NOTES

Isolation and Characterization of Three Distinct cDNAs for the Chicken *c-ski* Gene

PRAMOD SUTRAVE AND STEPHEN H. HUGHES*

National Cancer Institute-Frederick Cancer Research Facility, BRI-Basic Research Program,
P.O. Box B, Frederick, Maryland 21701

Received 16 February 1989/Accepted 4 May 1989

Three types of *c-ski* cDNAs have been isolated from two different chicken cDNA libraries. Sequence comparisons suggest that the cDNAs derive from alternatively spliced mRNAs. A short stretch of sequence homology that exists between *c-ski* and avian leukosis virus may have played a role in viral transduction.

Avian retroviruses containing the oncogene *v-ski* were isolated from transformed foci first seen in cultures of chicken embryo cells infected with the replication-competent avian leukosis virus (ALV) tdB77 (24). Viruses with genomes ranging from 3.0 to 8.9 kilobases (kb) in length were derived from these cultures. These isolates can transform chicken embryo fibroblasts in culture (12, 23, 24). *ski* sequences have been inserted into the viral *gag* gene; the different-size viral genomes differed from each other in the extent of deletions in viral genes. Analysis of *v-ski*-transformed fibroblasts indicates that the *gag-ski* fusion protein found in these cells is located principally in the nucleus (1).

The oncogenes of acutely transforming retroviruses have been acquired from their hosts (4); however, relatively little is known about the cellular gene *c-ski*, from which the viral oncogene *v-ski* was derived. Using a *v-ski* probe, genomic clones for *c-ski* have been isolated and partially sequenced (22). The recombination events that led to the incorporation of *ski* into the viral genome resulted in the truncation of *ski* sequences at both the 5' and 3' ends, making assignment of the boundaries of the gene difficult. In addition, λ clones that contain the portion of genomic *c-ski* corresponding to the 5' end of *v-ski* do not overlap with the λ clones containing the portion of genomic *c-ski* corresponding to the 3' end of *v-ski*. This raises the possibility that the *v-ski* gene present in the virus derives from two distinct cellular genes.

We have screened two chicken cDNA libraries (kindly provided by B. Vennstrom) with a *v-ski* probe by using standard protocols (13). One library was made from poly(A)⁺ mRNA isolated from the body wall of 10-day embryos, and the other was made from mRNA from avian erythroblastosis virus-transformed chicken erythroblasts. Four distinct *c-ski* cDNAs were isolated; three (designated FB27, FB28, and FB29) were from the body wall library, and one (designated CEL) was from the erythroblast library. These cDNAs include sequences extending both 5' and 3' of the portion of *ski* present in the virus. The cDNAs demonstrate that *v-ski* derives from a single cellular gene and suggest that multiple *c-ski* mRNAs, encoding distinct *ski* proteins, are produced from the *c-ski* locus by alternate splicing (10), adding to a growing list of oncogenes known to produce multiple mRNAs in this fashion (2, 11, 14, 15).

Structure of the cDNA clones. The structures of all of the *c-ski* cDNA clones that have been characterized by DNA sequencing (20) and their relationships to *v-ski* are presented in Fig. 1. Only FB28 and CEL have sequences 5' of the *v-ski* sequences; both FB27 and FB29 are truncated at the 5' end. Assuming that the missing sequences at the 5' end of FB29 are similar to those in FB28 and CEL, a composite nucleotide sequence and the deduced amino acid sequence for a cDNA of the FB29 type is shown in Fig. 2. The first ATG with a substantial downstream open reading frame is located at nucleotide position 168. Upstream of this ATG no reading frames are open, suggesting that these sequences represent the 5' untranslated region. On the basis of the cDNA sequence analyses and comparisons with the positions of the splice donor and acceptor sites known from the genomic sequence (22), we have derived exon boundaries (shown schematically in Fig. 3). *c-ski* sequences are distributed over seven exons (Fig. 3). FB29 contains all seven exons; FB28 and FB27 lack exons 2 and 6, respectively. This differential splicing affects the protein-coding potential of the three cDNAs. Differential splicing of exon 2 deletes 37 amino acids without affecting the coding potential of the open reading frame downstream. Differential splicing of exon 6, however, affects the coding potential of exon 7. If exon 5 is spliced to exon 7 (seen in FB27 cDNA), a translation termination codon is generated at the splice junction and exon 7 becomes a noncoding exon. However, if exon 5 is spliced to exon 6 and exon 6 is spliced to exon 7, as in FB28/29, then the open reading frame continues in exon 7 for 417 nucleotides encoding an additional 129 amino acids. Assuming that the missing 5' ends of the FB29 and FB27 mRNAs contain sequences identical to those present in CEL and FB28, then translation of mRNAs corresponding to the three different types of cDNAs would lead to the generation of three proteins, one containing 750 amino acids (from FB29), the second containing 713 amino acids (from FB28), and the third containing 510 amino acids (from FB27).

