lateral convexity of the lower shaft, and (iv) the anterior concavity of the upper portion of the shaft.
40. G. L. Issac, *Sci. Am.* **238**, 90 (April 1978); in *Emergence of Man*, J. Z. Young, E. M. Jope, E. P. Oakley, Eds. (Royal Society of London, London, 1981), pp. 177–188; R. Van Gelder, in *Early Hominids of Africa*, C. J. Jolly, Ed. (St. Martin's Press, New York, 1978), pp. 431–439.
41. J. T. Stern and R. L. Susman, *Am. J. Phys. Anthropol.* **57**, 232 (1982).

42. We thank M. Leakey and R. E. F. Leakey for allowing us to study fossils from Olduvai Gorge and East Turkana. We also thank F. J. Muruka of Kenyatta National Hospital and the staff of the Louis Leakey Memorial Institute for their hospitality. Supported by the University Awards Committee of the State University of New York and by NSF grants BNS 7924070, BNS 81119664, and BNS 7924162.

29 March 1982; revised 8 June 1982

Nucleotide Sequence of the p21 Transforming Protein of Harvey Murine Sarcoma Virus

Abstract. *Harvey murine sarcoma virus is a retrovirus which transforms cells by means of a single virally encoded protein called p21* has. *We have determined the nucleotide sequence of 1.0 kilobase in the 5′ half of the viral genome which encompasses the* has *coding sequences and its associated regulatory signals. The nucleotide sequence has identified the amino acid sequence of two additional overlapping polypeptides which share their reading frames and the carboxyl termini with p21 but which contain additional NH_2-terminal amino acids.*

The Harvey murine sarcoma virus (Ha-MuSV) is a replication-defective transforming retrovirus (*1*). The 5.5-kilobase genome contains a rat-derived sequence that is required for Ha-MuSV–mediated cellular transformation (*2, 3*). This conserved gene encodes a 21,000 dalton phosphoprotein [p21 *has*, formerly known as p21 Ha *ras* (*1*)] which has guanine nucleotide binding activity (*4*). The cellular rat DNA sequences coding for p21 *has* are represented by two genes, c-*has* I, which is composed of four exons and three introns, and c-*has* II, which is colinear (by heteroduplex analysis) with the viral *has* sequences (*5*). The p21 encoded by c-*has* I is closely related to, but distinguishable from, viral p21 *has* (v-*has*) (*5, 6*). Another related transforming retrovirus, Kirsten murine sarcoma virus (Ki-MuSV), also encodes a p21 which cross-reacts serologically with the p21 of Ha-MuSV; approximately two-thirds of the tryptic peptides of Ha and Ki viral p21 protein are the same (*7*). We have derived the nucleotide sequence for the transforming gene of Ha-MuSV from which the amino acid sequence of p21 *has* has been deduced. Oroszlan *et al.* have confirmed (*8*) the deduced amino acid sequence by direct sequencing of many peptides labeled with tritiated amino acids. Sequences important in the transcription and translation of this gene have been identified. Moreover, the structure of the gene p21 *has* has been compared (*9*) to the Ki-MuSV transforming gene (v-*kis*); (formerly known as Ki *ras*) (*7*).

The region between the restriction endonuclease cleavage sites Acc I and Pst I of Ha-MuSV for which we have obtained the nucleotide sequence is shown in Fig. 1. A larger fragment, extending from a Kpn I restriction endonuclease site 0.3 kilobase upstream from Acc I down to the same 3′ end (Pst I), previously had been cloned within the late region of SV40 (*10*). In studies on the messenger RNA (mRNA) transcribed from the chimeric SV40–Ha-MuSV molecule, the 5′ terminus of v-*has* mRNA was mapped to a location 160 bases upstream from the Hind III site in Fig. 1 (*10*). Accordingly, we analyzed the nucleotide sequences in this fragment for sequences related to the Goldberg-Hogness box TATAAA (*11*) (T, thymine; A, adenine). There are three such sets of sequences, around positions 45, 137, and 165. On the basis that the 5′ ends of RNA transcripts generally lie approximately 25 nucleotides downstream from the 3′ end of the consensus sequences (*11*), the putative 5′ ends of the transcripts (by RNA polymerase II) would map at approximately positions 70, 162, and 195. This is consistent with the approximate location observed for the primary transcript from the SV40–Ha-MuSV recombinant (*10*).

