Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 5291–5295, July 1989
Biochemistry

The v-*sea* oncogene of avian erythroblastosis retrovirus S13: Another member of the protein-tyrosine kinase gene family

(growth factor receptor)

DOUGLAS R. SMITH*†, PETER K. VOGT‡, AND MICHAEL J. HAYMAN§¶

*Imperial Cancer Research Fund Laboratories, Lincolns Inn Fields, P.O. Box 123, London WC2, England; ‡Department of Microbiology, University of Southern California School of Medicine, 2011 Zonal Avenue, HMR-401, Los Angeles, CA 90033-1054; and §Department of Microbiology, State University of New York, Stony Brook, NY 11794-8621

Contributed by Peter K. Vogt, April 20, 1989

ABSTRACT **The cloning and sequencing of the oncogene of the avian erythroblastosis virus S13 is described. The oncogene, termed v-*sea*, was found to be another member of the protein-tyrosine kinase gene family. The oncogene was fused in frame with the retrovirus S13 envelope gene, thus generating a fusion protein with a structure resembling that of a growth factor receptor. Sequence comparisons revealed that the v-*sea* gene was most closely related to the insulin receptor family of protein-tyrosine kinases, the greatest similarity being with the human *MET* oncogene.**

Study of retroviral oncogenes and their cellular counterparts has yielded many important clues concerning the causes and mechanisms underlying cancer development. There are now >40 known or putative oncogenes, which at present can be classified into five major groups according to the properties of the proteins they encode: protein kinases, GTP-binding proteins, nuclear proteins, growth factors, and proteins with multiple membrane-spanning domains (for review, see ref. 1). Of these groups the protein kinase family is, by far, the largest with most members being protein-tyrosine kinases. The protein-tyrosine kinases can be divided broadly into two structurally distinct groups: integral membrane proteins that span the membrane bilayer and those protein-tyrosine kinases that are extrinsic membrane proteins associated with the inner surface of the plasma membrane. The structural motif of the former group resembles that of certain growth factor receptors, and over the last few years it has become clear that certain oncogenes are, in fact, derived from cellular growth factor receptors [e.g., the *erbB* gene of avian erythroblastosis virus being derived from the epidermal growth factor receptor (2, 3) and the *fms* gene of feline sarcoma virus, McDonough strain, being derived from the macrophage colony-stimulating factor receptor (4)]. The physiological role of the cellular counterparts of the extrinsic tyrosine kinases remains unclear.

The demonstration that protein-tyrosine kinases could be oncogenic, together with the realization that certain growth factor receptors also had this enzymatic activity, has identified these enzymes as key players in various signal-transduction pathways. In addition, identification of additional members of the protein-tyrosine kinase family, by isolating and characterizing recently found oncogenes, led to the realization that this gene family is very large and that important information can be obtained when additional family members are identified (5, 6). In this report we describe the isolation of another member of this gene family, the *sea* gene of the S13 avian erythroblastosis retrovirus. The S13 virus causes sarcomas, erythroblastosis, and anemias on injection into young chicks (7). Recent characterization of this virus revealed that its genome length was 8.5 kilobases (kb), and this virus encoded normal-sized *gag* and *gag–pol* products but an abnormally sized *env* glycoprotein of 155 kDa, gp155 (8–10). Subsequent data revealed that this protein was cleaved into a normal-sized gp85 *env* product and a 70-kDa glycoprotein, gp70. *In vitro* kinase assays showed that the gp155 and gp70 proteins had an associated protein-tyrosine kinase activity that was temperature-sensitive in a mutant avian retrovirus S13 virus that was *ts* for transformation (11). These data indicated that avian erythroblastosis retrovirus S13 contained a protein-tyrosine kinase oncogene within its genome, positioned to encode a fusion protein between this oncogene, *sea*, and the *env* gene. We describe here the cloning and sequencing of the v-*sea* gene‖ and report information on transcription of the c-*sea* gene.

