

Nucleotide sequence analysis of the chicken *c-myc* gene reveals homologous and unique coding regions by comparison with the transforming gene of avian myelocytomatosis virus MC29, Δgag -*myc**

(DNA sequence analysis/identifying reading frames/*onc* genes and transforming proteins of retroviruses/cellular prototypes of viral *onc* genes)

DENNIS K. WATSON[†], E. PREMKUMAR REDDY[‡], PETER H. DUESBERG[§], AND TAKIS S. PAPAS^{†¶}

Laboratories of [†]Molecular Oncology and of [‡]Cellular and Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205; and [§]Department of Molecular Biology and Virus Laboratory, University of California, Berkeley, California 94720

Communicated by Robert J. Huebner, January 5, 1983

ABSTRACT Myelocytomatosis virus MC29 is a defective avian retrovirus with a hybrid transforming gene (Δgag -*myc*) consisting of a 1,358-base pair (bp) sequence from the retroviral *gag* gene and a 1,568-bp sequence (*v-myc*) shared with a cellular locus, termed *c-myc*. We have subjected to sequence analysis 2,735 bp of the cloned *c-myc* gene, which includes the *v-myc*-related region of 1,568 bp, an intervening sequence of 971 bp, and unique flanking sequences of 45 bp and 195 bp at the 5' and 3' ends, respectively. Analysis of the genetic information and alignment of the *c-myc* sequence with the known sequence of MC29 indicates that: (i) the two *myc* sequences share the same reading frame, including the translational termination signal; (ii) there are nine nucleotide changes between *c-myc* and *v-myc* that correspond to seven amino acid changes; (iii) the 971-bp intervening sequence of *c-myc* can be defined as an intron by consensus splice signals; (iv) the unique 5' sequence of *c-myc* could either extend its reading frame beyond the homology with *v-myc* or could be an intron because its junction with the *myc* region of the locus is a canonical 3' splice-acceptor site; (v) the *v-myc* contains 10 nucleotides at its 5' end not shared with the *c-myc* analyzed here and also not with known *gag* genes, probably derived from an upstream exon; and (vi) the *c-myc* locus can generate a mRNA whose termination signals have been identified to be located 83 bp and 119 bp from the point of divergence between the *v-myc* and *c-myc*. We conclude that the gene of the *c-myc* locus of the chicken and the *onc* gene of MC29 share homologous *myc* regions and differ in unique 5' coding regions and we speculate, on this basis, that their protein products may have different functions. The hybrid *onc* gene of MC29 must have been generated from the *c-myc* gene by deletion of the 5' cellular coding sequence, followed by substitution with the 5' region of the viral *gag* gene.

Myelocytomatosis virus MC29 is the prototype of a subgroup of replication-defective, avian retroviruses that also includes MH2, CMII, and OK10. The viruses of the MC29 subgroup have a broad oncogenic spectrum, inducing acute leukemias, carcinomas, and sarcomas, and transform fibroblasts and macrophages in cell culture (1, 2). The 5.7-kilobase (kb) RNA of MC29 contains an internal, specific sequence of 1.6 kb, termed *v-myc* (2), which is unrelated to the three essential virion genes of retroviruses (*gag*, *pol*, and *env*) (2) and which is the structural basis for the classification of avian acute leukemia viruses into the MC29 subgroup (1). In addition to *myc*, viruses of the MC29 subgroup contain a partial Δgag 5' to *myc* and a Δenv gene 3' to *myc*. The probable transforming *onc* gene of MC29 is a Δgag -*myc* gene defined by its product, a *gag*-related protein, termed p110 (3).

Recently we have molecularly cloned (4) and subjected to sequence analysis the 5' 3,782 nucleotides of proviral MC29 DNA, including the Δgag -*myc* transforming gene (5). The hybrid genome containing viral *gag* sequences and cell-derived *myc* sequences results in an open reading frame coding for a Δgag -*myc* hybrid protein containing 450 amino acids from *gag* and 425 from *myc*. This open reading frame encodes a 96,000-dalton protein and is terminated 301 bases upstream from the 3' *v-myc* helper-viral junction (5). This size estimate probably reflects the molecular mass of the primary polypeptide sequence of p110, which has an apparent molecular mass of 110,000 daltons. In MC29-transformed cells p110 is highly phosphorylated, lacks kinase activity (6, 7), and is localized in the nucleus (8, 9). *In vitro* translation of MC29 RNA also yields a protein with an apparent size of about 110,000 daltons (3).

In an effort to determine the functional relationship between the *onc* gene of MC29 and its cellular prototype, we have recently cloned the *c-myc* locus of chicken and found that it was interrupted relative to the viral counterpart by a 1-kb sequence of nonhomology (10).

