# Nucleotide sequence analysis of the proviral genome of avian myelocytomatosis virus (MC29)\*

(MC29 proviral genome/putative transforming protein)

E. PREMKUMAR REDDY<sup>†</sup>, R. K. REYNOLDS<sup>†</sup>, D. K. WATSON<sup>‡</sup>, R. A. SCHULTZ<sup>‡</sup>, J. LAUTENBERGER<sup>‡</sup>, and T. S. Papas<sup>‡</sup>

Laboratories of †Cellular and Molecular Biology and of ‡Molecular Oncology, National Cancer Institute, Bethesda, Maryland 20205

Communicated by Max Tishler, January 20, 1983

ABSTRACT The nucleotide sequence of the integrated proviral genome of avian myelocytomatosis virus (MC29) coding for gag-myc protein has been determined. By comparison of this nucleotide sequence with the helper virus as well as the c-myc region, it was possible to localize the junction points between helper viral and v-myc sequences. These studies demonstrate that (i) the large terminal repeat sequence of MC29 is very similar to that of Rous sarcoma virus, (ii) the viral genome has suffered extensive deletions in the gag, pol, and env genes, (iii) the gag region can code for p19, p10, and part of p27, (iv) the recombination between viral and cellular sequences occurred in the coding region of p27 such that the open reading frame extends for an additional stretch of 1,266 base pairs, resulting in a gag-myc hybrid protein, (v) the open reading frame terminated within the v-myc region 300 bases upstream of v-myc-helper viral junction, and (vi) the v-myc helperviral junction at the 3' end occurred in the middle of env gene, rendering it defective.

Myelocytomatosis virus (MC29) is a replication-defective avian retrovirus that induces a broad spectrum of malignant diseases, including myelocytomas, renal and liver tumors, and, less typically, carcinomas, sarcomas, and erythroblastosis (1, 2). The same virus induces morphological transformation of fibroblasts, epithelial cells, and macrophages in culture (3-5). This virus arose by recombination of the nondefective helper virus (MC29associated virus) and cellular sequences present within the normal chicken genome. These latter sequences termed "muc" appear to code for the transforming properties of the virus (6-8). The 5.7-kilobase (kb) RNA of MC29 has been shown to contain 1.6 kb of myc sequences that are flanked by partial  $\Delta gag$  gene at the 5' end and  $\Delta env$  gene at the 3' end (6). Nonproducer quail cell lines transformed by MC29 contain a 110,000-dalton protein with viral antigenic determinants (9). This protein appears to be a hybrid protein, the amino-terminal region of which is composed of helper virus gag gene products. The development of molecular cloning and DNA sequence analysis techniques has made the detailed analysis of the virus genome structure possible. In an attempt to better understand the structural organization and possible molecular mechanisms involved in transformation by MC29, we have undertaken primary DNA sequence analysis of the molecularly cloned integrated viral genome. Putative regulatory signals for transcription and translation of v-muc sequences have been identified. Sequence analysis has also demonstrated the occurrence of a long open reading frame within v-myc region that could code for the MC29 transforming protein.

## **MATERIALS AND METHODS**

Molecular Cloning. The integrated proviral genome of MC29 was initially cloned in  $\lambda$ gt WES· $\lambda$ B (10). In the present studies, two subclones of this DNA fragment were utilized for sequence analysis. A 1.4-kilobase-pair (kbp) Xho I fragment that contained the 5' large terminal repeat (LTR) and gag sequences was subcloned in M13. The 2.9-kbp BamHI fragment that contained the gag- and myc-specific sequence was subcloned in pBR322. The insert DNAs were purified by agarose gel electrophoresis and DEAE-cellulose (DE-52, Whatman) column chromatography after cleavage with appropriate restriction enzymes and were used in all subsequent analyses.

Nucleotide Sequence Analysis. Nucleotide sequence analysis of the 2.9-kbp BamHI fragment was carried out by the method of Maxam and Gilbert (11). Appropriate restriction fragments were prepared and labeled at their 5' ends by using  $[\gamma^{-32}P]$ ATP and polynucleotide kinase (P-L Biochemicals) as described by Maxam and Gilbert (11) or at their 3' end by using cordycepin 5'- $[\alpha^{-32}P]$ triphosphate and terminal deoxynucleotidyl transferase (P-L Biochemicals) according to Roychoudhury and Wu (12). The nucleotide sequence was determined by the procedure of Maxam and Gilbert (11).

The 1.4-kbp Xho I fragment was subjected to sequence analysis by the dideoxy chain-terminator DNA sequence analysis procedure of Sanger *et al.* (13). The Alu I and Sau3A fragments of the 1.4-kbp DNA were ligated into M13 mp7 DNA restricted with the appropriate enzyme. Subclones prepared with M13 mp7 as the vector were identified as clear plaques and amplified. Single-stranded DNA was isolated and the sequence reactions were carried out as suggested by the cloning or sequence analysis kit suppliers (Bethesda Research Laboratories).

## RESULTS

Strategy for Determining the Sequence of the MC29 Proviral Genome. The restriction map of molecularly cloned MC29 DNA was constructed by using both double-digestion analysis and the partial digestion technique of Smith and Birnstiel (14). Fig. 1 shows the restriction map of MC29 proviral DNA and the localization of its cell-derived sequences (v-myc). The sequence of both DNA strands was determined for >90% of the viral genome, and all restriction cleavage sites were confirmed by sequence analysis.

Sequence Organization of 5' Noncoding Sequences. The nucleotide sequence of the MC29 proviral genome along with its

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: MC29, avian myelocytomatosis virus strain MC29; kb, kilobase(s); kbp, kb pair; LTR, large terminal repeat; RSV, Rous sarcoma virus.

<sup>\*</sup> Presented at the RNA Tumor Virus Meeting, Cold Spring Harbor, NY, May 26–30, 1982.



FIG. 1. Summary of the major structural features of the MC29 proviral genome. Restriction map and the important features of the MC29 viral genome, including the open reading frames, possible signals for the initiation of transcription and translation, are illustrated.

5' noncoding and flanking quail cellular sequences is presented in Fig. 2. An important structural feature of the retroviral genome is the occurrence of two LTRs at both 5' and 3' ends of the proviral genome (for review, see ref. 15). Examination of the nucleotide sequence in Fig. 2 indicates the presence of such a LTR-like structure located at positions 114-426. Nucleotide sequence analysis of the LTRs of several avian and mammalian retroviruses has revealed that LTRs bear striking similarities to the terminal repeats of prokaryotic transposable elements and contain signals for the initiation and termination of transcription as well as mRNA capping (for review, see ref. 15). Thus, as with other known LTR sequences, we find the following features within the MC29 LTR: (i) inverted terminal repeats: the sequence T-G-T-A-G-T-C-T-T appears at the terminus of the LTR at positions 114-122 and in the inverted form at positions 418-426; (ii) a promoter-like sequence T-A-T-T-A-A-G was found at positions 298-305. This A+T-rich sequence precedes by 24 nucleotides the G-C-C-A sequence most likely to be the mRNA capping site; (iii) a polyadenylylation signal, A-A-T-A-A-A, was found in the LTR at positions 321-326. This signal preceded the dinucleotide C-A by 11 bases at position 338-339.

The DNA sequence of MC29 LTR region was compared to the Rous sarcoma virus (RSV) LTR sequence (16, 17) by the two-dimensional dot-matrix homology program (18). This analysis showed that there was 83% homology between the U<sub>5</sub> regions, whereas the homology between the U<sub>3</sub> regions was about 80%. This observation is consistent with earlier observations that U<sub>3</sub> regions of retroviral LTRs are most susceptible for sequence divergence (19, 20). Downstream from the 5' LTR, a 19-base sequence complementary to the 3' end of tRNA<sup>Trp</sup> (21) was localized at positions 428–446. Following this sequence was a stretch of 256 base pairs of noncoding sequences, followed by the gag gene coding sequences.

Sequence Organization of MC29 Transforming Gene. Downstream from the 5' LTR sequences, we observed two ATG codons at positions 519–521 and 704–706. The first ATG codon is unlikely to function as the initiator codon as it was followed

closely by in-phase termination codons. However, the second ATG codon at position 704-706 was followed by a large open reading frame of 2,625 bases ending with a TAG codon at position 3,329-3,331. We presumed that this sequence coded for the viral gag-myc hybrid protein. Fig. 2 shows the predicted amino acid sequence of this polypeptide. Earlier work on the biochemical and immunological analysis of the putative transforming protein of MC29 revealed that it is a 110,000-dalton polyprotein comprised of sequences derived both from the gag region and the cell-derived myc regions (9). To determine the exact point of recombinational event, we determined the sequence of the corresponding regions of the c-myc gene and also compared this with the known sequence for the RSV gag gene (17) (Fig. 3). Thus, it was possible to localize the junction points between the c-myc and helper viral sequences that are indicated in Figs. 2 and 3. It is interesting to note that there is a stretch of 10 bases in v-myc at the 5' junction that did not show correspondence either to c-myc or RSV gag sequences. The origin of these sequences at this point is unclear. On the other hand, at the 3' junction the v-myc, c-myc, and RSV env genes showed a sequence homology of five bases. Whether the occurrence of this homology is purely accidental or has a significance in the recombinational process between helper virus and host cellular sequences is speculative at this time.