MATERIALS AND METHODS

Cells and Viruses. Rat fibroblasts transformed with temperature-sensitive avian retrovirus S13 *ts1* (11) were grown in Dulbecco's modified Eagle's medium supplemented with 5% fetal calf serum. The chicken cell lines used were the macrophage cell line HD11 (12), the B-cell tumor line 301 (provided by N. Bumstead, Houghton Poultry Research Station, Cambridge, U.K.), and liver cell line DU249 (13). Primary chick embryo fibroblasts and the cell lines were grown in the same medium supplemented with 2% chicken serum.

Molecular Cloning Techniques. Genomic DNA was prepared from S13-transformed rat cells (see *Results and Discussion*) as described (14), and 200 μg was digested to completion with endonuclease *Bam*HI. After fractionation on a 10–40% sucrose gradient, samples of each fraction were used to prepare a Southern blot as described (14). The blot was then hybridized with a nick-translated probe derived from the Rous avian sarcoma virus *env* gene to detect fractions containing retrovirus S13 *env* sequences. DNA from the positive fractions was concentrated by precipitation and cloned into λ vector L47.1 by using bacterial strain LE392 (15), as described (14). After two rounds of plaque purification and DNA preparation, the cloned 7-kb *Bam*HI fragment was subcloned into pUC12 to form plasmid pS13-1, which was then mapped with restriction enzymes.

DNA Sequencing. A 2.5-kb *Hin*dIII–*Kpn* I fragment was purified from pS13-1, digested with *Hae* III or *Alu* I, and the resulting fragments were subcloned into M13mp10, -mp11, -mp18, or -mp19 (Pharmacia) by using bacterial strains JM83 and JM101 (16). Other fragments were similarly subcloned

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: LTR, long terminal repeat.

†Present address: Collaborative Research Incorporated, 2 Oak Park, Bedford, MA 01730.
¶To whom reprint requests should be addressed.

‖The sequence reported in this paper has been deposited in the EMBL/GenBank data base (accession no. M25158).

into M13: *Xba* I–*Kpn* I, *Sph* I–*Cla* I, *Hin*dIII–*Pst* I, *Sph* I–*Pst* I, *Xba* I–*Sac* I, *Pst* I–*Xba* I, and *Pst* I–*Pst* I. These clones were sequenced by the dideoxy chain-termination method (17). The resulting sequences were aligned, and sequence acquisition was continued until both DNA strands were completely represented. Both the DNA sequence and the derived protein sequence were used to search for homologies in EMBL/GenBank and other data bases.

Northern (RNA) Blot Analysis. Poly(A)$^+$ RNA samples from various chicken cell lines were prepared. Five micrograms of each cell line plus 2 μg of retrovirus S13-transformed rat cell poly(A)$^+$ RNA was subjected to denaturing gel electrophoresis with end-labeled λ *Hin*dIII size standards and transferred to nylon membranes as described (14) (except that the gel buffer was 20 mM Mops/5 mM $NaPO_4$/1 mM EDTA, pH 7). The resulting blots were hybridized with nick-translated *Pst* I/*Pvu* II fragment as described (18).

RESULTS AND DISCUSSION

To characterize the oncogene of the avian erythroblastosis retrovirus S13, we decided to clone and sequence the v-*sea* gene. The strategy used was to screen rat fibroblasts transformed with the mutant retrovirus S13 (*ts1*) by using Southern blotting to identify those cells containing only a single copy of the retrovirus S13 genome. The rationale behind this approach was as follows. A molecular probe specific for the retrovirus S13 genome was unavailable; therefore, probes against the replication genes of the virus had to be used. Thus, if we used chicken cells transformed by retrovirus S13, distinguishing the retrovirus S13 genome from those of the endogenous viruses would be difficult. By using mammalian cells we could avoid this problem because the mammalian cells contain no cross-hybridizing sequences. In addition, by using the temperature-sensitive mutant of retrovirus S13, *ts1* (11), we could confirm that the rat cells were, indeed, transformed by the S13 retrovirus and not by spontaneous transformation events.