The CEL clone is missing 3' sequences. cDNAs that derive from the body wall (FB) library have long 3' untranslated regions that contain a 95-base-pair (bp) A+T-rich region from nucleotides 2803 to 2898. Within this region there are two copies of a sequence, ATTTA, that has been implicated in mRNA destabilization in a variety of transiently induced mRNAs, including *c-myc*, interferon, *c-jun*,

* Corresponding author.

FIG. 1. Structures of the *c-ski* cDNA clones. The lengths of the cDNAs are drawn to scale, and the restriction sites are indicated. The sequence of *v-ski* is shown for comparison. The dotted boxes in the *v-ski* sequence represent the *gag* region of the *gag-ski* fusion in the acutely transforming Sloan-Kettering virus. A, B, C, and D represent the regions used for generating single-stranded probes for S1 nuclease protection analysis.

and *c-fos* (16, 19, 21). Addition or deletion of these sequences has also been shown to affect the transformation potential of *c-fos* (16). We are now attempting to determine what role (if any) these sequences play in the function or expression of the *c-ski* gene.

The *c-ski* cDNAs contain two potential poly(A) signals (AATAAA) located at positions 3348 and 4167. Although all three clones isolated from the FB library end at the same position, none has a poly(A) tail; therefore, it is likely that the 3' ends of the *c-ski* mRNAs are not contained in these clones. The 4.2-kb cDNAs we isolated and characterized are smaller than the 5.7-, 8.0-, and 10.0-kb mRNAs detected by Northern (RNA) transfer analysis (12; P. Suttrave and S. H. Hughes, unpublished observations). We have not explained this discrepancy; however, we suggest that the clones isolated so far, which lack poly(A) tails, also lack sequences from the 3' ends of the mRNA(s).

The 5' end of *c-ski* mRNA(s). Sequence comparisons at the 5' end revealed that both FB28 and CEL are colinear up to a position 89 bp upstream of the putative translational initiation ATG codon. FB28 has an additional 76 bp, whereas CEL has 25 bp that are different from those found in FB28. This region where the two sequences diverge has been compared with sequences from genomic clones (22), and the upstream sequences in FB28 are colinear with the genomic sequences. Examination of the sequences in the genomic DNA at the point of divergence of CEL and FB28 did not reveal consensus sequences for donor or acceptor splice

sites. To confirm the authenticity of the clones, S1 nuclease protection analysis was carried out. The results (data not shown) demonstrated that the sequences present in FB28 are expressed as mRNA in normal embryos. Similar S1 analyses have provided no evidence for the presence in mRNA of the 25 bp at the extreme 5' end of the CEL clone. We suggest that the first 25 bases of the CEL clone are the result of a cloning artifact and that the sequences found in the FB28 clone are expressed in *c-ski* mRNA(s).

In an attempt to determine how much of the 5' untranslated region of the *c-ski* mRNA(s) is contained in FB28, we carried out primer extension analysis. Copying the 5' segment present in FB28 should give an extension product of about 280 bases. However, a primer extension product of 220 bases was seen (data not shown). S1 analyses (data not shown) have shown that this segment is expressed in RNA. It is possible that the observed primer extension product is the result of premature termination; however, it is also possible that there are multiple 5' ends for the *c-ski* mRNAs.

Organization of the internal exons. Sequence comparisons in the central portion of FB27, FB28, FB29, and CEL cDNAs revealed that FB28 lacks a small region of 111 bp (exon 2) from positions 1079 to 1191 that would eliminate 37 amino acids. The genomic sequence (22) in the corresponding region reveals the existence of consensus splice donor and acceptor sequences at the boundaries of the deletion, suggesting that the cDNA derives from a differentially spliced mRNA. To confirm the existence of mRNAs of the

FIG. 2 (next page). Complete sequence and potential coding region of a cDNA of the FB29 type, assuming that the 5'-end sequences of FB29 are similar to the 5' ends of the FB28 and CEL clones. ↑, Site where FB28 and CEL sequences diverge. The 25 bases found only in the CEL clones are not shown. ↓, Boundary of *v-ski*. The exon boundaries are numbered, and the alternately spliced exons are boxed. The single base and amino acid change between *c-ski* and *v-ski* is also boxed with a dashed line. The translation termination codon is boxed in thick lines. The potential polyadenylation signals are densely underlined. The A+T-rich region containing ATTTA sequences that might be involved in mRNA stability is underlined with dashes.

TC CTC GTC GTC TGT GGA TTG CTA AAC CTG AGT GGG AAG GGG GGG GAA AAA AAA AAG GGT GGG TTG TTG TTT TGT TTA AAA AAA GAA AAA ATC CCT TAA 98

GTG GAT TTG TAC CAG CGT GGA AGA TAA CTG GGG ATT TTT GTT GTT TGT TTT GGG AAT AGA AAC TAA AAA ¹ATG GAG ACT GTA AGT AGA AGC AGC TTC CAG 197
Met Glu Thr Val Ser Arg Ser Ser Phe Gln 10