The sequence in Fig. 2 shows a fourth putative RNA polymerase II start site at position 16 downstream from a TATAAA box located around position 45; however, this transcript would be from the antisense strand. If this fourth RNA

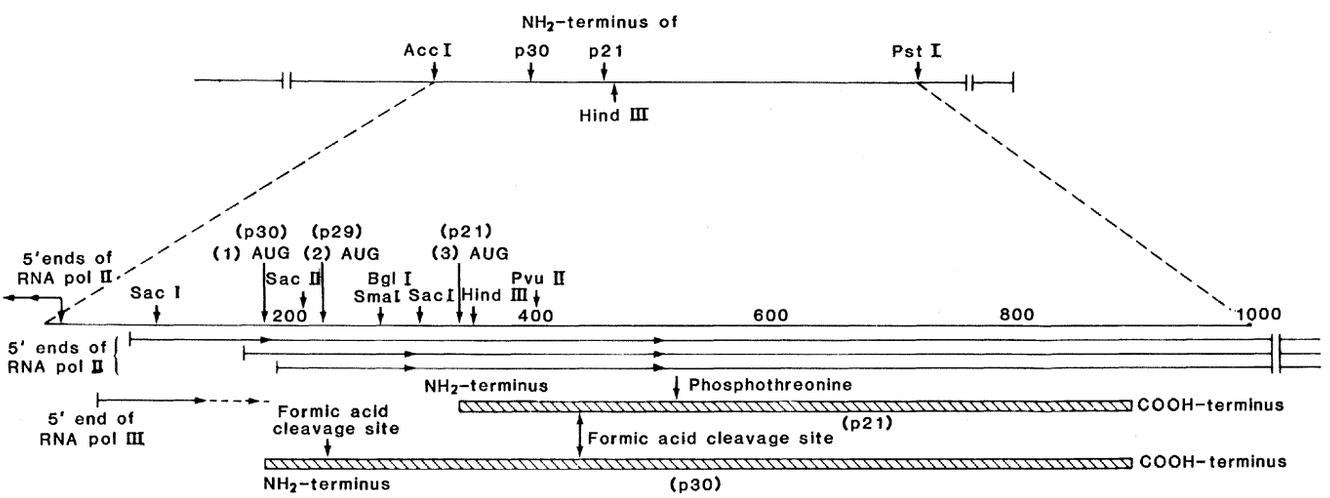

Fig. 1. Schematic representation of the 5′ half of the Ha-MuSV genome. The upper line denotes that part of the 5′ genomic half which contains the p21 *has* sequences. The locations of the NH_2-termini of p30 and p21 as well as of key restriction sites are marked. The region between the restriction sites Acc I and Pst I—map positions 3650 and 4710, respectively, as defined in (*2*)—is enlarged below. The numbers 0 to 1000 refer to the nucleotide positions in Fig. 2, and key restriction sites (Sac I, Sac II, Bgl I, Sma I, Hind III, Pvu II) are placed on this scale. The numbers (1), (2), and (3) next to AUG represent the initiation sites for the coding sequences of p30, p29, and p21, respectively. The putative 5′ ends and 5′ to 3′ directionality of transcripts initiated by RNA polymerase (pol) II and III are indicated. The hachured boxes represent the location of the p21 and p30 coding sequences; these proteins have a common COOH-terminus but distinct NH_2-termini. The locations of the formic acid cleavage sites and of the phosphorylation site (phosphothreonine) are denoted on the coding sequences by arrows.