In this study we have subjected to sequence analysis about 2.7 kb of the *c-myc* locus, and we compare it to the Δgag -*myc* gene of MC29. Alignment of the *c-myc* sequence with the hybrid *onc* gene of MC29 revealed that the two genes share homologous *myc* coding regions but that each gene contains a unique 5' coding region. Thus, structural and possible functional differences exist between the *onc* gene of MC29 and its cellular homolog.

METHODS AND MATERIALS

Molecular Cloning and Nucleotide Sequence Analysis. A normal chicken locus homologous to the *myc* sequence of the transforming gene of MC29 (*v-myc*) (10) was further subcloned in the plasmid vector, pBR322, and then was utilized for sequence analysis (Fig. 1B). Appropriate endonuclease-resistant DNA fragments were labeled (11) and all subsequent experiments were carried out as described (5, 12).

RESULTS

Genetic Information of *c-myc* Deduced from Its Nucleotide Sequence and by Alignment with the Sequence of MC29. Analysis of the genetic information of the 2.7-kb *c-myc* se-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: bp, base pair(s); kb, kilobase(s).

* Presented at the RNA Tumor Viruses Meeting, Cold Spring Harbor, NY, May 26–30, 1982.

¶ To whom reprint requests should be addressed.

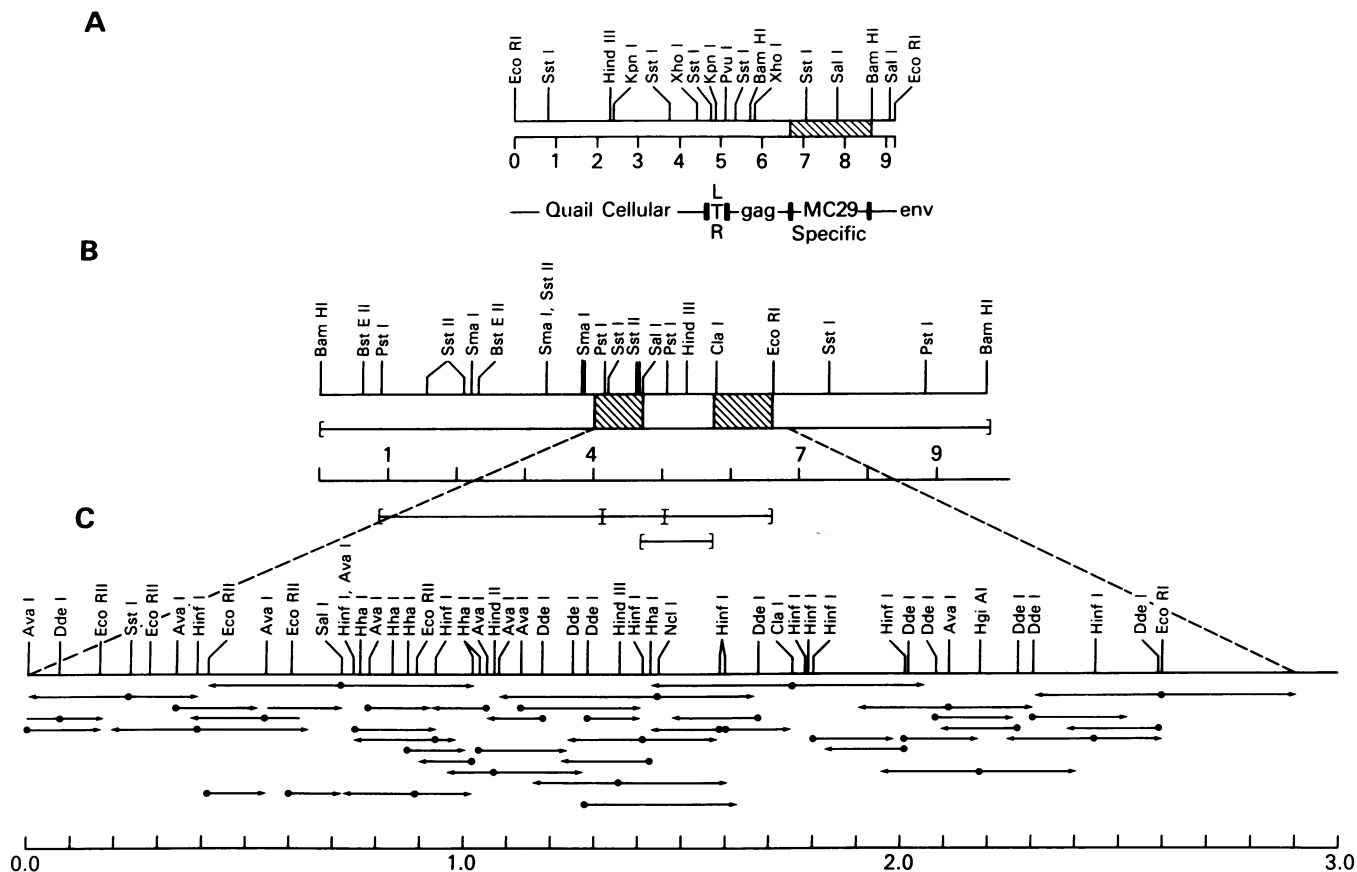


FIG. 1. Structural comparison between *v-myc* and *c-myc* and strategy for DNA sequence analysis of *c-myc*. Restriction site map of integrated MC29 proviral DNA (Δgag -*myc*) (A) and its cellular homolog (*c-myc*) (B). Shaded (hatched) areas represent *myc*-related sequences. Brackets below B define subclones used in this study. Cleavage sites of restriction enzymes used for sequence analysis are shown in C. Arrows detail orientation and extent of sequence analysis. Values are given in kb.

quence identified an open reading frame starting at position -45 and extending to position 706 and then continuing from position 707 to 1,266 with a noncoding, intervening sequence from position I¹ to I⁹⁷¹ (Fig. 2). A TAG termination signal of translation was identified at position 1,267 but we have not found an ATG translation start codon in the part of the *c-myc* locus that was subjected to sequence analysis here.

To determine how *c-myc* mRNA could be generated, we have looked in the *c-myc* sequence for signals that could be involved in the initiation and termination of mRNA. Consistent with the observation that the coding sequence of the *c-myc* locus appears to start at an undetermined location 5' to position -45, no initiation signal of transcription has been detected in the part of the *c-myc* locus that has been subjected to sequence analysis to date. However, two consensus signals involved in the termination of mRNA, A-A-T-A-A at position 1,651 and C-A-C-A at position 1,687, have been identified. Cellular sequences 5' to the overlap with *v-myc* must be subjected to sequence analysis to identify translation and mRNA initiation signals.