The open reading frame shown in Fig. 2 could code for a polypeptide of 875 amino acids with a molecular mass of 96,000 daltons. This is in close agreement to the estimated size of 110,000 daltons for *gag-myc* hybrid protein synthesized by MC29-infected quail nonproducer cell lines. The difference in molecular masses could be due to post-translational modifications such as glycosylation and phosphorylation of the protein molecule. This protein consists of 450 amino acids derived from the amino terminus of *gag* region, followed by 422 amino acids that are specific for the v-myc region. Thus, the *gag* region contained the entire sequences of p19 and p10 but only the first 211 amino acids of p27 followed by the c-myc-derived sequences. The coding region terminated within the cell-derived myc sequences,

Inverted reset - stert IR Internet de reset - stert IR ICITALICUTA ACAUMATATATATATATATATATATATATATATATATATATA	AGETGEGETÄÄTTEEGGETEGGTGAAG		50 SGGGAAATCGATATTTGCC	AAATGGAGAGAGAGAGAGAGAGAGAGA	001 0000000000000000000000000000000000
Providence - Polyadeny Law Cost Card Andre Card Card Card Card Card Card Andre Card Card Card Card Card Card Card Card	inverted repeat - start L	TR 150 GTAACGATGAAACAGCAAT	TACAGCCTTATAAGGAGAG	AAAGGGTACCGTGCATGATGATG	
<pre>ind of LTR LINK Ninding the index intervention of the interve</pre>	CCTTATTAGGAAGGTAACAGACGGGT		TCTACTTGATTCCGCATAG	• Promoter	• Polyadenylation signal
A GET MAGENA MAGTGET GEGCAAA BESCET GEGCAAT CET CCC TCAT CGET TECKET TAACGOARCAAGACAAT ACCTAAT TAAGAGGEGCAA GEGGAAT CHARGA BET CAAT CAA CEA AT TAAGAGEGGCAAT CHARGA BET CAAT AGA CEA AT TAAGAGEGGCAAT CHARGA BET CAAT AGA CEA AT TAAGAGEGGCAAT CHARGA BET CAAT AGA CEA AT TAAGAGEGGAAT CHARGA BET CAAT AGA CEA AT TAAGAGEGGAAT CHARGA BET CAAT AGA CEA AT TAAGAGEGGAAT CHARGA BET CAAT AGA AGA CEA AT TAAGAGEGAAT CHARGA BET CEA CEA AT AGA AGA CEA AT TAAGAGAGAAT CHARGA BET CEA CEA AT AGA AGA CEA AT TA GE CAA AGA CEA AT AGA CEA AT TAAGAGEGAAT CHARGA BET CEA CEA AT AGA CEA AT TA CEA CEA CEA TAA AT CE CEA CEA CT TAA CAA AGA CEA CEA CEA CEA CEA CEA CEA CEA CEA CE	350 TACCATCCACCACATTGGTCTGCACCT	GGGTAGATGGACAGACCGT	TGAGTCCCTAACGATTAC	end	of LTR tRNA binding site
<pre>1</pre>	450 ATCGTTAGGGAATAGTGGTCGGCCACA	GGCGTGGCGATCCTGCCCT	500 CATCCGTCTCGCTTAACG	GAGCAGGACGATGACCCTAGTAG	
CAGGGGTGATGCTGGGCGGGGGGGGGGGGGGGGGGGGGG	AAGCTGAGTGACGTCGGAGGGAGGGAGCTCC			GAGTEGTTGGAAGACGCGAAGGA	
The Giu Ara vi The up van The ser as a fail giu fau fur for Giu Giu y Thr fail and Giu Ara vi The up van The ser as a fail giu fau fur for Giu Giu y Thr fail and Giu Ara vi The up van The ser as a fail giu fau fur for CCC CCC Giu Giu Van Thr fail and Giu Giu Giu Giu Giu Yan Met Leu Ser Leu Leu Giu Liy Si Gu Giu Leu Ves Thr For Ser Asp Leu Tyr Ser Pro Giu y Ban HI The Tag Ara CC CCC ATT ACC GC GCA CTC ACC CAG GG GCA fails GTA CTT GGG AA TCG GGA GGT TA GGG GG GG GG CAT TA GG GG GT GG GG CAT TA GG GG GT GG GG CT CA ACC CAG CGA GG GCA fails GTA CTT GGG GAA GTA GTT AG GGG GG GG GG CAT GGA GTC GAG GA CA CA CA CA CA CA CA GCA AGC AGA CT TA TT GGT GG GAA TTA GGG GG GG GG GG AG GG CAT TAG GG GG GG GG GG CAT GGA GG GG GG GG GA CCA CCA CCA CGA GGA G		700 VStart p19	C GTC ATA AAG GTG	ATT TCC TCC CCC TCT AAA	750 ACC TAT TCC CCA AAA ACC TCT
cc T to Aug Aug Gua Aria Gug Guc Aria Tis Tis Tic C Ti Ti Aug Aug Gua Gua Gus Tis C Ti Aris Tic C C Gua		Met Glu Al	a Val Ile Lys Val	Ile Ser Ser Ala Cys Lys	Thr Tyr Cys Gly Lys Thr Ser
Pro Ser Use Lys chu life diy Ala Met Leu Ser Leu Sin Lys chu leu Sin yang Leu Leu And Lys Pro Jas Lys chu life diy Ala Met Leu Ser Leu Sin Cys chu and Ser Pro Jise Pro	CCT TCT AAG AAG GAA ATA GGG	GCC ATG TTG TCC CT	TC TTA CAA AAG GAA	GGG TTG CTT ATG TCT CCC	TCA GAC TTA TAT TCC CCG GGG
The tes off tee and tee and the area were to nee the decision of the decision	W Bam HI	ATA MEL LEU SEF LE		JIY LEU LEU MEL SER PRO	
GGG CAT TIG ANG GGT CAG AG GAT CAG GTT ACA TCT GAG CAA AG TIT TGG TTG GGA TA GGG GGA GGG AGG GTC TCT CGG   GJY ATA Leu Lys ATA ATA ATG GTU GTU GTU GTU ATT TS SET RTU GTU TATA CAG GGG GAG GGG AGG GTC TCT CGC GG   GCA GGT CGA TG GAT GAC AG CAG CGA CGA AGG GGC GGA GGG AGA GGG GAG GGG AGA CGA GAT GGT GGA GAA ACT GTG CAG   GCA GGT CGA TG GAT GAC GAC GAG CGA CGA GG GGA GGG CAA GGG GGA GGA	Ser Trp Asp Pro Ile Thr Ala	Ala Leu Thr Gln Ar	g Ala Met Val Leu (	Gly Lys Ser Gly Glu Leu	Lys Thr Trp Gly Leu Val Leu
Gi yan teu tys Ara Ara Ara G Grague sin varing ser site (in Ara Lys Prei trong and try of a Grague sin varing ser site (in Ara Lys Prei trong and try of a Grague site of the Ara Cara Grage Cara Ara Cara Grage Cara Cara Cara Cara Cara Cara Cara Car	GGG GCA TTG AAG GCG GCT CGA	GAG GAA CAG GTT AC	CA TCT GAG CAA GCA	AAG TTT TGG TTG GGA TTA	GGG GGA GGG AGG GTC TCT CCC
Uch dei Cus Guis viel Ale Ale Ale Cha Cus Ale Ale Cus Cus Ale Ale Cus Cus Ale Cus	GIN ATA LEU LYS ATA ATA ATA	1050	ir ser Giu Gin Ala	_ys Phe Trp Leu Giy Leu	1100 CAA ACA ACT CTC CAC CCA CAT
GCG AAG ATG GCG CCG GAG GAA ACT GCC TAA ACC TAA ACG GTT GGC ACA TCC TGC TAT CAT TGC GGA ACA GCT ATT GGC TGT ATT TGC ATA Lys Met ATA Pro Glu Glu Thr ATA Thr Pro Lys Thr Val Gly Thr Ser Cys Tyr Nis Cys Gly Thr ATA THE GTY Cys Astr tpl GC ACA GCC TCG GCC CT CCT CT TCT TAT GTG GGG AGT GGT TTG TAT CCT TCC CTG GGG GGG GTG GGA GAG CAC CAG GGC CAG GGG GTA AT Thr ATA SET ASTR Pro Pro Pro Pro Pro Pro Pro Yor AG US Pro AFT GTG TTG TAT CCT TCC CTG GGG GGG GTG GGA GAG CAC CAG GGC CAG GGG AGA CAT CT CGG GGG GGA ACAG CCA GGG GGG GGG GGG GG	Pro Gly Pro Glu Cys Ile Glu	Lys Pro Ala Thr Gl	u Arg Arg Ile Asp I	Lys Gly Glu Glu Val Gly	Glu Thr Thr Val Gln Arg Asp
Ala Lys Met Ala Pro Giu Giu Inr Ala Inr Pro Lys Inr Val Giy Inr Ser Uys Iry Mis Uys Giy Inr Ala IIE Giy Uys Ash Osy ecc AcA GCC TCG GCC CCT CCT CCT TAT GTG GGG ACT GGT TTG TAT CCT TCC CTG GCG GGG GTG GGA GAG CAG GGC CAG GGG Ala Thr Ala Ser Ala Pro	GCG AAG ATG GCG CCG GAG GAA	ACT GCC ACA CCT AA	A ACC GTT GGC ACA	ICC TGC TAT CAT TGC GGA	ACA GCT ATT GGC TGT AAT TGC
GCC ACA GCC TCG GCC CCT CCT CCT TAT GIG GGG ACT GGT TIG TAT CCT TCL CTG GGG GGG GGG GGG GGG GGG GGG GGG GGG	Ala Lys Met Ala Pro Glu Glu	. end pl9 ♥.	vs Inr Val Gly Inr : ▼ Start pl0	1250	Inr Ala Ile Giy Cys Ash Cys
