Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 6511–6515, July 1992
Cell Biology

Identification of murine homologues of the *Drosophila* Son of sevenless gene: Potential activators of *ras*

(tyrosine kinase/receptor/signal transduction/ras)

DAVID BOWTELL*†, PING FU*, MICHAEL SIMON‡, AND PAUL SENIOR*

*Howard Florey Institute of Experimental Physiology and Medicine, University of Melbourne, Parkville 3052, Australia; and ‡Howard Hughes Medical Institute and Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720

Communicated by Gerald M. Rubin, March 31, 1992

ABSTRACT **Several findings suggest that signals from tyrosine kinases are transduced, at least in part, through ras proteins. These findings include (*i*) blockage of the transforming activity of constitutively active tyrosine kinases by inhibiting *ras* function and (*ii*) genetic screens in *Caenorhabditis elegans* and in *Drosophila* that identified *ras* genes as downstream effectors of tyrosine kinases. The recently isolated *Drosophila* gene Son of sevenless (*Sos*) is postulated to act as a positive regulatory link between tyrosine kinase and ras proteins by catalyzing exchange of GDP for GTP on ras protein. Such exchange proteins have been reported in extracts of mammalian cells but have not been previously characterized at a molecular level. As *Sos* appears to function in this role in *Drosophila*, we sought to isolate a vertebrate counterpart(s). We have characterized two widely expressed murine genes with a high degree of homology to *Sos*. Hybridization with human DNA and RNA indicates a high degree of conservation of these genes in other vertebrates.**

Protein tyrosine kinases (PTKs) are a superfamily of genes whose protein products regulate many aspects of cellular proliferation, survival, and differentiation (1). Many of the extracellular ligands that modulate the activity of individual PTKs have been identified. However, less is known of the intracellular factors that interact either directly with the catalytic domain of tyrosine kinases or form more distal parts of the tyrosine kinase signaling pathway. There is compelling evidence (for review, see ref. 1) to suggest that proteins such as phospholipase C-γ and the c-*raf1* serine threonine kinase are phosphorylated, and their activity is modulated, as a result of the activation of tyrosine kinases. Ras proteins have also been identified as important components in PTK signal transduction in mammalian cells (2, 3), *Caenorhabditis elegans* (4), and *Drosophila* (5, 6). However, these proteins must represent only some of the components of the tyrosine kinase signaling pathway.

We have previously attempted to study the mechanism of signaling by PTKs by conducting a systematic genetic screen in *Drosophila*, involving the PTK receptor sevenless (5). Sevenless is required for the differentiation of R7 photoreceptor cells in the *Drosophila* eye. By creating circumstances where the sevenless signaling pathway was barely adequate for R7 formation, we were able to uncover mutations in several genes probably involved in signal transduction by the sevenless protein. Genetic analyses suggested that these genes are involved in additional developmental pathways in *Drosophila*. In particular, at least four genes appear to be also involved in signal transduction by another tyrosine kinase, the epidermal growth factor receptor. Of the four genes likely to be general components in tyrosine kinase signaling two have been characterized at a molecular level (5, 6). These are the *Drosophila Ras1* gene (7) and Son of sevenless (*Sos*, ref. 8).

A region of the *Sos* gene shows a high degree of homology with several yeast genes that appear to activate ras proteins by promoting GDP–GTP exchange (9–11), suggesting an equivalent role for *Sos*. Although *Sos* represents the first putative metazoan activator of ras protein to be molecularly characterized, its discovery was anticipated, as (*i*) the equilibrium dissociation rate of GDP from ras suggests an obligatory requirement for such exchange-promoting factors (12) and (*ii*) partially purified proteins that promote that addition of GTP to ras have been identified in mammalian cells by several groups (13–16). It seemed plausible that Sos-like proteins could account for, at least part of, this activity in mammalian cells. We, therefore, attempted to isolate mammalian homologues of the *Sos* gene by hybridizing the *Drosophila* gene to a mouse cDNA library under conditions of low stringency. We have identified two related, but distinct, murine genes each with extensive homology to *Sos*.§ The genes, designated mouse Son of sevenless 1 and 2 (m*Sos-1* and m*Sos-2*), are widely expressed during development and in adult tissues, consistent with a role as positive regulators of the ubiquitously expressed *ras* genes.

