Nucleotide sequence of testis-derived c-*abl* cDNAs: Implications for testis-specific transcription and *abl* oncogene activation

(Abelson virus/B-cell lymphoma/protooncogene/protein-tyrosine kinase)

CRISTINA OPPI*, SCOTT K. SHORE†, AND E. PREMKUMAR REDDY*‡

*Department of Molecular Oncology, Roche Institute of Molecular Biology, Roche Research Center, Nutley, NJ 07110; and †Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104

ABSTRACT   The c-*abl* gene codes for a protein-tyrosine kinase and is expressed in most examined murine cell types as two distinct mRNA species of 5.5 kilobases (kb) and 6.5 kb. In mouse testis, an additional species of 4.0 kb is expressed in very high levels. To study the interrelationship between various c-*abl* transcripts and to compare their sequence with the v-*abl* transcript, we prepared c-*abl*-specific cDNA clones from mouse testis and determined the complete nucleotide sequence of the 4.0-kb cDNA that appears to be the reverse transcript of the testis-specific mRNA. In addition, we have determined the 3′ sequence of an additional clone derived from the larger mRNA species that is expressed in somatic as well as germ-line cells. These cDNA sequences have been compared with the v-*abl* sequences to understand the mechanism of activation of this oncogene. The results demonstrate that (*i*) testis-specific c-*abl* mRNAs arise as a result of 3′ truncation, and (*ii*) the v-*abl* gene has arisen from its cellular homologue as a result of an extensive deletional/mutational process.

Abelson murine leukemia virus (Ab-MuLV) arose by recombination of nondefective helper virus [Moloney murine leukemia virus (Mo-MuLV)] sequences with cellular sequences present within the normal mouse genome (1–3). The latter sequences, termed *abl*, code for the transforming properties of this virus. In addition to its activation by viral transduction, the *ABL* gene is also activated in human chronic myelogenous leukemia, which is characterized by a reciprocal translocation between chromosomes 9 and 22 (4–6). This translocation results in the synthesis of an altered protein where the amino-terminal sequences of *ABL* are replaced by *BCR* gene sequences (4–7).

A comparative study of retroviral oncogenes and their cellular homologues has given important information regarding the structural changes that protooncogenes often undergo to acquire transforming potential. Toward this goal, we had earlier reported cloning and sequence analysis of the Ab-MuLV proviral genome (2, 8). To compare the structure of this viral oncogene with its normal cellular counterpart, we have undertaken the cDNA cloning and sequencing of the c-*abl*-specific mRNAs. The c-*abl* gene appears to be expressed in most examined murine cell types as two distinct mRNA species of 5.5 kilobases (kb) and 6.5 kb (9). In mouse testis, an additional species of 4.0 kb is expressed in very high levels along with the 5.5- and 6.5-kb species (10–12). To study the interrelationships between the various c-*abl* transcripts and to compare their sequence with the v-*abl* transcript, we prepared c-*abl*-specific cDNA clones from mouse testis and determined the complete nucleotide sequence of the 4.0-kb cDNA§ that appears to be the reverse transcript of testis-specific mRNA. In addition we have determined the 3′ sequence of an additional cDNA clone that is derived from

the larger mRNA species (5.5- to 6.5-kb size)§ that are present in somatic as well as germ-line cells. These cDNA sequences have been compared with the v-*abl* sequences to understand the mechanisms involved in the activation of this oncogene.

MATERIALS AND METHODS

RNA Preparation and cDNA Library Construction. RNA was prepared from adult mouse testes or from mouse thymuses by the method of Auffray and Rougeon (13). Poly(A)-containing RNA was selected by two cycles of purification on oligo(dT)-cellulose. Electrophoretic analysis of the RNA was carried out as described (14). The cDNA library was constructed as described previously (15). Fifty nanograms of cDNA was ligated to 1 μg of λ GT10 arms to obtain a phage library of 100,000 recombinants. Screening of the library was as described by Benton and Davis (16).

DNA Sequence Analysis. DNA sequence analysis was done by the method of Maxam and Gilbert (17).

RESULTS

Cloning of the Testis-Specific c-*abl* cDNA. The source of the mRNA used for the construction of the cDNA library was adult mouse testis (10–12). RNA blot analysis showed this preparation to contain three predominant species of 4.0-, 5.5- and 6.5-kb *abl*-specific mRNAs, whereas the RNA isolated from the thymus lacked the 4.0-kb species (Fig. 1). The cDNA preparation from testis was size selected on low-melting agarose gels, and only the DNA fragments that were longer than 2.5-kilobase pairs (kbp) were used for ligation to λ GT10 arms. Screening a library of 100,000 recombinants yielded 14 positive clones. To test whether 3′ heterogeneity could account for the difference in size between the 4.0-kb and larger mRNAs, we analyzed the cDNA clones for the presence of *Hind*III and *Bam*HI sites, which are known to be present towards the 3′ end of the v-*abl* genome. These results indicated that 12 of these positive clones did not contain these two restriction sites, whereas the other two clones did contain them. Fig. 2 shows a restriction map of two representative cDNA clones. Sequence analysis of the two clones containing the *Bam*HI and *Hind*III sites and four other clones lacking these sites revealed that all six clones contained a poly(A) tail at the 3′ end; the priming of the reverse transcription in all these cases was thus indicated to start
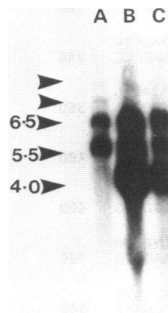
FIG. 1. RNA blot analysis of mouse testis and thymus mRNAs. RNA was isolated from pooled testes or thymuses as described (12) and analyzed on 1% agarose gels (13). Lanes: A, mouse thymus poly(A)$^+$ RNA; B, mouse testis total RNA; C, mouse testis poly(A)$^+$ RNA. Hybridization was done using a v-*abl*-specific probe. Numbers at left are in kb.

within the poly(A) tail of their cognate mRNAs. The predominance of clones that lacked these two restriction sites indicated that such clones might be derived from the testis-specific 4.0-kb species, whereas the other two might derive from the larger mRNAs because 4.0-kb mRNA is present at least five to ten times in larger abundance compared with the 5.5- and 6.5-kb species (see Fig. 1). Analysis of the clone sizes on agarose gels revealed that all the clones lacking the *Bam*HI and *Hin*dIII sites were between 3.8 and 4.0 kb in size, further suggesting that these clones derive from the 4.0-kb RNA species and represent nearly full-length cDNA clones.