GGT GÁC ACA CT CGG GGG GCG GAA CAG CCA AGG GCG AG CCA GGG CAC ACG GGT CAG GCT CGA GCT CT GAC GCC TGA ACT GAC TGG GCA GIY ASP Thr Pro Arg GIY AIA GIU GIN Pro Arg AIA GIU Pro GIY HIS AIA GIY GIN AIA Pro GIY Pro AIA Leu Thr ASP TFP AIA AGG GTC GĠG GAG GAA CTT CCA GAC TACT GGT CCG CCC GTG GTG GCC ATT CCT GTA GTG ATT ACA CAC GÁG GGA CCC GCC TGA CCC ATT GIY AIG IN GIU LEU AIA SET Thr GIY Pro Pro Val Val IA AM EM Pro Val Val IIE ASN THr GIU GIY Pro AIA TFP THR Pro 1500 15	GCC ACA GCC TCG GCC CCT CCT Ala Thr Ala Ser Ala Pro Pro	Pro Pro Tyr Val Gl	G AGI GGI IIG IAI y Ser Gly Leu Tyr I	Pro Ser Leu Ala Gly Val	GGA GAG CAG CAG GGC CAG GGG Gly Glu Gln Gln Gly Gln Gly
Gly Asp Inn Pro Arg Gly Ala Glu Gin Pro Arg Ala Glu Pro Gly Pro Ala Gly Gin Ala Pro Gly Pro Ala Leu Inn Asp Irp Ala AGG GTC GGG GAG GAG CTT GCA ACT GGT CCC CCC GT GTG GTG GCC ATT CCT GTA GTG ATT AAC ACA GÅG GGA CCC GCC TG ACC CCT AGG GTC GÅG GAG GAG CTT GCA ACT GGT CCC CCC GT GTG GCC GTC CT GTA GTG ATT AAC ACA GÅG GGA CCC GCC TG ACC CCT AGG GIL GIU Leu Ala Ser Thr Gly Pro Pro Val Val Ala Met Pro Val Val 11e Asn Thr GIU GIV Pro Ala GIU Val GIU ATA 1500 CTG GAG CCA AMA TTG ATC ACA AGA CTG GCT GAT ACG GCA CTA GGC TTA CGA TCC CG GATT ACT ATG GAA GTG GAA GTG GAA CLEU GIU Pro Lys Leu Ile Thr Arg Leu Ala Asp Thr Val Arg Thr Lys Gly Leu Arg Ser Pro Ile Thr Met Ala GIU Val GIU Ala 1550 GTT TAG TCC TCC CGG GTG CTG CCG CAT GAT GTC ACG AAT CTA ATG AGA GTT ATT TTA GGA CCT GCC GCA TAT GCC TTA GGA TCG Ala Trp GIV val GIn Leu GIN Thr Val 11e Ala Ala Thr Asn Leu Met Arg Val Ile Leu GIV Pro Ala Pro Tyr Ala Leu Trp Met Asp 160 GCT TGG GGA GTC CAA CTC CAG AGC GTT ATA GCG GCG GCC ACT CGC GAC ACC CCA GGA CCA GGA CGG GGG GAA CGG Ala Trp GIV val GIn Leu GIN Thr Val 11e Ala Ala Thr Arg Asp Pro Arg His Pro Ala Asn GIV GIV GUA GGA CGG GGG GAA TTG 180 ACT AAC TTG GAT CGC TTA AÅG GGC TTA GCT GAT GGG ATG GGG CGC GCA CCA CAG GGC GCA CTTA TA AGA GCG GGG GAA TTG Thr Asn Leu Asp Arg Leu Lys GIY Leu Ala Asp GIV GTI AGA GAA GTT GCC GGG GCA CCT GCA GGA CCT GCA GGA CCT CCA GG GCA CCT CCA GG GCA CCT CCA 180 GTT GCT ATT ACG GCG TCG GCT CTC AGA GCG TTT AAA GA GTT GCC GGG CTG CCG GGA CCT CCA GG GCA CCT CCA GG GCA CCT CCA GG GCA CCT CCC GCG GCG CTC CCC AAC CAG GCC TTA AGA GGA GTT GAG 190 GTG ATT ACG GCG GCG GCT CTC AGA GCC AAT TCG GCT TA AAG GCG GTA GAA CCT GCA GGA CCT CCC GCG CCG CCC CCC GCG GCG CCC GCC CCG CCC	GGT GAC ACA CCT CGG GGG GCG	GAA CAG CCA AGG GC	G GAG CCA GGG CAC	GCG GGT CAG GCT CCT GGG	CCG GCC CTG ACT GAC TGG GCA
AGG GIC GGG GAG CTI GCG AGT ALT GG CTG CCC GIG GIG GCC ALG CLI GLC IG ALT ALX GAG GAG CLU GLU GIGU PTO 1500 1500 1500 1500 1500 1500 CTG GAG CCA AAA TTG ATG AGA CTG GCT GAT AGG GTA AGA GGA CCA AGA GGA TTG CAA GAG CCC GCA TTA GGA AGA GTG GAA GGG Leu Glu PTO Lys Leu IIe Thr Arg Leu Ala Asp Thr Val Arg Thr Lys Gly Leu Arg Ser PTO IIe Thr Met Ala Glu Val Glu Ala 1550 CTT ATG TCC TCC CGG CTG CTG CCG CAT GAT GTC ACG AAT CTA ATG AGA GTT ATT TTA GGA CCT GCC CCA TAT GCC TTA GG ATG GAA Leu Met Ser Ser Arg Leu Leu PTO HIS ASP Val Thr Asn Leu Met Arg Val IIe Leu GIY PTO Ala PTO Tyr Ala Leu TTP Met ASp GCT TGG GGA GTC CAA CTC CAG AGG GTT ATA GCG GCG GCC ACT CGC GAC CCC GA CA CCA GG GAA C GGT CAA GGA CGG GGG GAA CGG 1700 ATT ACT TG GAT CCC TTA AGA GGC TTA GCC GCG GCC ACT CGC GAC CCC CGA CA CAG GGA CAG GT CAA GGA CGG GGA GAA TTG Thr Asn Leu Asp Arg Leu Lys Gly Leu Ala Asp Gly Met Val GJY Asn PTO GIn Gly GIn Ala Ala Leu Leu Arg PTO Gly Glu Leu 1750 GTT GG GGA GTC CTG CG GCG CTC CCAG GGC TTA GGA GTT AAA GAA GTT GCC CGG CTG GCG GAA CCT GCA GGT CCA TGG GGG GAA CTC A 1800 Thr Asn Leu Asp Arg Leu Lys Gly Leu Ala Asp Gly Met Val GJY Asn PTO GIn Gly GIn Ala Ala Leu Leu Arg PTO Gly Glu Leu 1801 1900 GTT GCT ATT AGG GGG TCG GCT CTC CAG GGC TTA AGA GAA GTT GCC CGG CTG GCG GAA CCT GCA GGT CCA TGG GGG GAC ATC AGG CAA 1900 GTT GG TATT AGG GGG TCG GCT GCC ATC CGC CTT ATA AGA GGG GTT GAA GGG GTT GAA GGG CCC TCC CCG GCG GGC GCC CAA GT AGA GAA 1900 GTT GG TATT AGG GGG TCG GCT TTA GAT GGC CTT ATA AGA GGG GTT GAA GGG GTG AGA CTC CCAA GG CCA CCT CC CGG CG GGG GCC ACC CCC CCAA GT CAG GGC GCC CCC CCC CCC AGA GT ATG CCAA GGG CTT ATA AGA GGG GTT GAA GGG CCC TCC CCG CCG CCG CCAA GT GCC AGA GT CT TG AGA CCA CC CC AGA GT CC AGA GT AA ASN AFG LEU IL LYS All VAI GIU GLY GAA SP LIV PTO TTP ALA ASP LIVE TTA GG CCC AGA GT CCC AGA GT CCC AGA GT CCAA GGC CCC CCC CCC AGA GCC AGA CCC CCC	GIY ASP INT Pro Arg GIY Ala	GIU GIN Pro Arg Al	a Giu Pro Giy His / end pl0 .V	start p27	Pro Ala Leu Inr Asp Irp Ala 1450
<pre>1500 CTG GAG CCA AÀA TTG ATC ACÀ AGA CTG GCT GAT ACG TC AGG ACC AAG GGC TTA CGA TCC CCG ATT ACT ATG GCA GAA GTG GAA GCG Leu Glu Pro Lys Leu IIe Thr Arg Leu Ala Asp Thr Val Arg Thr Lys Gly Leu Arg Ser Pro IIe Thr Met Ala Glu Val Glu Ala 1550 CTT ATG TCC TCC CGG CTG CTG CCG CAT GAT GTC ACG ATT CTA ATG AGA GTT ATT TTA GGA CCT GCC CCA TAT GCC TTA TGG ATG GAC Leu Met Ser Ser Arg Leu Leu Pro His Asp Val Thr Asn Leu Met Arg Val IIe Leu Gly Pro Ala Pro Tyr Ala Leu Trp Met Asp GCT TGG GGA GTC CAA CTC CAG ACG GTT ATA GCG GCG GCC ACT CGC GAC CCC GA CCC GC ACC GGT CAA GGA CGG GGA ACG AT TPG GTY Val GIn Leu GIn Thr Val IIe Ala Ala Ala Thr Arg Asp Pro Arg His Pro Ala Asn Gly GIn Gly Arg Gly Arg Gly GIU Arg 100 ACT AAC TTG GAT CGC TTA AGG GGC TTA GGG ATG GGG GC ACC CCC GA CC CA CA GG GC CA TTA TA AGA CCG GGG GAA TTG Thr Asn Leu Asp Arg Leu Lys Gly Leu Ala Asp Gly Met Val Gly Asn Pro GIn Gly GIn Ala Ala Leu Leu Arg Pro Gly Glu Leu 1550 GTT GG AG CT GG GC GG GC CT CCA GGG GTT AGA GAA GTT GCC CGG CTG GG GGAA CC TG CG GGT CAT GG GCG GAC ATC ACG CAG 1900 GTT AGT ACT TA GG GG TTA GT GAC GGG TTT AGA GAA GTT GCC CGG GTT GG GG GAC CT CGC GGG CCA TG GG GG GAC ATC ACG CAG 1900 GT ATC ATT AGG GC TC GAG GCC TTA CGG CTT AAGA GAA GTT GCC CGG GTT GG GG GAC CT CCG CCT TCC AGG GCG GAC CT CCG 1900 GT ATC ATT GAG CTC TTT GTT GAT TTC GCC AAT CGG CTT ATA AAG GCG GTT AAG GAG GT CA GAC CTC CCG CCT TCC AGG GCG GAC CT CCG 1900 Ser Ala Arg Jue GIN Ala Pro Pro Asp IIe Jus Ala Arg Leu Ala GiU GY Pro Trp Ala Asp IIe Thr GIN 1900 GTA ACT ATT GAC GCC TTT AGC GAG ATG CAG CCA CCC CCC CCC ACC AGA ATC CAG CAG CTT ACG CAG CTC CCC ACC AGG CAG CTC CCG ACC CCC CC CA CC AGG CAG CCC CCC C</pre>	AGG GIC GGG GAG GAG CII GCG Arg Val Gly Glu Glu Leu Ala	Ser Thr Gly Pro Pr	o Val Val Ala Met	Pro Val Val Ile Asn Thr	Glu Gly Pro Ala Trp Thr Pro