CCT CAT CCA GGA CTG CAG AAG ACC TTG GAA CAG TTT CAT CTG AGC ^{Vski}TCT ATG AGC TCC CTG GGT GGC CCT GCT GCT TTC TCA GCG CGA TGG GCA CAG GAG 296
Pro His Pro Gly Leu Gln Lys Thr Leu Glu Gln Phe His Leu Ser Ser Met Ser Ser Leu Gly Gly Pro Ala Ala Phe Ser Ala Arg Trp Ala Gln Glu 43

ATG TAC AAG AAA GAC AAT GGC AAA GAC CCA GCG GAA CCT GTA CTG CAT CTG CCC CCT ATC CAG CCC CCC CCG GTG ATG CCT GGT CCC TTC TTC ATG CCC 395
Met Tyr Lys Lys Asp Asn Gly Lys Asp Pro Ala Glu Pro Val Leu His Leu Pro Pro Ile Gln Pro Pro Pro Val Met Pro Gly Pro Phe Phe Met Pro 76

TCG GAC AGA TCC ACT GAG AGG TGC GAG ACC ATC CTG GAA GGG GAA ACC ATC TCC TGC TTC GTG GTG GGT GGG GAA AAG CGC CTT TGC TTG CCC CAG ATC 494
Ser Asp Arg Ser Thr Glu Arg Cys Glu Thr Ile Leu Glu Gly Glu Thr Ile Ser Cys Phe Val Val Gly Gly Glu Lys Arg Leu Cys Leu Pro Gln Ile 109

CTG AAC TCG GTG CTC AGG GAC TTC TCC CTG CAG CAG ATC AAT TCG GTG TGC GAT GAG CTA CAC ATT TAC TGC TCC AGA TGC ACC GCT GAC CAG CTG GAG 593
Leu Asn Ser Val Leu Arg Asp Phe Ser Leu Gln Gln Ile Asn Ser Val Cys Asp Glu Leu His Ile Tyr Cys Ser Arg Cys Thr Ala Asp Gln Leu Glu 142

ATC CTC AAA GTC ATG GGC ATC TTG CCC TTC TCT GCC CCC TCC TGC GGG CTG ATC ACT AAA ACT GAT GCT GAG AGG CTT TGC AAT GCC TTG CTT TAT GGT 692
Ile Leu Lys Val Met Gly Ile Leu Pro Phe Ser Ala Pro Ser Cys Gly Leu Ile Thr Lys Thr Asp Ala Glu Arg Leu Cys Asn Ala Leu Leu Tyr Gly 175

GGC ACC TAT CCT CCC CAC TGC AAG AAG GAA TTC TCT AGC ACG ATT GAG CTG GAG CTT ACA GAG AAG AGC TTC AAG GTG TAC CAC GAG TGC TTT GGG AAG 791
Gly Thr Tyr Pro Pro His Cys Lys Lys Glu Phe Ser Ser Thr Ile Glu Leu Glu Leu Thr Glu Lys Ser Phe Lys Val Tyr His Glu Cys Phe Gly Lys 208

TGT AAG GGA CTC CTG GTA CCA GAG CTT TAC AGT AAC CCC AGC GCA GCC TGC ATC CAG TGC TTG GAC TGC AGG CTC ATG TAC CCG CCT CAC AAA TTT GTG 890
Cys Lys Gly Leu Leu Val Pro Glu Leu Tyr Ser Asn Pro Ser Ala Ala Cys Ile Gln Cys Leu Asp Cys Arg Leu Met Tyr Pro Pro His Lys Phe Val 241

GTC CAC TCT CAC AAA TCC CTG GAA AAC AGG ACT TGC CAC TGG GGC TTT GAC TCT GCA AAC TGG AGG TCC TAC ATC CTC CTT AGC CAG GAT TAC ACT GGG 989
Val His Ser His Lys Ser Leu Glu Asn Arg Thr Cys His Trp Gly Phe Asp Ser Ala Asn Trp Arg Ser Tyr Ile Leu Leu Ser Gln Asp Tyr Thr Gly 274

AAA GAG GAG AAA GCT AGG CTG GGC CAG CTC TTA GAT GAA ATG AAA GAA AAA TTT GAC TAT AAC AAC AAA TAC AAG AGG AAA GCC CCC AGG ²AAC CGT CAG 1088
Lys Glu Glu Lys Ala Arg Leu Gly Gln Leu Leu Asp Glu Met Lys Glu Lys Phe Asp Tyr Asn Asn Lys Tyr Lys Arg Lys Ala Pro Arg Asn Arg Glu 307

TCT CCT AGA GTT CAG CTC CGC CGG AAC AAA ATG TTC AAG ACA ATG CTG TGG GAT CCA GCT GGA GGT TCA GCG GTA CTG CAG CGT CAG CCA GAT GGA AAT 1187
Ser Pro Arg Val Gln Leu Arg Arg Asn Lys Met Phe Lys Thr Met Leu Trp Asp Pro Ala Gly Gly Ser Ala Val Leu Gln Arg Gln Pro Asp Gly Asn 340