0036-8075/82/0903-0934$01.00/0

polymerase II start site did exist in vivo, it would not be able to code for any polypeptide, as the mRNA would have translational termination codons in all three reading frames downstream from position 16 (*12*). Another important transcriptional feature was the localization of a potential RNA polymerase III promotor. The first control region is generally located about ten nucleotides downstream from the 5′ end of the RNA having the concensus sequence of 5′-GUGGPyNNPuGUGG-3 (G, guanine; U, uracil; Py, a pyrimidine; Pu, a purine; and N, a nucleotide), and the second is about 50 nucleotides downstream from the 5′ end, with a consensus sequence of 5′-GGGUUCGAANCC-3′ (C, cytosine) (*13, 14*). The sequence between positions 53 and 64 in Ha-MuSV (5′-GUGUUUUGGGGG) has 10 out of 12 nucleotides homologous to the first (consensus) control region and the sequence between positions 113 to 124 (GGGUUCGGUCCC) has 10 out of 12 nucleotides homologous to the second (consensus) control region. This would place the 5′ end of a putative RNA polymerase III transcript around position 45 to 47 (Fig. 2), a location consistent with published data (*15*). We still cannot assess the physiological significance of this sequence in determining transcripts within the Ha-MuSV genome in vivo. Another important feature of this sequence is the presence of a short imperfect direct repeat sequence (15 out of 17 are identical): 5′-AAAACA-TAGTGTTTT-3′ and 5′-AAAACATA-TAGTGTTT-3′ located at 46 to 59 and 159 to 175 (Fig. 2), respectively. This repeat sequence overlaps the promotor sequences of the two RNA polymerase II starts.

The three potential AUG initiation codons are in the same reading frame (Figs. 2 and 3). Moreover, this reading frame is open over a stretch of 725 nucleotides in which the other two reading frames contain numerous stop codons. On this basis, amino acids have been assigned to the sequence between nucleotides 185 and 910. From the above transcriptional information, it appears that the v-*has* sequences (Fig. 2) have at least three potential translational initiation codons, located at amino acid positions −53, −39, and +1. This suggests that three distinct polypeptides may be specified by this viral gene. These polypeptides would be of sizes 30,000 (p30, initiating at amino acid position −53), 29,000 (p29, initiating at −39), and 21,000 daltons (p21, initiating at +1). Although studies of Ha-MuSV–transformed cells have indicated that p21 is the only protein definitely encoded by the viral genome, a minor species of p30 was found in addition to the major p21 species in African green monkey kidney cells infected with the SV40–Ha-MuSV recombinant (*10*). These two proteins (p30 and p21) shared most of their tryptic peptide maps (*16*). The third polypeptide (p29), which has a putative initiation codon at position −39, has yet to be detected. The nucleotide sequence in Fig. 2 suggests the presence of two formic acid cleavage sites (formic acid, which cleaves between aspartic acid and proline residues, should cleave twice within p30 and only once within p21), at amino acids −36 and +34. This prediction has been confirmed by the cleavage of [^{35}S]methionine labeled p30 and p21 with 70 percent formic acid (*8*).

It generally is recognized that in eukaryotic translational systems the first of several successive AUG codons is preferentially recognized by ribosomes for the initiation of translation (*17*). Our data suggest that the ATG at position 185 (Fig. 2) initiates p30, the one at position 224 initiates p29, and the one at position 341 initiates p21. The fact that the latter is downstream to the first two suggests that, if initiation of translation could occur internally, then all three RNA molecules (with 5′ ends located around positions 70, 165, and 195) could potentially encode p21, thereby explaining the relative abundance of p21 over p30 in transformed cells.