Leu Glu Pro Lys Leu Ile Thr Arg Leu Ala Asp Thr Val Arg Thr Lys Gly Leu Arg Ser Pro Ile Thr Met Ala Glu Val Glu Ala 1550 CTT ATG TCC TCC CGG CTG CTG CG CAT GAT GTC AGA GTC AGA GTA TTA FG AGA GTT ATT TTA GGA CCT GCC CCA TAT GC TTA FGG ATG GAG Leu Met Ser Ser Arg Leu Leu Pro His Asp Val Thr Asn Leu Met Arg Val Ile Leu Gly Pro Ala Pro Tyr Ala Leu Trp Met Asp 1700 GCT TGG GGA GTC CAA CTC CAG AGG GTI ATA GGG GG GCC ACT CGC GAC CCC GA CCC CAG GGG AAC GGT CAA GGA CGG GGA AC GG Ala Trp Gly Val Gln Leu GIn Thr Val 11e Ala Ala Ala Thr Arg Asp Pro Arg His Pro Ala Asn Gly GIn Gly Arg Gly Glu Arg 1701 ACT ACT TG GAT CGC TTA AGG GGC TA GGA TG GGA GGG GCA ACT CCA CGA GGC CAC TCA GGA CGA GGA CAG CGA GGA ATG Thr Asn Leu Asp Arg Leu Lys Gly Leu Ala Asp Gly Met Val Gly Asn Pro Gln Gly Gln Ala Ala Leu Leu Arg Pro Gly Glu Leu 1850 GTT GCT ATT ACG GCG TCG GCT CTC CAG GCG TTI AGA GAA GTT GCC CGG CTG GCG GAA CCT GCA GGC TCA TG GCG GAC ATC AGG CAG Val Ala Ile Thr Ala Ser Ala Leu Gln Ala Phe Arg Glu Val Ala Arg Leu Ala Glu Pro Ala Gly Pro Trp Ala Asp Ile Thr Gln 1900 GTG ATC ATT GAC TGC TTT GTT GAT TC GCC AAT CGG CTT ATA AGA GCG GTT GAG GGG TCA GAC CTC CGG CGT CCC GGG GGG GAC ATC AGG CGG Gly Pro Ser Glu Ser Phe Val Asp Phe Ala Asn Arg Leu Ile Lys Ala Val Glu Gly Ser Asp Leu Pro Pro Ser Ala Arg Ala Pro 2000 GTG ATC ATT GAC TGC TTT GGA GGC CAG CCA CCC CCC CAG CAG AGA CTA CGA CAC CCC CCA CAG GGC CAG GCC AG Val Ile Ile Asp Cys Phe Arg Gln Lys Ser Gln Pro Asp Ile Gln Gln Leu Ile Arg Ala Ala Pro Ser Thr Val His Gly Gln Ala 2100 GTG ATC ATT GAC TGC TTT CAG GCC CAG CCA CCC CCC CAG CAG CAG CAG	CTG GAG CCA AAA TTG ATC ACA	AGA CTG GCT GAT AC	G GTC AGG ACC AAG	GGC TTA CGA TCC CCG ATT	ACT ATG GCA GAA GTG GAA GCG
CTT ATG TCC TCC CGG CTG CCG CAT GAT GTC ACG AAT CTA ATG AGA GTT ATT TTA GGA CCT GCC CCA TAT GCC TTA TGG ATG GAC Leu Met Ser Ser Arg Leu Leu Pro Hts Asp Val Thr Asn Leu Met Arg Val IIe Leu Gly Pro Ala Pro Tyr Ala Leu Trp Met Asp 1700 GCT TGG GGA GTC CAA CTC CAG ACG GTT ATA GCG GGG GCC ACT CGC GAC CCC CGA CAC GGC AAC GGT CAA GGA CGG GAA CGG Ala Trp Gly Val Gln Leu Gln Thr Val IB Ala Ala Thr Arg Asp Pro Arg His Pro Ala Asn Gly Gln Gly Arg Gly Glu Arg 1800 ACT AAC TTG GAT CGC TTA AAG GGC TTA GGT GAT GGT GAG GGC AAC CCA CAG GGT CAG GCC GCA TTA TTA AGA CCG GGG GGA ATG Thr Asn Leu Asp Arg Leu Lys Gly Leu Ala Asp Gly Met Val Gly Asn Pro Gln Gly Gln Ala Ala La Lu Leu Arg Pro Gly Glu Leu 1850 GTT GCT ATT ACG GCG TCG GCT CTC CAG GCG TTT AGA GAA GTT GCC CGG CTG GCG GAA CCT GCA GGC CAA TCA AGA CGG Ala In Pro Ala Asp Arg Leu Lys Gly Leu Ala Asp Gly Met Val Gly Asn Pro Gln Gly Gln Ala Ala Leu Leu Arg Pro Gly Glu Leu 1850 GAA CCA TCT GAG TCC TTT GTT GAT TTC GCC AAT CGG CTT ATA AGG GCG GTG GGG GAA CCT GCG GGC GAC ATC ACG CAG Gly Pro Ser Glu Ser Phe Val Asp Phe Ala Asn Arg Leu IIe Lys Ala Val Glu Gly Ser Asp Leu Pro Pro Ser Ala Arg Ala Pro 2000 GTG ATC ATT GAC TGC TTT AGG CAG AAG CAA CAG CCA GAT ATC CAG CAG CTT ATA CGG GCA GCC CCC TCC ACA GTG CAC GGC CGG GCC ALI LUS Pro Leu Ser Phe Arg Gln Lys Ser Gln Pro Asp IIe Gln Gln Leu IIe Arg Ala Alro For Ser Thr Val His Gly Gln Ala 2100 GCA GCC GCG GTG GCG CTC CAC GCC GCC AGC CAG CAG CAG AAT TAC CAG CAG CTT AGA CTG GTG CAG CCC TCC TAC ACG CAG CTC TC TC TAC Ala Ala Ala Ala Met Pro Leu Ser Ala Ser Leu Pro Ser Lys Asn Tyr Asp Tyr Asp Tyr Asp Ser Val Gln Pro Tyr Phe Tyr Phe Tyr Phe 2200 AST II GCG GG GG GAG CAG CTC CTC GGC GGC CAG CAG CGC GCC CTC CTC GGC GCC CTC TCC CT TCC CCC GC CAC CAC CTC CG AGG CAG GAG GAG GAG CAG CTG CTC GGC GGC CAG CAG CGC CGC CTC CTC GCC CGC CCG CCC TCC TGC CCC GCC CTC TGC CCC CC	Leu Glu Pro Lys Leu Ile Thr	Arg Leu Ala Asp Th	ir Val Arg Thr Lys	aly Leu Arg Ser Pro Ile	Thr Met Ala Glu Val Glu Ala
GCT TGG GGA GTC CAG ACG GTT ATA GCG GCG GCC ACT CGC GAC CCC CGA CAC CGA GGA GAC GGT CAA GGA GGG GGG GAA GG Ala Trp Gly Val Gln Leu Gln Thr Val 11e Ala Ala Ala Thr Arg Asp Pro Arg His Pro Ala Asn Gly Gln Gly Arg Gly Glu Arg 1500 ACT AAC TTG GAT GGC TTA AGG GGC TTA GCT GAT GGG ATG GTG GGC AAC CCA CAG GGT CAG GCC GCA TTA TTA AGA CGG GGG GAA TTG Thr Asn Leu Asp Arg Leu Lys Gly Leu Ala Asp Gly Met Val Gly Asn Pro Gln Gly Gln Ala Ala Leu Leu Arg Pro Gly Glu Leu 1850 GTT GCT ATT ACG GCG TCG GCT CTC CAG GCG TTT AGA GAA GTT GCC CGG CTG GCG GAA CCT GCA TGG GGC CAA TG AC GGG Val Ala 11e Thr Ala Ser Ala Leu Gln Ala Phe Arg Glu Val Ala Arg Leu Ala Glu Pro Ala Gly Pro Trp Ala Asp 11e Thr Gln 1950 GGA CCA TCT GAG TCC TTT GTT GAT TTC GCC AAT CGG CTT ATA AAG GCG GTT GAG GGG TCA GAC CTC CGC GCT TCC GCG GGG GCA CCC Gly Pro Ser Glu Ser Phe Val Asp Phe Ala Asn Arg Leu 11e Lys Ala Val Glu Gly Ser Asp Leu Pro Pro Ser Ala Arg Ala Pro 2000 GTG ATC ATT GAC TGC TTT AGG CAG GAG AGT CA GCA GAT ATC CAG CAG TTT ATA CGG GCA GCA CCC TCC ACA GTG CAC GGC CAG Val 11e 11e Asp Cys Phe Arg Gln Lys Ser Gln Pro Asp 11e Gln Gln Leu 11e Arg Ala Ala Pro Ser Thr Val His Gly Gln Ala 2150 GCA GCC GCC GCG ATG CGC TCA GG GCC AGC CTC CCC AGA AGA ACT CC GAT TAC GAC TAG GAC CTG CG GTG CAG CCA GCC CAG Val 11e 11e Asp Cys Phe Arg Gln Lys Ser Gln Pro Asp 11e Gln Gln Leu 11e Arg Ala Ala Pro Ser Thr Val His Gly Gln Ala 2150 GCA GCC GCC GCG ATG CGC CTC AGC GCC AGC CCC CCC AGA CAG CGC GGC CAG CGA CCC CCC	CTT ATG TCC TCC CGG CTG CTG Leu Met Ser Ser Arg Leu Leu	CCG CAT GAT GTC AC Pro His Asp Val Th	CG AAT CTA ATG AGA ( Ir Asn Leu Met Arg	GTT ATT TTA GGA CCT GCC Val Ile Leu Gly Pro Ala	CCA TAT GCC TTA TGG ATG GAC Pro Tyr Ala Leu Trp Met Asp
Ala Trp Gly Val Gin Leu Gin Inr Val Tie Ala Ala Ala Thr Arg Asp Pro Arg His Pro Ala Asn Gly Gin Gly Arg Gly Glu Arg 150 ACT AAC TIG GAT CGC ITA AAG GGC TTA GCT GAT GGG ATG GTG GGC GAC CCA CAG GGT CAG GCC GCA TTA TTA AGA CCG GGG GAA TTG Thr Asn Leu Asp Arg Leu Lys Gly Leu Ala Asp Gly Met Val Gly Asn Pro Gln Gly Gin Ala Ala Leu Leu Arg Pro Gly Glu Leu 185 GTT GCT ATT ACG GCG TCG GCT CTC CAG GCG TTA GAA GAT GTG CCC GGC CTG GCG GAA CCT GCA GGT CCA TGG GCG GAC ATC ACG CAG Val Ala Tie Thr Ala Ser Ala Leu Gin Ala Phe Arg Glu Val Ala Arg Leu Ala Glu Pro Ala Gly Pro Trp Ala Asp Tie Thr Gin 1950 GGA CCA TCT GAG TCC TTT GTT GTT TTC GCC AAT CGG CTT ATA AAG GCG GTT GAG GGG TCA GAC CTC CCG CCT TCC GCG CGG GGT CCG Gly Pro Ser Glu Ser Phe val Asp Phe Ala Asn Arg Leu Ile Lys Ala Val Glu Gly Ser Asp Leu Pro Pro Ser Ala Arg Ala Pro 2000 GTG ATC ATT GAC TGC TTT AGG CAG AAG TCA CAG CCA GAT ATC CAG CAG CTT ATA CGG GCA GCA CCC TCC ACA GTG CAG GGC CAG GCC ACA Val Tie Ile Asp Cys Phe Arg Gin Lys Ser Gin Pro Asp Tie Gin Gin Ctt ATA CGAC TAC GAC TCC GGT CACA CTG CTA CTTC TAC TTC Ala Ala Ala Met Pro Leu Ser Ala Ser Leu Pro Ser Lsu Asn Tyr Asp Tyr Asp Ser Val Gin Pro Tyr Phe Tyr Phe 2200 AGG GAG GAG GAG GAG AAC TTC TAC CTG GGC GCA GCA GCA GCA GCA GCA CGC TGC GCC CCG TCC CGC GCC GCC GCC GCC	GCT TGG GGA GTC CAA CTC CAG	ACG GTT ATA GCG GC	G GCC ACT CGC GAC	CCC CGA CAC CCA GCG AAC	GGT CAA GGA CGG GGG GAA CGG