Nucleotide Sequence of Testis-Specific c-*abl* cDNA Clones. The primary nucleotide sequence of a 4.0-kb cDNA clone was determined according to the method of Maxam and Gilbert (17) and is shown in Fig. 3. A single large open reading frame of 3369 base pairs (bp) was detected in this sequence with an ATG codon at position 52–54. This open reading frame could code for a polypeptide of 1123 amino acids with an approximate molecular mass of 125 kDa, in reasonable agreement with the estimated size of 145–150 kDa for the normal c-*abl* protein. The difference between the estimated and observed sizes is probably due to the posttranslational modifications that c-*abl* protein is known to undergo (18). Computer analysis of the predicted protein sequence detected no extended hydrophobic regions that could serve as transmembrane signals. The tyrosine phosphorylating kinase region (19) spans amino acids 235–486 and includes an ATP binding site (20) composed of the conserved 9-amino acid peptide at position 248–256 and a lysine at position 271. The tyrosine residue that is phosphorylated *in vitro* appears at position 393 in the sequence presented. Comparison of the predicted amino acid sequence of the normal mouse protein with the analogous human sequence (21) shows that the two proteins share an overall homology of 90%.

Four hundred and sixty base pairs downstream to the translation-termination signal, a stretch of 20 adenines were seen that appear to constitute the poly(A) tail of its cognate mRNA. This poly(A) stretch appeared at an unusual locus in this sequence because no known polyadenylylation signals

(such as the AATAAA sequence known to occur in most eukaryotic genomes) were seen upstream to this sequence. The sequence that came closest to the consensus sequence was TACAAA present at position 3867–3873. To determine whether the short stretch of poly(A) sequence present at the 3' end of the 4.0-kb cDNA clones were due to poor polyadenylylation of this mRNA, we determined the 3' sequence of four of the 4.0-kb cDNA clones and found that all of them had a poly(A) tail that was 12–20 bases in size, suggesting that the short adenine stretch is characteristic of the 4.0-kb c-*abl* mRNA.

Nucleotide Sequence of the 3' End of the Larger c-*abl* mRNAs. To determine whether the 3' ends of the other testis-specific mRNAs have unique sequences, we analyzed the nucleotide sequence of the 3' end of one cDNA clone that contained the *Bam*HI and *Hin*dIII sites. Sequence analysis of this clone (Fig. 4) indicated that the cDNA has an identical translational terminator sequence to that seen in the 4.0-kb mRNA and the v-*abl* sequence. However, a second open reading frame that could code for a protein of 18 kDa was seen in this sequence. We had previously noticed this open reading frame in the v-*abl* genome (2). Because the sequence of the c-*abl* gene was unknown at that time, it had been unclear whether the two open reading frames seen in the v-*abl* genome originated from a single larger open reading frame as a result of point mutations and/or deletions. The present analysis demonstrates that this is not the case and that the larger c-*abl* mRNAs contain the second open reading frame. A large part of this sequence is deleted in the 4.0-kb testis-specific mRNA because of premature termination of transcription. These results also formally demonstrate that the larger mRNAs present in the testis have the same 3' end as is seen with c-*abl* mRNAs of somatic cells (1).

Comparison of v-*abl* and c-*abl* Sequences and Implications for *abl* Oncogene Activation. To understand the molecular mechanisms involved in the oncogenic activation of the *abl* gene, we compared the sequence of the viral genome with the cDNA sequences reported here. Two prototypes of Abelson virus, designated P120 and P160, are known to occur in nature, and both these virus strains have been cloned and their nucleotide sequence has been determined (2, 3, 8, 22). A comparison of the DNA sequence of the two strains revealed that strain P120 has probably arisen from P160 by an in-frame deletion of 789 bp from the middle of the coding sequences. A comparison of the c-*abl* sequence presented in Figs. 3 and 4 with the v-*abl* sequence of P120 strain (Fig. 5) reveals that the viral gene has replaced 114 codons from the 5' end of the c-*abl* gene with 240 codons derived from the *gag* gene of Mo-MuLV. Furthermore, the viral gene is deleted for an additional stretch of 263 codons from the middle of the coding sequences. In addition to the coding sequences, a stretch of 765 bp of noncoding sequences have been deleted from the 3' end. Further, in addition to these deletions, one point mutation was observed at position 2545 of the c-*abl*
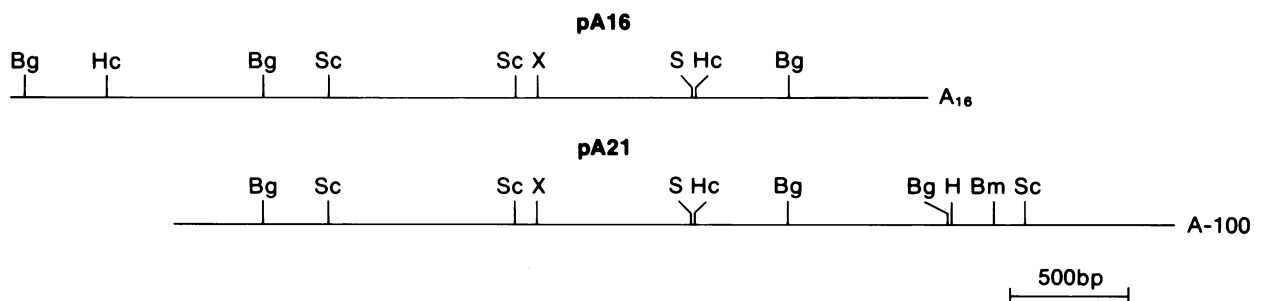

