Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 7862–7866, November 1988
Biochemistry

Mammalian *ets-1* and *ets-2* genes encode highly conserved proteins*

(protooncogene family)

DENNIS K. WATSON, MARY J. MCWILLIAMS, PETER LAPIS, JAMES A. LAUTENBERGER, CLIFFORD W. SCHWEINFEST, AND TAKIS S. PAPAS

Laboratory of Molecular Oncology, National Cancer Institute, Frederick, MD 21701-1013

Communicated by Peter Duesberg, June 24, 1988

ABSTRACT **Cellular *ets* sequences homologous to v-*ets* of the avian leukemia virus E26 are highly conserved. In mammals the *ets* sequences are dispersed on two separate chromosomal loci, called *ets-1* and *ets-2*. To determine the structure of these two genes and identify the open reading frames that code for the putative proteins, we have sequenced human *ets-1* cDNAs and *ets-2* cDNA clones obtained from both human and mouse. The human *ETS1* gene is capable of encoding a protein of 441 amino acids. This protein is >95% identical to the chicken c-*ets-1* gene product. Thus, the human *ETS1* gene is homologous to the chicken c-*ets-1* gene, the protooncogene that the E26 virus transduced. Human and mouse *ets-2* cDNA clones are closely related and contain open reading frames capable of encoding proteins of 469 and 468 residues, respectively. Direct comparison of these data with previously published findings indicates that *ets* is a family of genes whose members share distinct domains.**

Retroviral transforming genes originate by transduction of cellular sequences by retroviruses. Such an event is typically associated with the viral genome recombining with a normal cellular gene, termed a protooncogene, resulting in the truncation and damage of the cellular gene in such a fashion as to confer oncogenic potential to the newly recombined cell–retrovirus hybrid gene. The viral (v-)*ets* sequence was originally identified as a cell-derived sequence present in the genome of the avian leukemia virus E26. The *gag*, *myb*, and *ets* sequences of the E26 genome encode a single transforming protein of 135,000 daltons (p135) that is capable of inducing erythroblastosis and myeloblastosis in infected chickens (1–3).

To better understand the process of conversion of a protooncogene to a viral oncogene, the structure and function of the normal cellular gene must be determined and compared to those of the viral oncogene. For this purpose, we have analyzed the molecular structures and transcription patterns of the avian and mammalian *ets* genes. The chicken cellular (c-)*ets-1* gene is present on a single chromosomal locus of >60 kilobases (kb) of genomic DNA (4, 5). Nucleotide sequence analysis of chicken genomic DNA and cDNA clones and direct comparison to the v-*ets* sequence demonstrated that the chicken c-*ets-1* protooncogene has 27 unique amino acids at the amino terminus and 13 unique amino acids at the carboxyl terminus (6). Thus, the viral oncogene and the cellular protooncogenes are not identical. In humans, there are three *ets* genes, located on two different chromosomes, termed *ETS1*, *ETS2*, and *ERG*. All three genes are transcriptionally active and differentially regulated, yielding distinct RNAs (7–11). All are on chromosomal locations involved in translocations associated with specific malignancies (12). Also, the position of *ETS2* and *ERG* genes at 21q22.3 has implicated these genes in Down syndrome, and at least *ETS2* is triplicated in trisomy 21 and microduplications of chromosome 21 (partial trisomy) associated with Down syndrome (13). Recently, the equivalent of the human *ETS2* gene has been identified in chicken (14). The *ets*-related sequences in *Drosophila* (15), sea urchin (16), and *Xenopus* (Z. Q. Chen and L. A. Burdett, personal communication) have been isolated and characterized. In this paper we describe the predicted gene products of the human *ETS1* and *ETS2* genes and mouse *Ets-2* gene.† We will compare the conserved protein domains encoded by these genes to the chicken protooncogene and to the *ets*-related coding sequences of other species such as *Xenopus*, sea urchin, and *Drosophila*.