Rat cells that were temperature-sensitive for transformation were screened by Southern blotting, and a clone of rat cells was identified that contained a single provirus. In this clone a single 7-kb *Bam*HI fragment was identified that hybridized with both avian retrovirus *env* and long terminal repeat (LTR) probes but not to *gag*-specific probes (data not shown). Because we knew the *sea* oncogene was fused to *env*, this fragment was an ideal candidate to contain *sea*. This *Bam*HI fragment was cloned by using the λ phage vector L47 and then subcloned into vector pUC12 for further analysis, as described. After digestion with a variety of restriction enzymes and probing with the various replication gene probes, the restriction map shown in Fig. 1 was prepared. A comparison of this restriction map with that of Rous sarcoma virus from the *Bam*HI site in the polymerase gene (Fig. 1) reveals that the maps are very similar for the region spanning the *pol* and 5′ portion of the *env* gene up to the *Pst* I site, and then the sequences diverge. These data are consistent with the schematic representation of the 3′ portion of the retrovirus S13 genome shown in Fig. 1, in which *sea* is located between the *env* and LTR sequences. The *Kpn* I–*Cla* I fragment spanning this putative *env–sea* region (Fig. 1), cloned into the avian retrovirus vector pRCAS (19) could transform chicken embryo fibroblasts in a temperature-dependent fashion; this occurred with an efficiency of ≈1000 agar colony-forming units per μg of DNA transfected. Similarly, when the *Kpn* I–*Cla* I fragment was exchanged for *erbB* in an avian erythroblastosis virus vector, the resultant virus could transform erythroid cells in temperature-dependent fashion (20). These biological data, then, indicated that this region contained the *sea* oncogene.

Nucleotide Sequence of Retrovirus S13 v-*sea*. The 2538-base pair (bp) *Hin*dIII–*Kpn* I fragment that contains this region was sequenced using the dideoxy nucleotide chain-termination method (17); Fig. 2 shows the sequence of this fragment. The sequence begins near the 3′ end of the *env* gene and extends beyond the 3′ LTR, presumably into rat cellular sequences. The *env* sequences show 96% homology with those of the RAV 2 strain avian leukosis virus, and the region between the *env* gene and the 3′ LTR is 92% homologous to a similar region in Rous sarcoma virus. A nonviral sequence of ≈1085 bp, which contains a single open reading frame, is located between the *env* gene and this viral 3′ noncoding region; this open reading frame is the *sea* gene. The 3′ junction point (located at nucleotide 1775, Fig. 3) is clearly