FIG. 2. Restriction maps of two c-*abl* cDNA clones A16 and A21 used for nucleotide sequence analysis. Bg, *Bgl* II; Bm, *Bam*HI; H, *Hin*dIII; Hc, *Hin*cII; S, *Sal* I; Sc, *Sac* I; X, *Xho* I.

```
                                                        MetLeuGluIleCysLeuLysLeuValGlyCysLysSerLysLysGlyLeuSerSerSerSerSerCys
    1  GGCGGGGCTGGGGGCCCGGGGTGGTGGGGCCTTCGCCGTGGCCACGGGACCATGTTGGAGATCTGTTTGAAGTTGGTGGGCTGCAAATCCAAGAAGGGGCTCTCTTCGTCCTCCAGCTGC    120
                                                                     ▲ Bgl 11
       TyrLeuGluGluAlaLeuGlnArgProValAlaSerAspPheGluProGlnGlyLeuSerGluAlaAlaArgTrpAsnSerLysGluAsnLeuLeuAlaGlyProSerGluAsnAspPro
  121  TACCTGGAAGAAGCCCTGCAGAGGCCAGTGGCATCTGACTTTGAGCCCCAGGGTCTCAGCGAAGCAGCTCGATGGAACTCCAAGGAAAACCTTCTTGCTGGGCCCAGTGAAAATGACCCC    240

       AsnLeuPheValAlaLeuTyrAspPheValAlaSerGlyAspAsnThrLeuSerIleThrLysGlyGluLysLeuArgValLeuGlyTyrAsnHisAsnGlyGluTrpCysGluAlaGln
  241  AACCTTTTTGTGGCACTCTATGATTTTGTGGCCAGTGGAGATAACACTCTCAGCATCACTAAAGGTGAAAAGCTCCGGGTCTTGGGTTATAATCACAATGGGGAATGGTGTGAAGCCCAA    360

       ThrLysAsnGlyGlnGlyTrpValProSerAsnTyrIleThrProValAsnSerLeuGluLysHisSerTrpTyrHisGlyProValSerArgAsnAlaAlaGluTyrLeuLeuSerSer
  361  ACGAAAAATGGCCAAGGATGGGTCCCAAGCAACTACATCACCCCCGTCAACAGCCTGGAGAAACATTCCTGGTATCATGGCCCTGTATCTCGGAATGCTGCTGAGTATCTGCTGAGCAGC    480
                            Start v-abl ▲          ▲ Hinc 11
       GlyIleAsnGlySerPheLeuValArgGluSerGluSerSerProGlyGlnArgSerIleSerLeuArgTyrGluGlyArgValTyrHisTyrArgIleAsnThrAlaSerAspGlyLys
  481  GGAATCAACGGCAGCTTCTTAGTGCGGGAGAGTGAGAGTAGCCCTGGCCAGAGATCCATCTCGCTGCGGTATGAAGGGAGGTGTACCACTACAGGATCAACACTGCCTCTGATGGCAAG    600

       LeuTyrValSerSerGluSerArgPheAsnThrLeuAlaGluLeuValHisHisHisSerThrValAlaAspGlyLeuIleThrThrLeuHisTyrProAlaProLysArgAsnLysPro
  601  CTGTACGTGTCCTCCGAGAGCCGCTTCAACACTCTGGCTGAGTTAGTTCACCATCACTCCACGGTGGCTGATGGCCTCATCACCACACTCCACTACCCAGCTCCCAAGCGCAACAAGCCC    720

       ThrIleTyrGlyValSerProAsnTyrAspLysTrpGluMetGluArgThrAspIleThrMetLysHisLysLeuGlyGlyGlyGlnTyrGlyGluValTyrGluGlyValTrpLysLys
  721  ACTATCTACGGTGTGTCCCCCAACTACGACAAGTGGGAAATGGAGCGCACCGACATCACCATGAAGCACAAGTTGGGTGGAGGCCAGTACGGGGAGGTGTACGAGGGCGTTTGGAAGAAG    840

       TyrSerLeuThrValAlaValLysThrLeuLysGluAspThrMetGluValGluGluPheLeuLysGluAlaAlaValMetLysGluIleLysHisProAsnLeuValGlnLeuLeuGly
  841  TACAGCCTCACTGTGGCCGTGAAGACCTTGAAGGAGGACACCATGGAGGTGGAGGAGTTCCTGAAGGAAGCGGCGGTGATGAAGGAGATCAAACACCCTAACCTGGTGCAGCTGCTAGGG    960

       ValCysThrArgGluProProPheTyrIleIleThrGluPheMetThrTyrGlyAsnLeuLeuAspTyrLeuArgGluCysAsnArgGlnGluValSerAlaValValLeuLeuTyrMet
  961  GTGTGTACCCGGGAACCACCATTCTACATAATCACTGAGTTCATGACCTATGGGAACCTGCTGGACTACCTGAGGGAGTGTAACCGCCAGGAAGTGTCAGCTGTCGTCCTCCTATACATG   1080

       AlaThrGlnIleSerSerAlaMetGluTyrLeuGluLysLysAsnPheIleHisArgAspLeuAlaAlaArgAsnCysLeuValGlyGluAsnHisLeuValLysValAlaAspPheGly
 1081  GCCACACAGATCTCATCAGCCATGGAGTACTTGGAGAAGAAGAACTTCATCCACAGAGACCTTGCTGCCCGGAACTGCCTGGTAGGGGAAAACCACTTGGTGAAGGTGGCTGATTTTGGC   1200
              ▲ Bgl 11
       LeuSerArgLeuMetThrGlyAspThrTyrThrAlaHisAlaGlyAlaLysPheProIleLysTrpThrAlaProGluSerLeuAlaTyrAsnLysPheSerIleLysSerAspValTrp