To determine structural and functional relationship between the *c-myc* locus and the genome of MC29, we have aligned, on the basis of nucleotide sequence homology, the sequence of *c-myc* with the known sequence of MC29 (5) (Fig. 2). This alignment located the 5' point of overlap with *v-myc* at position 1 of *c-myc* and position 2,063 of MC29 (Fig. 2). The 3' boundary of overlap with MC29 was identified at position 1,568 of *c-myc* and at position 3,629 of MC29 (5), which are located 83 and 119 base pairs (bp) upstream from the putative mRNA termination signal of *c-myc*. By comparing the complete sequence of *c-myc* to that of *v-myc*, we confirm that the *c-myc* locus is interrupted

by an intervening sequence. The 5' *c-myc* region is 706 bp in length and extends from position 1 to 706. The 3' *c-myc* region is 862 bp in length and extends from position 707 to 1,568. The intervening sequence measures 971 bp. As shown in Fig. 2, the 5' junction of the intervening sequence C-T-C-G-G/G-T-G-A-G is very similar to the common 5' splice-site sequence A-G/G-T-N-A-G (13). The 3' junction sequence C-C-T-G-T-G-C-A-G/ is similar to the 3' splice-site signal Y-N-Y-Y-N-C-A-G/ (Y = pyrimidine). Therefore, it seems that splice signals are present at the *myc* boundaries of this sequence and that this portion of the cellular sequence is probably an intron that is spliced out upon transcription into a mRNA.

This alignment also revealed that the two exons of *c-myc* and *v-myc* have a common reading frame and another unique feature of sequence conservation between *c-myc* and *v-myc*—namely, a common TAG termination signal of translation located at position 1,267 in the 3' exon of *c-myc*, or 301 bp upstream from the point where *c-myc* and *v-myc* cease to overlap, and at position 3,329 of MC29. If compared at the sequence level, the two exons of *c-myc* differ from *v-myc* in a total of nine base changes, which correspond to seven amino acid changes. Two of these seven changes are in the 5' exon and the remaining seven, in the 3' exon of *c-myc* (Table 1). Although many of these substitutions could result in minor conformational changes, the substitution of threonine for methionine is highly significant and could generate a different protein conformation. The available *c-myc* sequence information already suggested that the reading frame of *c-myc* may extend at least 15 codons 5' to the boundary of overlap with *v-myc*. However, close examination of the *v-myc* and *c-myc* 5' junction also reveals the