MATERIALS AND METHODS

Cloning of Murine Son of Sevenless Homologues. An embryonic mouse eye cDNA library (random-bred Swiss at day 17 of gestation, D.B., unpublished work) was screened with a ^{32}P-labeled fragment of the *Drosophila Sos* gene that corresponded to amino acids 841–1303 (5). Approximately 1 $\times$ 10^6 recombinant phage were hybridized in 5$\times$ standard saline citrate (SSC)/5$\times$ Denhardt's solution/5 mM EDTA/herring testes DNA at 100 μg/ml/0.1% SDS at 65°C for 18 hr and then washed in 2$\times$ SSC/0.1% SDS at 50°C. cDNA clones that weakly hybridized were purified, and their cDNA inserts were subcloned into Bluescript II (Stratagene). Partial DNA sequence was obtained and confirmed the isolation of two *Sos*-related genes, m*Sos-1* and m*Sos-2*. Subsequent screening of the embryonic-eye cDNA library with these clones identified cDNA of $\approx$4.9 kilobases (kb) (clone 2.1) and 5.3 kb (clone 10) that corresponded to m*Sos-1* and m*Sos-2*, respectively. cDNA corresponding to m*Sos-1* and m*Sos-2* were present in the cDNA library at $\approx$1:150,000 recombinants.

Cloning of 5′ Sequences by PCR. Amplification of the 5′ end of the mSos-1 mRNA was done as described (17) by using polyadenylylated RNA derived from an adult mouse brain. PCR products were subcloned into M13 and phage containing m*Sos-1* sequences identified by hybridization to a ^{32}P-labeled oligonucleotide probe.

Abbreviations: PTK, protein tyrosine kinase; m*Sos-1* and m*Sos-2*, murine homologues 1 and 2, respectively, of *Drosophila Sos*.
†To whom reprint requests should be addressed.
§The sequences reported in this paper have been deposited in the GenBank data base (accession nos. Z11574, Z11578, and Z11664).

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.

DNA Sequencing. Double-stranded dideoxynucleotide chain-termination DNA sequencing was done either on nested deletions of cDNA clones 2.1 (m*Sos-1*) and 10 (m*Sos-2*) or by using specific oligonucleotide primers and standard methods (18). The sequence of both DNA strands was assembled with the IntelliGenetics suite of programs. Alignment of the predicted amino acid sequences was made using the CLUSTAL V program (19) and displayed using the BOXSHADE program (K. Hoffman, obtained from the European Molecular Biology Library).

RNA Isolation and Analysis. Polyadenylylated RNA was isolated from tissues and cell lines by disruption in proteinase