 1201  CTGAGCAGGTTGATGACAGGGGACACCTACACGGCCCATGCTGGAGCCAAATTCCCCATCAAATGGACCGCACCTGAGAGCCTGGCCTACAACAAGTTCTCCATCAAGTCGGACGTGTGG   1320

       AlaPheGlyValLeuLeuTrpGluIleAlaThrTyrGlyMetSerProTyrProGlyIleAspLeuSerGlnValTyrGluLeuLeuGluLysAspTyrArgMetGluArgProGluGly
 1321  GCATTTGGAGTATTGCTCTGGGAGATTGCTACCTATGGCATGTCACCTTACCCGGGAATTGACCTGTCTCAGGTTTATGAGCTGCTGGAAAAAGACTACCGCATGGAGCGCCCTGAAGGC   1440

       CysProGluLysValTyrGluLeuMetArgAlaCysTrpGlnTrpAsnProSerAspArgProSerPheAlaGluIleHisGlnAlaPheGluThrMetPheGlnGluSerSerIleSer
 1441  TGCCCGGAGAAGGTCTACGAGCTCATGCGAGCATGTTGGCAGTGGAACCCCTCTGACCGGCCCTCCTTTGCTGAAATCCACCAAGCCTTTGAAACCATGTTCCAGGAATCCAGTATCTCA   1560
                                  ▲ Sac 1
       AspGluValGluLysGluLeuGlyLysArgGlyThrArgGlyGlyAlaGlySerMetLeuGlnAlaProGluLeuProThrLysThrArgThrCysArgArgAlaAlaGluGlnLysAsp
 1561  GATGAGGTGGAGAAGGAGCTGGGGAAACGAGGCACGAGGAGGAGGTGCTGGGAGTATGCTGCAGGCCCCAGAGCTGCCCACCAAGACCAGAACCTGCAGGAGAGCAGCTGAGCAGAAAGAT   1680
                                                                                     Start of deletion in p120 ▲
       AlaProAspThrProGluLeuLeuHisThrLysGlyLeuGlyGluSerAspAlaLeuAspSerGluProAlaValSerProLeuLeuProArgLysGluArgGlyProProAspGlySer
 1681  GCGCCTGACACCCCTGAGCTGCTCCACACGAAGGGCCTGGGAGAAAGCGATGCACTGGACAGTGAGCCTGCTGTATCGCCACCTGCTTCCTCGGAAAGAGCGCGGGCCCCCAGACGGCAGC   1800

       LeuAsnGluAspGluArgLeuLeuProArgAspArgLysThrAsnLeuPheSerAlaLeuIleLysLysLysLysLysMetAlaProThrProProLysArgSerSerSerPheArgGlu
 1801  CTAAATGAAGATGAGCGCCTTCTCCCCAGAGACAGAAAGACCAACCTGTTCAGCGCTTTGATCAAGAAGAAGAAGAAATGGCGCCGACGCCCCCTAAGCGCAGCAGTTCCTTCCGAGAG   1920

       MetAspGlyGlnProAspArgArgGlyAlaSerGluAspAspSerArgGluLeuCysAsnGlyProProAlaLeuThrSerAspAlaAlaGluProThrLysSerProLysAlaSerAsn
 1921  ATGGATGGCCAGCCAGACCGCAGAGGGGCTAGTGAGGATGACAGCAGGGAACTCTGCAATGGACCCACCTGCACTCACAGCGCAGCAGAGCCTACCAAGTCCCCAAAGGCCAGCAAT   2040
                                                    * Frame shift mutation in p160
       GlyAlaGlyValProAsnGlyAlaPheArgGluProGlyAsnSerGlyPheArgSerProHisMetTrpLysLysSerSerThrLeuThrGlySerArgLeuAlaAlaAlaAlaGluGluGlu
 2041  GGGGCTGGCGTCCCTAATGGAGCCTTCCGGGAGCCGGGCAACTCAGGCTTCCGTTCTCCCCACATGTGGAAAAAGTCCAGCACACTGACCGGGAGCCGCCTGGCTGCTGCCGAAGAGGAG   2160
             * Frame shift mutation in p160
       SerGlyMetSerSerSerLysArgPheLeuArgSerCysSerAlaSerCysMetProHisGlyValArgAspThrGluTrpArgSerValThrLeuProArgAspLeuProSerAlaGly
 2161  AGCGGCATGAGCTCCAGTAAGCGCTTCCTGCGTTCTTGTTCGGCCTCCTGCATGCCCCATGGGGCAAGGGACACAGAGTGGCGGTCGGTCACGCTGCCTCGAGACCTGCCGTCTGCTGGC   2280
             ▲ Sac 1                              ▲ Xho 1
       LysGlnPheAspSerSerThrPheGlyGlyHisLysSerGluLysProAlaLeuProArgLysArgThrSerGluSerArgSerGluGlnValAlaLysSerThrAlaMetProLeuPro
 2281  AAGCAGTTTGACTCATCCACCTTTGGAGGGCACAAAAGCGAAAAGCCAGCTCTGCCTCGGAAACGCACCAGTGAGAGCAGGTCTGAGCAGGTGGCCAAAAGCACGGCGATGCCCCTCCCC   2400

       GlyTrpLeuLysLysAsnGluGluAlaAlaGluGluGlyPheLysAspThrGluSerSerProGlySerSerProProSerLeuThrProLysLeuLeuArgArgGlnValThrAlaSer
 2401  GGCTGGTTGAAGAAGAACGAGGAGGCTGCTGAAGAAGGCTTCAAAGACACAGAATCCAGCCCTGGCTCCAGCCCTCCCAGCTTGACTCCCAAACTCCTCCGCAGGCAGGTCACTGCCTCT   2520
                                                         ▲ End of deletion in p120