We have considered two possibilities for the translational controls. First, on theoretical grounds the AUG signals are

```
                                                          |-----------|
     GCTCTAGTGGCAGTGTGTTGGTTGATAGCCAAAGTTAATTTTTAAAACATA[GTGTTTTGGGG]GTTGGGGATTTA   (74)

GCTCAGTGATAGAGCTCTTGCCTAGCACGCAAGCCCT[GGGTTCGGTCCC]CCCAGCTCTGAAAAAAAGGAAAGAGAACAA  (153)
                                                                              -41
  |-----------------|                    met pro ala ala arg ala ala pro ala ala asp glu
AACAAAAACATATAGTGTTTTATCTGTGCTT ATG CCC GCA GCC CGA GCC GCA CCC GCC GCG GAC GAG  (220)
-40                                                                           -21
pro met arg asp pro val ala pro val arg ala pro ala leu pro arg pro ala pro gly
CCC ATG CGC GAC CCA GTC GCA CCC GTC CGC GCC CCC GCC CTG CCC CGC CCC GCC CCG GGG  (280)
-20                                                                            -1
ala val ala pro ala ser gly gly ala arg ala pro gly leu ala ala pro val glu ala
GCA GTC GCG CCA GCA AGC GGT GGG GCA AGA GCT CCT GGT TTG GCA GCC CCT GTA GAA GCG  (340)
  1                                                                            20
met thr glu tyr lys leu val val val gly ala arg gly val gly lys ser ala leu thr
ATG ACA GAA TAC AAG CTT GTG GTG GTG GGC GCT AGA GGC GTG GGA AAG AGT GCC CTG ACC  (400)
 21              ↑HindIII                                                      40
ile gln leu ile gln asn his phe val asp glu tyr asp pro thr ile glu asp ser tyr
ATC CAG CTG ATC CAG AAC CAT TTT GTG GAC GAG TAT GAT CCC ACT ATA GAG GAC TCC TAC  (460)
 41                                                                            60
arg lys gln val val ile asp gly glu thr cys leu leu asp ile leu asp thr[thr]gly
CGG AAA CAG GTA GTC ATT GAT GGG GAG ACG TGT TTA CTG GAC ATC TTA GAC ACA[ACA]GGT  (520)
 61                                                                            80
gln glu glu tyr ser ala met arg asp gln tyr met arg thr gly glu gly phe leu cys
CAA GAA GAG TAT AGT GCC ATG CGG GAC CAG TAC ATG CGC ACA GGG GAG GGC TTC CTC TGT  (580)
 81                                                                           100
val phe ala ile asn asn thr lys ser phe glu asp ile his gln tyr arg glu gln ile
GTA TTT GCC ATC AAC AAC ACC AAG TCC TTT GAA GAC ATC CAT CAG TAC AGG GAG CAG ATC  (640)
101                                                                           120
lys arg val lys asp ser asp asp val pro met val leu val gly asn lys cys asp leu
AAG CGG GTG AAA GAT TCA GAT GAT GTG CCA ATG GTG CTG GTG GGC AAC AAG TGT GAC CTG  (700)
121                                                                           140
ala gly arg thr val glu ser arg gln ala gln asp leu ala arg ser tyr gly ile pro
GCT GGT CGC ACT GTT GAG TCT CGG CAG GCC CAG GAC CTT GCT CGC AGC TAT GGC ATC CCC  (760)
141                                                                           160
tyr ile glu thr ser ala lys thr arg gln gly val glu asp ala phe tyr thr leu val
TAC ATT GAA ACA TCA GCC AAG ACC CGG CAG GGT GTA GAG GAT GCC TTC TAC ACA CTA GTA  (820)
161                                                                           180
arg glu ile arg gln his lys leu arg lys leu asn pro pro asp glu ser gly pro gly
CGT GAG ATT CGG CAG CAT AAA CTG CGG AAA CTG AAC CCG CCT GAT GAG AGT GGC CCT GGC  (880)
181                                                                           200
cys met ser cys lys cys val leu ser ter
TGC ATG AGC TGC AAG TGT GTG CTG TCC TGA CACCAGGTGAGGCAGGGACCAGCAAGACATCTGGGGCAG  (949)

TGGCCTCAGCTAGCCAGATGAACTTCATATCCACTTTGATGTCGCTCG                                 (997)
```

Fig. 2. The nucleotide sequence of the v-*has* gene. The nucleotide sequence, determined from both the strands (*26, 27*), corresponds to the enlarged portion of Fig. 1. This sequence does not include approximately 20 nucleotides adjacent to both the Acc I and Pst I restriction sites. The bracketed numbers (74 to 997) at the right of each line refer to the nucleotide sequence. From the single large open reading frame and ATG triplets in this sequence, we have deduced an amino acid sequence between nucleotides 185 and 910. The numbers at the right and left of each line above the amino acids refer to this deduced amino acid sequence. The positions of the NH_2-terminal amino acids for p30, p29, and p21 are −53, −39, and +1, respectively. The dashes underline the putative Hogness-Goldberg box (*11*). The two sequences within boxes represent the first and second sets of concensus sequences for RNA polymerase III (*13, 14*). The solid bars highlight the repeat sequence described in the text. The phosphothreonine residue at amino acid position +59 is marked by a solid box.