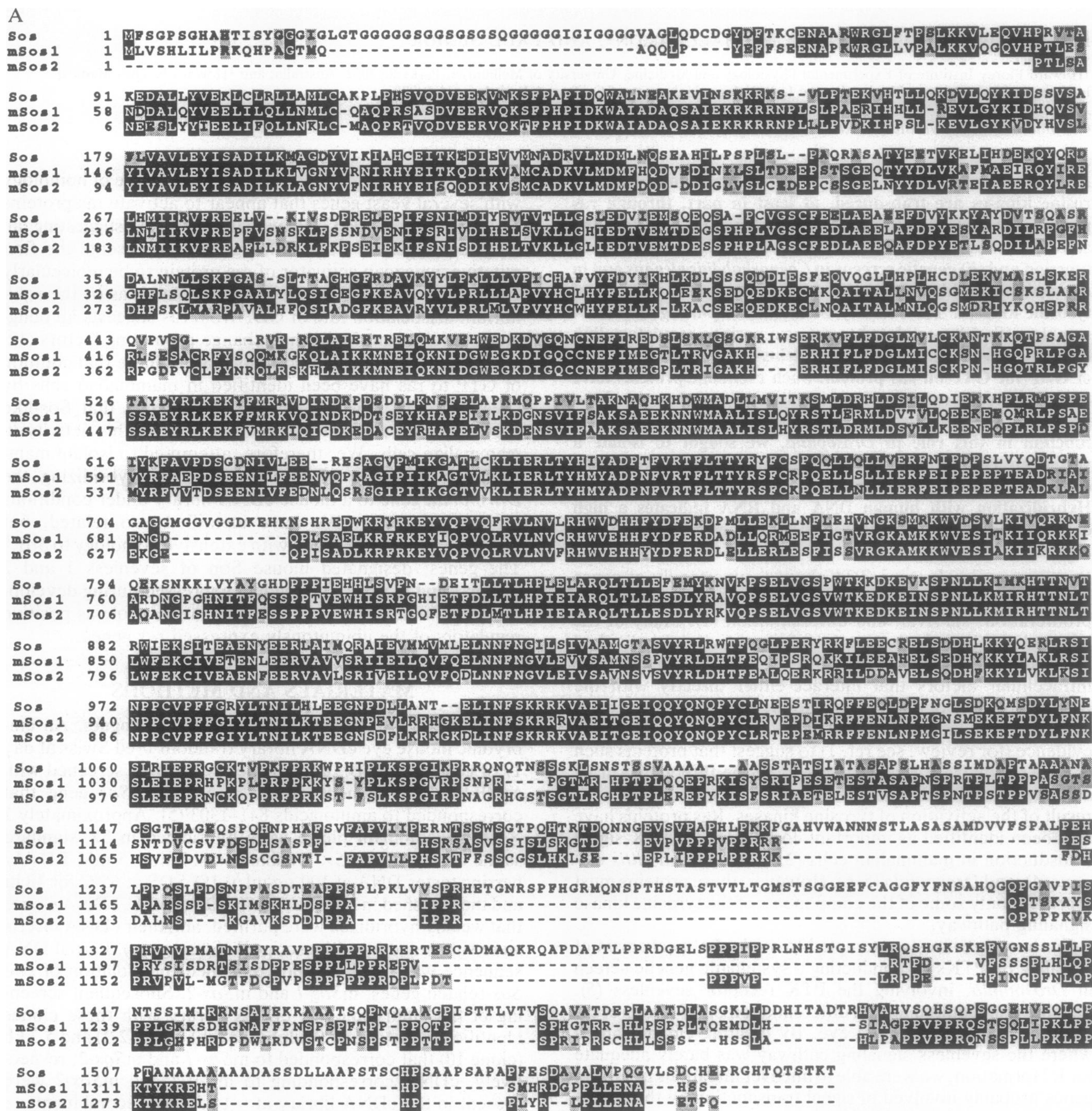

FIG. 1. (*A*) Alignment of the predicted amino acid sequences of the m*Sos-1*, m*Sòs-2*, and *Drosophila Sos* genes. Identical residues are in black boxes; conservative substitutions are in grey boxes. The available m*Sos-2* sequence extended to within 85-amino acid residues of the beginning of the *Drosophila Sos* sequence. The molecular mass of the predicted mSos-1 protein is ≈150 kDa. (*B*) Alternative amino-terminal coding regions of the m*Sos-1* gene identified by PCR. Most PCR products (type 2) terminated before reaching a potential initiating methionine in a highly G+C region that could not be processed by reverse transcriptase. An alternative PCR product was identified (type 1), which diverged from this *sequence* as *in*dicated and extended to a potential methionine initiation codon and 5′ stop codon.