       ProSerSerGlyLeuSerHisLysGluGluAlaThrLysGlySerAlaSerGlyMetGlyThrProAlaThrAlaGluProAlaProProSerAsnLysValGlyLeuSerLysAlaSer
 2521  CCTTCCTCTGGCCTCTCTCACAAGGAAGAGGCCACCAAGGGCAGTGCCTCAGGCATGGGGACTCCGGCCACTGCAGAGCCAGCACCCCCCAGCAACAAAGTGGGCCTCAGCAAGGCCTCC   2640
                * mutation in v-abl
       SerGluGluMetArgValArgArgHisLysHisSerSerSerGluSerProGlyArgAspLysArgLeuLysLeuProAlaProProProProProAlaCysThrGlyLysAla
 2641  TCTGAGGAGATGCGCGTAAGGAGGCACAAGCACAGCTCGGAGTCCCCAGGGAGGAGACAAGGGGCGACTGGCTAAGCTCAAGCCTGCCCCGCCGCCTCCTCCTGCCTGCACAGGAAAAGCA   2760

       GlyLysProAlaGlnSerProSerGlnGluAlaGlyGluAlaGlyGlyProThrLysThrLysCysThrSerLeuAlaMetAspAlaValAsnThrAspProThrLysAlaGlyProPro
 2761  GGCAAGCCCGCACAGAGCCCCAGCCAAGAGGCCGGGGAGGCAGGGGGGCCCACAAAGACAAAATGCACGAGTCTGGCTATGGATGCTGTGAACACTGACCCCACCAAGGCCGGCCCACCT   2880

       GlyGluGlyLeuArgLysProValProProSerValProLysProGlnSerThrAlaLysProProGlyThrProThrSerProValSerThrProSerThrAlaProAlaProSerPro
 2881  GGAGAAGGACTGAGAAAGCCTGTGCCCCCATCTGTGCCAAAGCCCCAGTCGACGGCTAAGCCTCCAGGGACTCCCACCAGCCCGGTCTCCACCCCCTCCACAGCACCAGCTCCTTCACCC   3000
                                                                                 Sal I ▲▲ Hinc II
       LeuAlaGlyAspGlnGlnProSerSerAlaAlaPheIleProLeuIleSerThrArgValSerLeuArgLysThrArgGlnProProGluArgIleAlaSerGlyThrIleThrLysGly
 3001  CTGGCTGGGGACCAGCAGCCATCTTCTGCCGCCTTCATCCCCCTCATATCAACCCGTGTGTCTCTTAGGAAGACCCGCCAGCCGCCAGAGCGCATTGCCAGTGGCACCATCACCAAGGGT   3120

       ValValLeuAspSerThrGluAlaLeuCysLeuAlaIleSerArgAsnSerGluGlnMetAlaSerHisSerAlaValLeuGluAlaGlyLysAsnLeuTyrThrPheCysValSerTyr
 3121  GTGGTTCTGGACAGTACTGAGGCCCTGTGCCTTGCCATCTCCCGGAACTCAGAGCAGATGGCCAGCCACAGTGCTGTACTGGAGGCTGGCAAGAACCTGTACACTTTCTGTGTGAGCTAT   3240

       ValAspSerIleGlnGlnMetArgAsnLysPheAlaPheArgGluAlaIleAsnLysLeuGluSerAsnLeuArgGluLeuGlnIleCysProAlaThrAlaSerSerGlyProAlaAla
 3241  GTGGACTCTATCCAGCAGATGAGGAACAAGTTTGCCTTCCGTGAGGCTATCAACAAGCTGGAGAGCAACCTCCGAGAGCTGCAGATCTGCCCTGCCACAGCCTCCAGTGGGCCAGCTGCC   3360
                                                                        ▲ Bgl 11
       ThrGlnAspPheSerLysLeuLeuSerSerValLysGluIleSerAspIleValArgArg***
 3361  ACCCAAGACTTCAGCAAGCTGCTCAGCTCTGTGAAGGAGATCAGCGACATTGTCCGGAGGTAGCAGCAACCAGTGTATGTCAGCAAGAGATGTTGCAGTTCACAGGGCTCTTGTGCCTAT   3480
 3481  AAAGATGGGGACAGGGGACTGGGGAGCTGGCGTCTTTCCCCAGGAGCTTTAAAGAGAGACAAGCAGAGCCTGAGGGAGACCTGGATGGAGCCTGGTGGAGTTGGCTCTTCCTCCTGTGTT   3600
 3601  GTGCACCAGCTGCCCTGCACCTTTCCTGTCCAGCCCAGGCGTCAGCCACCTCTCCTCACTGCCTGTGGATGGGTCTCCTGCTCTGAAGACTACATCTGGCCTGCCTGGCCACCAGGCTTC   3720
 3721  TCACTCCCCGGTGCCTCAGACCCAGCTCCCAGGTCAGCCTGGAGTGCTCTTCCCTGTCCTTGCAGAACGACCTCCTCTGATGGACCTTCTTGTCACCAAGGCATGGGAGCCCCTGTGCTT   3840
 3841  ACTGTACCTGCACCTTTGATGCTTACAAACTGTCCCCGAGAAAAAAAAAAAAAAAAAAAA   3900
```
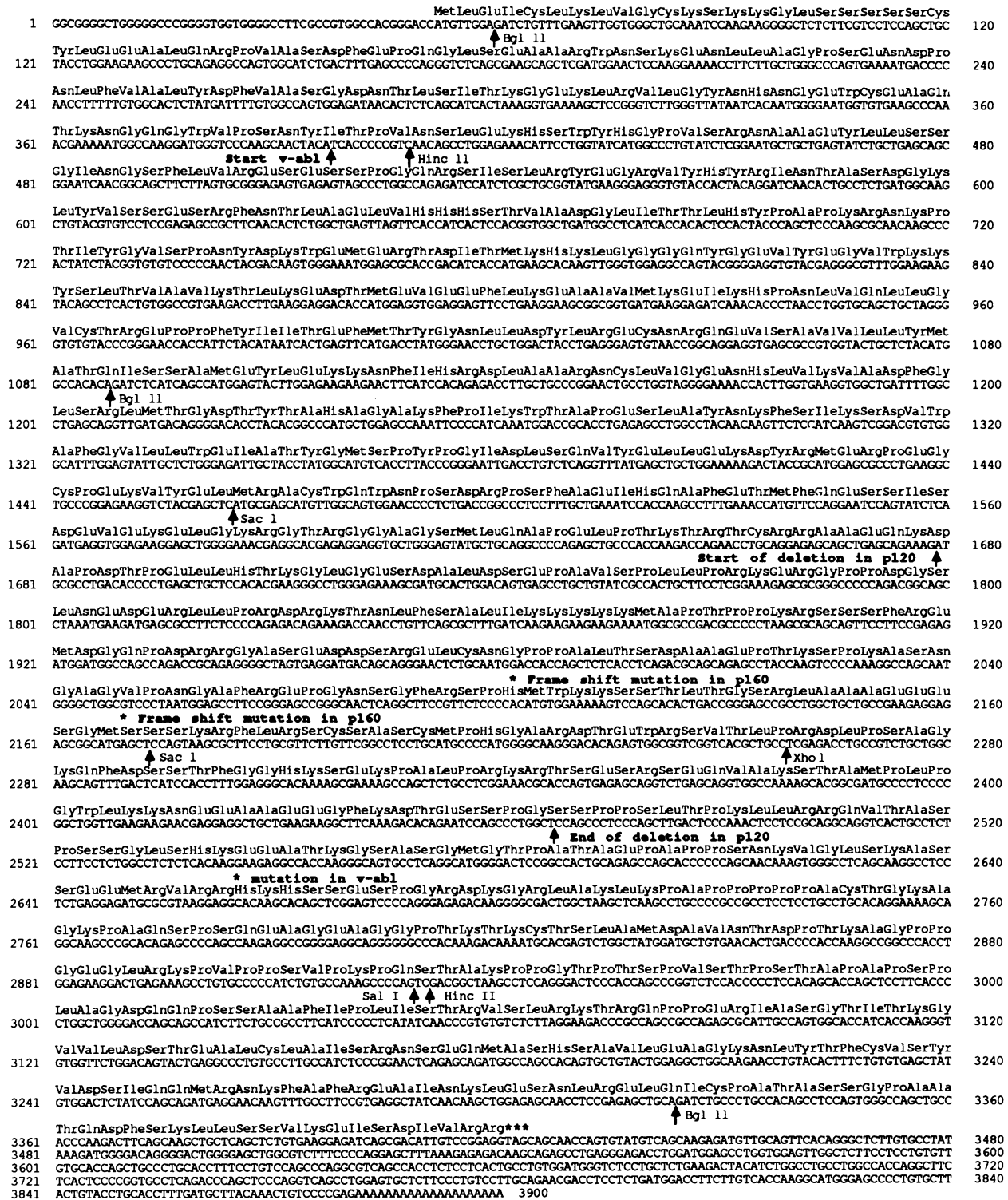
FIG. 3. Nucleotide sequence of mouse testis-specific c-*abl* cDNA clone A16. The amino acid sequence predicted for the protooncogene product is shown above the nucleotide sequence. Pertinent restriction enzyme sites and v-*onc* homologous regions are shown. The position of frame-shift mutations and point mutations that are observed in the v-*abl* gene are indicated by an asterisk. Translation terminator codon is denoted by ***.