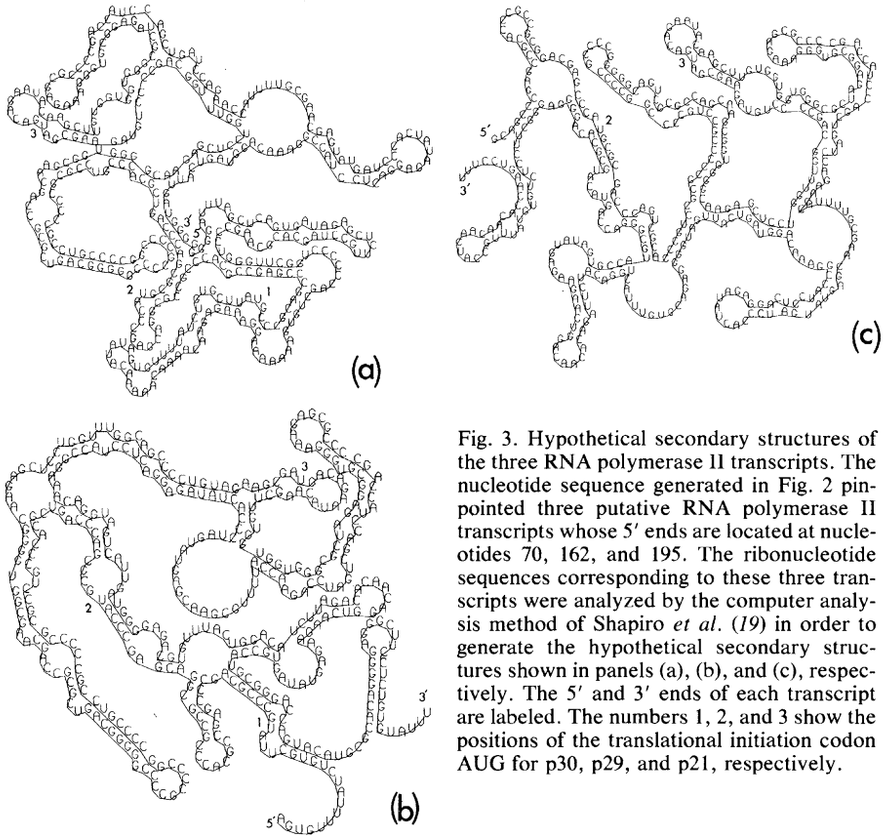

Fig. 3. Hypothetical secondary structures of the three RNA polymerase II transcripts. The nucleotide sequence generated in Fig. 2 pinpointed three putative RNA polymerase II transcripts whose 5′ ends are located at nucleotides 70, 162, and 195. The ribonucleotide sequences corresponding to these three transcripts were analyzed by the computer analysis method of Shapiro *et al.* (*19*) in order to generate the hypothetical secondary structures shown in panels (a), (b), and (c), respectively. The 5′ and 3′ ends of each transcript are labeled. The numbers 1, 2, and 3 show the positions of the translational initiation codon AUG for p30, p29, and p21, respectively.

preceded at 341 by GCG, at 224 by CCC, and at 185 by CTT. Only the sequence at position 341 would fit the model for translation proposed by Kozak (*18*) whereby translation is initiated at the sequence PuNNAUG. Second, the secondary structure of the three different RNA species putatively transcribed by RNA polymerase II can be predicted by computer analysis of the most stable conformation of the RNA's (*19*). These computer models would predict the availability of any particular AUG codon for the initiation of translation. The AUG codon at position 341 is in a less stable configuration within all three RNA structures (Fig. 3). This would render the AUG codon into an open loop configuration so that the p21 translational initiation codon might be recognized in all three different RNA molecules. Nevertheless, we have obtained evidence (*20*) that only information downstream from the p21 AUG codon at position 341 is required for cell transformation; a DNA fragment from Sac I (positions 309 to 314) to Pst I (positions 1017 to 1022) cloned downstream to the Ha-MuSV long terminal repeat (LTR) sequences is capable of efficient cellular transformation and p21 expression.

In the three models, the first AUG codon (position 185 initiating p30) is in an open configuration while the second AUG codon (position 224, initiating p29) is in a closed configuration. This might explain the abundance of p30 and the absence of p29 in Ha-MuSV transformed cells (*1*). The p21 protein is synthesized as a precursor which is cleaved after translation (*21*) and phosphorylated at a threonine residue at amino acid position +59 (*22*). The posttranslational cleavage occurs at the carboxyl terminus, as deduced by formic acid cleavage of both the precursor and the mature p21 polypeptide. The nature of the processing at the carboxyl terminus of the polypeptide has not been determined.