GAG ³GTC CCT TCA GAT CCT CCT GCT TCC AAG AAA ACC AAA ATA GAC GAC TCC GCT TCC CAA TCT CCA GCT TCT ACT GAG AAG GAA AAG CAG TCC AGT TGG 1286
Glu Val Pro Ser Asp Pro Pro Ala Ser Lys Lys Thr Lys Ile Asp Asp Ser Ala Ser Gln Ser Pro Ala Ser Thr Glu Lys Glu Lys Gln Ser Ser¹ Trp¹ 373

TTA CGG TCC TTA TCC AGT TCA TCT AAT AAG AGC ATT GGC TGT GTC CAT CCC CGT CAG CGT CTC TCA GCT TTC CGG CCC TGG TCC CCT GCT GTA TCA GCA 1385
Leu Arg Lys His Ala Ala Glu Thr Pro Ala Val Pro Glu Pro Val Ala Thr Val Pro Arg Gln Arg Leu Ser Ala Phe Arg Pro Trp Ser Pro Ala Val Ser Ala 406

AAT GAG AAA GAG CTC TCA ACC CAT CTT CCT GCA TTG ATC CGA GAC AGC ⁴AGT TTT TAC TCC TAC AAA AGC TTT GAG AAT GCT GTG GCC CCC AAC GTG GCA 1484
Asn Glu Lys Glu Leu Ser Thr His Leu Pro Ala Leu Ile Arg Asp Ser Ser Phe Tyr Ser Tyr Lys Ser Phe Glu Asn Ala Val Ala Pro Asn Val Ala 439

CTC GCA CCT CCT GCC CAA CAG AAA GTT GTG AGC AAC CCA CCC TGT GCC ACA GTG ^{Vski}GTG TCC CGG AGC AGC GAA CCG CCG AGC AGC GCT GCG CAG CCA CGG 1583
Leu Ala Pro Pro Ala Gln Gln Lys Val Val Ser Asn Pro Pro Cys Ala Thr Val Val Ser Arg Ser Ser Glu Pro Pro Ser Ser Ala Ala Gln Pro Arg 472

AAA AGA AAA CAT GCT GCA GAA ACC CCG GCT GTC CCA GAG CCA GTG GCC ACG GTT ACT GCC CCT GAA GAG GAT AAG GAA TCA GAA GCA GAA ATT GAA GTA 1682
Lys Arg Lys His Ala Ala Glu Thr Pro Ala Val Pro Glu Pro Val Ala Thr Val Pro Thr Ala Pro Glu Glu Asp Lys Glu Ser Glu Leu Ile Glu Val 505

GAG ACC AGG GAG GAA ⁵TTC ACC TCC TCC TTA TCC TCG CTC TCC TCC CCA TCC TTT ACT TCA TCC AGC TCT GCA AAG GAC ATG AGC TCA CCT GGG ATG CAA 1781
Glu Thr Arg Glu Glu Thr Thr Ser Ser Leu Ser Ser Leu Ser Ser Pro Ser Phe Thr Ser Ser Ser Ser Ala Lys Asp Met Ser Ser Pro Gly Met Gln 538

GCC CCA GTC CCA GTC AAC AGT TCA TAT GAG GTT GCA GCA CAT TCT GAC TCT CAC AGC AGT GGG TTG GAA GCT GAG CTG GAG CAC TCA AGG CAG GCC CTG 1880
Ala Pro Val Pro Val Asn Ser Ser Tyr Glu Val Ala Ala His Ser Asp Ser His Ser Ser Gly Leu Glu Ala Glu Leu Glu His Leu Arg Gln Ala Leu 571

GAC AGT GGC CTA GAT ACA AAA GAA GCC AAA GAA AAA TTC CTC CAT GAA GTT GTT AAA ATG AGA GTG AAG CAG GAA GAG AAG CTA AAT GCT GCC TTG CAA 1979
Asp Ser Gly Leu Asp Thr Lys Glu Ala Lys Glu Lys Phe Leu His Glu Val Val Lys Met Arg Val Lys Gln Glu Glu Lys Leu Asn Ala Ala Leu Gln 604

GCC AAA CGC AGC CTA CAT CAG ⁶GAG CTG GAG TTC CTC AGA GTG GCA AAG AAG GAG AAA CTG AGA GAA GCA ACG GAG GCA AAA CGC AAC TTA AGG AAA GAG 2078
Ala Lys Arg Ser Leu His Gln Glu Leu Glu Phe Leu Arg Val Ala Lys Lys Glu Lys Leu Arg Glu Ala Thr Glu Ala Lys Arg Asn Leu Arg Lys Glu 637

ATT GAG CGT CTG AGA GCT GAG AAT GAG AAG AAA ATG AAG GAA GCA AAC GAG TCT CGG ATA CGG CTA AAG AGG GAA CTG GAA CAA GCC AGG CAG ATC CGG 2177
Ile Glu Arg Leu Arg Ala Glu Asn Glu Lys Lys Met Lys Glu Ala Asn Glu Ser Arg Ile Arg Leu Lys Arg Glu Leu Glu Gln Arg Gln Ile Arg 670