MATERIALS AND METHODS

Isolation of *ets-1* and *ets-2* cDNA Clones. A cDNA library prepared from human K562 cells in λgt10 was the generous gift of E. Cananni (17). A mouse cDNA library was constructed from BALB/c 3T3 fibroblast RNA by C.W.S., and a human cDNA library was constructed from CEM T-lymphoblast RNA by D.K.W., using λgt10 as a vector according to published procedures (17). The libraries were propagated in *Escherichia coli* strain C600 *hfl* and 5×10^5 plaques were screened (18). The restriction map of a partial human *ETS2* cDNA clone, designated cDNA14, has been described (8). To identify *ETS2* cDNA clones with larger insert DNA, a 240-base-pair *Hin*fI fragment from the 5′ end of cDNA14 was used as a probe for screening the human library. The v-*ets* probe E1.28, used for analysis of the mouse and human libraries, consists of a 1.28-kb *Bgl* I fragment of v-*ets* DNA subcloned into the *Eco*RI site of pBR322 (8).

Analysis of Cloned DNA. Initially, 5 human *ETS2* clones, 2 human *ETS1* clones, and 10 mouse *Ets-2* clones were plaque-purified. The phage DNA from these clones was digested with *Eco*RI under standard conditions, and the restriction fragments were resolved by electrophoresis in 1% agarose gels. Immobilized DNA (19) was hybridized under stringent conditions [50% (vol/vol) formamide/5× SSC at 42°C; 1× SSC = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0] with E1.28 and various *ets-1* and *ets-2* probes. In addition to the 240-base-pair *Hin*fI fragment described above, human *ETS2* clones were further characterized by using H33, a genomic *ets-2*-specific probe from the 3′ end of human *ETS2* (7, 8). Human *ETS1* clones were distinguished by *ets-1*-specific probe pRD6K (7, 8). Mouse *Ets-2* cDNA clones were distinguished from mouse *Ets-1* clones by hybridization with probes, derived from a mouse genomic library, specific for *ets-1* [a 0.87-kb *Bgl* II fragment homologous to human clone pRD700] or *ets-2* [a 1.27-kb *Pst* I fragment homologous to

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.

*Presented in part at the Third Annual Meeting on Oncogenes, Hood College, Frederick, MD, July 7–11, 1987.

†The sequences reported in this paper are being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession nos. J04101, J04102, and J04103).

H33] loci (8). After hybridization, the filters were washed initially at room temperature for 30 min with 2× SSC/0.1% $NaDodSO_4$ and then twice at 42°C for 15 min with 0.1× SSC/0.1% $NaDodSO_4$. DNA from one human *ETS2* clone, λK3A, with *Eco*RI inserts of 2.3 kb and 0.4 kb, was subcloned and the plasmid with the *ets-2* insert of 2.3 kb was designated pK3A. A 2.2-kb *Hin*dIII fragment was isolated from the human *ETS1* phage λJ10, subcloned into pUC18, and designated pJ10-2. λJ10 contains a total insert of ≈5 kb. Of the 8 mouse *Ets-2* clones, 4 contained inserts of similar size. One of these clones, pA3, had an insert of 3.4 kb and was subcloned and used for further analysis. Restriction fragments of these subclones and clone cDNA14 were obtained, end-labeled, and sequenced by the Maxam and Gilbert technique (20). In addition, some of the human clone, pK3A, was sequenced by the dideoxy method of Sanger *et al.* (21).

RESULTS

Nucleotide Sequence Analysis of the Human *ETS1* Gene. The partial nucleotide sequence of the human *ETS1* cDNA clone is shown in Fig. 1. Alignment of the human *ETS1* cDNA sequence with that of the chicken *ets-1* cDNA (6) reveals common open reading frames with identical coding capacity of 441 amino acids each. These genes are highly homologous at the nucleotide (85%) and amino acid (95%) levels. The methionine at position 1 of the human ETS1 protein is identical to that of the chicken. In both cases, this methionine is preceded by an in-frame terminator (TAA at position −75