We have compared the amino acid and the nucleotide sequence of v-*kis* (*9*) to v-*has*. The NH_2-terminal amino acids of both p21 *has* and p21 *kis* are Met-Thr-Glu-Tyr-Lys-Leu-Val-Val-Val-Gly-Ala (*23*). The NH_2-terminal region half of the polypeptides shows very extensive homology (110 of 120 amino acids). There is very little homology (3 of 22 amino acids) at the respective COOH-termini. Nevertheless, overall the p21's share approximatedly 81 percent of their amino acids. Interestingly, 19 of the 33 changes between the two are conservative (with the exception of the final 22 amino acids of the polypeptides, 12 of 16 changes are conservative). Furthermore, of the 168 nucleotide changes, 105 do not alter the specified amino acid. This highlights the strong evolutionary conservation pressure upon the polypeptide and suggests that the two proteins may transform cells through similar mechanisms. Nevertheless, the fact that these two distinct p21-coding genes have been conserved in a wide range of species (*7, 24*) suggests that their physiological functions, although perhaps similar, may be distinct. The final 22 amino acids of the polypeptides may mediate major structural changes in the proteins that cause v-*has* and v-*kis* to have distinct functions. Alternatively, the regulation of expression of the two genes may be distinct. In this connection, p21 *has* expression has been found to be associated with human bladder carcinoma (*24*) and p21 *kis* with human colon and lung carcinomas (*25*). Having available the *has* and *kis* coding sequences will facilitate structure-function studies through site-directed mutagenesis.

RAVI DHAR
Laboratory of Molecular Virology, National Cancer Institute, Bethesda, Maryland 20205
RONALD W. ELLIS
THOMAS Y. SHIH
Laboratory of Tumor Virus Genetics, National Cancer Institute
STEPHEN OROSZLAN
Biological Carcinogenesis Program, Frederick Cancer Research Facility, Frederick, Maryland 21701
BRUCE SHAPIRO
Laboratory of Pathology, Image Processing Section, National Cancer Institute
JACOB MAIZEL
Laboratory of Molecular Genetics, National Institute of Child Health and Development, Bethesda, Maryland 20205
DOUGLAS LOWY
Laboratory of Dermatology, National Cancer Institute
EDWARD SCOLNICK
Laboratory of Tumor Virus Genetics, National Cancer Institute

References and Notes

1. R. W. Ellis, R. D. Lowy, E. M. Scolnick, *Advances in Viral Oncology*, vol. 1, *Cell-derived Oncogenes*, G. Klein, Ed. (Raven, New York, in press).
2. R. W. Ellis *et al.*, *J. Virol.* **36**, 408 (1980).
3. T. Y. Shih, M. O. Weeks, H. A. Young, E. M. Scolnick, *Virology* **96**, 64 (1979).
4. T. Y. Shih, A. G. Papageorge, P. E. Stokes, M. O. Weeks, E. M. Scolnick, *Nature (London)* **287**, 686 (1980).
5. D. DeFeo, M. A. Gonda, H. A. Young, E. H. Chang, D. R. Lowy, E. M. Scolnick, R. W. Ellis, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3328 (1981).
6. H. Langbeheim, T. Y. Shih, E. M. Scolnick, *Virology* **106**, 292 (1980).
7. R. W. Ellis *et al.*, *Nature (London)* **292**, 506 (1981).
8. Ta. A. Oroszlan *et al.*, in preparation.
9. N. Tsuchida, T. Ryder, E. Ohtsubo, *Science* **217**, 937 (1982).
10. P. Gruss, R. W. Ellis, T. Y. Shih, M. Koenig, E. M. Scolnick, G. Khoury, *Nature (London)* **293**, 486 (1981).
11. P. Goldberg, thesis, Stanford University (1979).

