Structure and Expression of Canary *myc* Family Genes

ROBERT G. COLLUM,¹ DAVID F. CLAYTON, AND FREDERICK W. ALT^{1,2,3*}

Howard Hughes Medical Institute^{1*} and Departments of Biochemistry and Molecular Biophysics² and Microbiology,³
College of Physicians and Surgeons, Columbia University, 701 West 168th Street, New York, New York 10032,
and The Rockefeller University, 1230 York Avenue, New York, New York 10021⁴

Received 13 August 1990/Accepted 10 December 1990

We found that the canary N-*myc* gene is highly related to mammalian N-*myc* genes in both the protein-coding region and the long 3' untranslated region. Examined coding regions of the canary c-*myc* gene were also highly related to their mammalian counterparts, but in contrast to N-*myc*, the canary and mammalian c-*myc* genes were quite divergent in their 3' untranslated regions. We readily detected N-*myc* and c-*myc* expression in the adult canary brain and found N-*myc* expression both at sites of proliferating neuronal precursors and in mature neurons.

The *myc* family of cellular oncogenes is composed of three well-characterized members: c-*myc*, N-*myc*, and L-*myc* (1). The three genes encode highly related proteins that share regions of substantial similarity separated by regions of divergent sequence. c-*myc* is expressed in a broad range of cell lineages in association with cell proliferation; its expression is also activated in a wide variety of transformed cells (for a review, see reference 6). N-*myc* was isolated based on amplification in neuroblastomas and similarity to c-*myc* (11, 12, 21, 24). N-*myc* is expressed in a more restricted set of cell lineages than c-*myc* (30). Based on in situ analyses of developing murine tissues, N-*myc* expression was found to often correlate better with cell differentiative state rather than proliferative state (reviewed by Zimmerman and Alt [28]). For example, in the developing mouse brain, N-*myc* is expressed in postmitotic but undifferentiated neuroblasts and declines upon their differentiation (19).

Canaries, and other songbirds, offer a unique biological system for studying neuronal development during adulthood (reviewed by Nottebohm [22] and Konishi [14]). Young neuroblasts are generated continuously from the lining of the lateral ventricles in the adult forebrain and then migrate to sites throughout the forebrain, where they differentiate into mature functioning neurons. Neuronal migration, differentiation, and survival all appear to be regulated within one specific anatomical region of the forebrain, the higher vocal center (HVC), by seasonal changes in circulating levels of gonadal steroid hormones. Thus, it may be possible in the songbird system to link the differential expression of potentially involved genes such as N-*myc* to specific stages in neuronal development and function. As a first step, we isolated and characterized the canary N-*myc* gene and exon 3 of the canary c-*myc* gene and assayed the expression of N-*myc* and c-*myc* in adult canary brains.

Identification and isolation of canary N-*myc* gene. To identify the canary homolog of the mammalian N-*myc* gene, we screened a genomic *Eco*RI library with a murine N-*myc* cDNA probe. A clone containing a 7.2-kb N-*myc*-hybridizing insert was identified and subcloned, and the nucleotide sequences of regions homologous to the murine N-*myc* gene were determined (Fig. 1). The exon/intron borders and reading frame of the canary N-*myc* protein were determined by comparison with mammalian (9, 13) and chicken (23) N-*myc* genes. The protein-coding sequences of canary

N-*myc* are contained entirely within exons 2 and 3 and encode a protein of predicted molecular size of 47 kDa. The predicted protein is most closely related to that of the chicken (81% identity) but is also highly related to human and murine N-*myc* genes (69 and 71% identity, respectively). The canary N-*myc* gene also contains a large and evolutionarily conserved 3' untranslated region (see below). The 5' untranslated region of N-*myc* comprising exon 1 and the beginning of exon 2 is conserved between canary and chicken but not between avian and mammalian species.