K and SDS and subsequent oligo(dT) affinity chromatography, as described (20). Two micrograms of RNA from each source was subjected to formaldehyde/agarose gel electrophoresis (18), transferred overnight to Hybond-C super (Amersham) in 20× SSC and then baked at 80°C for 2 hr. Filters were prehybridized for 4–6 hr in 50% formamide/5× SSC/5× Denhardt's solution/5 mM EDTA/herring testes DNA at 100 $\mu g \cdot ml^{-1}$/0.5% SDS at 42°C and then hybridized for 18 hr. The ^{32}P-labeled probes were from nucleotides 299 to 4464 of m*Sos-1* and from nucleotides 1 to 3801 of the available m*Sos-2* sequence. Washes were performed in 0.2× SSC/0.3% SDS at 65°C, and the filters were autoradiographed for 3–7 days at −70°C in the presence of an intensifying screen. Molecular weights were estimated by using RNA standards (BRL).

Southern Analysis. DNA was isolated from either human peripheral blood leukocytes, BALB/c mouse liver, or Oregon R strain *Drosophila melanogaster* by proteinase K digestion and phenol extraction (18). Ten to fifteen micrograms was restricted with either *Eco*RI or *Hin*dIII and separated on a 1% agarose gel. The DNA was transferred overnight onto a Hybond-N^+ filter (Amersham) in 0.5 M NaOH/1.5 M NaCl, rinsed in 2× SSC, and baked for 2 hr. The prehybridization solution and the probes used for hybridization were as described above for RNA analysis, except that formamide was not used, and the hybridizations were done at 65°C. Washes were performed in 0.2× SSC/0.3% SDS at 65°C, and the filters were autoradiographed for 3 days at −70°C in the presence of an intensifying screen.

RESULTS AND DISCUSSION

Identification of Two Murine Homologues of Son of Sevenless. A fragment of the *Drosophila Sos* gene was used to screen a mouse embryonic-eye cDNA library at low stringency. Complementation analysis of weakly hybridizing clones identified two cDNA families, and partial DNA sequence confirmed that two distinct genes, termed m*Sos-1* and m*Sos-2*, with homology to the *Drosophila* gene had been isolated. Further screening of the eye and an adult-brain cDNA library using these clones resulted in the isolation of several nearly full-length cDNA but failed to identify additional genes. A complete open reading frame was obtained for m*Sos-1*, and the predicted amino acid sequence obtained for m*Sos-2* extended to within 85 amino acids of the amino-terminal end of Sos (Fig. 1*A*). Substantial difficulty was encountered in the cloning of the 5′ ends of both m*Sos-1* and m*Sos-2*. A total of 24 independent cDNAs were isolated for both genes, and although most exceeded 3 kb in length, none contained sequences corresponding to the amino-terminal end of the two genes. Amplification of the 5′ end of the mSos-1 mRNA by PCR was done to obtain a complete open reading frame. Clones corresponding to two alternative amino-terminal ends were obtained, one of which (type 1) extended to an in-frame methionine codon that was preceded by a stop codon (Fig. 1*B*). The predicted molecular mass of this protein is 150 kDa. The other class (type 2) invariably terminated in a highly G+C-rich region that could not be processed by reverse transcriptase, even when thermostable reverse transcriptase or prior denaturation of RNA with methyl mercury were used. This G+C-rich sequence corresponds to a glycine-rich region close to the initiating methionine codon in the *Drosophila Sos*-predicted protein (5, 6).