ACT AAC IIG GAT CGC TIA AAG GGC TIA GCT GAT GGG AIG GIG GGC GAC CCA CAG GGT CAG GCC GCA TIA TIA AGA CCG GGG GAA TIG Thr Asn Leu Asp Arg Leu Lys GIy Leu Ala Asp GIy Met Val GIy Asn Pro Gin GIy Gin Ala Ala Leu Leu Arg Pro GIy GIu Leu 1850 GTT GCT ATT ACG GCG TCG GCT CTC CAG GCG TTT AGA GAA GTT GCC CGG CTG GCG GAA CCT GCA GGT CCA TGG GCG GAC ATC ACG CAG Val Ala Ile Thr Ala Ser Ala Leu Gin Ala Phe Arg Giu Val Ala Arg Leu Ala Giu Pro Ala Giy Pro Trp Ala Asp Ile Thr Gin 1900 GGA CCA TCT GAG TCC TTT GTT GAT TC GCC AAT CGG CTT ATA AAG GCG GTT GAG GGG TCA GAC CTC CCG CCT TCC GCG CGG GCT CCG GIy Pro Ser Glu Ser Phe Val Asp Phe Ala Asn Arg Leu Ile Lys Ala Val Giu Giy Ser Asp Leu Pro Pro Ser Ala Arg Ala Pro 2000 GTG ATC ATT GAC TGC TTT AGG CAG AG TCA CAG CAG GAT ATC CAG CAG CTT ATA CGG GCA GCA CCC TCC CACA GTG CA GGC CAG GCA Val Ile Ile Asp Cys Phe Arg Gin Lys Ser Gin Pro Asp Ile Gin Gin Leu Ile Arg Ala Ala Pro Ser Thr Val His Gly Gin Ala 2100 GCA GCC GCC GCG ATG CCG CTC AGC CCC CCC AGC AGA AAC AGA CAC TAC GAT TAC GAC TAC GG CAG CCC TAC TTC TAC TTC Ala Ala Ala Ala Ala Met Pro Leu Ser Ala Ser Leu Pro Ser Lys Asn Tyr Asp Tyr Asp Tyr Asp Ser Val Gin Pro Tyr Phe Tyr Phe 2200 40 Giu Giu Giu Giu Giu Giu Giu Giu Asn Phe Tyr Leu Ala Ala Gin Gin Arg Giy Ser Giu Leu Gin Pro Pro Pro Ala Pro Ser Thr Ala Asp Ile Trp Lys 2250 40 Set I 40 TT AGG CTG CTG CCC ATG CGG CCC CTC TGC GCG CAG CAG CGG CGC CCG CCC CCC CCC C	Ala Trp Gly Val Gin Leu Gin	Thr Val Ile Ala Al 1750	a Ala Thr Arg Asp	Pro Arg His Pro Ala Asn	Gly Gin Gly Arg Gly Glu Arg 1800
GTT GCT ATT ACG GCG TCG GCT CTC CAG GCG TTT AGA GAA GTT GCC CGG CTG GCG GAA CCT GCG CCG GCG GCC GAA GTT AGG GCG GCC ATC ACG CAG Val Ala 11e Thr Ala Ser Ala Leu Gin Ala Phe Arg Glu Val Ala Arg Leu Ala Glu Pro Ala Gly Pro Trp Ala Asp I1e Thr Gin 1950 GGA CCA TCT GAG TCC TTT GTT GAT TTC GCC AAT CGG CTT ATA AAG GCG GTT GAG GGG TCA GAC CTC CCG CCT TCC GCG CGG GGT CCG Gly Pro Ser Glu Ser Phe Val Asp Phe Ala Asn Arg Leu I1e Lys Ala Val Glu Gly Ser Asp Leu Pro Pro Ser Ala Arg Ala Pro 2000 GTG ATC ATT GAC TGC TTT AGG CAG AAG TCA CAG CCA GAT ATC CAG CAG CTT ATA CGG GCA GCA CCC TCC ACA GTG CAC GGC CAG GCA Val 11e I1e Asp Cys Phe Arg Gln Lys Ser Gln Pro Asp I1e Gln Gln Leu I1e Arg Ala Ala Pro Ser Thr Val His Gly Gln Ala 2100 GCA GCC GCC GCG GCG GCG CTC CCC AGC CTC CCC AGC AAG AAC TAC GAT TAC GAC TAC GAC TCG GTG CAG CCC TAC TTC TAC TTC Ala Ala Ala Ala Ala Met Pro Leu Ser Ala Ser Leu Pro Ser Lys Asn Tyr Asp Tyr Asp Tyr Asp Ser Val Gln Pro Tyr Phe Tyr Phe 2200 Glu Glu Glu Glu Glu Glu Asn Phe Tyr Leu Ala Ala Gln Gln Arg Gly Ser Glu Leu Gln Pro Pro Ala Pro Ser Glu Asp I1e Trp Lys 2200 AGG GAG GAG GAG GAG AAC TTC TAC CTG GCC CCC CCC AGC CGC CGC CGC CGC CGC CCC GCC CGC TCC TC	ACT AAC IIG GAT CGC ITA AAG Thr Asn Leu Asp Arg Leu Lys	GGC TTA GCT GAT GG Gly Leu Ala Asp Gl	GG ATG GTG GGC AAC ( y Met Val Gly Asn )	CCA CAG GGT CAG GCC GCA Pro Gln Gly Gln Ala Ala	ITA ITA AGA CCG GGG GAA TTG Leu Leu Arg Pro Gly Glu Leu
Val Ala 11e inr Ala Ser Ala Leu Gin Ala Phe Arg Giu Val Ala Arg Leu Ala Giu Pro Ala Giy Pro Trp Ala Asp Iie Thr Gin 1950 GGA CCA TCT GAG TCC TTT GTT GAT TTC GCC AAT CGG CTT ATA AAG GCG GTT GAG GGG TCA GAC CTC CCG CCT TCC GCG CGG GGT CCG Gly Pro Ser Giu Ser Phe Val Asp Phe Ala Asn Arg Leu Iie Lys Ala Val Giu Gly Ser Asp Leu Pro Pro Ser Ala Arg Ala Pro 200 GTG ATC ATT GAC TGC TTT ÅGG CAG AAG TCA CAG CCA GAT ATC CAG CAG CTT ATA CGG GCA GCA CCT CC CCA GTG CAC GGC CAG Val Iie Iie Asp Cys Phe Arg Gin Lys Ser Gin Pro Asp Iie Gin Gin Leu Iie Arg Ala Ala Pro Ser Thr Val His Giy Gin Ala 2100 GCA GCC GCC GCG ATG CCG CTC AGC GCC CC CC CAGC AAG AAC TAC GAT TAC GAC TAC GAC TCG GTG CAG CCC TAC TTC TAC TTC Ala Ala Ala Ala Met Pro Leu Ser Ala Ser Leu Pro Ser Lys Asn Tyr Asp Tyr Asp Tyr Asp Ser Val Gin Pro Tyr Phe Tyr Phe 2200 GAG GAG GAG GAG GAC ATC TCA CTG GCG GCG CAG CAG CAG CGG GGC GAC GAG CTG CAG CTC CCG CCC GCC CCG TCC GAG GAC ATC TGG AAG Glu Glu Glu Glu Gu Asn Phe Tyr Leu Ala Ala Gin Gin Arg Giy Ser Giu Leu Gin Pro Pro Ala Pro Ser Giu Asp Iie Trp Lys 2300 2350 CAG CTG CTG CCC ATG CCG CCC CTC TCG CCC AGC CGC GCC CC AGC CTG GCC GCC TCC TGC TTC CTTC ACC GCC GAC GTG GAG GAG ATG GTG ACG GAG CTG CTC GGG GGG GAC ATG GTC AAC CAG AGC TTC ATC TGC GAC GAC GAA TCC TTC GC ACC 410 Ala Ala Ala Ala Ala Ser Cys Phe Pro Ser Thr Ala Asp 2350 CAG CTG GAG ATG GTG ACG GAG CTG CTC GGG GGG GAC ATG GTC AAC CAG AGC TTC ATC TGC GAC CGA CAA TCC TTC GTC AAA Gin Leu Giu Met Val Thr Glu Leu Leu Gly Gly Asp Met Val Asn Gin Ser Phe Iie Cys Asp Pro Asp Asp Glu Ser Phe Val Lys 2550 TCC ATC ATC CAG GAC TGC ATG TG AGG GGC TTC TCC GCC GCC GCC AAG CTG GAG AAG GTG GTG TCG GAG AAG CTC GCC ACC TAC Ser Iie Iie Gin Asp Cys Met Trp Ser Giy Phe Ser Ala Ala Ala Ala Lys Leu Glu Lys Val Val Ser Glu Lys Leu Ala Thr Tyr 2550 CAA GCC TCC CGC CAG GAG GGG GGC CCC GCC TCC CGA CCC TCC CGC CCG CCC TCC CGC CCG CCC CCC	GTT GCT ATT ACG GCG TCG GCT	CTC CAG GCG TTT AG	GA GAA GTT GCC CGG	CTG GCG GAA CCT GCA GGT	CCA TGG GCG GAC ATC ACG CAG
GGA CCA TCT GAG TCC TTT GAT TTC GCC AAT CGG CTT ATA AAG GCG GTT GAG GGG TCA GAC CTC CCG CCT TCC GCG CGG GG GCT CCG GIY Pro Ser Glu Ser Phe Val Asp Phe Ala Asn Arg Leu Ile Lys Ala Val Glu Gly Ser Asp Leu Pro Pro Ser Ala Arg Ala Pro 2000 GTG ATC ATT GAC TGC TTT ÅGG CAG AAG TCA CAG CCA GAT ATC CAG CAG CTT ATA CGG GCA GCA CCC TCC ACA GTG CAC GGC CAG GCA Val Ile Ile Asp Cys Phe Arg Gln Lys Ser Gln Pro Asp Ile Gln Gln Leu Ile Arg Ala Ala Pro Ser Thr Val His Gly Gln Ala 2100 GCA GCC GCC GCG ATG CCG CTC AGC GCC AGC CTC CCC AGC AAG AAC TAC GAT TAC GAC TAC GAC TCG GTG CAG CCC TAC TTC TAC TTC Ala Ala Ala Ala Met Pro Leu Ser Ala Ser Leu Pro Ser Lys Asn Tyr Asp Tyr Asp Tyr Asp Ser Val Gln Pro Tyr Phe Tyr Phe 2200 GAG GAG GAG GAG GAG AAC TTC TAC CTG GCG GCG CAG CGA GCG GG GGC AGC GAC CGC GCC CCG TCC GAG GAC ATC TGG AAG Glu Glu Glu Glu Glu Glu Asn Phe Tyr Leu Ala Ala Gln Gln Arg Gly Ser Glu Leu Gln Pro Pro Ala Pro Ser Glu Asp Ile Trp Lys 2300 AAG TTT GAG CTC CTG CCC ATG CCG CCC CTC TCG CCC AGC CGC TCC AGC CGC TCC GCC GCC TCC TCC ACC GCC GAC CAG CTG GAG ATG GTG ACG GAG CTG CTC GGG GGG GAC ATG GTC AAC GAG CTG GCC GCC TCC TGC TTC CCT TCC ACC GCC GAC CAG CTG GAG ATG GTG ACG GAG CTG CTC GGG GGG GAC ATG GTC AAC GAG CTG CTG GCG GCC GCC GCC GAC GAC GAC GAC TCT TC GTC AAA Gln Leu Glu Met Val Thr Glu Leu Leu Gly Gly Asp Met Val Asn Gln Ser Phe Ile Cys Asp Pro Asp Asp Glu Ser Phe Val Lys 2250 TCC ATC ATC CAG GAC TGC ATG TGG AGC GGC TTC TCC GCC GCC GCC GCC AAG CTG GAG AAG GTG GTG TGG GAG AAG CTC GCC ACC TAC Ser Ile Ile Gln Asp Cys Met Trp Ser Gly Phe Ser Ala Ala Ala Ala Lys Leu Glu Lys Val Val Ser Glu Lys Leu Ala Thr Tyr CAA GCC TCC CGC CAG GAG GGG GCC CCC GCC TCC CGC CCC CCC CCC CC	Val Ala Ile Ihr Ala Ser Ala 1900	Leu Gin Ala Phe Ar	g Glu Val Ala Arg	Leu Ala Glu Pro Ala Gly 1950	Pro Trp Ala Asp Ile Thr Gln