GTT TGC GAC AAG GGT TGT GAA GCT GGC AGG CTT CGG GCC AAG TAC TCT GCC CAG ATT GAG GAC CTA CAG GTT AAG CTT CAG CAT GCA GAG GCT GAC AGG 2276
Val Cys Asp Lys Gly Cys Glu Ala Gly Arg Leu Arg Ala Lys Tyr Ser Ala Gln Ile Glu Asp Leu Gln Val Lys Leu Gln His Ala Glu Ala Asp Arg 703

GAG CAG CTC CGA GCT GAC CTG ATG CAT GAG AGG GAG GCT CGA GAA CAC TTG GAA AAA GTA GTC AAG GAA CTT CAG GAA CAG CTG TGG CCT AAA TCA AGC 2375
Glu Gln Leu Ala Asp Leu Met His Glu Arg Glu Ala Arg Glu His Lys Val Val Lys Glu Leu Gln Glu Gln Leu Trp Pro Ser Arg 736

AGT CAA TCC AGC AGT GAA AAC ACA ACG AGC AAC ATG GAG AAT ⁷TAA ACC ACG TCG TCT AAT ACA ACA GAA TGA CAT ATA TGC ACA GTA AGG GAG GAT GGG 2474
Ser Gln Ser Ser Ser Glu Asn Thr Thr Ser Asn Met Glu Asn End 751

TGGGGTACGT GTGTAAGTGC ATGTGTGAGT AGTTGTGTCT TAACACACAG ATCTAGGAAT ATGGATTCTT ATTAGTTGGA AGGCAATGT TACTCTTAT AACAGAAGCA CTGAATTACG 2594

CCTCTTTTTT TTTCCAATCC ATATAGCACA ACATCTTACT ⁸GGCCTATAA AACACAAATG TGTTTATAAA CAAATACCTT TTAAGTCCAC AGCAATTTTT CTACTGGCAA ACTCCAAGCA 2714

AGCAGCATCC TCCAAC TAGA ATCAGAGTAA AAGGCAAGCA TGGCAGTGT TTTCATGTTGC CCTCTGCCCT GTCGGAACAT TTGGAAATTT AAAACAACAA TTTCTTATA AGCTATTATA 2834

AGTAATTCAT TACACAGACT TGGTATTAAA AAAAATTAA AAGATTITTT ATAACGAACC TTTAAAAGCA AAACAAAAAC CTTCGATGCA CAATTTTTAC GACTTGTAA AGGCTTTGGG 2954

ATTCITACTG CAGAAGCCCT TTGGTGATGA TGCCATTICA TTAGCAGTTT TTTTATAATC TGTCCTGTGG TTGTATGAGA ATTTCAGAGT GCTTTTCAAA GTTGATTTTT TTCTTAGAA 3074

ACAATCACCCT TCATTTCCTG TCCTGAACAC AAGAAGAAAG GAAGATGCAG GACTGTAAGG CGGTGGGGGA GGGCAGGAAG AGAAGATGGA CGCTTTGGAA TTATAAACCC AGCCTTACAG 3194

ACTTCAGTGT TTCAATACAC GCCATTGTTT CTAAGACGT CTCATTAAAT CGATGTGTTT AAAAGACTCA CTTCATCCAA GAGCATTCTA GCTTTAGGAA AAGAAGAAG GAAGTAAAGG 3314

AAGGAAATGG ATGACCTGTT AAGTTGGTTG AGAAATAAAG CAGAAGATGT GTTTTGAAGT CATCTGAAA TCCTCGCGTC AGCTTCAGT TCTCTGGAAA ACTCATCTTT GTTGCAACAT 3434

CTTACCATAG AATTTCAGTAT TTACCTACTT CTATTCTGAA CTGTTTGTCG GGATTTCGTG GCCCAAGGAG AGTGCAACAC CGCATTATTG GATACATACAG AAAAGAAAAA CCACGTTTTT 3554

GCTGCTGTA ATAAGCCTAC ATCTTTTTTA AAAGAAAAAC TTCTGTTTTT AAGAATAGAA ATTACTTTAA TTTTGGGATC CGAGCCGCAG CCTTGGAATA GAAATGCAGC CTACCATCAC 3674

TCTGCTTAC TACCATTGTT AGCGTCGTCG TTACTTTTTT TTTAACTGAC ACITTTGTCAG AACCTCAGCT TGCAATTIAT TCCATATTTT GGAAGTTTAC AAGTTCAGCA TTCTCGATTG 3794

TGCTCTCGAG ATGTAAAAAT CATCACCACC ATTTTCCACC ACGGCACACC TCGGCGCTCA TTTCCATGTA TGCAAAAGAA GAACTCAGTG GGTACAGAAT GCTACCAAT ACAAGGCAG 3914

CAGAGCAGCG TGCTGTGTT TGGGTTTACC AGCTGCGCTG CACGCTGTG GCTGTGAGG CTGGGAAGTG CTCAAATACA GTTGGTGCTT TACTGAATGA GAGAGGAGTT ATTTTACACC 4034

ACACACACTC ACCTCTGATA CACTCAAGCT CAGTGAAAAG TTGATCTGGG GCTGCAGTTG TGCCITCCAG CTCATTITTC CTCTCAGCAT CTCTATAGG CAATGCTGAC ACTTTTTTTT 4154

TAAACCTTAA AGAATAAAAA G 4175

FIG. 3. Diagram of the alternate mRNAs generated for the *c-ski* locus as deduced from the cDNA sequence analysis. The exons are not drawn to scale. The *c-ski* mRNAs are shown in relation to *v-ski* viral RNAs. Symbols: ■, noncoding regions; □, protein-coding regions of the cDNA; ▨, *gag* regions of the *gag-ski* fusion in transforming *ski* viruses. The relative positions of the putative translational initiation codon and the translation termination codons are also shown.