Alignment of the two murine genes showed that they share ≈67% amino acid identity, with the lowest degree of similarity residing in the final 270 amino acids (41% identity, Fig. 1*A*). No significant areas of homology were identified when

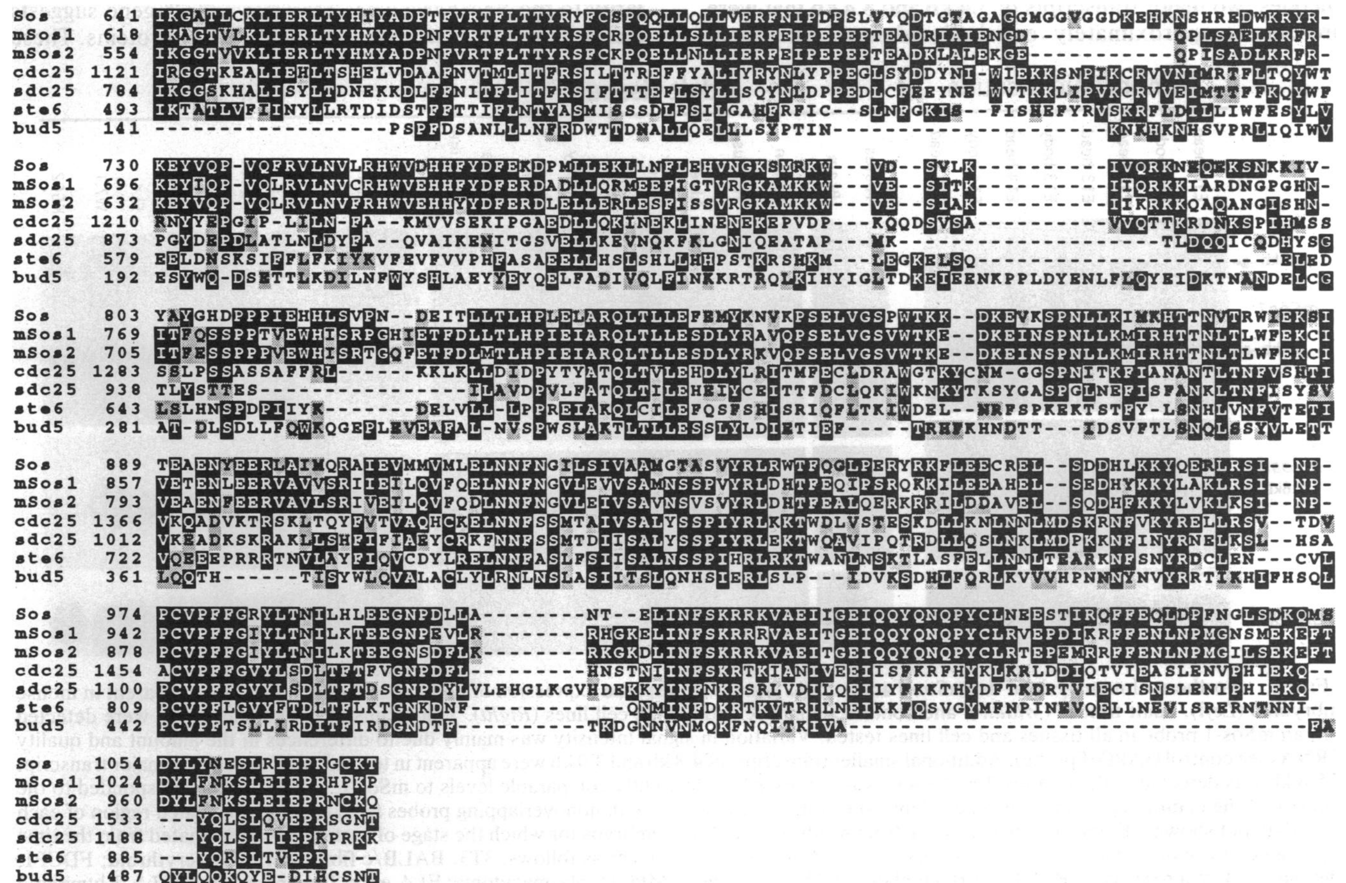

FIG. 2. Alignment of the predicted *Drosophila* and mouse Sos proteins with four related yeast proteins CDC25, SDC25, STE6, and BUD5.

the 3'-untranslated regions of the two genes were compared. Comparison with Sos showed that both mSos-1 and mSos-2 have an overall amino acid identity of 45% with the *Drosophila* gene product. Both mSos-1 and mSos-2 remain colinear with Sos over their coding regions, except at their carboxyl-terminal ends, where homology between Sos and the murine gene products is more scattered.