GTG ATC ATT GAC TGC TTT ÅGG CAG AAG TCA CAG CCA GAT ATC CAG CAG CAG CTT ATA CGG GCA GCA CCC TCC ACA GTG CAC GGC CAG GCA GCA GCA GGC CAG GCA GCA	GGA CCA ICI GAG ICC III GII Gly Pro Ser Glu Ser Phe Val	Asp Phe Ala Asn Ar	G CII AIA AAG GCG °g Leu Ile Lys Ala	GTT GAG GGG TCA GAC CTC Val Glu Gly Ser Asp Leu	CCG CCT TCC GCG CGG GCT CCG Pro Pro Ser Ala Arg Ala Pro
Val lie lie Asp Cys Phe Arg Gin Lys Ser Gin Pro Asp Ile Gin Gin Leu Ile Arg Ala Ala Pro Ser Thr Val His Gly Gin Ala 2100 GCA GCC GCG GCG GCG GCG CTC AGC GCC AGC CTC CCC AGC AAG AAC TAC GAT TAC GAC TAC GAC TCG GTG CAG CCC TAC TTC TAC TTC Ala Ala Ala Ala Met Pro Leu Ser Ala Ser Leu Pro Ser Lys Asn Tyr Asp Tyr Asp Tyr Asp Ser Val Gin Pro Tyr Phe Tyr Phe 2200 AV Pst I GAG GAG GAG GAG GAG AAC TTC TAC CTG GCG GCG CAG CAG CGG GGC AGC GAG CTG CAG CTC CCC GCC CCG TCC GAG GAC ATC TGG AAG Glu Glu Glu Glu Glu Asn Phe Tyr Leu Ala Ala Gin Gin Arg Gly Ser Glu Leu Gin Pro Pro Ala Pro Ser Glu Asp Ile Trp Lys 2250 2250 AAG TTT GAG CTC CTG CCC ATG CCG CCC CTC TCG CCC AGC CGC CGC TCC AGC CTG GCC GCC TCC TGC TTC CTT CC ACC GCC GAC Lys Phe Glu Leu Leu Pro Met Pro Pro Leu Ser Pro Ser Arg Arg Ser Ser Leu Ala Ala Ala Ser Cys Phe Pro Ser Thr Ala Asp 2400 CAG CTG GAG ATG GTG ACG GAG CTG CTC GGG GGG GAC ATG GTC AAC CAG AGC TTC ATC TGC GAC CCG GAC GAC GAA TCC TTC GTC AAA Gin Leu Glu Met Val Thr Glu Leu Leu Gly Gly Asp Met Val Asn Gin Ser Phe Ile Cys Asp Pro Asp Asp Glu Ser Phe Val Lys 2450 TCC ATC ATC CAG GAC TGC ATG TGG AGC GGC TTC TCC GCC GCC GCC GCC AAG CTG GAG AAG CTC GCC ACC TAC Ser Ile Ile Gin Asp Cys Met Trp Ser Gly Phe Ser Ala Ala Ala Lys Leu Glu Lys Val Val Ser Glu Lys Leu Ala Thr Tyr CAA GCC TCC CGC CAG GAG GGG GGC CCC GCC GCC TCC CGA CCC GCC CCG CCG CCG CCG CCG CCG C	GTG ATC ATT GAC TGC TTT AGG	CAG AAG TCA CAG CO	CA GAT ATC CAG CAG	• •-myc-he CTT ATA CGG GCA GCA CCC	Iper virus junction TCC ACA GTG CAC GGC CAG GCA
GCA GCC GCC GCG ATG CCG CTC AGC GCC AGC CTC CCC AGC AAG AAC TAC GAT TAC GAC TAC GAC TCG GTG CAG CCC TAC TTC TAC TTC Ala Ala Ala Ala Ala Met Pro Leu Ser Ala Ser Leu Pro Ser Lys Asn Tyr Asp Tyr Asp Tyr Asp Ser Val Gln Pro Tyr Phe Tyr Phe 2200 AGG GAG GAG GAG GAG AAC TTC TAC CTG GCG GCG CAG CAG CAG CAG CGG GGC AGC GAG CTG CAG CCT CCC GCC CCG TCC GAG GAC ATC TGG AAG Glu Glu Glu Glu Glu Asn Phe Tyr Leu Ala Ala Gln Gln Arg Gly Ser Glu Leu Gln Pro Pro Ala Pro Ser Glu Asp Ile Trp Lys 2250 AAG TTT GAG CTC CTG CCC ATG CCG CCC CTC TCG CCC AGC CGC CGC CGC TCC AGC CTG GCC GCC TCC TGC TTC CCT TCC ACC GCC GAC Lys Phe Glu Leu Leu Pro Met Pro Pro Leu Ser Pro Ser Arg Arg Ser Ser Leu Ala Ala Ala Ser Cys Phe Pro Ser Thr Ala Asp 2350 CAG CTG GAG ATG GTG ACG GAG CTG CTC GGG GGG GAC ATG GTC AAC CAG AGC TTC ATC TGC GAC CCG GAC GAC GAA TCC TTC GTC AAA Gln Leu Glu Met Val Thr Glu Leu Leu Gly Gly Asp Met Val Asn Gln Ser Phe Ile Cys Asp Pro Asp Asp Glu Ser Phe Val Lys 2450 TCC ATC ATC CAG GAC TGC ATG CGC ATG GGG GGC TTC TCC GCC GCC GCC AAG CTG GAG AAG GTG GTG TCG GAG AAG CTC GCC ACC TAC Ser Ile Ile Ile Gln Asp Cys Met Trp Ser Gly Phe Ser Ala Ala Ala Lys Leu Glu Lys Val Val Ser Glu Lys Leu Ala Thr Tyr 2550 CAA GCC TCC CGC CAG GAG GGG GGC CCC GCC GCC GCC G	Val Ile Ile Asp Cys Phe Arg	Gln Lys Ser Gln Pr 2100	ro Asp Ile Gln Gln	Leu Ile Arg Ala Ala Pro	Ser Thr Val His Gly Gln Ala 2150
GAG GAG GAG GAG GAG AAC TTC TAC CTG GCG GCG CAG CAG CAG CAG CAG CAG CAG CA	GCA GCC GCC GCG AIG CCG CIC Ala Ala Ala Ala Met Pro Leu	AGC GCC AGC CTC CC Ser Ala Ser Leu Pr	CC AGC AAG AAC TAC o Ser Lys Asn Tyr .	GAT TAC GAC TAC GAC TCG Asp Tyr Asp Tyr Asp Ser	GTG CAG CCC TAC TTC TAC TTC Val Gln Pro Tyr Phe Tyr Phe
AG TTT GAG CTC CTG CCC ATG CCG CCC CTC TGG GGG GAC ATG GTC AAC CAG AGC TTC ATC TGC GAC CCG GAC GAC TTC ATC CCG GAC GAC GAC TTC ATC CCG GAC GAC GAC TTC GTC AAC CAG AGC TTC ATC TGC GAC CCG GAC GAC TTC ATC CCG CCAG CCG CCG CCC AGC CCC AGC CCC AGC CCC AGC CCC AGC CCC CC	GAG GAG GAG GAG GAG AAC TTC	TAC CTG GCG GCG CA	G CAG CGG GGC AGC	GAG CTG CAG CCT CCC GCC	CCG TCC GAG GAC ATC TGG AAG
And TIT GAG CTC CTG CCC ATG CCG CCC TCC TCG CCC AGC CGC CGC CGC CCC AGC CGC GCC GC	2250 V Sst I.	lyr Leu Ala Ala Gi	In Gin Arg Giy Ser	Glu Leu Gin Pro Pro Ala 2300	Pro Ser Glu Asp Ile Trp Lys
CAG CTG GAG ATG GTG ACG GAG CTG CTC GGG GGG GAC ATG GTC AAC CAG AGC TTC ATC TGC GAC CCG GAC GAC GAA TCC TTC GTC AAA Gin Leu Giu Met Val Thr Giu Leu Leu Giy Giy Asp Met Val Asn Gin Ser Phe Ile Cys Asp Pro Asp Asp Giu Ser Phe Val Lys TCC ATC ATC CAG GAC TGC ATG TGG AGC GGC TTC TCC GCC GCC GCC AAG CTG GAG AAG GTG GTG TCG GAG AAG CTC GCC ACC TAC Ser Ile Ile Gin Asp Cys Met Trp Ser Giy Phe Ser Ala Ala Ala Lys Leu Giu Lys Val Val Ser Giu Lys Leu Ala Thr Tyr CAA GCC TCC CGC CAG GAG GGG GGC CCC GCC GCC GCC TCC CGA CCC GCC CCG CCG CCG CCG CCG CCG C	Lys Phe Glu Leu Leu Pro Met	Pro Pro Leu Ser Pr	o Ser Arg Arg Ser	AGC CIG GCC GCC GCC ICC Ser Leu Ala Ala Ala Ser	IGC TTC CCT TCC ACC GCC GAC Cys Phe Pro Ser Thr Ala Asp
TCC ATC ATC ATC CAG GAC TGC ATG TGG AGC GGC GCC GCC GCC GCC GCC GCC GCC AAG CTG GAG AAG GTG GTG TCG GAG AAG CTC GCC ACC TAC Ser Ile Ile Gln Asp Cys Met Trp Ser Gly Phe Ser Ala Ala Ala Lys Leu Glu Lys Val Val Ser Glu Lys Leu Ala Thr Tyr CAA GCC TCC CGC CAG GAG GGG GGC CCC GCC GCC GCC G	CAG CTG GAG ATG GTG ACG GAG	CTG CTC GGG GGG GA	AC ATG GTC AAC CAG	AGC TTC ATC TGC GAC CCG	GAC GAC GAA TCC TTC GTC AAA
Ser Ile Ile Gin Asp Cys Met Trp Ser Giy Phe Ser Ala Ala Ala Ala Lys Leu Giu Lys Val Val Ser Giu Lys Leu Ala Thr Tyr CAA GCC TCC CGC CAG GAG GGG GGC CCC GCC GCC GCC G	TCC ATC ATC ATC CAC CAC TCC			AAC CTC CAC AAC CTC CTC	ASP ASP GIU SER Phe Val Lys
CAA GCC TCC CGC CAG GAG GGG GGC CCC GCC GCC GCC TCC CGA CCC GGC CCG CCG CCC TCG GGG CCG CCG CCT CCT	Ser Ile Ile Ile Gin Asp Cys	Met Trp Ser Gly Ph	ne Ser Ala Ala Ala Ala 2550	Lys Leu Glu Lys Val Val	Ser Glu Lys Leu Ala Thr Tyr
Gin Ala Ser Arg Gin Giu Giy Giy Pro Ala Ala Ala Ser Arg Pro Giy Pro Pro Pro Ser Giy Pro Pro Pro Pro Pro Ala Giy Pro	CAA GCC TCC CGC CAG GAG GGG Gln Ala Ser Arg Gln Glu Gly	GGC CCC GCC GCC GC Gly Pro Ala Ala Al	CC TCC CGA CCC GGC	CCG CCG CCC TCG GGG CCG Pro Pro Pro Ser Gly Pro	CCG CCT CCT CCC GCC GGC CCC Pro Pro Pro Pro Ala Gly Pro

