviations used are: kb, kilobase(s); SCLC, small cell lung cancer; cDNA, complementary DNA; bp, base pair(s); PCR, polymerase chain reaction; HLH, helix-loop-helix; RACE, rapid amplification of cDNA ends.
Results
Cloning of N-cym cDNA. Our initial efforts focused on designing a probe that would be relatively specific for cDNA clones derived from antisense mRNA. Mapping of polyadenylated antisense transcripts by RNase protection (data not shown) predicted that a BgII site located at the 3' end of exon 1B of N-myc (Fig. 1A) was in the midst of an antisense exon which extended approximately 50 bases upstream of the BgII site and 75 bases downstream of it (referring to the N-myc orientation). We therefore used a 374-bp genomic Smal fragment encompassing this sequence (Fig. 1A) to probe a Dp-primed λgt10 cDNA library constructed from the Kelly neuroblastoma cell line (27). Approximately 1 × 10^6 plaques were screened. Secondary plaques were discarded as sense clones if they hybridized to an N-myc exon 2-specific probe (which fails to detect antisense transcripts by RNase protection). Three independent partial cDNA clones were plaque purified through quaternary platings, and the inserts were amplified by PCR and subcloned into pGEM 3; two of the clones were sequenced completely in both orientations by the dideoxynucleotide technique, and the third was determined to be very similar to the other two by restriction enzyme analysis. The sequence generated was compared to previously published genomic and cDNA sequences derived from the N-myc locus (27–30). The 5' end of all of the clones was what appeared to be an internal EcoRI site (because it was present in the genomic sequence), located in mid-intron 1 of the N-myc locus. The 3' end of the sequenced clones stopped just short of a redundant polyadenylation signal at a different site in each clone, indicating that they were independently derived (Fig. 1B). The sequence of these partial cDNA clones was extended using the RACE protocol for amplification of the 5' and 3' ends of rare mRNAs (31). The 3' RACE procedure yielded clones which identified two polyadenylation sites separated by a single nucleotide downstream of the redundant AA-TAAA sequence (Fig. 1B). A modified 5' RACE protocol identified multiple 5' ends grouped in three regions approximately 60–90 bases 5' of the EcoRI site (Fig. 1B), which was consistent with the results of RNA protection experiments probing for expressed sequences beyond the EcoRI site (data not shown). Inspection of the gen-
omic sequence did not reveal a TATA box in the vicinity of the 5' ends, consistent with the existence of multiple ends. The size of the composite cDNA sequence, 746-769 bases plus the poly(A) tail, was consistent with the approximately 0.9-kb RNA previously detected on Northern blot (Ref. 17; Fig. 2). Comparison with the genomic sequence revealed the presence of three exons, with exon 1 located completely within intron 1 of N-myc, exon 2 overlapping with exon 1B of N-myc, and exon 3 located 5' of all N-myc transcriptional initiation sites (Fig. 1A). No divergence from the published genomic sequence was noted except for occasional base changes in PCR-generated clones, which could be accounted for as in vitro polymerase errors.

Inspection of the cDNA sequence revealed a long open reading frame extending from the multiple transcriptional initiation sites to a stop codon within the redundant polyadenylation signal. However, the first ATG, which is in relatively good context to serve as a translational initiation codon [with a purine at position -3 (32)] appears approximately halfway down the mRNA, leaving a long 5' untranslated region and a very short 3' untranslated region, with the predicted protein consisting of 109 amino acids. Translation of in vitro synthesized RNA in a wheat germ lysate demonstrated the protein-coding potential of the mRNA and confirmed the predicted molecular weight of approximately 12,000 (Fig. 1C). We have called this gene product N-cym.

Analysis of the predicted amino acid sequence of N-cym using the algorithm of Garnier et al. (33) reveals interesting secondary structural features. At the amino terminus, there are two adjacent amphipathic α-helices separated by a loop. Adjacent to the second helix is a highly basic region, with 30% of the 43 carboxy-terminal amino acids consisting of arginine and lysine. These structural motifs (underlined in Fig. 1B) are shared by a class of DNA-binding transcriptional activators (21, 34), with the helix-loop-helix serving as a dimerization motif and the basic region functioning as the domain which interacts with DNA. Other than possessing these two secondary structural motifs, however, there is no significant homology between N-cym and the previously described class of DNA-binding proteins, nor does a search of GenBank reveal any other gene with significant homology to N-cym.