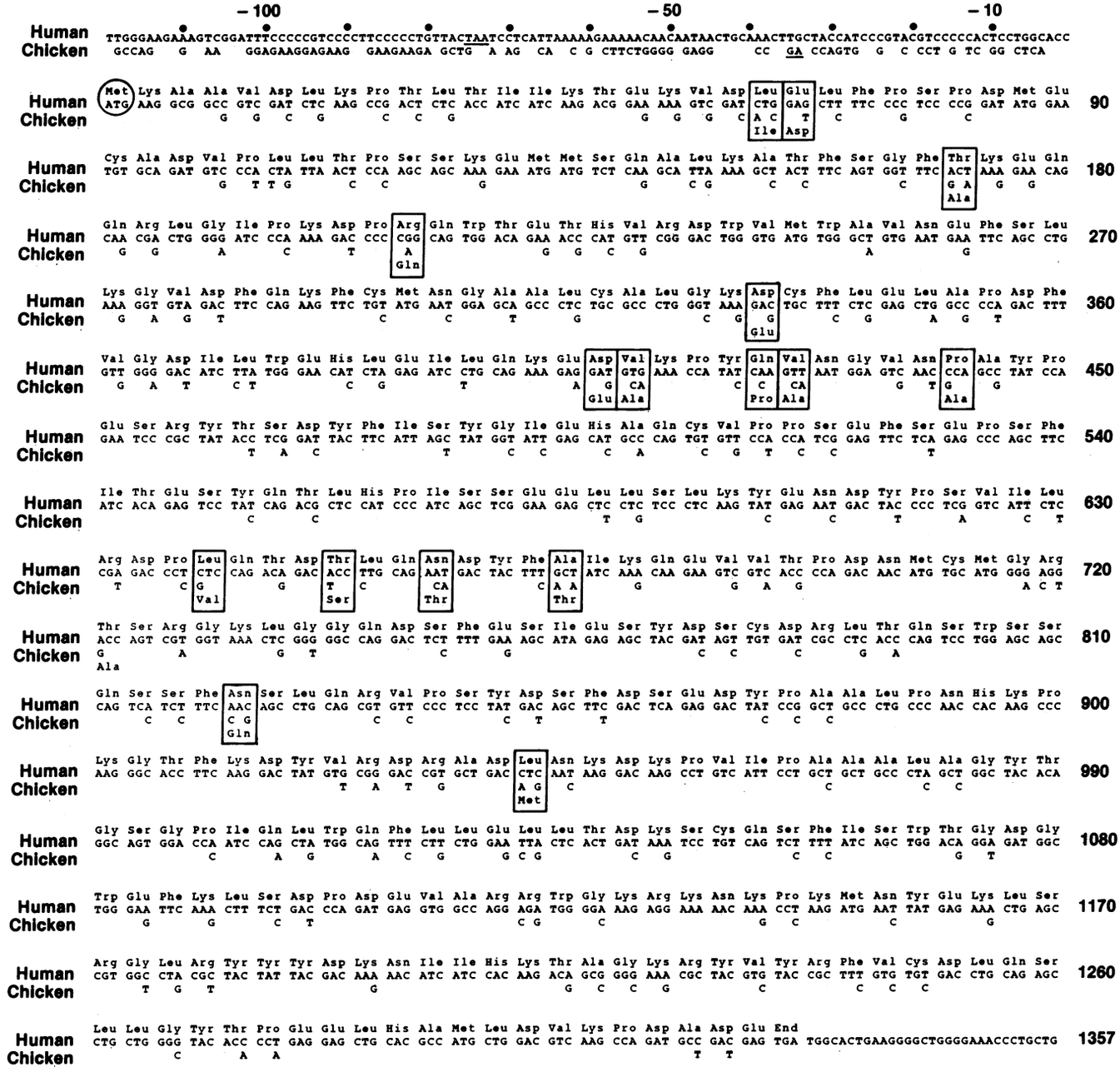

FIG. 1. Alignment of human and chicken *ets-1* sequences. The nucleotide sequence of the human *ETS1* sequence and the predicted amino acids are presented. Nucleotide changes occurring in chicken c-*ets-1* are shown and those affecting the amino acid sequence are highlighted by boxes. The presumptive start methionine is circled and upstream in-frame termination codons are underlined. The chicken c-*ets-1* sequence used for comparison stops at an *Eco*RI site (GAATTC) immediately following the termination codon at 1324–1326 (TGA).

