JOURNAL OF VIROLOGY, Mar. 1985, p. 984–987
0022-538X/85/030984-04$02.00/0

Vol. 53, No. 3

Nucleotide Sequence Analysis of the BALB/c Murine Sarcoma Virus Transforming Gene

E. PREMKUMAR REDDY,[1†] DAVID LIPMAN,[2] PHILIP R. ANDERSEN,[3] STEVEN R. TRONICK,[1*] AND STUART A. AARONSON[1]

Laboratory of Cellular and Molecular Biology, National Cancer Institute,[1] and National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases,[2] Bethesda, Maryland 20205, and Abbott Laboratories, North Chicago, Illinois 60064[3]

Received 18 September 1984/Accepted 16 November 1984

We determined the nucleotide sequence of the v-H-*ras*-related oncogene of BALB/c murine sarcoma virus. This oncogene contains an open reading frame of 189 amino acids that initiates and terminates entirely within the mouse cell-derived *ras* sequence. The protein encoded by this open reading frame matches the sequence predicted for the T24 human bladder carcinoma oncogene product, p21, in all but two positions. The presence of a lysine residue in position 12 of BALB/c murine sarcoma virus p21 likely accounts for its oncogenic properties.

BALB/c murine sarcoma virus (BALB-MSV) is a replication-defective transforming retrovirus of mouse origin (1, 15) that was isolated from a spontaneously occurring BALB/c mouse tumor. Recent studies have demonstrated that BALB-MSV arose by recombination between a murine leukemia virus and a mouse cell DNA sequence (4). Molecular cloning and detailed structural analysis of the proviral genome of BALB-MSV have revealed that large portions of helper virus *pol* and *env* genes were deleted as a result of this recombination and were replaced by a stretch of cellular sequences ca. 0.6 kilobase long (4). The molecularly cloned BALB-MSV cell-derived sequence, v-*bas*, exhibited a high degree of sequence homology with the transforming gene (H-*ras*) of Harvey murine sarcoma virus, as determined by heteroduplex and Southern blot hybridization analyses (3). The importance of the *ras* family of oncogenes has been highlighted by recent discoveries that the majority of the dominant transforming genes isolated from human tumors are closely related to these retroviral oncogenes (7, 10, 14, 16, 24, 26). In fact, a series of recent advances in this area has demonstrated that single point mutations in either of two specific codons frequently lead to the activation of *ras* proto-oncogenes as oncogenes (5, 6, 18, 19, 25, 27–29, 35, 36).

In an effort to understand the structural relationships between the oncogenes of BALB-MSV and Harvey murine sarcoma virus and their human cellular counterparts, we carried out primary nucleotide sequence analysis of these genes. In this paper we present the nucleotide sequence of the BALB-MSV transforming gene.

Molecular cloning and preliminary characterization of the BALB-MSV proviral genome were described earlier (4). The clone (λ9EB-18) contained the entire proviral genome and adjacent host cellular sequences. Portions of the proviral genome were subcloned in the plasmid vector pBR322 and used for sequencing. One subclone, designated pHBI (4), contained almost the entire v-*bas* sequence. Another subclone used contained the viral sequences immediately upstream of pHBI, located between the two *Hin*dIII sites (2.9 kilobase pairs [kbp] to 4.4 kbp on the BALB-MSV map [Fig. 1]).

To map in greater detail the regions of the helper virus that underwent deletions or substitutions or both during the recombinational events that generated BALB-MSV, we carried out heteroduplex analysis between the BALB-MSV proviral genome and that of a mouse helper virus, Rauscher murine leukemia virus (R-MuLV) (11), which is closely related to the natural helper virus present in the original BALB-MSV isolate (R. P. Reddy, unpublished data). A representative heteroduplex structure is shown in Fig. 2. Such an analysis revealed that the two viral genomes shared 3.7 ± 0.2 kbp of homologous sequences at their 5′ ends. This region of homology represents the long terminal repeat, the entire *gag* sequence, and the amino-terminal region of the polymerase gene of the R-MuLV genome (11). Following this region, a deletion loop of ca. 1.7 ± 0.2 kbp was observed and was deduced to represent the polymerase gene sequences of the helper virus deleted in BALB-MSV. This feature was followed by a second region of homology 0.62 ± 0.05 kbp long which could have been derived from either the carboxy-terminal region of the polymerase gene of R-MuLV or the amino-terminal region of the envelope gene of R-MuLV or both. Downstream of this, a sequence substitution bubble was observed. From the restriction maps of R-MuLV and BALB-MSV, we could conclude that 0.85 ± 0.1 kbp of *env* sequences were replaced by the 0.70 ± 0.1-kbp v-*bas* of host cellular origin. At the 3′ end, the substitution was followed by a homologous region of 2.2 ± 0.2 kbp which must represent the carboxy-terminal region of the *env* gene and the 3′ long terminal repeat. Thus, the generation of BALB-MSV involved a large deletion in the helper virus *pol* gene and a substitution of the *env* gene with v-*bas*.