Alignment of the canary N-*myc* protein with previously characterized *myc* proteins located nine blocks of evolutionarily conserved amino acids (Fig. 2). One of these blocks, VIII, was specific for N-*myc* proteins. Six blocks, I, II, III, V, VII, and IX, were also conserved in L-*myc* (17) and c-*myc* (5), while region IV was conserved only among N-*myc* and L-*myc*. These blocks of conserved sequence were separated by regions of protein with little or no sequence conservation from one *myc* protein to another. This striking lack of similarity in these spacer regions is particularly well illustrated by comparing the otherwise highly related canary and chicken proteins. Mismatches in three spacer regions denoted S1, S2, and S3 (Fig. 2) account for 63% of the mismatches between these two proteins and all the gaps that must be introduced to align these two sequences, even though S1, S2, and S3 compose only 20% of the total protein. In S1, there are 16 mismatches in 24 residues, while in S2 and S3, any alignment of sequences is difficult. These spacer regions are similar, however, in that all contain a large number of glycine, alanine, and proline residues. The observation that regions of remarkable evolutionary conservation are separated by regions of little or no conservation suggests that *Myc* proteins are composed of several discrete functional domains tethered to each other by linking sequences. These linking regions most likely would not be as well conserved as the functional domains and could vary in both size and amino acid composition. The spacer regions S1, S2, and S3 fit these criteria. Moreover, the high percentage of alanine, glycine, and proline residues in these spacers is reminiscent of collagen proteins, suggesting that the spacer regions function as intramolecular bridges between the different conserved functional domains.

The largest of the conserved *myc* domains (region IX) contains two protein motifs, the helix-loop-helix and the

* Corresponding author.

[illegible]

FIG. 2. Comparison of canary *N-myc* protein sequence (Can N) with the chicken *N-myc* (Ch N), murine *N-myc* (Mo N), human *N-myc* (Hu N), human *c-myc* (Hu C), and murine *L-myc* (Mo L) proteins. Nine conserved sequence motifs (I to IX) are boxed. Three highly variable spacer regions (S1 to S3) are overlined. Vertical lines indicate amino acid identity. Dashes indicate gaps in the alignment of sequences. Computer analyses were performed on an IBM-AT personal computer with the Beckman-MicroGenie Sequence Analysis Program (SciSoft, Inc.).

leucine zipper, found in a wide variety of known or putative transcription factors (16, 20). The observation that *myc* proteins contain these two motifs suggests a function in regulation of gene expression (reviewed by Collum and Alt [7]). Little is known about the functional role of the other evolutionarily conserved regions of *myc* proteins. However, even though specific functions have not been demonstrated for homology blocks I through VIII, mutational studies of *c-myc* indicate that blocks I, II, III, and IX are essential for

transformation of primary rat fibroblasts. In addition, homology region V is required for transformation of rat-1a cells (26). These five regions are conserved among *c-myc*, *N-myc*, and *L-myc*, consistent with the similar transformation properties of these three genes in cell culture (8, 15, 25, 27). It is interesting that regions IV and VII, although conserved in *c-myc* and *N-myc*, are dispensable for transformation of primary fibroblasts.

The evolutionary conservation of untranslated regions of

FIG. 3. (A) Comparisons of murine and canary N-myc exon 3 sequences. The comparison was performed by using the Beckman-MicroGenie Program with a window of 12 and a stringency of 75%. (B) Results of an identical dot-matrix analysis of nucleotide homology between murine and canary c-myc exon 3 sequences. (C) Nucleotide and deduced amino acid sequence of canary c-myc exon 3. The protein-coding portion is highly conserved among avian and mammalian c-myc genes, while the 3' untranslated portion is quite divergent.