The apparent duplication of the *Sos* genes to create two mammalian homologues suggests the potential for some separation of function of the mSos proteins. One possibility is that the mSos proteins interact with different targets. In this respect it is interesting to note that there are three mammalian genes (H-*ras*, Ki-*ras*, and N-*ras*, ref. 21) related to the single *Dras1* gene, the proposed target of *Sos* in *Drosophila*. (Although two other *Drosophila* genes with ras protein homology have been identified, *Dras2* and *Dras3*, these are more closely related to the divergent human *RRAS* and *RAP1A* genes, respectively; refs. 7, and 21–24.)

The most conserved region of the *Sos* and m*Sos* genes is a central domain of ≈430 amino acids, which shows a high degree of homology to several yeast guanine nucleotide-exchange factor-encoding genes, including *CDC25* (9, 10, 25), *SDC25* (26), *STE6* (27), and *BUD5* (28). This domain is present in a fragment of the *SDC25* gene that can catalyze nucleotide exchange by either the yeast RAS2 protein or human c-HRAS (29). Inclusion of the mSos1 and mSos2 proteins in an alignment with these yeast proteins further highlights residues that are highly conserved between members of this gene family (Fig. 2). Similarity of the *Sos* and m*Sos* genes to those from yeast is limited to this domain.

Both m*Sos-1* and m*Sos-2* Are Widely Expressed. Northern analysis of RNA isolated from various developmental stages and adult tissues revealed that both m*Sos-1* and m*Sos-2* have a broad pattern of expression (Fig. 3). The m*Sos-1* gene encodes two major transcripts of 5.4 kb and 8.4 kb that were present in approximately equal abundance. Additional smaller transcripts of 4.8 kb and 3.9 kb were detected in RNA derived from testes (Fig. 3, adult tissues). The m*Sos-2* gene appears to encode a single transcript of 5.6 kb (Fig. 3). As the mSos-2 cDNA that was isolated and sequenced was 5.3 kb in length, it is likely that most of the coding region has been obtained for this gene. Expression of both m*Sos-1* and m*Sos-2* was apparent in all the embryonic stages and adult tissues examined.

To investigate whether there was any lineage restriction in expression of either gene within a given tissue, we examined RNA expression in hemopoietic cell lines representative of early and late lymphoid, myeloid, and erythroid lineages (Fig. 3, cell lines). Expression of both mSos-1 and mSos-2 was detected in all hemopoietic lineages tested and was comparable to that in BALB/c 3T3 fibroblasts. The broad pattern of expression observed with the m*Sos-1* and m*Sos-2* genes, in a range of tissues and hemopoietic lineages, is consistent with their postulated role in regulating the widely expressed ras proteins (33, 34).

Northern and Southern blot analyses were also done by using human RNA and DNA. Transcripts corresponding in size to those seen in the mouse with the m*Sos-1* and m*Sos-2* genes were present in RNA from the human erythroleukemic cell line K-562 (ref. 32, Fig. 3, Cell lines), and mSos-1 and mSos-2 probes hybridized with human genomic DNA, even when washed at high stringency (Fig. 4). These findings indicate the presence of separate human genes closely related to both m*Sos-1* and m*Sos-2*.