#### Biochemistry: Reddy et al.

2650 2600 GCC GCC TCG GCC GGC CTC TAC CTG CAC GAC CTG GGA GCC GCG GCC GCC GAC TGC ATC GAC CCC TCG GTG GTC TTC CCC TAC CCG CTC Ala Ala Ser Ala Gly Leu Tyr Leu His Asp Leu Gly Ala Ala Ala Ala Asp Cys Ile Asp Pro Ser Val Val Phe Pro Tyr Pro Leu 2750 W Sal I Ser Glu Arg Ala Pro Arg Ala Ala Pro Pro Gly Ala Asn Pro Ala Ala Leu Leu Gly Val Asp Thr Pro Pro Thr Thr Ser Ser Asp WCla T 2850 ACA GAA GCÀ TCA GAG GAG CAC TGT AAG CCC CAC CAC AGT CCG CTG GTC CTC AAG CGG TGT CAC GTC AAC ATC CAC CAA CAC AAC TAC 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 Asn Tyr 2950 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 AGC AAC 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 Ser Asn 3100 AAC CGA AAA TGC TCC AGT CCC CGC ACG TTA GAC TCA GAG GAG AAC GAC AAG AGG CGA ACG CAC AAC GTC TTG GAG CGC CAG CGA AGG Asn Arg Lys Cys Ser Ser Pro Arg Thr Leu Asp Ser Glu Glu Asn Asp Lys Arg Arg Thr His Asn Val Leu Glu Arg Gln Arg Arg 3150 AAT GAG CTG AAG CTG CGT TTC TTT GCC CTG CGT GAC CAG ATA CCC GAG GTG GCC AAC AAC GAG AGG GCG CCC AAG GTT GTC ATC CTG Asn Glu Leu Lys Leu Arg Phe Phe Ala Leu Arg Asp Gln Ile Pro Glu Val Ala Asn Asn Glu Lys Ala Pro Lys Val Val Ile Leu 3250 AAA AAA GCC ACG GAG TAC GTT CTG TCT CTC CAA TCG GAC GAG CAC AAA CTG ATC GCA GAG AAA GAG CAG TTG AGG CGG AGG AGA GAA Lys Lys Ala Thr Glu Tyr Val Leu Ser Leu Gln Ser Asp Glu His Lys Leu Ile Ala Glu Lys Glu Gln Leu Arg Arg Arg Arg Glu CAG TTG AAA CAC AAC CTT GAG CAG CTA AGG AAC TCT CGT GCA TAG GAACTCTTGGACATCACTTAGAATACCCCCAAACTAGACTAGAACTATGATAAAAAT GIn Leu Lys His Asn Leu Giu Gin Leu Arg Asn Ser Arg Ala End 3450 3450 3500 ATTAGTGTTTCTAATATCACTCATGAACTACATCAGTCCATTGAGTATGGAACTATTGCAACTGCATGCTGTGCGACTTAACTTGAGACTACACAACCTTGGCCGAATCTCCGAA CGGTTTGGCCAGAACCTCAAAACTGCCTCATAATTGATACTTTGGGCATAAGGGATGATGGGACATTCTTCATGCTTGGGGATGAACTCTTCAACTTTTTTCTTTTAAAATTTTG ¥ v-myc-helper viral junction, start ∆ env TATTTAAGGCATT CCT GGT GGC CCT GAT AAC AGC ACA ACC CTC ACC TAT CGG AAG GTT TCG TGC TTG TTA AAG CTG AAC GTT TCT CT Pro Gly Gly Pro Asp Asn Ser Thr Thr Leu Thr Tyr Arg Lys Val Ser Cys Leu Leu Lys Leu Asn Val Ser Le Bam HI TTĂ GAC GAG CCA TCA GAA CTA CĂA CTA TTĂ GGT TCC CAG TCT CTC CCC ATT ATĂ ACT AAT ATT ACT CGG ATC C Leu Asp Glu Pro Ser Glu Leu Gln Leu Leu Gly Ser Gln Ser Leu Pro Ile Ile Thr Asn Ile Thr Arg Ile

FIG. 2. Nucleotide sequence of the MC29 proviral genome. The upper line shows the sequence proceeding in the 5' to 3' direction and has the same polarity on the MC29 genomic RNA. The amino acid sequence deduced from the open reading frames is given in the bottom line. The major structural features of the genome are indicated.