Start of v-myc homology

-40
CCC GAG CGC GGC TCA CCC GGC CCC CCC GTG TCC CCC TCC CGC CCG CAG GCA GCA GCC GCC GCG ATG CCG CTC AGC GCC AGC CTC CCC
Pro Glu Arg Gly Ser Pro Gly Pro Pro Val Ser Pro Ser Arg Pro Gln Ala Ala Ala Ala Ala Met Pro Leu Ser Ala Ser Leu Pro
50
AGC AAG AAC TAC GAT TAC GAC TAC GAC TCG GTG CAG CCC TAC TTC TAC TTC GAG GAG GAG AAC TTC TAC CTG GCG GCG CAG
Ser Lys Asn Tyr Asp Tyr Asp Tyr Asp Ser Val Gln Pro Tyr Phe Tyr Phe Glu Glu Glu Glu Glu Asn Phe Tyr Leu Ala Ala Gln
PstI SstI
CAG CGG GGC AGC GAG CTG CAG CCT CCC GCC CCG TCC GAG GAC ATC TGG AAG AAG TTT GAG CTC CTG CCC ACC CCG CCC CTC TCG CCC
Gln Arg Gly Ser Glu Leu Gln Pro Pro Ala Pro Ser Glu Asp Ile Trp Lys Lys Phe Glu Leu Leu Pro Thr Pro Pro Leu Ser Pro
250 300
AGC GCG CGC TCC AGC CTG GCC GCC GCC TCC TGC TTC CCT TCC ACC GCC GAC CAG CTG GAG ATG GTG ACG GAG CTG CTC GGG GGG GAC
Ser Arg Arg Ser Ser Leu Ala Ala Ala Ser Cys Phe Pro Ser Thr Ala Asp Gln Leu Glu Met Val Thr Glu Leu Leu Gly Gly Asp
350
ATG GTC AAC CAG AGC TTC ATC TGC GAC CCG GAC GAC GAA TCC TTC GTC AAA TCC ATC ATC ATC CAG GAC TGC ATG TGG AGC GGC TTC
Met Val Asn Gln Ser Phe Ile Cys Asp Pro Asp Asp Glu Ser Phe Val Lys Ser Ile Ile Ile Gln Asp Cys Met Trp Ser Gly Phe
400 450
TCC GCC GCC GCC AAG CTG GAG AAG GTG GTG TCG GAG AAG CTC GCC ACC TAC CAA GCC TCC CCG CCG GAG GGG GGC CCC GCC GCC GCC
Ser Ala Ala Ala Lys Leu Glu Lys Val Val Ser Glu Lys Leu Ala Thr Tyr Gln Ala Ser Arg Arg Glu Gly Gly Pro Ala Ala Ala
500 550
TCC CGA CCC GGC CCG CCG CCC TCG GGG CCG CCG CCT CCT CCC GCC GGC CCC GCC TCG GCC GGC CTC TAC CTG CAC GAC CTG GGA
Ser Arg Pro Gly Pro Pro Pro Ser Gly Pro Pro Pro Pro Ala Gly Pro Ala Ala Ser Ala Gly Leu Tyr Leu His Asp Leu Gly
600 650
GCC GCG GCC GCC GAC TGC ATC GAC CCC TCG GTG GTC TTC CCC TAC CCG CTC AGC GAG CGC GCC CCG CGG GCC GCC CCG CCC GGC GCC
Ala Ala Ala Ala Asp Cys Ile Asp Pro Ser Val Val Phe Pro Tyr Pro Leu Ser Glu Arg Ala Pro Arg Ala Ala Pro Pro Gly Ala
SalI
AAC CCC GCG GCT CTG CTG GGG GTC GAC ACG CCG CCC ACG ACC AGC AGC GAC TCG G GTGAGCGGCGCCCACTCGTGTGGGGACCCGAGCGTTCT
Asn Pro Ala Ala Leu Leu Gly Val Asp Thr Pro Pro Thr Thr Ser Ser Asp Ser
150 700 110
CCTTCCCGCTGCTCGGGACCTCACCTGCTTCTCCCCCTTGC GCGCACCGCTCTGTTAAATGCCAGTTAGCGAGGCGCTCTGCGAGTTTATTCTGGAAAAGCAGTTTTC
1200 1250
CCTCCACCCTTTGGAGGAAAAAGGAGTCAAAAGGGGAAAAAAGGGAGGGAGAGAACTTTATAACGGGGT CAGAGCCAGAGGTCCCCAGCGCG
1300 1350
GGAGCGGAGGAGCGCTTCCCGTGCACCCGAGGGTCCGGAAGGGTTAACTGCAGCCCGAGAACCGCGGGGAGAATCGGGCGGGGAGGTGCCAGTGAAGTGCTCGCAGG
1400 1450
AGCGAGTGAATGGACTCGGAAAGTTTAGAAATGCTTCTTAGGATTTGACAGCGTGTCCGCTCGCTATAAATTTGGGTTGAAAGCGAGTTGCTACCTTCGGGTTTCTCAG
1550 1600