The nucleotide sequence of the transforming region of BALB-MSV is shown in Fig. 3. To precisely localize the points of recombination, we compared the sequences shown in Fig. 3 with those of AKR murine leukemia virus (12) by using the computer program developed by Wilbur and Lipman (34). This analysis revealed that the BALB-MSV sequence (Fig. 3) was identical to the AKR murine leukemia virus genome from positions 1 through 44 at the 5′ end and from position 750 to the end.

The region from positions 45 through 749, corresponding to the location of v-*bas*, was not significantly related to

* Corresponding author.
† Present address: Department of Molecular Genetics, Roche Research Center, Nutley, NJ 07110.

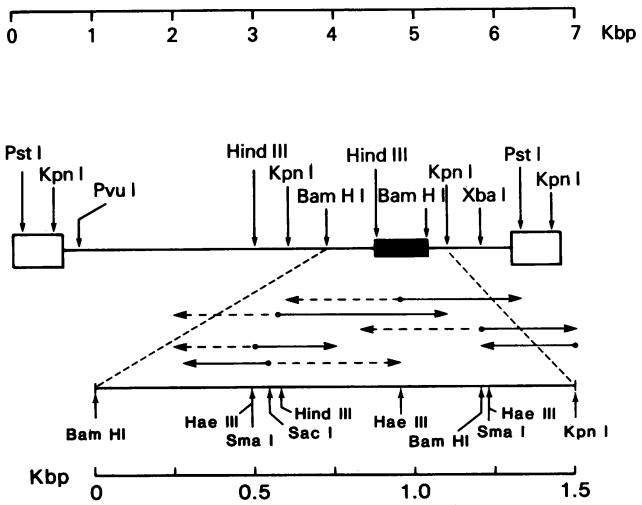

FIG. 1. Restriction enzyme map of BALB-MSV and sequencing strategy. The map shows the helper virus-derived (open boxes [long terminal repeats] and thin solid line) and cell-derived (black box) sequences. Shown beneath this map is the strategy used for the nucleotide sequence analysis; the solid and broken arrows represent regions sequenced on the sense and complementary strands, respectively.

murine leukemia virus sequences. The sequence comparison also demonstrated that the host-derived sequences (positions 45 through 749) underwent recombination with the helper virus such that they were flanked by the *pol* gene carboxy-terminal sequences at the 5′ end and by the *env* gene carboxy-terminal sequences at the 3′ end. The recombination resulted in the deletion of 63 codons from the 3′ end of the *pol* gene and a deletion of the amino-terminal 735 nucleotides coding for the first 245 amino acids of the *env* gene. However, the putative acceptor splice signal used in the generation of subgenomic *env* mRNA was retained. It is possible that a similar splicing mechanism may be used to generate the mRNA of the BALB-MSV transforming gene. Examination of the cell-derived sequences upstream of the first ATG codon did not reveal any apparent promoter-like sequences in this region, suggesting transcription via a spliced mRNA.

The transforming gene product of BALB-MSV has been identified as a 21,000-dalton protein (3), and examination of the sequence revealed the presence of an open reading frame starting at positions 144 through 146 and extending for 567 nucleotides, that could encode a protein of 189 amino acids with a molecular weight of ca. 21,000. The predicted amino acid sequence for this protein is shown in Fig. 3.

The high degree (97%) of nucleotide sequence similarlity between the v-*bas* open reading frame and that of v-H-*ras* (9) formally established that v-*bas* indeed arose from the mouse homologue of the H-*ras*-1 proto-oncogene. Furthermore, there were only three differences in the predicted p21 amino acid sequences of v-*bas* and v-H-*ras*; these were, respectively, lysine versus arginine at position 12, alanine versus threonine at position 59, and lysine versus glutamic acid at position 143.