12. R. Dhar *et al.*, unpublished data.
13. D. F. Bogenhagen, S. Sakonju, D. D. Brown, *Cell* **19**, 27 (1980).
14. D. M. Fowlkes and T. Shenk, *ibid.* **22**, 405 (1980).
15. M. C. Ostrowski, D. Berard, G. L. Hager, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4485 (1981).
16. T. Y. Shih *et al.*, in preparation.
17. M. Kozak, *Curr. Top. Microbiol. Immunol.* **93**, 81 (1981).
18. ———, *J. Virol.*, in press.
19. B. Shapiro, J. Maizel, L. Lipkin, *Proceedings of the 4th Annual WAMI Meeting*, April 1981, p. 93.
20. B. Williamson *et al.*, unpublished data.
21. T. Y. Shih, M. O. Weeks, S. Oroszlan, R. Dhar, E. M. Scolnick, *J. Virol.*, in press.
22. T. Y. Shih, P. E. Stokes, G. W. Smythers, R. Dhar, S. Oroszlan, *J. Biol. Chem.*, in press.
23. Abbreviations for amino acid residues: Ala, alanine; Arg, arginine; Asp, aspartic acid; Asn, asparagine; Cys, cysteine; Glu, glutamic acid; Gln, glutamine; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Tyr, tyrosine; Trp, tryptophan; Val, valine.
24. E. H. Chang, M. A. Gonda, R. W. Ellis, E. M. Scolnick, D. Lowy, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
25. M. Barbacid, G. Cooper, R. Weinberg, M. Wigler, personal communications.
26. I. Seif, G. Khoury, R. Dhar, *Nucleic Acids Res.* **8**, 2225 (1982).
27. A. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
28. We thank George Khoury for reviewing the manuscript.

5 April 1982; revised 8 June 1982

Nucleotide Sequence of the Oncogene Encoding the p21 Transforming Protein of Kirsten Murine Sarcoma Virus

Abstract. *The transforming protein of Kirsten murine sarcoma virus (Ki-MuSV) is a virally encoded 21-kilodalton protein called p21* kis. *The sequences encoding p21* kis *were genetically localized to a 1.3-kilobase segment near the 5′ end of the viral genome by assaying the capacity of a series of defined deletion mutants of molecularly cloned Ki-MuSV DNA to induce focal transformation of mouse cells. Nucleotide sequencing of a portion of this region has led to the identification of an open reading frame of 567 nucleotides coding for p21* kis *protein.*

There is accumulating evidence that homologs of the viral *onc* (v-*onc*) genes of the transforming sequences of retroviruses present in the cellular genome (*1*) are incorporated into the viral genome of replication-competent leukemia viruses (*2, 3*). This process often involves substitution of a portion of the leukemia virus genome with the cellular *onc* (c-*onc*) gene (*4*). Thus, the mammalian transforming viruses are defective in their replication. Kirsten murine sarcoma virus (Ki-MuSV) (*5*) and Harvey murine sarcoma virus (Ha-MuSV) (*6*) are two such replication-defective transforming viruses. Each of the transforming sequences, derived from rat cellular DNA, was incorporated into its respective replication-competent murine leukemia virus (MuLV) genome while the MuLV's had been passaged in rats (*2*). The transforming proteins of Ki-MuSV and Ha-MuSV are virally encoded 21-kilodalton proteins called p21 *kis* and p21 *has*, respectively. The p21's of both viruses share certain antigenic determinants and the capacity to bind guanine nucleotides. However, molecular hybridization studies suggest that the p21 coding sequences of Ki-MuSV and Ha-MuSV were derived from different c-*onc* genes. We now report the genetic mapping of the transforming region of the Ki-MuSV genome and the complete nucleotide sequence of the open reading frame that encodes p21 *kis* viral protein.

The viral RNA genomes of the rat-derived Ki-MuSV [6.5 kilobases (kb) in length] and of Ha-MuSV (5.5 kb) (*7, 8*) are both composed of three distinct types of nucleotide sequences: (i) sequences homologous to MuLV, which are located in the first 0.2 kb at the 5′ end