Expression of N-cym in Tumor Cell Lines. Following the structural characterization of the cDNA clones, we sought to define the pattern of N-cym expression in a panel of human tumor cell lines. We subcloned a segment of exon 3 from the Smal site (Fig. 1A) to just 5' of the polyadenylation signal for use as a double-stranded probe which would specifically detect N-cym mRNA and not cross-react with N-myc mRNA. Probing a Northern blot of polyadenylated RNA with this DNA segment (Fig. 2A), we detected high-level N-cym expression in neuroblastoma cell lines LA-N-5 and KCNR and SCLC cell line H249, all three of which contain amplified N-myc loci and express high levels of N-myc (24, 35). Cell line H187, which expresses N-myc without amplification (24), expressed a very small amount of N-cym, which is not readily apparent after photographic reproduction of the autoradiograms. Cell lines which failed to express N-myc, including some SCLC lines (which express either c- or L-myc), a hepatoma, a breast cancer, and a chronic myelogenous leukemia cell line, failed to express N-cym. Thus, there is a correlation between N-myc and N-cym expression, especially in those lines containing amplified N-myc loci. To strengthen this correlation, we decided to modulate N-myc expression and observe the effects on N-cym expression. Retinoic acid has been shown to differentiate neuroblastoma cells in culture and cause down-regulation of N-myc expression (35). Treatment of LA-N-5 cells with retinoic acid for 4 days caused a marked parallel drop in both N-myc and N-cym expression, whereas actin expression did not change (Fig. 2B). Therefore, N-myc and N-cym expression appear to be coregulated, at least in response to retinoic acid. Since the 5' ends of the two transcription units overlap, one possibility is that the genes share some common transcriptional regulatory elements.

Developmental Regulation of N-cym Expression. Our data indicate that significant N-cym expression is seen only in tumor cell lines that contain amplified N-myc loci. This could mean that N-cym expression is simply a pathological consequence of N-myc amplification. On the other hand, if the gene is developmentally regulated, as is N-myc (36, 37), significant normal expression may be seen only in specific tissues at specific developmental
stages. Since N-cym is coregulated with N-myc under some circumstances, and N-myc is normally expressed predominantly during fetal development, we screened several fetal tissues for N-cym expression. Northern blot analysis failed to detect N-cym mRNA in polyadenylated RNA samples isolated from human fetal brain, lung, liver, and kidney; N-myc was detected only in the fetal brain sample at relatively low levels, possibly because of the moderately degraded condition of the RNA (data not shown). To enable us to detect the processed N-cym mRNA with greater sensitivity, we devised a PCR assay with the 5' primer located in exon 2 and the 3' primer located in exon 3 to give a 306-bp amplification product specific for the spliced mRNA. To determine the sensitivity of our PCR assay, we performed the reaction using 1 μg of polyadenylated RNA isolated from cell lines characterized for N-myc and N-cym expression by Northern blot (Fig. 2) and for N-myc expression by PCR (data not shown). Fig. 3A illustrates that H187, a cell line with barely detectable N-cym expression by Northern blot, gave a strong signal using the PCR assay. H146, which had undetectable N-myc expression by Northern blot and PCR, also had undetectable N-cym expression by Northern blot and gave no signal in the N-cym PCR assay. K562, which had no detectable N-myc or N-cym on Northern blot, had detectable N-myc expression by PCR and gave a weak signal in our PCR assay for N-cym. The PCR analysis of fetal RNAs (Fig. 3B) revealed that N-cym mRNA was present in fetal brain, lung, liver, and kidney at varying low levels. Kidney consistently expressed the highest levels (among two different isolates and several assays), comparable to that expressed by SCLC line H187, which expresses N-myc and N-cym from an unamplified locus. Lung consistently expressed the lowest levels, with one isolate being completely negative. This PCR assay was not designed to be quantitative, however, and we do not wish to draw any conclusions about the relative levels of N-myc and N-cym in these tissues. These expression studies simply illustrate that expression of N-cym is not unique to tumor cells and suggest that it may have a functional role during normal fetal development.

Discussion
We have demonstrated that a polyadenylated RNA encoding a 109-amino acid protein is transcribed from the opposite strand of the N-myc locus. The overlap of N-myc and N-cym at their 5' ends is another example of the growing number of instances in which a common DNA segment is shared by two transcription units oriented in opposite directions (11–15). What is unique about this example, however, is that N-cym and N-myc expression appear to be coregulated, both in tumor cell lines under basal growth conditions and in neuroblastoma cell lines differentiated with retinoic acid. Part of this regulation is certainly at the transcriptional level, since we have already demonstrated that the ratio of sense to antisense nuclear run-on transcription across exon 1 of N-myc is relatively constant and independent of the varying levels of basal expression among different cell lines (17). Considering the close structural arrangement of the two transcription units, it is possible that they are regulated by a common bidirectional promoter or by independent promoters oriented in opposite directions but sharing some important elements. The relative contribution of transcriptional and posttranscriptional mechanisms in regulating N-myc and N-cym expression following retinoic acid exposure needs to be investigated further.