given *myc* genes has suggested a potential regulatory role (reviewed by Zimmerman and Alt [28]). Such conservation is also evident upon examination of the canary N-myc 3' untranslated region; this sequence is 75% identical to the corresponding murine region over a length of 900 nucleotides (Fig. 3A). This evolutionary conservation of untranslated sequences is specific for the 3' untranslated sequences, as the 5' untranslated sequences composing exon 1 are not conserved between canary and mammalian genes. To determine whether the conservation of the 3' untranslated region seen for avian and mammalian N-myc genes was a feature shared by the c-myc gene, we determined the nucleotide

sequence of exon 3 of canary c-myc (Fig. 3C). The protein-coding portion of exon 3 is very well conserved, whereas the 3' untranslated region is quite divergent (Fig. 3B). Therefore, the evolutionary conservation between birds and mammals of the 3' N-myc untranslated sequences is not a property common to all *myc* genes.

myc expression in adult brain. To assay for N-myc and c-myc expression, we isolated specific canary N-myc and c-myc probes and used them to probe Northern (RNA) blots of total RNA from adult canary forebrain and hindbrain. Expression of a 2.9-kb N-myc RNA and 2.5-kb c-myc RNA was readily observed in both RNA preparations (data not

FIG. 4. *N-myc* expression in canary brain. (a to c) *N-myc* gene expression in ventricular zone detected by in situ hybridization. Labeling over the ventricular zone (V) appears as a local concentration of grains (a and b); individual cells along the ventricle are typically not well resolved by ^{35}S -labeled probes due to the small cell size. (a) Along the forebrain lateral ventricle; N indicates a group of labeled cells nearby with neuronal morphologies. (b) Along the third ventricle (nonneurogenic). (c) Negligible background levels generated by control hybridization with a sense-strand probe generated from the canary *N-myc* gene (both ventricular zone to the left and nearby neurons to the right are visible). All sections shown were taken from a single animal and hybridized in parallel. Autoradiographic exposure, 52 days; $100\times$ objective; bar = $20\text{ }\mu\text{m}$. (d to f) *N-myc* gene expression in song control nucleus HVC detected by in situ hybridization. Adjacent sections were hybridized to an antisense probe (d) or a sense-strand control (e and f). The plane of microscopic focus is on the silver grains in panels d and e; note significant labeling of cells in clusters in panel d compared with negligible background levels in panel e. Panel F shows the same field as panel E, but focused to demonstrate cellular morphologies most clearly; note large pale cells (neurons) and neuronal clusters. ($63\times$ objective; bar = $20\text{ }\mu\text{m}$). In situ experiments with cryostat sections and ^{35}S -labeled single-stranded RNA probes were conducted as described previously (4) with high-stringency washing (11 mM sodium at 65°C) and emulsion autoradiographic exposure times of 30 to 60 days; cells were counterstained with cresyl violet.

shown). To determine whether *N-myc* is expressed at sites of dividing or differentiating neuronal precursors in the canary brain, RNA probes specific for *N-myc* were hybridized in situ to adult canary brain tissue sections and bound probe was detected by emulsion autoradiography (4). In all experiments, negligible background levels were generated by control sense-strand probes (Fig. 4c and e). Concentrations of silver grains were seen over the lateral ventricular walls where neuroblasts proliferate (Fig. 4a). However, similar labeling was also detected along the third ventricle (Fig. 4b), indicating that *N-myc* expression is not restricted to active neurogenic regions of the ventricular wall. Finally, labeling also was detected in the HVC (Fig. 4d), a region where new neurons differentiate. The HVC contains many large neurons, typically organized into clusters with other smaller neurons (3, 10). These neuronal clusters in the HVC were clearly labeled (Fig. 4d), strongly suggesting *N-myc* expression in differentiated neurons. Although labeling of differentiated neurons in the HVC may represent retained *N-myc* expression in newly differentiated neurons, labeling of cells with mature neuronal morphology (large soma, pale nucleus) was also evident in other parts of the brain (e.g., see Fig. 4a), including brain regions, such as the thalamus and brain stem, that do not receive neurons in adulthood.