TTTTTAAGGTTTTTTTTTTTTTTCTGCTTAGATTTTCCACAAATAGCAGATCAAGAGAGATTGAGAAAATGTCTCTGAATACACATGTGTGTTAAGTAAGCTTTCTTTGAG
1650 1700
GCGTGGCCAAAGCCTTAAATCCTTAATTAAGGGCAGCTTGAGTCTGGGAGCTTTATTGCGCTGCGTGTGACAGCCGAGGTTAACTTTGCTGCTATCTCTGTGCTCGCAGT
1750 1800
TACGCAAGAGCTGTGCACTTTTGAATCAGGGAGCCAGAGTTCGAGGGCAGTCTTGAAACGCTGCTTCTTAAGGGAAGGGCTCTGGGTAGGGAATGGGGAGTCTTGCGA
1850 1900 1950
GTCCGATGGTTGAAAGTAGTTTGTCAATGGGTTCTTGCTTCTGCTGCTGGAGTAAACCTTGTGGGCTGTCTGAGAGTTTTTGCACAGATCAGCTCATGATGCTATTTTT
2000
TTCCTGTGCAG AA GAA GAA CAA GAA GAA GAT GAG GAA ATC GAT GTC GTT ACA TTA GCT GAA GCG AAC GAG TCT GAA TCC AGC ACA GAG
Glu Glu Glu Gln Glu Glu Asp Glu Glu Ile Asp Val Val Thr Leu Ala Glu Ala Asn Glu Ser Glu Ser Ser Thr Glu
800 850
TCC AGC ACA GAA GCA TCA GAG GAG CAC TGT AAG CCC CAC CAC AGT CCG CTG GTC CTC AAG CGG TGT CAC GTC AAC ATC CAC CAA CAC
Ser Ser Thr Glu Ala Ser Glu Glu His Cys Lys Pro His His Ser Pro Leu Val Leu Lys Arg Cys His Val Asn Ile His Gln His
900 950
AAC TAC GCT GCT CCT CCC TCC ACC AAG GTG GAA TAC CCA GCC GCC AAG AGG CTA AAG TTG GAC AGT GGC AGG GTC CTC AAA CAG ATC
Asn Tyr Ala Ala Pro Pro Ser Thr Lys Val Glu Tyr Pro Ala Ala Lys Arg Leu Lys Leu Asp Ser Gly Arg Val Leu Lys Gln Ile
1000
AGC AAC AAC CGA AAA TGC TCC AGT CCC CGC ACG TCA GAC TCA GAG GAG AAC GAC AAG AGG CGA ACG CAC AAC GTC TTG GAG CGC CAG
Ser Asn Asn Arg Lys Cys Ser Ser Pro Arg Thr Ser Asp Ser Glu Glu Asn Asp Lys Arg Arg Thr His Asn Val Leu Glu Arg Gln
1050 1100
CGA AGG AAT GAG CTG AAG CTG AGT TTC TTT GCC CTG CGT GAC CAG ATA CCC GAG GTG GCC AAC AAC GAG AAG GCG CCC AAG GTT GTC
Arg Arg Asn Glu Leu Lys Leu Ser Phe Phe Ala Leu Arg Asp Gln Ile Pro Glu Val Ala Asn Asn Glu Lys Ala Pro Lys Val Val
1150 1200
ATC CTG AAA AAA GCC ACG GAG TAC GTT CTG TCT ATC CAA TCG GAC GAG CAC AGA CTA ATC GCA GAG AAA GAG CAG TTG AGG CGG AGG
Ile Leu Lys Lys Ala Thr Glu Tyr Val Leu Ser Ile Gln Ser Asp Glu His Arg Leu Ile Ala Glu Lys Glu Gln Leu Arg Arg Arg
1250
AGA GAA CAG TTG AAA CAC AAA CTT GAG CAG CTA AGG AAC TCT CGT GCA TAG GAACTCTGGACATCACTTAGAATACCCAACTAGACTGAACTAT
Arg Glu Gln Leu Lys His Lys Leu Glu Gln Leu Arg Asn Ser Arg Ala End
1350 1400
GATAAAATATTAGTGTCTTAATATCACTCATGAACATACATCAGTCCATTGAGTATGGAACATTTGCAACTGCATGCTGTGCGACTTAAGTGTGAGACTACACAACCTTGGCCGAA
1450 1500
TCTCCGAAACGTTTGGCCAGAACCTCAAACTGCTCATAATTGATACTTTGGGCATAAGGGATGATGGGACATTCTTCATGCTTGGGGATGAACTTTCAACTTTTTCTTTTCTTTT
1550 End of v-myc homology 1600 Polyadenylation signal
AAAATTTTGTATTAAAGGCATTTTTCTTAGCGAGAATCCAAATAGAGTTGTCAGATTGCTGTATATTTACACATCTTCTGGCATGAAATACCTTTAATAAAGCTTT
Poly A addition site 1700 1750
TATAGAAAAATGTGCAACATTAATACACAGCAGTTGTGGGAAGTGGATTATACTGTCTTGAACCTGTGTCCATAACATTTACAGTTTTGTTTTTATTT