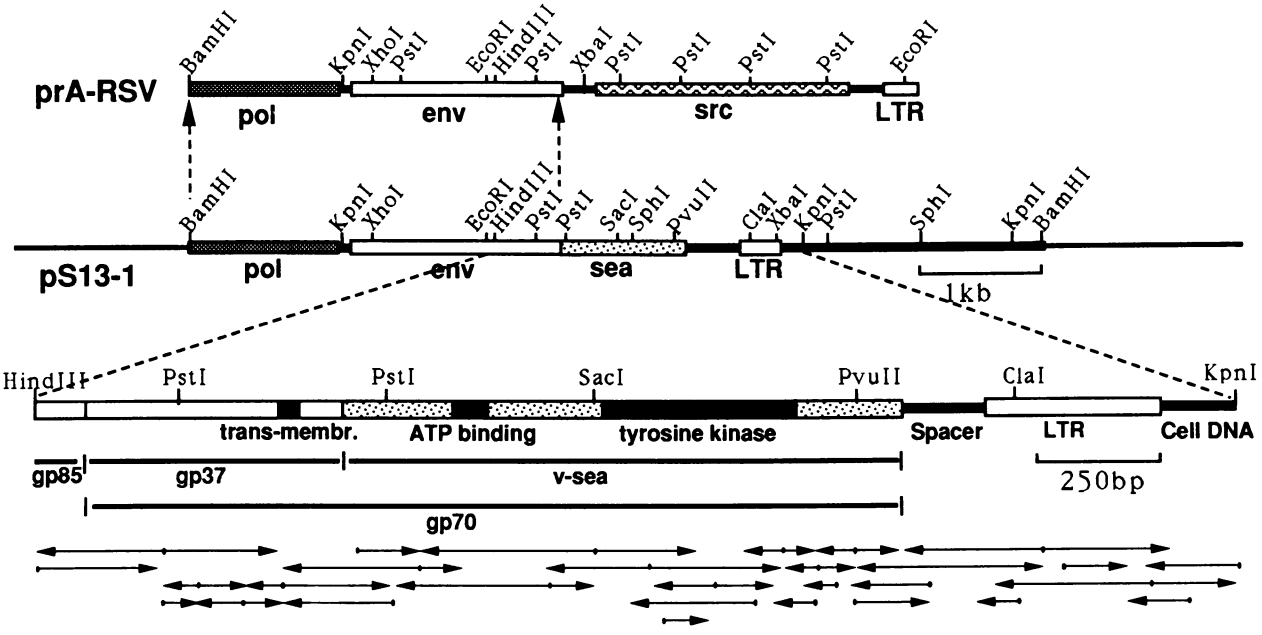

FIG. 1. Restriction map of clone pS13-1 and DNA sequencing strategy. Restriction sites unique to the cloned insert are shown. The *pol* and *env* region (left) contains a number of restriction sites that correspond in position to sites on the Rous avian sarcoma virus map. Specific fragments sequenced are indicated at bottom.