sequence presented in Fig. 2. This results in the substitution of a glutamic acid residue with lysine in the v-*abl*-encoded transforming protein. Other differences that were seen between the c-*abl* and v-*abl* sequences were a point mutation at position 3629 where a thymine in v-*abl* is replaced by a cytosine in c-*abl*. This results in the substitution of an alanine present in the second open reading frame (coding for a putative 18-kDa protein) of v-*abl* by a valine in c-*abl*.

A comparison of the cDNA sequences with the P160 strain of Ab-MuLV (22) revealed that this virus has identical recombinational points at the 5' and the 3' ends as that of P120 strain. However, the internal deletion of 0.789 kb that characterized the structure of P120 strain is not seen. Interestingly, however, two frame-shift mutations were seen in the viral genome within the sequences that are deleted in the P120 strain. Thus, a guanine present at position 1982 was found to

```
                SerAspIleValArgArg***
        3401    TCAGCGACATTGTCCGGAGGTAGCAGCAACCAGTGTATGTCAGCAAGAGATGTTGCAGTTCACAGGGCTCTTGTGCCTATAAAGATGGGGACAGGGGACTGGGGAGCTGGCGTCTTTCCC    3520


                                                                    MetGluProGlyGlyValGlySerSerSerCysValValHisGlnLeuProCysThrPheProAlaGlnProArgA
        3521    CAGGAGCTTTAAAGAGAGACAAGCAGAGCCTGAGGGAGACCTGGATGGAGCCTGGTGGAGTTGGCTCTTCCTCCTGTGTTGTGCACCAGCTGCCCTGCACCTTTCCTGCCCAGCCCAGGC    3640


                rgGlnProProLeuLeuThrAlaCysGlyTrpValSerCysSerGluAspTyrIleTrpProAlaTrpProProGlyPheSerLeuProGlyAlaSerAspProAlaProArgSerAlaT
        3641    GTCAGCCACCTCTCCTCACTGCCTGTGGATGGGTCTCCTGCTCTGAAGACTACATCTGGCCTGCCTGGCCACCAGGCTTCTCACTCCCCGGTGCCTCAGACCCAGCTCCCAGGTCAGCCT    3760


                rpSerAlaLeuProCysProCysArgThrThrSerSerAspGlyProSerCysHisGlnGlyMetGlyAlaProValLeuThrValProAlaProLeuMetLeuThrAsnCysProArgG
        3761    GGAGTGCTCTTCCCTGTCCTTGCAGAACGACCTCCTCTGATGGACCTTCTTGTCACCAAGGCATGGGAGCCCCTGTGCTTACTGTACCTGCACCTTTGATGCTTACAAACTGTCCCCGAG    3880


                luProValLeuThrValPheSerLeuGluGlySerCysArgPheLysGlyHisGluValLeuLysProGlyAlaGlnMetGlyGlyHisTrpLysGlnGluLeuGlySerValValCysH
        3881    AGCCTGTGCTCACTGTGTTTTCATTGGAAGGAAGCTGTCGCTTTAAGGGTCATGAGGTGCTAAAGCCAGGGGCCCAGATGGGTGGGCACTGGAAACAGGAGCTGGGCAGTGTGGTCTGTC    4000
                ↑ end of 4 kb clone
                isLeuLeuSerValSerSerAlaValCysProAlaAspLeuGlyGlnGlnAla***
        4001    ACCTGCTCTCAGTATCTTCAGCAGTGTGCCCGGCAGATCTTGGACAGCAAGCTTGAGTTTTATGGGTGGCAGTCACTGGCTGGCTAGGCACATAGCCAGGCCAAACCTAGGCCTCCAAGG    4120
                                                    ↑ BglII        ↑ HindIII                         ↓ BamHI
        4121    GCTCCCCAAAATCTGAATTTCTGAGTAGTCTTCATCCCCTCTCCTGCTCTAAGGTTAGGTCCATCCTCTCTGGTCCTTACCTTGATGACAAGGATCCAGCCTTCTGGTGTTTTTGAGCAT    4240
        4241    TTCAAAGGTCTGCATAGAAAGGAACAGCCACTAGGTACAAACAGGCAGGAATCTGGGCTCTCCCGAGCCCTGACTGCTTCTCAGAACTGCTGCTTAACCACTGAGAACAGTGCTCAGCCC    4360
        4361    AGAGCTCCTGAGAGCATGCACATCACGCTGGGGTTGGGTGTCCCTTGCAGACTCAGCCAGGGGGGCCATATCTTCCTCCTAAGGCAGGCAGGACCAGAGCTGAGCACAGGCACAGGTGTG    4480
        4481    GGCGACTCGTCTGACTAACCCAGAGCCGCCAGCCTAGGCCTCGTGCCAGGCTGACCCAAAACAGGGCGCATGCCTAGCCTTGTACCAGGCTGGAAGACACCTCCACACTCCCTGCTTCTC    4600
        4601    CACTGCAGGTGTGCTGTGCGTCCCCAGGCCCACACTGATCACCTAGATGCGGGCCTGGTAGAGATGATTTCCTCTGGATGGTCTTCCATGGTCCTCGTAGCACCTCACCTCGTTACCTCC    4720
        4721    AAAGGCTGCTCTGCAGAGGGAGGCTGTGCACACTGCCAGCGACCTTGAATAGCCAGACACCGCCATCCCCTGACACTGTGCTCACAAGGCCCCTGCTCTGTTGCTGCCGTGGGTGCCACT    4840
        4841    ATATTTTGCATTACACTTTGATACTACACTCTTTTTTTAAAATAGACTTGCGCTCTTATAAATGACATGTAAATAGTTCACCATCTGCTCCCTCTGGCACACCAACAGCATTCCCCTTTT    4960
        4961    TTCTCAGTCCAGTACATTTTGTTTCTGTATATGACTGTTTTTTGAATCCAAATCTCTCTGTAGTATTTTTTAAATAAATGTTTACAGAACTAAAAAACAAAAAAAAAAAAAAAAAAAAAA    5080
        5081    AAAAAAAAAAAAAAAAAAAAAAAAAAAA    5108
```