FB28 and of the FB27/29 type, S1 analysis was carried out on total RNA isolated from 8- or 12-day-old chicken embryos. A uniformly labeled single-stranded probe spanning the region between the *Kpn*I and *Hind*III sites of FB29/27 of FB28 (probes A and B, Fig. 1) was hybridized with total cellular RNA. Hybridization to mRNA and subsequent S1 digestion of a probe derived from FB29 produced protected fragments of 645 bp (Fig. 4A), indicating hybridization to mRNA of the FB27/29 type, and the 262- and 272-bp fragments expected if the probe hybridized with mRNA of the FB28 type. FB28 (probe B) protected a fragment of 534 bp (Fig. 4B). We did not observe smaller fragments from the hybridization of the FB28 probe to mRNAs of the FB27/29 type; however, S1 does not always cleave a DNA probe efficiently opposite a looped-out region of RNA. These experiments indicate that mRNAs corresponding to both the FB27/29 type, containing the second exon, and the FB28 type, missing the second exon, are expressed in normal cells.

Organization of the 3' coding exons. Sequence comparison of the three fibroblast clones demonstrated that FB28 and FB29 contain a segment (exon 6) that is absent in FB27. Examination of the genomic DNA at a position corresponding to the position where the cDNAs diverge reveals a consensus splice donor (22). The two cDNAs appear to derive from splicing events that either include or exclude exon 6. Downstream of this alternate exon, the 3' portions of all three cDNAs are identical.

To investigate whether both types of cDNAs represent normal cellular mRNAs, we carried out S1 analysis. Uniformly labeled single-stranded probes were generated from a 799-bp *Hind*III fragment (probe D, Fig. 1) from the 3' end of FB28 and a 497-bp *Hind*III fragment from the 3' end of FB27 (probe C) subcloned in M13mp18. Single-stranded probes were prepared and hybridized to 20 μ g of total RNA prepared from whole chicken embryos of different ages. Hybridization of the RNA to probe D (FB28) followed by S1 treatment produced a 799-bp fragment that would be expected if mRNA of FB28/29 type were present (Fig. 4B). We have failed to detect smaller fragments that would be generated by the FB28 probe hybridizing with mRNA of the FB27 type. S1 digestion of the FB27 probe (probe C) produced a

FIG. 4. S1 nuclease protection analysis of total RNA. Total RNA was prepared from 8-, 10-, 12-, 15-, and 17-day-old embryos by using standard protocols (6). Approximately 20 to 30 μ g of total RNA was used for nuclear S1 analysis according to standard procedures (15). Uniformly labeled single-stranded probes used for hybridization are shown schematically below each picture; thick lines represent the cDNA sequence, and thin lines represent the M13 sequence. The overall lengths of probes and expected lengths of protected fragments are also shown. The RNA used in the hybridization is indicated at the top of each lane. The number above each lane indicates the age (in days) of the embryo from which the RNA was prepared. (A) Results for probe A (see Fig. 1), containing a *Kpn*I-*Hind*III fragment of the FB27/29 type. This probe produced a fragment of 645 bp and two smaller fragments of 262 and 272 bp (→). (B) Results for probe B (see Fig. 1), containing a *Kpn*I-*Hind*III fragment of the FB28 type. This probe produced a fragment of 534 bp (←). Smaller fragments were not detected. (C) Results for probe C (see Fig. 1), containing a 497-bp *Hind*III fragment of the FB27 type linked to M13 sequences. The probe yielded a 497-bp fragment and two smaller fragments of 243 and 254 bp. Only the 479-bp fragment is marked (←). (D) Results for probe D (see Fig. 1), which is 1,116 bp long and contains a *Hind*III fragment of the FB28/29 type. Probe D produced a 799-bp fragment (→).

ALV	GAGA	GGAACAG	G	TT	A	CATCTGAGC	AA
c-ski	CCTT	GGAACAG	T	TT	-	CATCTGAGC	TC

FIG. 5. Sequence homology between *c-ski* and the p19 region of *gag* from ALV. The *c-ski* sequences are from positions 218 to 242, and the p19 sequences of the *gag* region are from positions 633 to 658. The homologous regions are boxed.

fragment of 497 bp as well as two smaller fragments of 243 to 254 bp (Fig. 4C). It is likely that these smaller fragments derived from S1 cleavage of the FB27 probe at the site where the probe differs from mRNAs of the FB28/29 type. These data confirm the existence of the FB28/29 mRNA but suggest that if FB27 mRNA exists, it is likely to be present at a lower level than FB28/29.