Recent studies with several acute transforming viruses have revealed that the proto-oncogenes have undergone extensive deletions in their amino- or carboxy-terminal regions or both during recombination with the helper virus. Thus, in the case of Moloney murine sarcoma virus (21, 32) and simian sarcoma virus (8), it was demonstrated that the amino-terminal regions of the *mos* and *sis* sequences, respectively, have been replaced by *env*-derived sequences. In the case of MC29 virus (2, 20, 33) and Abelson murine leukemia virus (22) genomes, the 5′-terminal sequences have been replaced by *gag* gene sequences. With avian myeloblastosis virus (13, 23) and Finkel-Biskis-Jenkins murine osteosarcoma virus (31), the carboxy-terminal regions of c-*myb* and c-*fos*, respectively, have been substituted for by sequences derived from the helper virus. The present studies indicate that, in contrast, BALB-MSV has incorporated the entire coding region of the c-H-*ras*-1 (mouse) gene. This appears to be the case with all members of the v-*ras* family (9, 17, 30). In this respect, this family of retroviruses differs from all other known transforming viruses analyzed to date.

The human *ras* proto-oncogenes have been shown to be activated as transforming genes in human tumors by single point mutations within codon 12 or codon 61 that result in the replacement of the normally occurring glycine (codon 12) or glutamine (codon 61) residues by other amino acids. In positions 12 and 61 of the predicted amino acid sequence of BALB-MSV p21, lysine and glutamine, respectively, are

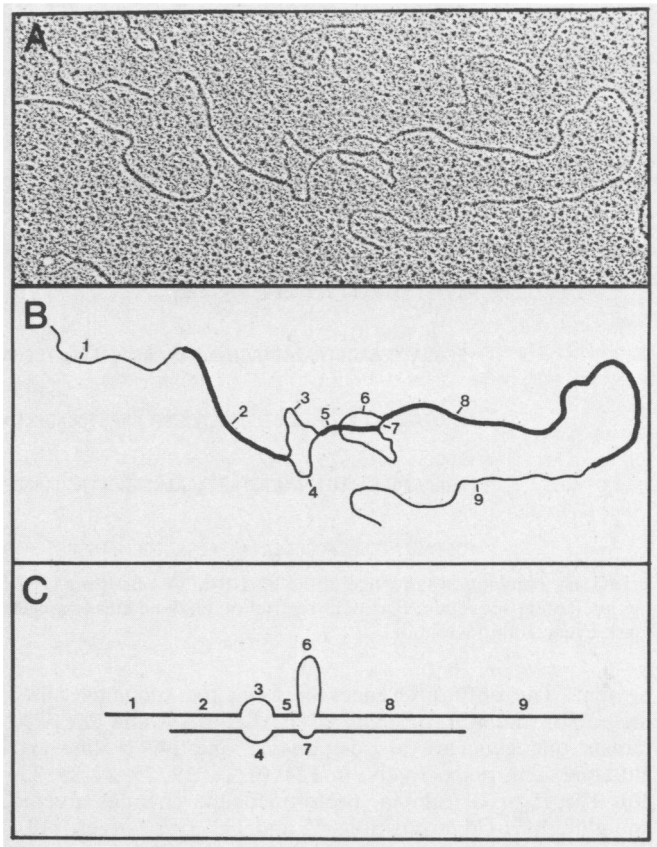

FIG. 2. Electron micrograph of (A) the heteroduplex formed between BALB-MSV and R-MuLV. A diagrammatic representation (B) and interpretive sketch (C) are shown below the micrograph. In panel B, the heavy lines represent double strands and the light lines represent single strands. In panel C, the top and bottom lines represent R-MuLV DNA and BALB-MSV DNA, respectively. Features 1 (1,475 ± 196 base pairs) and 9 (2,798 ± 300 base pairs) are cellular sequences flanking the R-MuLV proviral DNA. Contour lengths of other features are as follows (in base pairs): 2, 1,659 ± 183; 3, 857 ± 100; 4, 712 ± 80; 5, 615 ± 40; 6, 1,670 ± 183; 7, 82 ± 23; and 8, 3,518 ± 214. Features 1, 2, 3, 5, 6, 8, and 9 represent R-MuLV; features 2, 4, 5, 7, and 8 represent BALB/MSV. Transcriptional orientation of the molecules (5′ → 3′) is from right to left.