N-myc is expressed at high levels in the developing fetal mouse brain, and expression declines dramatically shortly after birth (29). Furthermore, in situ studies of developing mouse brain demonstrated *N-myc* expression in undifferentiated neuroblasts with a decline in expression upon differentiation. We found a number of potential differences with respect to *N-myc* expression in adult canary brains. *N-myc* expression was readily detectable in adult brains, although we have not ruled out the possibility of even higher levels in fetal brain. Furthermore, canary *N-myc* expression did not appear to be limited to undifferentiated neurons. However, the differences in *N-myc* expression in adult mouse and canary brains did not appear to correlate directly with the capability for neurogenesis in the latter. Expression of *N-myc* was not restricted to those regions of the canary brain where neurons born during adulthood differentiate. Yet, no other RNA transcripts with such a restricted pattern of expression in adult canary brain have been characterized. Indeed, several transcripts identified based on enrichment in regions of the brain contributing to song control are expressed in both mature and developing neurons in regions of the brain involved in song control as well as in regions unrelated to song control (4). Therefore, it remains possible that expression of *N-myc* is a general property of the songbird brain that contributes to the increased levels of neuronal development and plasticity but that other factors are also required for the highest levels of plasticity evident in the forebrain. Alternatively, *N-myc* expression may be involved in multiple processes in the adult avian brain, perhaps including but not limited to neurogenesis. Analysis of how *N-myc* gene expression may vary during brain development and during the different steroid-dependent phases of the song control circuit should help elucidate any potential role.

We are grateful to F. Nottebohm for providing canaries and enthusiastic support and to M. Huecas and E. Sinclair-Thompson for expert technical assistance.

This work was supported by the Howard Hughes Medical Institute and by Public Health Service grant CA 42335 to F.W.A. and NIH training grant GM-07367 to R.G.C. from the National Institutes of Health, and by grants to the Whitehall Foundation to D.F.C. and F. Nottebohm.

REFERENCES

- Alt, F. W., R. A. DePinho, K. Zimmerman, E. Legouy, K. Hatton, P. Ferrier, A. Tesfaye, G. D. Yancopoulos, and P. Nisen. 1986. The human *myc* gene family. Cold Spring Harbor Symp. Quant. Biol. 51:931-941.
- Alt, F. W., N. Rosenberg, V. Enea, E. Siden, and D. Baltimore. 1982. Multiple immunoglobulin heavy-chain gene transcripts in Abelson murine leukemia virus-transformed cell lines. Mol. Cell. Biol. 2:386-400.
- Burd, G. D., and F. Nottebohm. 1985. Ultrastructural characterization of synaptic terminals formed on newly generated neurons in a song control nucleus of the adult canary forebrain. J. Comp. Neurol. 240:143-152.
- Clayton, D. F., M. E. Huecas, E. Y. Sinclair-Thompson, K. L. Nastiuk, and F. Nottebohm. 1988. Probes for rare mRNAs reveal distributed cell subsets in canary brain. Neuron 1:249-261.
- Colby, W. W., E. Y. Chen, D. H. Smith, and A. D. Levinson. 1983. Identification and nucleotide sequence of a human locus homologous to the *v-myc* oncogene of avian myelocytomatosis virus MC29. Nature (London) 301:722-725.
- Cole, M. D. 1986. The *myc* oncogene: its role in transformation and differentiation. Annu. Rev. Genet. 20:361-364.
- Collum, R. G., and F. W. Alt. 1990. Are *myc* proteins transcription factors? Cancer Cells 2:69-75.
- DePinho, R. A., K. S. Hatton, A. Tesfaye, G. D. Yancopoulos, and F. W. Alt. 1987. The human *myc* gene family: structure and activity of *L-myc* and an *L-myc* pseudogene. Genes Dev. 1:1311-1326.
- DePinho, R. A., E. Legouy, L. B. Feldman, N. E. Kohl, G. D. Yancopoulos, and F. W. Alt. 1986. Structure and expression of the murine *N-myc* gene. Proc. Natl. Acad. Sci. USA 83:1827-1831.
- Kirn, J. R., and F. Nottebohm. 1990. Neuronal clusters: incorporation sites for neurons born in adult canary forebrain. Abstr. Soc. Neurosci. 16:1249.
- Kohl, N. E., C. E. Gee, and F. W. Alt. 1984. Activated expression of the *N-myc* gene in human neuroblastomas and related tumors. Science 226:1335-1337.
- Kohl, N. E., N. Katoshi, R. R. Schreck, G. Bruns, S. A. Latt, F. Gilbert, and F. W. Alt. 1983. Transposition and amplification of oncogene-related sequences in human neuroblastomas. Cell 35:359-367.
- Kohl, N. E., E. Legouy, R. A. DePinho, P. D. Nisen, R. K. Smith, C. E. Gee, J. D. Minna, and F. W. Alt. 1986. Human *N-myc* is closely related in organization and nucleotide sequence to *c-myc*. Nature (London) 319:73-77.
- Konishi, M. 1989. Birdsong for neurobiologists. Neuron 3:541-549.
- Land, H., L. F. Parada, and R. A. Weinberg. 1983. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Nature (London) 304:596-602.
- Landschulz, W. H., P. F. Johnson, and S. L. McKnight. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240:1759-1764.
- Legouy, E., R. DePinho, K. Zimmerman, R. Collum, G. Yancopoulos, L. Mitsock, R. Kriz, and F. W. Alt. 1987. Structure and expression of the murine *L-myc* gene. EMBO J. 6:3359-3366.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Mugrauer, G., F. W. Alt, and P. Ekblom. 1988. *N-myc* expression in mouse embryos correlates with early cell determination. J. Cell Biol. 107:1325-1335.
- Murre, C., P. Schonleber-McCaw, and D. Baltimore. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and *myc* proteins. Cell 56:777-783.
- Nau, M. M., B. J. Brooks, J. Battey, E. Sausville, A. F. Gazdar, I. R. Kirsch, O. W. McBride, V. Bertness, G. F. Hollis, and J. D. Minna. 1985. *L-myc*, a new *myc*-related gene amplified and expressed in human small cell lung cancer. Nature (London) 318:69-73.
- Nottebohm, F. 1985. Neuronal replacement in adulthood. Ann.