300 bases upstream from the v-myc helper-viral junction at the 3' end.

Examination of the amino acid sequence of the putative transforming protein revealed that the carboxyl terminus of the *v-myc*-encoded protein was highly hydrophilic, containing a large number of glutamic acid, arginine, and lysine residues, whereas the amino terminus of the *myc*-encoded protein was more hydrophobic. This observation is further confirmed by the computer analysis of the *v-myc*-encoded protein by using the method of Hopp and Woods (22). The two regions are joined by a stretch of sequence that contained a large number of proline residues, indicating that these two regions of the polypeptide are linked by a highly flexible region. This structure is reminiscent of the hinge region present in immunoglobulin molecule at the junction of the variable and constant regions. It is interesting to note that these two biochemically distinct domains of the *v-myc* protein are derived from two different exons of the *c-myc* gene (7).

As mentioned earlier, the v-myc gene contains 300 base pairs of noncoding sequences at the 3' end which did not contain any signals for termination of transcription or polyadenylylation of mRNAs. Comparison of our sequence presented here with that of c-myc (23) reveals that the polyadenylylation signal indeed occurs in the c-myc gene beyond the point of recombination. It is interesting to note that the recombination between the proviral and myc sequences occurs in the middle of envelope gene, deleting the first 75 codons, thus rendering it defective.

#### DISCUSSION

Nucleotide sequence analysis of the MC29 transforming region has revealed several important features of its molecular organization. Examination of the sequence data presented here reveals a single open reading frame on the viral RNA strand that could code for a protein of 875 amino acids with a molecular mass of 96,000 daltons.

Like many of the transforming retroviruses, MC29 appears to synthesize its transforming protein by means of a gag-myc polyprotein, the amino-terminal region of which is composed of helper virus gag gene products. In the case of MC29, the gag-myc hybrid protein contains the entire sequence of p19, p10, and the first 211 amino acids of p27 and 422 amino acids of v-myc. Thus, the transforming protein utilizes helper viral sequences for the initiation of its synthesis.

Like all other transforming genes of retroviruses, DNA sequences homologous to myc (termed c-myc) are found in normal chicken DNA (6-8). Thus, the viral oncogene represents a transduced cellular gene. Avian retroviruses of the MC29 group that carry v-myc genes cause an abnormally large variety of neoplasms, such as myelocytomas, endotheliomas, mesotheliomas, renal and hepatic carcinomas, and sarcomas (1, 2). MC29 virus also transforms both fibroblasts and macrophages in culture but not erythroblasts, even though p110 is synthesized in the infected erythroblasts (3-5). These experiments indicate that MC29 virus has certain target-cell specificity in spite of its wide spectrum of target cells. Furthermore, the disease spectrum can be restricted by creating deletions *in vitro* within the *onc* gene (24, 25). Such deletion mutants have been shown to transform fibroblasts but not macrophages *in vitro* (24, 25).

Lymphoid neoplasms are not usually caused by MC29 virus, yet the activation of c-myc by avian leukosis virus integration in the vicinity of the cellular gene is observed in a vast majority

C-MYC	CCCGTGTCCCCCTCCCGCCCGCAGGCAGCAGCCGCC	TTAAGGCATTTTTTCTTAGCGAGAATTCCAAATA
v-myc	GCACCCTCCACAGTGCACGGCCAGGCAGCAGCCGCC	TTAAGGCATTCCTGGTGGCCCTGATAACAGCACA
RSV	GCACCCTCCACGCTGACCACCCCAGAGAGATAATTA	AGCCAGCATTACCGGCGGCCCTGACAACAGCACA

FIG. 3. The 5' and 3' recombination junctions between v-myc and c-myc. The 5' and 3' boundaries between helper virus and c-myc leading to the formation of MC29 virus are shown. The vertical lines indicate the areas of sequence homology. The RSV gag and env sequences are from Schwartz et al. (17). Note the presence of 10 bases in v-myc at the 5' junction that has no apparent homology with RSV gag or c-myc.

of the bursal lymphomas induced by avian leukosis virus (26, 27). These observations led to the "downstream" promotion hypothesis that implicated the enhanced expression of c-muc protein in the transformation process induced by the avian leukosis virus. Recently, Collins and Groudine (28) and Dalla-Favera et al. (29) have reported an alternative mechanism for the increased transcription of c-myc gene in HL-60 cells. In these cells it appears that the increased transcription occurs as a result of amplification of c-myc genes. It is interesting to note that comparison of our v-myc sequence with that of c-myc (unpublished data) has revealed that v-myc in MC29 virus is an incomplete gene that has lost at least 300 base pairs of leader sequences at the amino-terminal region during the integration process. Therefore, it can be considered a deletion mutant of the normal c-myc gene and therefore has a decreased ability to transform lymphoid cells. The data on deletion mutants and revertants taken together suggest that mutations and deletions affect the tissue specificity of this onc gene.

Unlike all other known onc gene products, the p110 of MC29 is not a protein kinase and is not associated with cell membrane structure. Subcellular fractionation and immunofluorescence studies have shown that p110 is located in the nucleus of the transformed quail nonproducer cells (30, 31). Furthermore, purified p110 behaves as a DNA binding protein, suggesting that this protein may be involved in gene regulation. Because the putative amino acid sequence presented in Fig. 2 for v-myc region exhibits two biochemically distinct domains, it would be interesting to study the location of the sequences that are associated with the DNA binding properties of the molecule. Such studies may not only provide insights into the transformation process induced by MC29 but also shed light on the mechanisms involved in gene regulation of eukaryotic cells.

We are grateful to Richard Feldman for his help in computer analysis and Stuart Aaronson for helpful advice and support.

- Beard, J. W. (1980) in Viral Oncology, ed. Klein, G. (Raven, New 1. York), pp. 55-87.
- Mladenov, Z., Heine, U., Beard, D. & Beard, J. W. (1967) J. Natl. 2 Cancer Inst. 38, 251–285.
- Langlois, A. J., Sankaram, S., Hsiung, P.-H. L. & Beard, J. W. 3. (1967) J. Virol. 1, 1082-1084.
- Bolognesi, D. P., Langlois, A. J., Sverak, L., Bonar, R. A. & Beard, 4.
- J. S. (1968) J. Virol. 2, 576–586. Langlois, A. J., Fritz, R. B., Heine, U., Beard, D., Bolognesi, D. P. & Beard, J. W. (1969) Cancer Res. 29, 2056–2074. 5.

- 6. Mellon, P., Pawson, A., Bister, K., Martin, G. S. & Duesberg, P. H. (1978) Proc. Natl. Acad. Sci. USA 75, 5874-5878.
- Robins, T., Bister, K., Garon, C., Papas, T. & Duesberg, P. (1982) 7. I. Virol. 41, 635-642.
- Sheiness, D., Fanshier, L. & Bishop, J. M. (1978) J. Virol. 28, 600-8. 610.
- Bister, K., Hayman, M. J. & Vogt, P. K. (1977) Virology 82, 431-9. 448.
- Lautenberger, J. A., Schultz, R. A., Garon, C. F., Tsichlis, P. N. & Papas, T. S. (1981) Proc. Natl. Acad. Sci. USA 78, 1518-1522. 10.
- Maxam, A. & Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 11. 560-564
- Roychoudhury, R. & Wu, R. (1980) Methods Enzymol. 65, 43-62. 12
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. 13. Sci. USA 74, 5463-5468.
- Smith, H. O. & Birnstiel, M. L. (1976) Nucleic Acids Res. 5, 4537-14. 4545
- Varmus, H. E. (1982) Science 216, 812-820. 15.
- Swanstrum, R., Varmus, H. E. & Bishop, J. H. (1982) J. Virol. 41, 16. 535-541
- Schwartz, D., Tizard, R. & Gilbert, W. (1982) in Molecular Bi-17. ology of Tumor Viruses: RNA Tumor Viruses, eds. Weiss, R., Teich, N., Varmus, H. & Coffin, J. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 1340-1348.
- Maizel, J. V., Jr., & Lenk, R. P. (1981) Proc. Natl. Acad. Sci. USA 18. 78, 7665-7669.
- Rushlow, K., Lautenberger, J., Reddy, E. P. & Papas, T. (1982) 19. I. Virol. 42, 840–846
- Devare, S. G., Reddy, E. P., Law, J. D. & Aaronson, S. A. (1982) 20. J. Virol. 42, 1108–1113.
- Harada, F., Sawyer, R. C. & Dahlberg, J. E. (1975) J. Biol. Chem. 21. 250, 3487-3497
- Hopp, T. P. & Woods, K. R. (1981) Proc. Natl. Acad. Sci. USA 78, 22. 3824-3828
- 23. Watson, D. K., Reddy, E. P., Duesberg, P. H. & Papas, T. S. (1983) Proc. Natl. Acad. Sci. USA 80, 2146-2150.
- Bister, K., Ramsay, G. M. & Hayman, M. J. (1982) J. Virol. 41, 24.
- 754-766. Ramsay, G. M. & Hayman, M. J. (1982) J. Virol. 41, 745-753. 25.
- Hayward, W. S., Neel, B. G. & Astrin, S. M. (1981) Nature (Lon-26.
- don) 290, 475-480. Payne, G. S., Bishop, J. M. & Varmus, H. E. (1982) Nature (Lon-27.
- don) 295, 209-214. 28 Collins, S. J. & Groudine, M. (1982) Nature (London) 298, 679-
- 682 Dalla-Favera, R., Wong-Staal, F. & Gallo, R. C. (1982) Nature 29. (London) 299, 61-63.
- Donner, P., Greiser-Wilkie, I. & Moelling, K. (1982) Nature 30 (London) 296, 262-266.
- Abrams, H. D., Rohrschneider, L. R. & Eisenman, R. N. (1982) 31. Cell 29, 427-439.