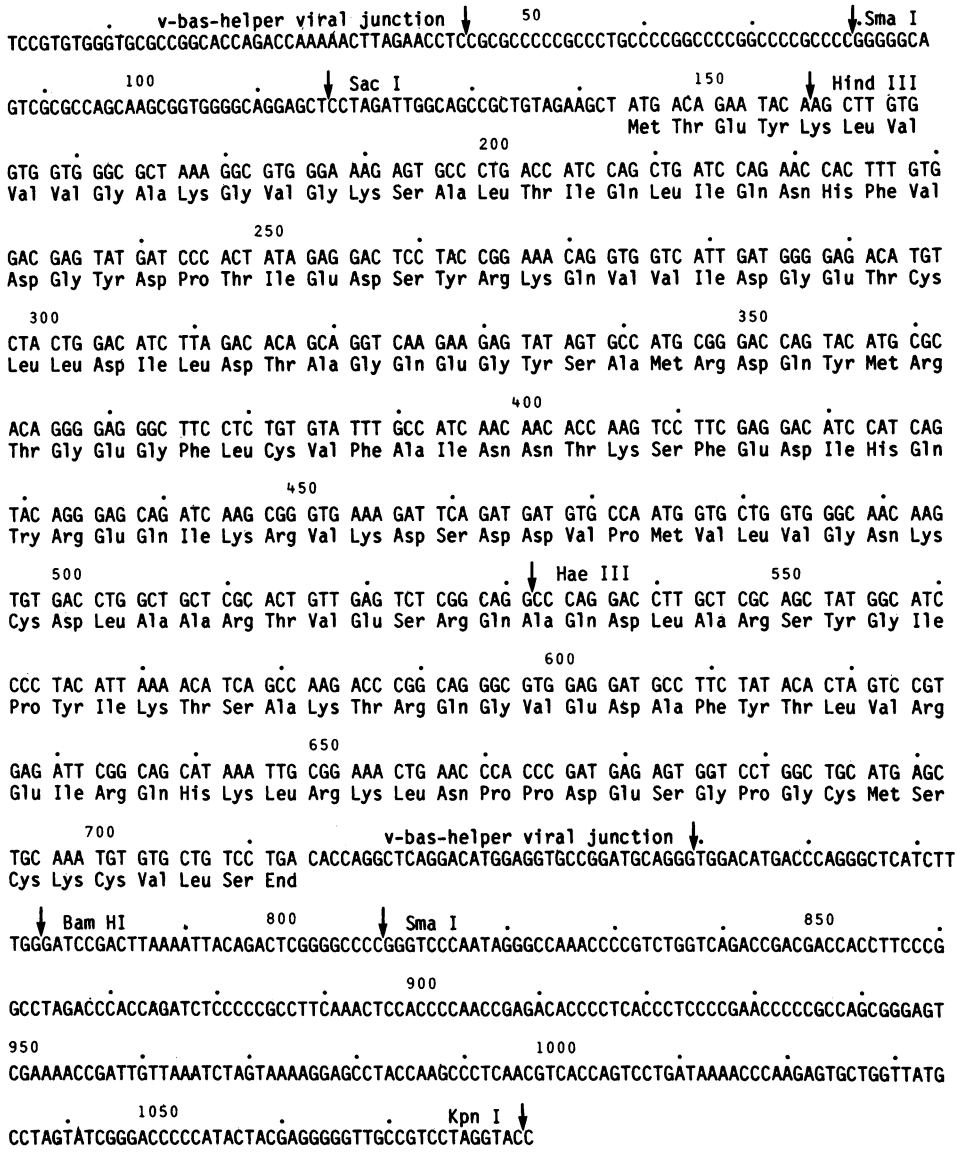

FIG. 3. Nucleotide sequence of the BALB-MSV oncogene. The amino acid sequence predicted for the oncogene product is shown below the nucleotide sequence. Pertinent restriction enzyme cleavage sites are shown, as are the v-*onc*–helper virus junctions (see text). The dots mark every 10th nucleotide.

present. The only differences between the sequences predicted for the p21 proteins of BALB-MSV and the T24 human oncogene are at positions 12 and 143 (valine and glutamic acid, respectively, in T24) (6, 18, 19, 25, 27–29, 35, 36). The H-*ras*-1 (human) proto-oncogene encodes glycine and glutamic acid at positions 12 and 143, respectively (18). The nucleotide sequence of H-*ras*-1 (mouse) has not yet been determined. Thus, whether the lysine at position 143 in the BALB-MSV transforming gene plays any role in its activation is not known. However, extensive evidence for the acquisition of transforming properties by *ras* oncogenes as a result of mutations at position 12 strongly implies that this lesion alone is responsible for the oncogenic activity of BALB-MSV p21.