FIG. 2. (Legend appears at the bottom of the following page.)

Table 1. Nucleotide and amino acid differences between *c-myc* and *v-myc*

c-myc position	c-myc		v-myc	
	Nucleotide	Amino acid	Nucleotide	Amino acid
5' exon				
200	C	Thr	T	Met
455	G	Arg	A	Gln
3' exon				
992	C	Ser	T	Leu
1,066	A	Ser	C	Arg
1,165	A	Ile	C	Leu
1,184	G	Arg	A	Lys
1,188	A	Leu	G	Leu
1,239	A	Lys	C	Asn
1,535	+T	—	—	—

+, Inserted nucleotide.

possible splice-acceptor signal (C-G-C-C-C-G-C-A-G) between positions -6 and 3. If this is indeed an acceptor, the C-A-G (Gln) and the adjacent 10 unique nucleotides seen in *v-myc* (5) are possibly derived from an additional exon. Hence, the reading frame of *c-myc* may continue into another as yet unidentified exon.

Finally, alignment of the MC29 sequence with the known *gag* sequence of Rous sarcoma virus (2) demonstrated that the two viral *gag* sequences overlap up to position 2,052 of MC29 (not shown) (5). Because the 5' boundary of overlap with *c-myc* is at position 2,062, this leaves a MC29 sequence of 10 nucleotides between positions 2,052 and 2,062 that is MC29-specific and possibly defines additional nucleotides in the upstream exon of *c-myc*.

DISCUSSION

Comparison of the Structures of the Cellular Gene of the *c-myc* Locus and the *onc* Gene of MC29. The nucleotide sequence demonstrates that the *c-myc* locus has all of the structural characteristics of a typical cellular gene. It carries mRNA termination signals and its coding regions (exons) are interrupted by at least one intervening sequence with consensus splice signals (13). Further sequence analysis upstream from the overlap with *v-myc* is expected to identify 5' initiation signals. Analysis of mRNA from normal chicken embryo fibroblasts revealed *myc*-related RNA species ranging from 6.5 to 2.7 kb in the nucleus and a 2.5-kb RNA species in the cytoplasm, indicating that *c-myc* is transcriptionally active (14). The existence of the large (possible) nuclear precursors of the 2.5-kb cytoplasmic RNA species implies that the *c-myc* may contain exons outside the regions defined here by homology with *v-myc*, which is consistent with our evidence that the 5' ATG start codon of translation is not within the 1,568 coding bases of *c-myc* that have been subjected to sequence analysis to date. The human *c-myc* locus, which has a structure similar to that of chicken *c-myc* (15), also expresses a cytoplasmic mRNA of 2.7 kb (16, 17).

Comparison of the gene of the *c-myc* locus and the *onc* gene of MC29 at the sequence level revealed that the two genes share homologous 3' *myc* coding regions and a common translation termination signal. The common *myc* coding regions differ by

nine nucleotides that do not introduce translation termination codons into the *c-myc* sequence (Table 1). These nucleotide differences correspond to seven amino acid changes between the respective proteins (Table 1). Therefore, the protein sequence encoded by this region of *c-myc* is not identical with the viral counterpart.