- N.Y. Acad. Sci. **457**:143–161.
23. Sawai, S., K. Kato, Y. Wakamatsu, and H. Kondoh. 1990. Organization and expression of the chicken *N-myc* gene. *Mol. Cell. Biol.* **10**:2017–2026.
24. Schwab, M., K. Altialo, L. Klempnauer, H. Varmus, J. Bishop, F. Gilbert, G. Brodeur, M. Goldstein, and J. Trent. 1983. Amplified DNA with limited homology to *myc* cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumor. *Nature (London)* **305**:245–247.
25. Schwab, M., H. E. Varmus, and J. M. Bishop. 1985. Human *N-myc* gene contributes to neoplastic transformation of mammalian cells in culture. *Nature (London)* **316**:160–162.
26. Stone, J., T. DeLange, G. Ramsay, E. Jakobovits, J. M. Bishop, H. E. Varmus, and W. Lee. 1987. Definition of regions in human *c-myc* involved in transformation and nuclear localization. *Mol. Cell. Biol.* **7**:1697–1709.
27. Yancopolous, G. D., P. D. Nisen, A. Tesfaye, N. E. Kohl, M. P. Goldfarb, and F. W. Alt. 1985. *N-myc* can cooperate with *ras* to transform normal cells in culture. *Proc. Natl. Acad. Sci. USA* **82**:5455–5459.
28. Zimmerman, K. A., and F. W. Alt. 1990. Expression and function of *myc* family genes. *Crit. Rev. Oncogen.* **2**:75–95.
29. Zimmerman, K. A., G. D. Yancopolous, R. G. Collum, R. K. Smith, N. E. Kohl, K. A. Denis, M. M. Nau, O. N. Witte, D. Toran-Allerand, C. E. Gee, J. D. Minna, and F. W. Alt. 1986. Differential expression of *myc* family genes during murine development. *Nature (London)* **319**:780–783.