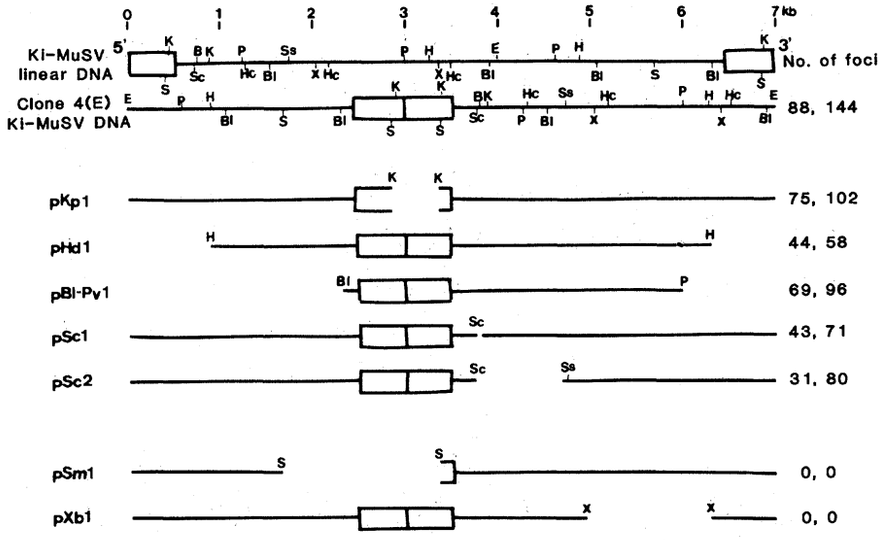

Fig. 1. Focus-forming activities of Ki-MuSV DNA clone 4(E) and deletion mutants derived from it. Recombinant DNA (1.5 μg) was used to transfect BALB/3T3 cells on duplicate plates, and foci were counted as described (*7*). The restriction enzyme sites on Ki-MuSV DNA are shown for Eco RI (E), Kpn I (K), Sma I (S), Sac I (Sc), Bam HI (B), Hinc II (HC), Pvu II (P), Bal I (Bl), Sst II (Ss), Xba I (X), and Hind III (H). The box represents the LTR sequence of Ki-MuSV linear DNA. The left end and the right end of the Ki-MuSV DNA correspond to 5′ and 3′ sides of the viral RNA, respectively. Preparation of recombinant DNA, digestion with restriction enzymes, purification of DNA fragments, ligation, transformation, and selection of *Escherichia coli* carrying a plasmid DNA were as described (*7*). For pKp1, clone 4(E) DNA, which is a circularly permuted form of the viral DNA genome cloned at the Eco RI site in pBR322, was digested with Kpn I, heated at 70°C for 5 minutes, and recircularized. The ligated DNA's were then used to transform *E. coli* χ1776. For pHd1, clone 4(E) DNA was digested with Hind III, and the resulting 5.4-kb fragment was ligated to pBR322 DNA, which had been digested with Hind III and treated with bacterial alkaline phosphatase (BAP) (*16*). The ligated DNA was used to transform *E. coli*, and a bacterial clone was selected by colony hybridization (*17*) with a ^{32}P-labeled Ki-MuSV RNA probe, prepared as described (*7*). For pBl–Pv1, clone 4(E) DNA was digested with Bal I and the 2.25-kb Bal I fragment containing two LTR's was purified and ligated to pBR322 DNA, which had been digested with Bal I and BAP. When this step was completed, DNA's were prepared from each clone and digested with Pvu II. A recombinant clone that produced 0.9- and 5.8-kb fragments after digestion was selected. The selected clone was digested with Pvu II, and the large fragment was isolated, treated with BAP, and ligated to the 1.7-kb Pvu II fragment that had been digested from clone 4(E) DNA. After transformation, a recombinant DNA clone was selected which produced 1.45-, 2.25-, and 3.75-kb fragments after Bal I restriction. For pSc1 and pSc2, clone 4(E) DNA was digested with Sac I, heated at 60°C for 5 minutes, and treated with 250 units of S1 nuclease (BRL) at 37°C for 10 minutes in 30 m*M* sodium acetate buffer (*p*H 4.6) containing 50 m*M* NaCl, 1 m*M* $ZnSO_4$ and 5 percent glycerol. DNA was extracted with a mixture of phenol, chloroform, and isoamyl alcohol, precipitated with ethanol, self-ligated, and used to transform *E. coli*. Bacterial clones that carried plasmid DNA were selected, and the plasmid DNA's were prepared. The pSc1 and pSc2 recombinant DNA's could not be restricted by Sac I, but produced Ki-MuSV DNA fragments of 7.0 and 6.15 kb, respectively, after the digestion with Eco RI (*18*). For pSm1, clone 4(E) DNA was digested with Xma I (which cleaves CCC GGG, as does Sma I) heated at 70°C for 5 minutes, circularized, and used to transform *E. coli*. For pXb1, clone 4(E) DNA was digested with Xba I, heated at 70°C for 5 minutes, recircularized, and used to transform *E. coli*.

0036-8075/82/0903-0937$01.00/0