Differential splicing. A schematic illustration of the differential splicing of three *c-ski* mRNAs as deduced from the nucleic acid sequence analysis of the cDNA clones is shown in Fig. 3. S1 analysis of total RNA derived from chicken embryos has confirmed the existence of two classes of mRNAs, those that do and those that do not contain exon 2. Using a similar protocol, we were able to confirm the existence of mRNA that contains exon 6. We were unable to demonstrate by this technique the existence of mRNA lacking exon 6, which would correspond to FB27 cDNA; however, several lines of argument suggest that the FB27 cDNA is not a simple cloning artifact. The sequences absent from FB27 are bounded by apparent splice junctions (as judged by an examination of both the cDNA sequences and the available genomic DNA sequence). Although cDNAs are occasionally obtained that contain one or more introns, presumably because a partially processed mRNA was reverse transcribed, isolating cDNAs artifactually missing an exon is less likely on theoretical grounds and seems to occur rarely, if at all, in the manufacture of a cDNA library. For these reasons, we currently favor the interpretation that FB27 represents a real, if relatively rare, *c-ski* mRNA.

Comparison of *c-ski* with *v-ski*. Transduction of *c-onc* genes by retroviruses often results in truncation, deletion, or point mutation(s) in the genes (4). Comparison of *c-ski* and *v-ski* sequences showed that the *v-ski* gene is truncated at both 5' and 3' ends and represents only part of the *c-ski* coding region (Fig. 1 and 3). *v-ski* sequences begin at position 244 (the putative initiator ATG is at position 168) and end at position 1541 (Fig. 3). The biological significance of the 5' and 3' truncations is unknown. There is only one base change in the *v-ski* relative to *c-ski*, at position 1284, where *v-ski* has a C and *c-ski* has T. This base change alters the amino acid from Trp in *c-ski* to Arg in *v-ski*. Deletion analysis of *v-ski* implies that this amino acid change does not play a significant role in the transforming potential of *v-ski* (E. Stavnezer, personal communication). The biological activities of the cDNAs are being evaluated by expressing the coding regions from the cDNAs, using replication-competent retroviral vectors (9).

As shown in Fig. 5, 18 of 20 bp are identical between *c-ski* and the p19 region of *gag* in the parental ALV. This region of homology contains the 5' junction between viral sequences and *v-ski*. With such a long stretch of homology, it is impossible to assign the 5' recombinational joint precisely. No substantial homology can be seen precisely at the 3' *ski*/ALV joint. However, just downstream of the 3' junction there is an ALV sequence that is closely homologous to a segment of *c-ski* found just 3' of the *v-ski*/ALV junction. It is possible to invoke this region of homology in the alignment of the nucleic acids involved in the recombination event (22).

Transduction mechanism. The exact mechanism(s) by which retroviruses acquire cellular oncogenes remains uncertain. It has been suggested that as a first step, the viral DNA is integrated next to or within a cellular oncogene (4). The retroviral and cellular sequences can then be fused by DNA deletion events (4, 7) or RNA readthrough (8, 17); however, homology is not known to be involved in either of these processes. In the case of *ski*, a comparison of *c-ski* sequences with the p19-coding region implies that some event in the generation of the *ski* viruses involved homologous recombination. However, the homologous recombination event may be secondary. If the original event were nonhomologous and generated a viral genome that contained a direct repeat, it would be expected, as has been observed for Rous sarcoma virus, that sequences between the repeats could be lost by recombination during viral passage, presumably through copy choice during reverse transcription. It is possible, however, to propose models in which the oncogene is acquired by homologous recombination between the *c-onc* and the replication-competent viruses. Short stretches of homologies have been observed at both ends of viral oncogenes (5, 25) or only at the left-hand or 5' recombination joint (3, 18). The data we have do not allow us to determine whether the 5' homologous recombination event was primary or secondary.

We are grateful to Edward Stavnezer for sharing his unpublished data on the sequence of the genomic copy of *c-ski*. We also thank N. Nomura for sharing human *c-ski* sequence data, M. Powers for oligonucleotide synthesis, and H. Marusiodis for preparing the manuscript. P.S. thanks all of his colleagues in the laboratory for their help and constant encouragement.

This research was sponsored by the National Cancer Institute under Public Health Service contract NO1-CO-74101 with BRI.