Coregulation of structurally integrated genes is the hallmark of prokaryotic operons and by this analogy could indicate a potential functional relationship between the N-cym and N-myc gene products. The presence of a HLH structure and a highly basic region suggests that N-cym may be a DNA-binding protein. However, as opposed to previously described HLH proteins, the basic region is carboxyl to the HLH region, and attempts at alignment of hydrophobic and hydrophilic residues within the α-helices of N-cym to those in other HLH proteins are unfulfilling. This does not necessarily rule out direct or indirect interactions between N-myc and N-cym, however. Further characterization of the function of N-cym will require the generation of antibodies to the protein. Nevertheless, it is interesting to note that, because of the size of the amplicon (26) and the close structural relationship between the two genes, N-cym should be invariably amplified in tumors containing N-myc amplification. The role that N-cym may play in the pathogenesis of these tumors requires further prospective investigation.

Materials and Methods
Cell Culture. SCLC cell lines and the LA-N-5 neuroblastoma cell line were maintained in RPMI 1640 supplemented with 2 mM L-glutamine and 10% fetal bovine serum, as previously described (16).

RNA Analysis. RNA isolation and separation into polyadenylated and nonpolyadenylated fractions, as well as Northern blots and RNase protection analyses, were performed exactly as previously described (17). All
probes were derived from a genomic phage clone (CH N-myc 5) isolated from a human placental library (24).

**cDNA Cloning.** A λgt10 cDNA library derived from the Kelly neuroblastoma cell line (27), purchased from Clontech Laboratories (Palo Alto, CA), was screened with the probes described in the text through quaternary platings by routine methods (38). The inserts were amplified from single positive plaques by PCR using primers complementary to sequences in the λgt10 arms (Clontech Laboratories) and the protocol supplied by the manufacturer. Inserts were cloned into pGEM 3 and sequenced manually in both orientations by the dideoxy technique as previously described (17); all sequence analysis was performed on a minimum of two independently derived PCR-generated clones to guard against polymerase errors. The 3’ RACE procedure (31) was accomplished essentially as previously described, using 3 μg of polyadenylated RNA from cell line H249 and an oligo(dT) primer-adapter (17) for the reverse transcription and 3’ amplification primer. The specific 5’-nested primers used for amplification consisted of sequences −161 to −142 and 128 to 144 in Fig. 2B. The 5’ RACE procedure, using 3 μg of polyadenylated RNA from H249, was performed as previously described (17), except that the first strand cDNA was linearly amplified using a thermostable reverse transcriptase (rTh; Cetus Corp.) to compensate for low abundance and inefficient recovery after the dA-tailing step. Two independent 3’ reverse transcription primers, representing the complement of sequences −158 to −193 and −295 to −313 in Fig. 2, were used and cycled for 25 cycles with rTh in the manufacturer’s buffer for reverse transcription. Each cycle consisted of 1.5 min at 95°C, 2 min at 62°C, and 2 min at 72°C. The dA-tailing and amplification steps were then accomplished as in Ref. 17, using the oligo(dT) primer-adapter as the 5’ primer and the complement of sequences −325 to −342 (Fig. 28) as the nested 3’ primer. Amplified products were cloned into pGEM 3 for sequencing.

**In Vitro Translation of N-cym.** A portion of the composite cDNA including the entire coding sequence, extending from the Smal site in exon 3 (−50) to the second polyadenylation site, was cloned into the pGEM 3 vector (Promega), and capped transcripts were synthesized using the cap analogue 7 meGpppG according to the supplier’s protocol. One μg of RNA was translated in a wheat germ lysate (Promega) containing 35S-labeled methionine and cysteine according to the manufacturer’s directions. A portion of the translation reaction was electrophoretically separated on a 15% sodium dodecyl sulfate-polyacrylamide gel which was impregnated with fluor (Enlightening; NEN) and autoradiographed on Kodak XAR-2 film.

**PCR Analysis of N-cym Expression.** N-cym expression was analyzed by reverse transcription-PCR as previously described (17), using 1 μg of polyadenylated RNA as template. The 3’ reverse transcription-amplification primer was located in exon 3 (Fig. 28; complement of bases 128 to 145), and the 5’ amplification primer was located in exon 2 (Fig. 2B, −161 to −142), giving a 306-bp amplification product specific for the processed mRNA. The reaction products were analyzed on a 3% low melting point-0.5% agarose gel, blotted onto nitrocellulose, probed with an N-cym-specific probe (Fig. 28, −50 to 310 bp), and autoradiographed.

**Acknowledgments**

We wish to thank Eric Westin, I. David Goldman, and Debbie Lebman for critical review of the manuscript.

**References**