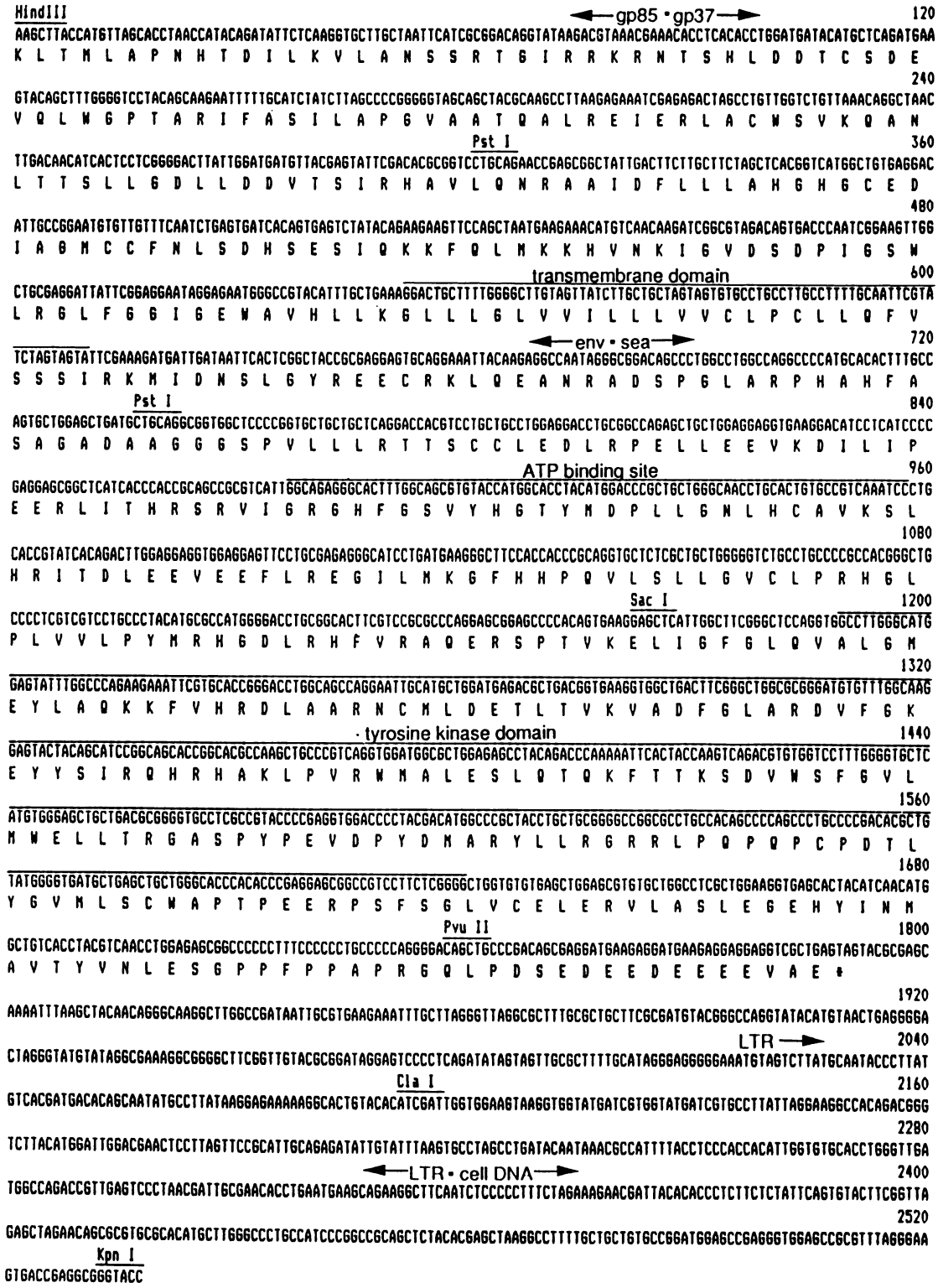

FIG. 2. DNA sequence of the 3′ portion of avian retrovirus S13 including the *sea* oncogene. Various significant regions are indicated above the line containing the DNA sequence (extending in the 5′–3′ direction from the *Hin*dIII site in the *env* gene). Below the DNA sequence is the protein sequence predicted from the reading frame shown (continuing from *env*; the other two reading frames contain multiple stop codons). Important restriction sites are also shown.

defined by comparison with avian viral sequences, and the termination codon for *sea* is located within the viral 3′ noncoding region. Position of the 5′ junction is less distinct and probably occurs between nucleotides 680 and 690, using RAV 2 strain for comparison. This region of the *env* gene is not highly conserved, making precise allocation of the junction point difficult. The alignment shown in Fig. 3 places the 5′ junction point just 3′ of the *env* stop codon, with readthrough into the *sea* gene allowed by a single-base deletion 14 bp upstream of the normal *env* stop codon. Interestingly, both junction points occur within regions of the viral sequence that are especially purine-rich and that contain numerous small repeats. These regions may be of significance concerning the mechanism of transduction of the c-*sea* se-

```
      635                                                  5' Junction region
S13   TCGGCTACCGCGAGGAGTGCAGGAAA-TTACAAGAGGCCAATAGGGCGGACAGCCCTGGCCTGGCCAGGCCCCA
      ::::::: :::::::: :  :  ::: ::::: ::::::  ::: :     ::::::       :   :
RAV2  TCGGCTATCGCGAGGAATATAAAAAAATTACAGGAGGCTTATAAG-----CAGCCCGAAAGAAGAGCGTAGGCG
                                                RAV2 env Stop Codon

                                  3' Junction Point
      1743                                 *            S13 Stop Codon
S13   GCCCGACAGCGAGGATGAAGAGGATGAAGAGGAGGAGGTCGCTGAGTAGTACGCGAGCAAAATTTAAGCTACAA
      :   :    : ::          : :   : :  :::::::::::::::::: :::::: ::::::::::::::::
ASV   GGGAGTAGCCCAGCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGTAAAATTTAAGCTACAA
```