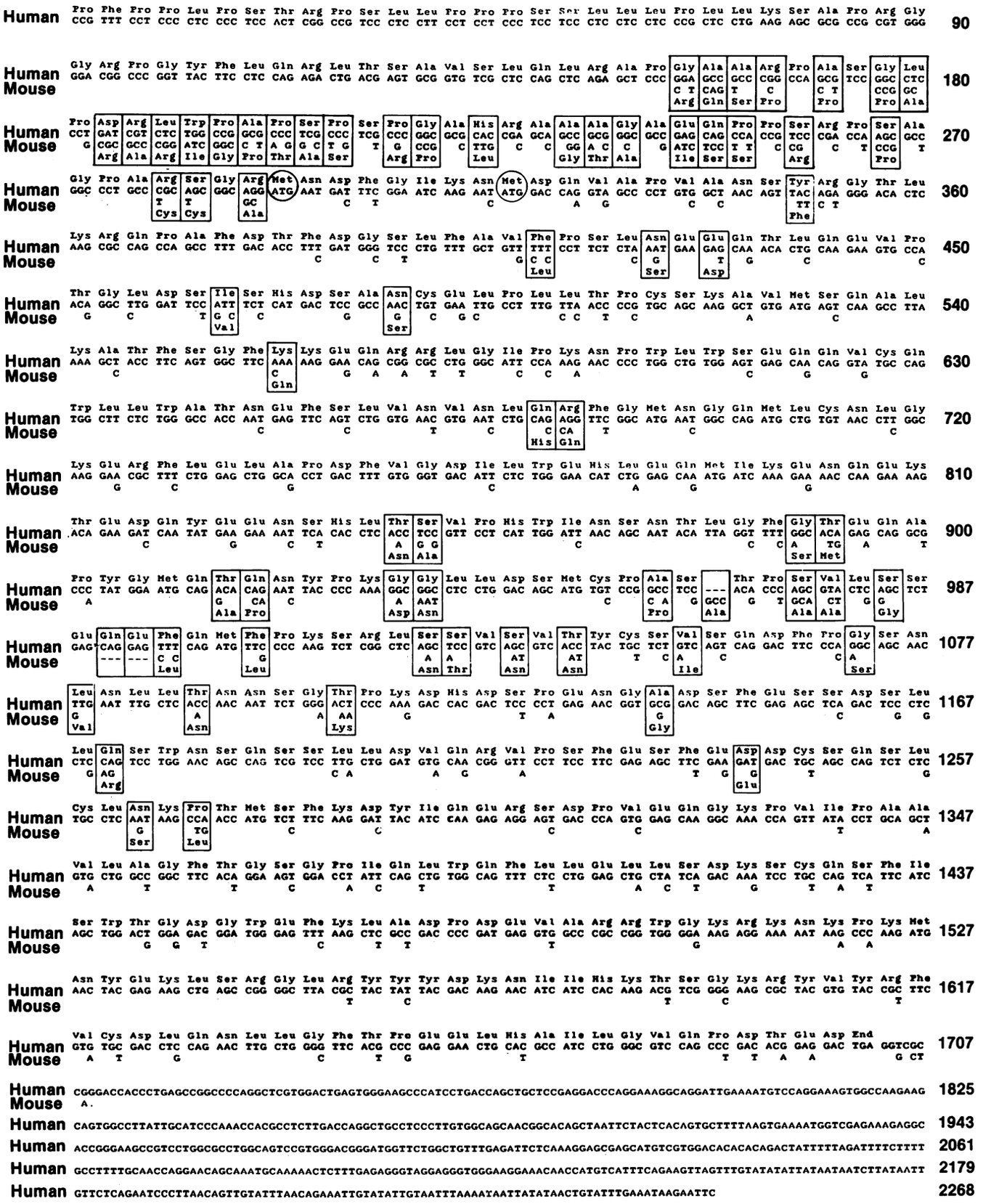

FIG. 2. Comparison of human and mouse *ets-2* gene sequences. Alignment required insertion of an in-frame gap in both human and mouse sequences. Amino acid differences between human and mouse are boxed. Presumptive start methionines are circled.

in human and TGA at position −36 in chicken). The sequence preceding the methionine codon (human), GGCACCATGA, is similar to the Kozak consensus sequence (22, 23); thus, the methionine at position 1 is the putative protein initiator. The open reading frame terminates at the TGA codon at position 1324. The size of the putative human ETS1 protein initiating at

position 1 and terminating at position 1324 is 50,407 daltons, similar to that identified in human cells (24).