FIG. 4.   Nucleotide sequence of the 3' end of a c-*abl* cDNA clone derived from larger (5.5-kb or 6.5-kb) c-*abl* mRNAs. The sequence derived from the 3' end of A21 is shown. The position where the testis-specific c-*abl* cDNA sequences end is shown by an arrow. The position of *Hind*III and *Bam*HI sites that are absent in testis-specific c-*abl* cDNA are shown. The translated sequence of the open reading frame that could code for the putative 18-kDa protein is also shown. For comparative purposes, the same numbering was used at the start of the sequence as that in the clone A16 described in Fig. 3.

be missing in the viral sequence, whereas an additional guanine has been inserted into the viral genome at the position 2050 of the corresponding cDNA sequence (see Fig. 3). These two frame-shift mutations render a stretch of 69 bp, coding for 23 internal amino acids, out-of-frame, changing entirely the amino acid sequence of the viral protein in this region. Thus, in the P120 strain of Ab-MuLV, this entire region has been deleted, whereas in the P160 strain, the sequence of the viral genome is altered in the corresponding region as a result of two frame-shift mutations. These results, taken together, indicate that the deletion of amino-terminal sequence, as well as the alteration of the middle region of the c-*abl* gene, might play a crucial role in its oncogenic activation.

DISCUSSION

Earlier studies had indicated that the c-*abl* gene of mouse is expressed in virtually all tissues examined, usually as two distinct mRNA species of 5.5 kb and 6.5 kb (9). Expression of the c-*abl* gene appears to be maximal in testis, where both the 5.5-kb and 6.5-kb species are made, along with another species of 4.0 kb that is expressed at 5- to 10-fold higher levels compared with the two other species (10–12, Fig. 1). The

FIG. 5.   A schematic comparison of v-*abl* and c-*abl* sequences. The position of the cellular sequences incorporated into the v-*abl* (P120) genome is shown by boxes. The position of frame-shift mutations observed in the v-*abl* sequence of P160 strain is indicated by asterisks. The length of each portion of v-*abl* and c-*abl* sequence is indicated in kb. LTR, long terminal repeats.

molecular basis for the variation in transcript size was recently studied by Ben-Neriah *et al.* (23), who examined the 5' sequence of cDNA clones derived from the 5.5-kb and 6.5-kb mRNAs. These studies indicated that this variation arises as a result of differential splicing of 5' exons, which results in the generation of mRNAs that are divergent at the 5' end but contain a common sequence at the 3' end. Because these cDNAs were prepared from a B-lymphoid cell line, the molecular mechanisms involved in the generation of the 4.0-kb *abl*-specific mRNA could not be studied. In this communication, we have isolated the cDNA clones for the cognate 4.0-kb mRNA and determined its complete nucleotide sequence. The cDNA clones isolated by us seem to represent near full-length clones because they closely correspond to the size of mRNA and virtually all the clones derived from the testis-specific mRNA terminate within 20 bp of each other at the 5' end. Our sequence analysis of the testis-specific cDNA clones indicates that the 4.0-kb c-*abl* RNA is produced in the testis as a result of premature termination of transcription. Polyadenylylation of this RNA was found to occur at an unusual place where no consensus polyadenylylation signal was seen. The 5' sequence of this cDNA clone corresponds to the sequence of type 1 c-*abl* cDNA described by Ben-Neriah *et al.* (23). Thus, it appears that the variation in the 4.0-kb mRNA is primarily at the 3' end. The mechanism by which this short mRNA is produced is at present unknown. It is likely that testis cells contain special enzymes that have different sequence specificity for transcription termination and polyadenylylation compared with their somatic cell counterparts. This agrees with the observation that several mRNA species produced in testis have smaller sizes compared with their somatic cell counterparts (24). For mouse c-*abl*-specific mRNAs, the truncation results in the deletion of 3' sequences that contain an open reading frame that could code for an 18-kDa protein. In addition, a stretch of 1.0 kb of noncoding sequences that are otherwise present in the two larger mRNAs is deleted in the 4.0-kb mRNA. The 4.0-kb mRNA could be generated specifically to avoid the synthesis of 18-kDa protein in large quantities in testis. However, there is no experimental evidence indicating that this short reading frame codes for a polypeptide. Alternatively, the 3' sequences might code for control signals that determine the message stability and translational efficiency of the mRNAs (25–27). The deletion of the 3' sequences in such an instance would be expected to result in the decreased
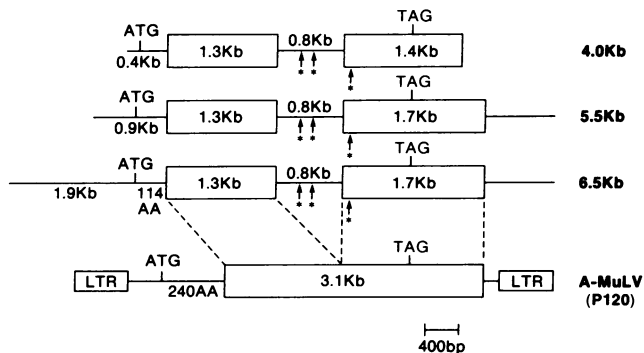
turnover and increased translatability of the mRNA, consistent with the presence of high levels of 4.0-kb mRNA in testis cells compared with 5.5-kb and 6.5-kb species. The transcriptional promoter sequence that is used for the synthesis of the 4.0-kb RNA is at present unknown. Though our testis-specific c-*abl* cDNA clones have an identical 5' end as that of type 1 cDNA of Ben-Neriah (which is derived from either 5.5-kb or 6.5-kb mRNA), the 5.5-kb mRNA would be predicted to have at least 500 bp of additional sequences at its 5' end to account for the size difference. Should this prediction prove accurate, the 4.0-kb mRNA might be produced from a promoter distinct from those used for the synthesis of the 5.5-kb and 6.5-kb mRNAs. This situation is consistent with the finding of Shtivelman *et al.* (21), who showed that the 6.0-kb and 7.0-kb c-*abl* mRNAs in the human cells use different promoters for their transcription.