FIG. 3. Junction regions where avian retroviral S13 sequences are joined to nonviral sequences of the *sea* oncogene. *, Nucleotide 1776 at 3' junction. Both these regions contain numerous copies of the sequence GAAGA or related sequences that differ from this by one nucleotide (the 3' region contains 11 copies, many of which overlap). Both ends of the transduced sequence contain perfect tandem repeats (5' region: TGGCC and 3' region: TGAAGAAGA). ASV, Rous avian sarcoma virus; RAV2, RAV2 strain of avian leukosis virus.

quences because this situation is reminiscent of transduction of *kit* sequences into the *gag* gene (21). More definitive information on the mechanism of transduction awaits the cloning and sequencing of the c-*sea* sequences.

The structure of the *env–sea* protein predicted from the above sequence agrees with previous data on the retrovirus S13 protein (22). The *env* gene is fused in frame with the *sea* sequence, generating a protein in which the external glycosylated domain and the transmembrane domain would all be encoded by the 95 kDa of *env* sequences. The *sea* sequence would encode ≈42 kDa of protein with similarities to known protein-tyrosine kinases (see below). Thus, the *env–sea* protein would have a predicted molecular mass of 137 kDa and a structure similar to a growth factor receptor. However, because *env* is providing the signal sequence and the transmembrane sequence, it is unclear whether the c-*sea*-encoded protein would be a transmembrane protein or would instead belong to the extrinsic class of protein-tyrosine kinases.

A comparison of the predicted amino acid sequence of v-*sea* with that of other protein-tyrosine kinases reveals similarities within the ATP-binding site and the catalytic domains (Fig. 4), the greatest similarity being with those kinases belonging to the insulin receptor family (6). The human *MET* oncogene had the highest similarity of ≈72%. However, this similarity drops to <25% both 5' and 3' to the tyrosine kinase domain, which would indicate that *sea* is probably not the avian homolog of the human *MET* gene. Similarities between v-*sea* and the other tyrosine kinases (Fig. 4) reflect the conservation of residues necessary for kinase function. v-*sea* has two tyrosine residues at the tyrosine position that is conserved in the catalytic domain of protein-tyrosine kinases. This feature is conserved in human insulin receptor, insulin-like growth factor 1 receptor, *Drosophila* sevenless gene, and *met*, *ros*, and *trk* oncogenes (6). All these genes are either known growth factor receptors or protein-tyrosine kinases, containing membrane-spanning regions characteristic of growth factor receptors. Therefore, from the sequence similarities, the cellular *sea* gene will probably be a growth factor receptor. Definitive proof for this identification will require isolation and characterization of the cellular gene.

A striking feature of the predicted sequence of the v-*sea* protein is the large number of acidic residues at the C terminus of the protein; eleven acidic residues occur in a stretch of fourteen residues, of which three acidic residues are encoded by the viral sequences (Fig. 3) and, therefore, their presence would not be expected in c-*sea*. This run of acidic amino acids is reminiscent of a similar stretch of acidic residues in the v-*erbB* oncogene. In both cases, these acidic residues are located ≈40 residues C-terminal of the end of the tyrosine-kinase domain. Both these viruses can cause erythroblastosis, and it is known for v-*erbB* that a region C-terminal to the kinase domain is important for erythroid cell transformation (for review, see ref. 3). One could thus speculate that this highly acidic region may play a role in erythroblast transformation.