Nucleotide Sequence Analysis of the Human and Mouse *ets-2* Genes. Several overlapping human and mouse *ets-2* cDNA clones were sequenced and compared (Fig. 2). Direct comparison of these two *ets-2* genes reveals a high level of homology at the nucleotide (85%) and amino acid (91%) levels. There are two methionine codons, at nucleotide positions 292 and 316, that are potential initiators in accordance with Kozak's consensus. Both genes have a common terminator located at position 1699. It is likely that one of these methionine residues is the true initiator, based upon the size of the protein detected in mammalian cells. This protein has been identified in human (24) and mouse (9) cells as a 56,000-dalton product, similar to that predicted by the human (53,001 daltons) and mouse (53,827 daltons) *ets-2* DNA sequences when methionine encoded by nucleotides 292–294 is the first amino acid.

DISCUSSION

The human *ETS1* gene is highly conserved, with >95% of its predicted amino acids identical to those of the chicken *ets-1* protooncogene. Of the 16 amino acid differences, only four are nonconservative. Thus, the *ets-1* genes from human and chicken code for proteins having a 99% conserved amino acid homology.

The predicted products of the *ets-2* genes in human and mouse are highly conserved and found to be >91% identical. This conservation is further supported by the observation that *ets*-specific antibody is capable of recognizing the 56,000-dalton nuclear protein of both human (24) and mouse (9). The *ets-2* coding sequences of mouse and human encode a consensus glycosylation site (Asn-Xaa-Ser/Thr) beginning at amino acid 268 (nucleotide 1093); such a site is absent in human *ETS1*. A common feature of the *ets-1* and *ets-2* products is the sequence of basic amino acid residues present at positions 377–383 (nucleotides 1129–1149) and 405–411 (nucleotides 1504–1524) in *ets-1* and *ets-2*, respectively; these resemble the nuclear-transit signal found in proteins such as the simian virus 40 large tumor antigen (25), consistent with our data that the *ets-2* gene product is a nuclear protein (24).

A diagrammatic representation of the homology of the predicted amino acid sequences of the *ets* genes characterized in this laboratory, from human to *Drosophila*, is shown in Fig. 3. The predicted proteins are compared to the chicken c-*ets-1* protooncogene product, since this gene was the one transduced by the E26 virus. The black areas represent regions of amino acid identity and the white areas are the regions of divergence. In the protooncogene we can identify three distinct domains. One domain, C, is located at the carboxyl terminus of the protooncogene product. This domain is highly conserved in all the genes we have characterized, with >90% amino acid sequence identities between diverse species ranging from human to *Drosophila*. A second domain, A, is located near the amino terminus and is less homologous than domain C (e.g., 66% sequence identity exists between *Xenopus* and human). We have not as yet been able to identify this domain in *Drosophila* or sea urchin, and it may indeed be absent in these species. A third domain, B, is present and highly conserved between the chicken c-*ets-1* and the human *ETS1*. This region is absent in human *ERG*, conserved within *ets-2* (84% between human and mouse; 55% between human and *Xenopus*), and divergent between *ets-1* and *ets-2*.