The major difference between the coding sequences of the two genes is that the *onc* gene of MC29 carries about 1,358 (*gag*) nucleotides at its 5' end, which are not shared with the gene encoded by the *c-myc* locus. The unique sequence at the 5' end of the *c-myc* could be viewed in two ways: (i) as an extension of the open reading frame for at least 45 nucleotides, which are not shared with the viral transforming gene, or (ii) as an upstream exon. Evidence for the latter comes from the presence of a perfect splicing signal at position -6 to 3 and from unique 10-bp sequence in *v-myc* 5' to the border with *c-myc*, which is neither homologous with *c-myc* nor with helper virus-related sequences. Assuming that the 2.5-kb cytoplasmic *myc*-related mRNA of normal cells is transcribed from the *c-myc* locus analyzed here, we estimate that the gene of the *c-myc* locus may carry at its 5' end a coding sequence of up to 600 nucleotides that is absent from *v-myc*, coding a putative polypeptide of about 80,000 daltons. [This estimate subtracts about 300 nucleotides from the 2.5-kb mRNA for poly(A) and 5' noncoding sequences.] We conclude that the *onc* gene of MC29 and the gene encoded by the cellular *myc* locus share closely related 3' coding regions, but each contains a unique 5' coding region.

How Was MC29 Generated from *c-myc*? Based upon the lack of sequence homology between the *c-myc* locus of the chicken and the *gag* and *env* genes of Rous sarcoma virus and other avian retroviruses (2), the generation of MC29 must have been the net product of illegitimate recombination between the *gag* and *env* genes of an avian retrovirus and the *c-myc* locus of the cell. The low probability and the necessity for all steps of such an event to occur within the lifespan of an infected animal for a transmissible virus to emerge are consistent with the infrequent occurrence of MC29-like viruses in nature (1, 2).

Conceivably, the primary recombination between the *gag* gene of a hypothetical retroviral progenitor of MC29 could have occurred at many possible sites within a noncoding sequence 5' of the coding region of *c-myc* to generate a splice-donor site. Subsequent splicing would have fused the coding *gag* sequences with the *myc* sequences shared by the virus and the cell. Such a mechanism has been favored because the resulting viral hybrid *onc* gene would contain a complete exon of a putative cellular *onc* gene and could then be functionally related to its cellular prototype. However, our data suggest that the hypothetical 5' recombination between *gag* and the *c-myc* locus must have occurred within the coding region of the cellular gene. This process would have generated a new hybrid gene, the *onc* gene of MC29, by linking the 5' region of the viral *gag* gene with the 3' coding region of the cellular gene.

Because the 3' *v-myc-env* interface is outside the coding region of *c-myc* and *v-myc*, the primary recombination at the 3' end of *c-myc* must have occurred in a noncoding region. However, this recombination would have to be limited to the 400 nucleotides between the translation termination codon of *c-myc*, at position 1,267, and the transcription termination codons at about position 1,650, in order for viable RNA virus to emerge.

FIG. 2 (on preceding page). Translated nucleotide sequence of the chicken *c-myc* locus. Nucleotide sequence of *c-myc* details position of exons showing viral homology (1-706 and 707-1,568) and the intron (I¹-I⁹⁷¹) and flanking cellular sequences (-45 to -1 and 1,570-2,735) with no MC29 viral homology. Dots mark every 10th nucleotide. The amino acid sequence deduced from the open reading frame is given in the bottom line. Several restriction sites are shown to further facilitate comparison with the *v-myc* sequence (5). Nucleotide and amino acid changes (listed in Table 1) between *c-myc* and *v-myc* are shown in boxes. Also highlighted are the donor (∩) and acceptor (∩) splice signals, polyadenylation signals, and protein translation termination site.

The presence of a transcription termination signal within the viral genome would obviously generate a viral RNA without the 3' terminal regulatory sequences necessary to synthesize a viral large terminal repeat (2).

Are *c-myc* and *v-myc* Parts of Functionally Related Genes?

The functional relationship between retroviral *onc* genes and cellular prototypes has been viewed in terms of two models (1): the qualitative model, which holds that viral *onc* genes and cellular prototypes have related structures but different functions (1), and the quantitative model, which holds that they are functionally the same (18). With regard to the relationship between the *c-myc* locus and the *onc* gene of MC29, our data favor the qualitative model, because it is probable that the hybrid gene (Δgag -*myc*) and the cellular gene [5' unique coding exon(s)-*myc*] would encode functionally different proteins. By contrast, it has been claimed in accord with the quantitative model that *c-myc* is a functional homolog of the *onc* gene of MC29 and that enhanced expression of *c-myc*, due to adjacent integration of a retrovirus, is the cause of B-cell lymphoma (19, 20). These integration points have been clearly defined (21). Yet, this hypothesis fails to explain why *c-myc* would cause lymphomas rather than the acute myelocytomatoses, carcinomas, or sarcomas typical of MC29 virus infection. This could be readily explained in view of our results that the *onc* gene of MC29 and the gene of the *c-myc* locus each have unique domains.