LITERATURE CITED

1. **Aaronson, S. A., and M. Barbacid.** 1978. Origin and biological properties of a new BALB/c mouse sarcoma virus. J. Virol. **27**:366–373.
2. **Alitalo, K., J. M. Bishop, D. H. Smith, E. Y. Chen, W. W. Colby, and A. D. Levinson.** 1983. Nucleotide sequence of the v-myc oncogene of avian retrovirus MC29. Proc. Natl. Acad. Sci. U.S.A. **80**:100–104.
3. **Andersen, P. R., S. G. Devare, S. R. Tronick, R. W. Ellis, S. A. Aaronson, and E.M. Scolnick.** 1981. Generation of BALB-MSV and HA-MuSV by type C virus transduction of homologous transforming genes from different species. Cell **26**:129–134.
4. **Andersen, P. R., S. R. Tronick, and S. A. Aaronson.** 1981. Structural organization and biological activity of molecular clones of the integrated genome of a BALB/c mouse sarcoma virus. J. Virol. **40**:431–439.
5. **Capon, D. J., E. Y. Chen, A. D. Levinson, P. H. Seeburg, and D. V. Goeddel.** 1983. Complete nucleotide sequences of the T24 human bladder carcinoma oncogene and its normal homologue. Nature (London) **302**:33–37.
6. **Capon, D. J., P. H. Seeburg, J. P. McGrath, J. S. Hayflick, U. Edman, A. D. Levinson, and D. V. Goeddel.** 1983. Activation of Ki-*ras* 2 gene in human colon and lung carcinomas by two