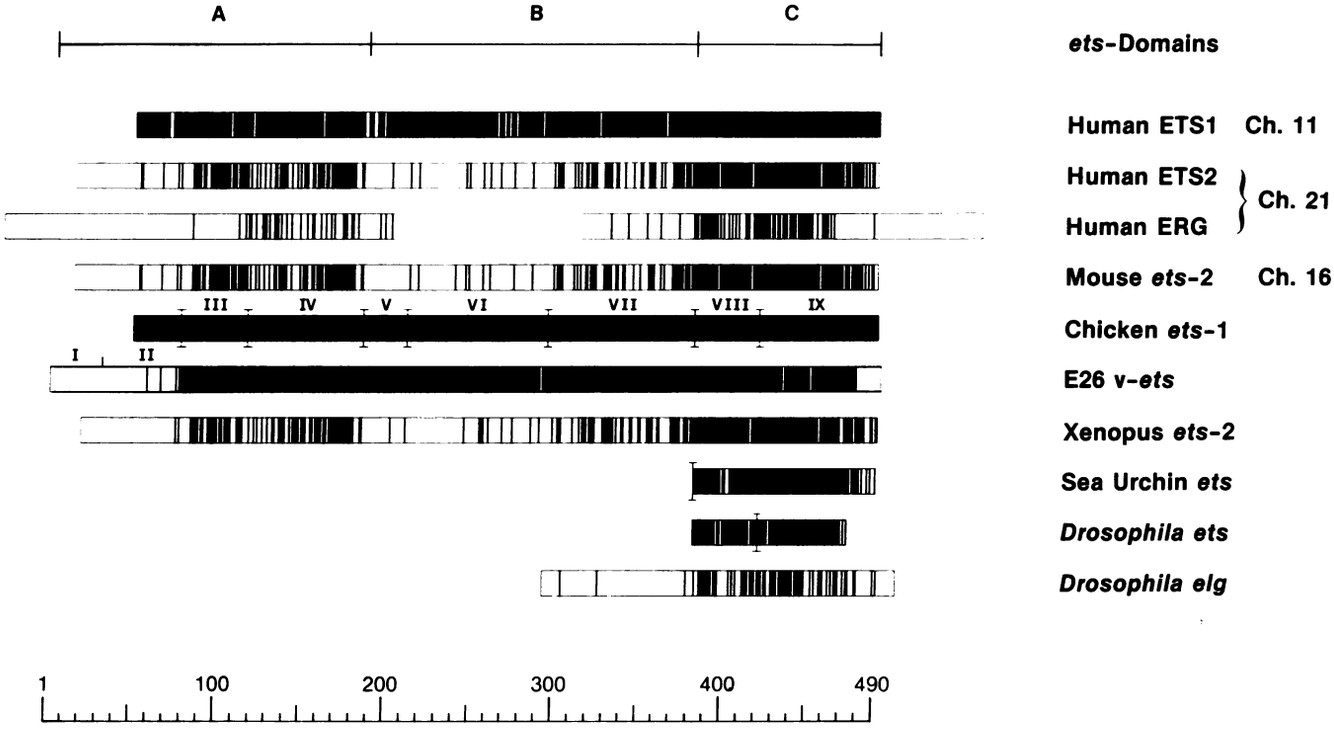

FIG. 3. Comparison of *ets*-related deduced amino acid sequences. The sequences displayed are human *ETS1* (this paper), human *ETS2* (this paper), human *ERG* (10, 11), mouse *Ets-2* (this paper), chicken c-*ets-1* (5, 6), E26 v-*ets* (1), *Xenopus ets* (Z. Q. Chen and L. A. Burdett, personal communication), sea urchin *ets* (16), *Drosophila ets* (15), and *Drosophila elg* (26). These sequences were compared to the chicken c-*ets* by a graphics program we have developed. Prior to display the sequences were aligned by the program LINEUP of the University of Wisconsin Genetics Computer Group (UWGCG) software package (27, 28). Each sequence is displayed as a box, and positions identical with the chicken sequence are displayed as black vertical lines. Single unmatched residues between matches thus appear as thin white lines. The dotted lines in the *ERG* sequence represent a large gap introduced to maximize homology. The brackets represent known exon boundaries; Roman numerals denote the v-*ets*-homologous domains of chicken c-*ets-1*. Uppercase letters (A, B, C) define the hypothetical *ets* domains (see *Discussion*). Scale at bottom represents number of amino acid residues. Ch, chromosome.

We can conclude from the above studies that *ets* is a family of genes that can be divided into two distinct classes. Class I consists of genes that contain all three chicken c-*ets-1*-homologous regions, domains A, B, and C (Fig. 3). This category is best exemplified by the human *ETS1* and the chicken c-*ets-1* genes. Class II consists of genes that contain only two v-*ets*-homologous regions, domains A and C. These are best exemplified by the human *ETS2*, mouse *Ets-2*, *Xenopus ets-2*, and human *ERG* genes. A third group of *ets* genes, which cannot be classified at this point, contain only domain C (*Drosophila* and sea urchin). These could be a unique class having only this domain conserved and the others diverged, or they could be members of the other classes. This determination will have to await the cDNA isolation and sequencing for correct class assignment. This category includes the *Drosophila ets*, the sea urchin *ets*, and the *Drosophila elg* genes (Fig. 3). It is interesting that the two classes map to different chromosomes in mammals. Class I *ets* genes map to human chromosome 11 and mouse chromosome 9, and class II *ets* genes map to human chromosome 21 and mouse chromosome 16.