The availability of full-length cDNA clones for the c-*abl* mRNA has allowed us to determine its complete nucleotide sequence and compare it with v-*abl* sequence (see Fig. 5). Both the P120 and P160 strains have deleted 114 codons at the 5' end and have replaced these sequences with 240 codons derived from the *gag* gene of Mo-MuLV. In addition, the P120 strain has 263 codons deleted from the middle of the coding region. The P160 strain, while retaining these sequences, has suffered two frame-shift mutations in this region that render a stretch of 69 bp (codon for 23 internal amino acids) out of reading frame, changing the amino acid sequence of this stretch of coding sequences. Recently a feline sarcoma virus that also transduced the *abl* gene has been isolated, and its sequence has been determined (28). During the transduction process, this virus has eliminated 51 codons from the 5' end and all the sequences downstream to the 490th codon of the c-*abl* sequences. Thus, this virus also has been deleted in both 5' and 3' sequences close to the region where the Ab-MuLV (P120) sequences have undergone deletion. These results lend further support to the argument that the amino- and carboxyl-terminal deletions play an important role in the oncogenic activation of c-*abl*. In addition to the frame-shift mutations/deletions, the v-*abl* gene has suffered one point mutation at position 2545 (of c-*abl* sequence shown in Fig. 3), which results in the substitution of a glutamic acid residue with lysine in the v-*abl* gene product (2).

In the human system, Philadelphia chromosome-positive chronic myelogenous leukemia cells have been shown to undergo translocations in the *ABL* locus that result in the formation of *BCR/ABL* fusion product (4–7, 18, 21, 29). In this case, 26 codons derived from the 5' end of the *ABL* gene are replaced with 927 codons from the *BCR* gene. However, in this system no mutations or deletions have been observed near the 3' end. Thus amino-terminal deletions alone seem to be involved in the activation of this gene. Recent reports (29) indicate that recombinant retroviruses coding for *BCR/ABL* product do not transform NIH 3T3 cells or B lymphocytes *in vitro*, suggesting that *BCR/ABL* gene product is less potent in its oncogenic activity than the Ab-MuLV-encoded *GAG/ABL* gene product. It is possible that the 3' deletions and mutations play an important role in potentiating the oncogenic activity of this transforming gene. The availability of full-length c-*abl* cDNA clones should facilitate the construction of recombinants between v-*abl* and c-*abl* se-

quences, as well as v-*abl* and *bcr/abl* sequences that should resolve the role of 5' and 3' deletions and mutations in the activation of the c-*abl* gene.

1. Wang, J. Y. J., Ledley, F., Goff, S., Lee, R., Groner, Y. & Baltimore, D. (1984) *Cell* **36**, 349–356.
2. Reddy, E. P., Smith, M. J. & Srinivasan, A. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3623–3627.
3. Goff, S. P., Gilboa, E., Witte, O. N. & Baltimore, D. (1980) *Cell* **22**, 777–785.
4. Heistercamp, N., Stephenson, J. R., Groffen, J., Hansen, P. F., de Klein, A., Bartram, C. R. & Grosveld, G. (1983) *Nature (London)* **306**, 239–242.
5. Canaani, E., Gale, R. P., Stein-Saltz, D., Berrebi, A., Aghai, E. & Januszewicz, E. (1984) *Lancet* **i**, 593–595.
6. Shtivelman, E., Lifshitz, B., Gale, R. B. & Canaani, E. (1985) *Nature (London)* **315**, 550–554.
7. Konopka, J. B., Watanabe, S. M., Singer, J. W., Collins, S. J. & Witte, O. N. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1810–1814.
8. Srinivasan, A., Reddy, E. P. & Aaronson, S. A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2077–2081.
9. Wang, J. Y. J. & Baltimore, D. (1983) *Mol. Cell. Biol.* **3**, 773–779.
10. Muller, R., Salmon, D. J., Tremblay, J. M., Cline, M. J. & Verma, I. (1982) *Nature (London)* **299**, 640–643.
11. Ponzetto, C. & Wolgemuth, D. J. (1985) *Mol. Cell. Biol.* **5**, 1791–1794.
12. Propst, F., Rosenberg, M. P., Iyer, A., Kaul, K. & Vande Woude, G. F. (1987) *Mol. Cell. Biol.* **7**, 1629–1637.
13. Auffray, C. & Rougeon, F. (1980) *Eur. J. Biochem.* **107**, 303–308.
14. Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977) *Biochemistry* **16**, 4743–4748.
15. Rosson, D. & Reddy, E. P. (1986) *Nature (London)* **319**, 604–606.
16. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
17. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
18. Konopka, J. B. & Witte, O. N. (1985) *Biochim. Biophys. Acta* **823**, 1–17.
19. Feng, D. F., Johnson, M. S. & Doolittle, R. F. (1985) *J. Mol. Evol.* **21**, 112–125.
20. Kamps, M. P., Taylor, S. S. & Sefton, B. P. (1984) *Nature (London)* **310**, 589–592.
21. Shtivelman, E., Lifshitz, B., Gale, R. P., Roe, B. A. & Canaani, E. (1986) *Cell* **47**, 277–284.
22. Lee, R., Paskind, M., Wang, J. Y. J. & Baltimore, D. (1985) in *RNA Tumor Viruses*, eds. Weiss, R., Teich, N., Varmus, H. & Coffin, J. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 861–868.
23. Ben-Neriah, Y., Bernards, A., Paskind, M., Daley, G. Q. & Baltimore, D. (1986) *Cell* **44**, 577–586.
24. Propst, F. & Vande Woude, G. (1985) *Nature (London)* **315**, 516–518.
25. Shaw, G. & Kamen, R. (1986) *Cell* **46**, 659–667.
26. Meijlink, F., Curran, T., Miller, A. D. & Verma, I. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4987–4991.
27. Rahmsdorf, H. J., Schonthal, A., Angel, P., Litfin, M., Ruther, U. & Herrlich, P. (1987) *Nucleic Acids Res.* **15**, 1643–1659.
28. Bergold, P. J., Blumenthol, J. A., D'Andrea, E., Snyder, H. W., Lederman, L., Silverstone, A., Nguyen, H. & Besmer, P. (1987) *J. Virol.* **61**, 1193–1202.
29. Daley, G. Q., McLaughlin, J., Witte, O. N. & Baltimore, D. (1987) *Science* **237**, 532–535.