It is recognized that this explanation implies that the *c-myc* locus may be a potential cellular *onc* gene, albeit different from the *onc* gene of MC29. However, because activation of *c-myc* is not a necessary condition for chicken lymphoma (19, 20) and because another DNA locus has since been identified in chicken lymphoma that may be the transforming gene (22), the *c-myc* locus may not serve a maintenance function in transformation. This view is compatible with the lack of correlation between *c-myc* expression and the transformed phenotype of a number of normal and neoplastic human cells (16, 17). The high degree of conservation of *c-myc* from *Drosophila* to vertebrates (10, 14, 15, 23) also suggests a basic *c-myc* function, rather than a function that specifically affects lymphoblasts.

A direct comparison between the functions of the proteins encoded by the two related genes would be necessary for a conclusive decision between the two models.

It may be relevant for the function of the transforming protein of MC29 that its *v-myc* domain is linked to the viral *gag* protein instead of to the amino acid-terminal region of the *c-myc* product. Conceivably, the *gag* domain directs the protein to a specific cellular target, perhaps in the nucleus, that would not be reached by the protein encoded by the *c-myc* locus.

We thank M. Nunn for critical review of this manuscript. This research was supported in part by National Institutes of Health Research Grant CA 11426 from the National Cancer Institute to P.H.D.

1. Bister, K. & Duesberg, P. H. (1982) in *Advances in Viral Oncology*, ed. Klein, G. (Raven, New York), pp. 3-43.
2. Weiss, R. A., Teich, N. M., Varmus, H. E. & Coffin, J. M., eds. (1982) *RNA Tumor Viruses*, Molecular Biology of Tumor Viruses (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), 2nd Ed.
3. Mellon, P., Pawson, A., Bister, K., Martin, G. S. & Duesberg, P. H. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5874-5878.
4. Lautenberger, J. A., Schulz, R. A., Garon, C. F., Tschlis, P. N. & Papas, T. S. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1518-1522.
5. Reddy, E. P., Reynolds, K., Watson, D. K., Schulz, R., Lautenberger, J. & Papas, T. S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, in press.
6. Bister, K., Hayman, M. & Vogt, P. K. (1977) *Virology* **82**, 431-448.
7. Bister, K., Lee, W.-H. & Duesberg, P. H. (1980) *J. Virol.* **36**, 617-621.
8. Abrams, H. D., Rohrschneider, L. R. & Eisenman, R. N. (1982) *Cell* **29**, 427-439.
9. Donner, P., Greiser-Wilke, I. & Moelling, K. (1982) *Nature (London)* **296**, 262-266.
10. Robins, T., Bister, K., Garon, C., Papas, T. & Duesberg, P. (1982) *J. Virol.* **41**, 635-642.
11. Rushlow, K. E., Lautenberger, J. A., Papas, T. S., Baluda, M. A., Perbal, B., Chirikjian, J. G. & Reddy, E. P. (1982) *Science* **216**, 1421-1423.
12. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499-560.
13. Sharp, P. A. (1981) *Cell* **23**, 643-646.
14. Vennstrom, B., Sheiness, D., Zabiolski, J. & Bishop, J. M. (1982) *J. Virol.* **42**, 773-779.
15. Dalla-Favera, R., Gelmann, E. P., Martinotti, S., Franchini, G., Papas, T. S., Gallo, R. C. & Wong-Staal, F. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6497-6501.
16. Eva, A., Robbins, K. C., Andersen, P. R., Srinivasan, A., Tronick, S. R., Reddy, E. P., Ellmore, N. W., Galen, A. T., Lautenberger, J. A., Papas, T. S., Westin, E. H., Wong-Staal, F., Gallo, R. C. & Aaronson, S. A. (1982) *Nature (London)* **295**, 116-119.
17. Westin, E. H., Wong-Staal, F., Gelmann, E. P., Dalla-Favera, R., Papas, T. S., Lautenberger, J. A., Eva, A., Reddy, E. P., Tronick, S. T., Aaronson, S. A. & Gallo, R. C. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2490-2492.
18. Bishop, J. M. (1981) *Cell* **23**, 5-6.
19. Hayward, W. S., Neel, B. G. & Astrin, S. M. (1981) *Nature (London)* **290**, 475-480.
20. Payne, G. S., Bishop, J. M. & Varmus, H. E. (1982) *Nature (London)* **295**, 203-214.
21. Neel, B. G., Gasic, G. P., Rogler, C. E., Skalka, A. M., Ju, G., Hishinuma, F., Papas, T., Astrin, S. M. & Hayward, W. S. (1982) *J. Virol.* **44**, 158-166.
22. Cooper, G. A. & Neiman, P. E. (1981) *Nature (London)* **292**, 857-858.
23. Shilo, B.-Z. & Weinberg, R. A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6789-6792.