We wish to thank Dr. R. Ascione, Dr. A. Seth, and L. Pribyl for critical reading of this manuscript; G. Smythers for assistance with computer analysis; and K. Cannon for typing the manuscript.

1. Nunn, M. F., Seeburg, P. H., Moscovici, C. & Duesberg, P. H. (1983) *Nature (London)* **306,** 391–395.
2. Leprince, D., Gegonne, A., Coll, J., de Taisne, C., Schneeberger, A., Lagrou, C. & Stehelin, D. (1983) *Nature (London)* **306,** 395–397.
3. Nunn, M., Weiher, H., Bullock, P. & Duesberg, P. (1984) *Virology* **139,** 330–339.
4. Gegonne, A., Leprince, D., Duterque-Coquilloud, M., Vandenbunder, B., Flourens, A., Ghysdael, J., Debuire, B. & Stehelin, D. (1987) *Mol. Cell. Biol.* **7,** 806–812.
5. Watson, D. K., McWilliams, M. J. & Papas, T. S. (1988) *Virology* **164,** 99–105.
6. Watson, D. K., McWilliams, M. J. & Papas, T. S. (1988) *Virology,* in press.
7. Watson, D. K., McWilliams-Smith, M. J., Nunn, M. F., Duesberg, P. H., O'Brien, S. J. & Papas, T. S. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7294–7298.
8. Watson, D. K., McWilliams-Smith, M. J., Kozak, C., Reeves, R., Gearhart, J., Nunn, M. F., Nash, W., Fowle, J. R., III, Duesberg, P., Papas, T. S. & O'Brien, S. J. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1792–1796.
9. Bhat, N. K., Fisher, R. J., Fujiwara, S., Ascione, R. & Papas, T. S. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3161–3165.
10. Reddy, E. S. P., Rao, V. N. & Papas, T. S. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6131–6135.
11. Rao, V. N., Papas, T. S. & Reddy, E. S. P. (1987) *Science* **237,** 635–639.
12. Sacchi, N., Watson, D. K., Guerts van Kessel, A. H. M., Hagemeijer, A., Drabkin, H. D., Patterson, D. & Papas, T. S. (1986) *Science* **231,** 379–382.
13. Sacchi, N., Nalbantoglu, J., Sergovich, F. R. & Papas, T. S. (1988) *Proc. Natl. Acad. Sci. USA,* **85,** 7675–7679.
14. Boulukos, K. E., Pognonec, P., Begue, A., Galibert, F., Gesquiere, J. C., Stehelin, D. & Ghysdael, J. (1988) *EMBO J.* **7,** 697–705.
15. Pribyl, L. J., Watson, D. K., McWilliams, M. J., Ascione, R. & Papas, T. S. (1988) *Dev. Biol.* **127,** 45–53.
16. Chen, Z.-Q., Kan, N. C., Pribyl, L., Lautenberger, J. A., Moudrianakis, E. & Papas, T. S. (1988) *Dev. Biol.* **125,** 432–440.
17. Shtivelman, E., Lifshitz, B., Gale, R. P. & Canaani, E. (1985) *Nature (London)* **315,** 550–554.
18. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–192.
19. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
20. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
21. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
22. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
23. Kozak, M. (1987) *J. Mol. Biol.* **196,** 947–950.
24. Fujiwara, S., Fisher, R. J., Seth, A., Bhat, N. K., Showalter, S. D., Zweig, M. & Papas, T. S. (1988) *Oncogene* **2,** 99–103.
25. Dingwall, C. & Laskey, R. A. (1986) *Annu. Rev. Cell Biol.* **2,** 367–390.
26. Pribyl, L. J., Watson, D. K., McWilliams, M. J., Ascione, R. & Papas, T. S. (1988) in *Advances in Gene Technology: Protein Engineering and Production,* eds. Brew, K., Ahmad, F., Bialy, H., Black, S., Fenna, R., Puett, D., Scott, W.A., Van Brunt, J., Voellmy, R. W., Whelan, W. J. & Woessner, J. F. (Cambridge Univ. Press, Cambridge, U.K.), p. 111.
27. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
28. Devereux, J., Haeberli, P. & Marquess, P. (1988) *Sequence Analysis Software Package User Manual: Version 5* (Univ. of Wisconsin Biotechnology Center, Madison, WI).