Chicken c-*sea* Transcription. To determine the size of the RNA transcripts of the c-*sea* gene, Northern blot analysis was performed on size-fractionated RNA isolated from normal chicken fibroblasts and various transformed chicken cells. Fig. 5 shows that, in all cells tested, we detected two mRNA species of sizes 7.0 kb and 3.0 kb, although the levels did vary among the different cell types. In addition, in fibroblasts, weakly hybridizing bands of ≈4.2 kb and 5.0 kb were also detected; these may represent *sea*-related genes. Whether the two major *sea* transcripts represent the products of two different genes or spliced variants of an initial transcript is not yet clear; cloning of c-*sea* cDNA and then using this sequence to analyze the structure of these two transcripts should resolve this question.

```
          * *  *                *
v-sea     GRGHFGSVYHGTYMDPLL--GNLHCAVKSLHRITDLEEVEEFLREGILMKGFHHPQVLSLLGVCLPRHGLPLVVLPYMRHGDLRHFVRAQ
  met     GRGHFGCVYHGTLLDNDG--KKIHCAVKSLNRITDIGEVSQFLTEGIIMKDFSHPNVLSLLGICLRSEGSPLVVLPYMKHGDLRNFIRNE
 insR     GQGSFGMVYEGNARDIIKGEAETRVAVKTVNESASLRERIEFLNEASVMKGFTCHHVVRLLGVVSKGQPTLVVMELMAH-GDLKSYLRSL
  abl     GGGQYGEVYEGVWKKYSLT-----VAVKTLKE-DTM-EVEEFLKEAAVMKEIKHPNLVQLLGVCTREPPFYIITEFMTY-GNLLDYLREC

                                                                                    ++
v-sea     ERS---------PTVKELIGFGLQVALGMEYLAQKKFVHRDLAARNCMLDETLTVKVADFGLARDVFGKEYYSIRQHRHAKLPVRWMALE
  met     TMN---------PTVKDLIGFGLQVAKGMKYLASKKFVHRDLAARNCMLDEKFTVKVADFGLARDMYDKEYYSVHNKTGAKLPVKWMALE
 insR     RPEAENNPGRPPPTLQEMIQMAAEIADGMAYLNAKKFVHRDLAARNCMVAHDFTVKIGDFGMTRDIYETDYY--RKGGKGLLPVRWMAPE
  abl     NRQEV--------SAVVLLYMATQISSAMEYLEKKNFIHRDLAARNCLVGENHLVKVADFGLSRLMTGDTYTAH---AGAKFPIKWTAPE

v-sea     SLQTQKFTTKSDVWSFGVLMWELLTRGASPYPEVDPYDMARYLLRGRRLPQPQPCPDTLYGVMLSCWAPTPEERPSFS
  met     SLQTQKFTTKSDVWSFGVVLWELMTRGAPPYPDVNTFDITVYLLQGRRLLQPEYCPDPLYEVMLKCWHPKAEMRPSFS
 insR     SLKDGVFTTSSDMWSFGVVLWEITSLAEQPYQGLSNEQVLKFVMDGGYLDQPDNCPERVTDLMRMCWQFNPNMRPTFL
  abl     SLAYNKFSIKSDVWAFGVLLWEIATYGMSPYPGIDPSQVYELLEKDYRMKRPEGCPEKVYELMRACWQWNPSDRPSFA
```

FIG. 4. Comparison of the DNA-encoded protein sequence of v-*sea* with other oncogenes in the region of the tyrosine kinase domain. Correspondences with the *sea* sequence are indicated by the use of bold characters: *, important conserved amino acids in the ATP-binding region; ++, important conserved tyrosine residues in the catalytic domain. One-letter amino acid code is used.