different point mutations. Nature (London) **304**:507–513.
7. **Der, C. J., T. G. Krontiris, and G. M. Cooper.** 1982. Transforming genes of human bladder and lung carcinoma cell lines are homologous to the *ras* genes of Harvey and Kirsten sarcoma viruses. Proc. Natl. Acad. Sci. U.S.A. **11**:3637–3640.
8. **Devare, S. G., E. P. Reddy, J. D. Law, K. C. Robbins, and S. A. Aaronson.** 1983. Nucleotide sequence of the simian sarcoma virus genome: demonstration that its acquired cellular sequences encode the transforming gene product $p28^{sis}$. Proc. Natl. Acad. Sci. U.S.A. **80**:731–735.
9. **Dhar, R., R. W. Ellis, T. Y. Shih, S. Oroszlan, B. Shapiro, J. Maizel, D. Lowy, and E. Scolnick.** 1982. Nucleotide sequence of the p21 transforming protein of Harvey murine sarcoma virus. Science **217**:934–936.
10. **Eva, A., S. R. Tronick, R. A. Gol, J. H. Pierce, and S. A. Aaronson.** 1983. Transforming genes of human hematopoietic tumors: frequent detection of ras-related oncogenes whose activation appears to be independent of tumor phenotype. Proc. Natl. Acad. Sci. U.S.A. **80**:4926–4930.
11. **Habara, A., E. P. Reddy, and S. A. Aaronson.** 1982. Rauscher murine leukemia virus: molecular cloning of infectious integrated proviral DNA. J. Virol. **44**:731–735.
12. **Herr, W., V. Corbin, and W. Gilbert.** 1982. Nucleotide sequence of the 3′ half of AKV. Nucleic Acids Res. **10**:6931–6944.
13. **Klempnauer, K. H., T. J. Gonda, and M. Bishop.** 1982. Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-myb: the architecture of a transduced oncogene. Cell **31**:453–463.
14. **Parada, L. F., C. J. Tabin, C. Shih, and R. A. Weinberg.** 1982. Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma virus *ras* gene. Nature (London) **297**:474–478.
15. **Peters, R. L., L. S. Rabstein, S. Louise, R. Van Vleck, G. J. Kelloff, and R. J. Huebner.** 1974. Naturally occurring sarcoma virus of the BALB/cCr mouse. J. Natl. Cancer Inst. **53**:1725–1729.
16. **Pulciani, S., E. Santos, A. V. Lauver, L. K. Long, S. A. Aaronson, and M. Barbacid.** 1982. Oncogenes in solid human tumors. Nature (London) **300**:539–542.
17. **Rasheed, S., G. L. Norman, and G. Heidecker.** 1983. Nucleotide sequence of the Rasheed rat sarcoma virus oncogene: new mutations. Science **221**:155–157.
18. **Reddy, E. P.** 1983. Nucleotide sequence analysis of the T24 human bladder carcinoma oncogene. Science **220**:1061–1063.
19. **Reddy, E. P., R. K. Reynolds, E. Santos, and M. Barbacid.** 1982. A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. Nature (London) **300**:149–152.
20. **Reddy, E. P., R. K. Reynolds, D. K. Watson, R. A. Schultz, J. Lautenberger, and T. S. Papas.** 1983. Nucleotide sequence analysis of the proviral genome of avian myelocytomatosis virus (MC29). Proc. Natl. Acad. Sci. U.S.A. **80**:2500–2504.
21. **Reddy, E. P., M. J. Smith, and S. A. Aaronson.** 1981. Complete nucleotide sequence and organization of the Moloney murine sarcoma virus genome. Science **214**:445–450.
22. **Reddy, E. P., M. J. Smith, and A. Srinivasan.** 1983. Nucleotide sequence of Abelson murine leukemia virus genome: structural similarity of its transforming gene product to other *onc* gene products with tyrosine-specific kinase activity. Proc. Natl. Acad. Sci. U.S.A. **80**:3623–3627.
23. **Rushlow, K. E., J. A. Lautenberger, T. S. Papas, M. A. Baluda, B. Perbal, J. G. Chirikjian, and E. P. Reddy.** 1982. Nucleotide sequence of the transforming gene of avian myeloblastosis virus. Science **216**:1421–1423.
24. **Santos, E., S. R. Tronick, S. A. Aaronson, S. Pulciani, and M. Barbacid.** 1982. T24 human bladder carcinoma oncogene is an activated form of the normal human homologue of BALB- and Harvey-MSV transforming genes. Nature (London) **298**: 343–347.
25. **Shimizu, K., D. Birnbaum, M. A. Ruley, O. Fasano, Y. Suard, L. Edlund, E. Taparowsky, M. Goldfarb, and M. Wigler.** 1983. Structure of the Ki-*ras* gene of the human lung carcinoma cell line Calu-1. Nature (London) **304**:497–500.
26. **Shimizu, K., M. Goldfarb, M. Perucho, and M. Wigler.** 1983. Isolation and preliminary characterization of the transforming gene of a human neuroblastoma cell line. Proc. Natl. Acad. Sci. U.S.A. **80**:383–387.
27. **Tabin, C. J., S. M. Bradley, C. I. Bargmann, R. A. Weinberg, A. G. Papageorge, E. M. Scolnick, R. Dhar, D. R. Lowy, and E. H. Chang.** 1982. Mechanism of activation of a human oncogene. Nature (London) **300**:143–149.
28. **Taparowsky, E., K. Shimizu, M. Goldfarb, and M. Wigler.** 1983. Structure and activation of the human N-ras gene. Cell **34**:581–586.
29. **Taparowsky, E., Y. Suard, O. Fasano, K. Shimizu, M. Goldfarb, and M. Wigler.** 1982. Activation of the T24 bladder carcinoma transforming gene is linked to a single amino acid change. Nature (London) **300**:762–765.
30. **Tsuchida, N., T. Ryder, and E. Ohtsubo.** 1982. Nucleotide sequence of the oncogene encoding the p21 transforming protein of Kirsten murine sarcoma virus. Science **217**:937–938.
31. **Van Beveren, C., F. van Straaten, T. Curran, R. Muller, and I. M. Verma.** 1983. Analysis of FBJ-MuSV provirus and c-*fos* (mouse) gene reveals that viral and cellular *fos* gene products have different carboxy termini. Cell **32**:1241–1255.
32. **Van Beveren, C., F. van Straaten, J. A. Galleshaw, and I. M. Verma.** 1981. Nucleotide sequence of the genome of a murine sarcoma virus. Cell **27**:97–108.
33. **Watson, D. K., E. P. Reddy, P. H. Duesberg, and T. S. Papas.** 1983. 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, delta *gag-myc*. Proc. Natl. Acad. Sci. U.S.A. **80**: 2146–2150.
34. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. U.S.A. **80**:726–730.
35. **Yuasa, Y., R. A. Gol, A. Chang, I. M. Chiu, E. P. Reddy, S. R. Tronick, and S. A. Aaronson.** 1984. Mechanism of activation of an N-*ras* oncogene of SW-1271 human lung carcinoma cells. Proc. Natl. Acad. Sci. U.S.A. **81**:3670–3674.
36. **Yuasa, Y., S. K. Srivastava, C. Y. Dunn, J. S. Rhim, E. P. Reddy, and S. A. Aaronson.** 1983. Acquisition of transforming properties by alternative point mutations within c-*bas/has* human proto-oncogene. Nature (London) **303**:775–779.