LITERATURE CITED

1. Barkas, A. E., D. Brodeur, and E. Stavnezer. 1986. Polypeptides containing a domain encoded by the *v-ski* oncogene are located in the nuclei of SKV transformed cells. *Virology* **151**:131-138.
2. Ben-Neriah, Y., A. Bernardo, M. Paskind, G. Q. Daley, and D. Baltimore. 1986. Alternative 5' exons in *c-abl* RNA. *Cell* **44**:577-586.
3. Besmer, P., J. E. Murphy, P. C. George, F. Qui, P. J. Bergold, L. Lederman, H. W. Snyder, D. Bradeur, E. E. Zuckerman, and W. D. Hardy. 1986. A new acute transforming feline retrovirus and relationship of its oncogene *v-kit* with the protein kinase gene family. *Nature (London)* **320**:415-421.
4. Bishop, J. M. 1983. Cellular oncogenes and retroviruses. *Annu. Rev. Biochem.* **52**:301-354.
5. Bonner, T. I., S. B. Kerby, P. Suttrave, M. A. Gunnell, G. M. Mark, and U. R. Rapp. 1985. Structure and biological activity of human homologs of the *raf1* oncogene. *Mol. Cell. Biol.* **5**:1400-1407.
6. Chirgwin, J. M., A. M. Przbyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonucleases. *Biochemistry* **18**:5294-5299.
7. Czernilofsky, A. P., A. D. Levinson, H. E. Varmus, J. M. Bishop, E. Tischler, and H. M. Goodman. 1983. Corrections to the nucleotide sequence of the *src* gene of Rous sarcoma virus. *Nature (London)* **301**:736-738.
8. Herman, S. A., and J. M. Coffin. 1987. Efficient packaging of readthrough RNA in ALV: implications for oncogene transduction. *Science* **236**:845-848.
9. Hughes, S. H., J. J. Greenhouse, C. J. Petropoulos, and P. Suttrave. 1987. Adaptor plasmids simplify the insertion of foreign DNA into helper-independent retroviral vectors. *J. Virol.* **61**:3004-3012.
10. Leff, S. E., M. G. Rosenfeld, and R. M. Evans. 1986. Complex transcriptional units: diversity in gene expression by alternative

- RNA processing. *Annu. Rev. Biochem.* 55:1091–1117.
11. Levy, J. B., T. Dorai, L. H. Wang, and J. S. Brugge. 1987. The structurally different form of pp^{60c-src} detected in neuronal cells is encoded by a unique c-src mRNA. *Mol. Cell. Biol.* 7: 4142–4145.
 12. Li, Y., C. M. Turck, J. K. Teumer, and E. Stavnezer. 1986. Unique sequence, *ski*, in Sloan-Kettering avian retrovirus with properties of a new cell derived oncogene. *J. Virol.* 57:1065–1072.
 13. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 14. Martinez, R., B. M. Prevot, A. Bernardo, and D. Baltimore. 1987. Neuronal pp^{60c-src} contains a six amino acid insertion relative to its non-neuronal counterpart. *Science* 237:411–415.
 15. McGrath, J. P., D. J. Capon, D. H. Smith, E. Y. Chen, P. H. Seeburg, D. V. Goeddel, and A. D. Levinson. 1983. Structure and organization of the human kinase protooncogene and a related processed pseudogene. *Nature (London)* 304:501–506.
 16. Meijlink, F., T. Cunen, A. D. Miller, and I. M. Verma. 1985. Removal of a 67-base-pair sequence in the noncoding region of protooncogene *fos* converts it to a transforming gene. *Proc. Natl. Acad. Sci. USA* 82:4987–4991.
 17. Nilsen, T. W., P. A. Moroney, R. W. Goodwin, F. M. Rottman, L. B. Crittenden, M. A. Raines, and H.-J. Kung. 1985. c-erbB activities in ALV-induced erythroblastosis: novel RNA processing and promoter insertion result in expression of an amino truncated EGF receptor. *Cell* 41:719–726.
 18. Roebroek, A. J. M., J. A. Schalken, C. Onnekink, H. P. J. Bloemers, and W. J. M. Van de Ven. 1987. Structure of the feline c-fes/fps proto-oncogene: genesis of a retroviral oncogene. *J. Virol.* 61:2009–2016.
 19. Ryder, K., L. F. Lau, and D. Nathans. 1988. A gene activated by growth factors is related to the oncogene v-jun. *Proc. Natl. Acad. Sci. USA* 85:1487–1491.
 20. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain termination inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463–5467.
 21. Shaw, G., and R. Kamen. 1986. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* 46:659–667.
 22. Stavnezer, E., D. Brodeur, and L. A. Brennan. 1989. The v-ski oncogene encodes a truncated set of c-ski coding exons with limited sequence and structural relatedness to v-myc. *Mol. Cell. Biol.* 9:4038–4045.
 23. Stavnezer, E., A. E. Barkas, L. A. Brennan, D. Brodeur, and Y. Li. 1986. Transforming Sloan-Kettering viruses generated from the cloned v-ski oncogene by in vitro and in vivo recombination. *J. Virol.* 57:1073–1083.
 24. Stavnezer, E., D. S. Gerhard, R. C. Binari, and I. Balzas. 1981. Generation of transforming viruses in culture of chicken fibroblasts infected with avian leukosis virus. *J. Virol.* 39:920–934.
 25. Van Beveren, C., F. Van Straaten, T. Curren, R. Muller, and I. M. Verma. 1983. Analysis of FBJ-MuSV provirus and c-fos (mouse) gene reveals that the viral and cellular fos gene products have different carboxy termini. *Cell* 32:1241–1255.