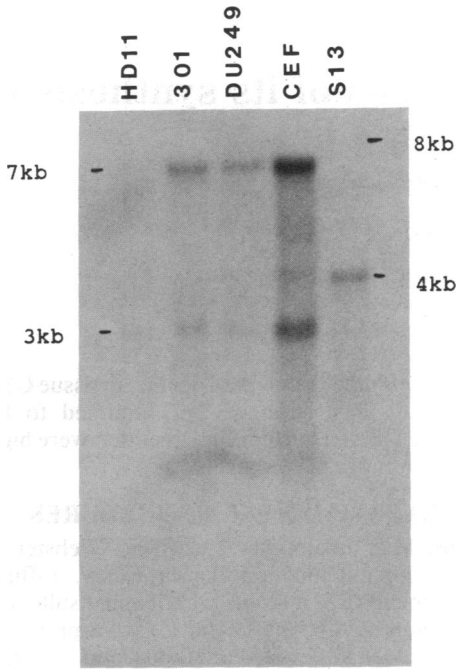

FIG. 5. Northern blot of chicken cell lines hybridized to a v-*sea*-specific probe to show expression of the c-*sea* gene in different cell types. Sizes of visible poly(A)$^+$ transcripts are indicated. The 8-kb and 4-kb transcripts present in S13 cells correspond to full-length genomic and putative spliced 5′ LTR–*env*/*sea* transcripts, respectively, found in S13-transformed rat cells.

This work was supported, in part, by grants from the National Cancer Institute to M.J.H. and to P.K.V.

1. Reddy, E. P., Skalka, A. M. & Curran, T., eds. (1988) *The Oncogene Handbook* (Elsevier, Amsterdam).
2. Downward, J., Yarden, Y., Mayes, E., Scrace, G., Totty, N., Stockwell, P., Ullrich, A., Schlessinger, J. & Waterfield, M. (1984) *Nature (London)* **307,** 521–527.
3. Hayman, M. J. (1986) *Trends Genet.* **2,** 260–263.
4. Sherr, C. J., Rettenmier, C. W., Sacca, R., Roussel, M. F., Look, A. T. & Stanley, E. R. (1985) *Cell* **41,** 665–676.
5. Hunter, T. (1987) *Cell* **50,** 823–829.
6. Hanks, S. K., Quinn, A. M. & Hunter, T. (1988) *Science* **241,** 42–52.
7. Stubbs, E. L. & Furth, J. (1935) *J. Exp. Med.* **61,** 593–616.
8. Benedict, S. H., Maki, Y. & Vogt, P. K. (1985) *Virology* **145,** 154–164.
9. Beug, H., Hayman, M. J., Graf, T., Benedict, S. H., Wallbank, A. M. & Vogt, P. K. (1985) *Virology* **145,** 141–153.
10. Hayman, M. J., Kitchener, G., Vogt, P. K. & Beug, H. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 8237–8241.
11. Knight, J., Zenke, M., Disela, C., Kowenz, E., Vogt, P. K., Engel, J. D., Hayman, M. J. & Beug, H. (1988) *Genes Dev.* **2,** 247–258.
12. Leutz, A., Beug, H., Walter, C. & Graf, T. (1988) *J. Biol. Chem.* **263,** 3905–3911.
13. Langlois, A. J., Lapis, K., Ishisaki, R., Beard, J. W. & Bolognesi, D. P. (1974) *Cancer Res.* **34,** 1457–1464.
14. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1981) in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
15. Murray, N. E., Brammer, W. J. & Murray, K. (1977) *Mol. Gen. Genet.* **150,** 53–61.
16. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9,** 309–318.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
18. Singh, L. & Jones, K. W. (1984) *Nucleic Acids Res.* **12,** 5627–5637.
19. Hughes, S. H., Greenhouse, J. J., Petropoulos, C. J. & Sutrave, P. (1987) *J. Virol.* **61,** 3004–3012.
20. Zenke, M., Kahn, P., Disela, C., Vennstrom, B., Leutz, A., Keegan, K., Hayman, M. J., Choi, H.-R., Yew, N., Engel, J. D. & Beug, H. (1988) *Cell* **52,** 107–119.
21. Besmer, P., Murphy, J. E., George, P. C., Qiu, F., Bergold, P. J., Lederman, L., Snyder, H. W., Jr., Brodeur, D., Zuckerman, E. E. & Hardy, W. D. (1986) *Nature (London)* **320,** 415–421.
22. Hayman, M. J. (1987) *Pathol. Immunopathol. Res.* **6,** 390–400.