Concluding Remarks. We have isolated two genes present in the mouse, which show a high degree of similarity to the *Drosophila Sos* gene. Genetic evidence indicates that *Sos* participates in the signal-transduction pathway of at least two tyrosine kinases, sevenless and the *Drosophila* epidermal growth factor receptor (5, 6, 8). Amino acid sequence similarity to the *Saccharomyces cerevisiae cdc25* gene suggests that *Sos* may act as a positive regulator of ras proteins. These

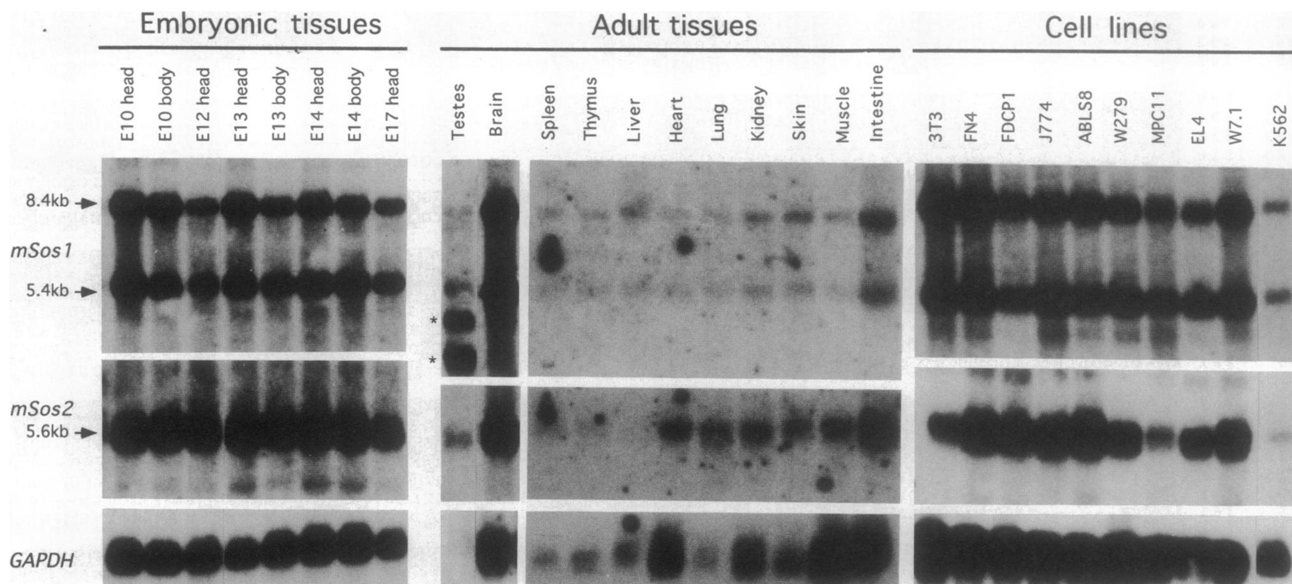

FIG. 3. Northern blot analysis of the m*Sos-1* and m*Sos-2* genes. Widespread expression of both m*Sos-1* and m*Sos-2* was apparent in mouse embryonic (*Left*), adult tissues (*Middle*), and continuous mouse and human cell lines (*Right*). Transcripts of 5.4 kb and 8.4 kb were detected with an mSos-1 probe in all tissues and cell lines tested. Variation in signal intensity was mainly due to differences in the amount and quality of RNA (see control GAPDH probe). Additional smaller transcripts of 4.8 kb and 3.9 kb were apparent in testes RNA (*). A single major transcript of 5.6 kb was detected with an mSos-2 probe and was expressed at apparently comparable levels to mSos-1. Probes used corresponded to the majority of the coding region of each gene. Identical results were obtained with nonoverlapping probes from the 3' untranslated region of each gene (data not shown). Embryonic tissues were from heads and bodies of embryos for which the stage of gestation was estimated from the time of plugging of donor females. Adult tissues are indicated. Mouse cell lines are as follows: 3T3, BALB/c fibroblast; FN4, erythroid; FDCP1, myeloid; J774, macrophage; ABLS8, pre-B lymphoid; W279, B lymphoid; MPC11, plasmacytoma; EL4, early T lymphoid, and W7.1, T lymphoid (30, 31). Transcripts of the same size were also detected in RNA from the human erythroleukemic cell line K-562 (32). GAPDH was used as a control probe for the amount and quality of RNA.

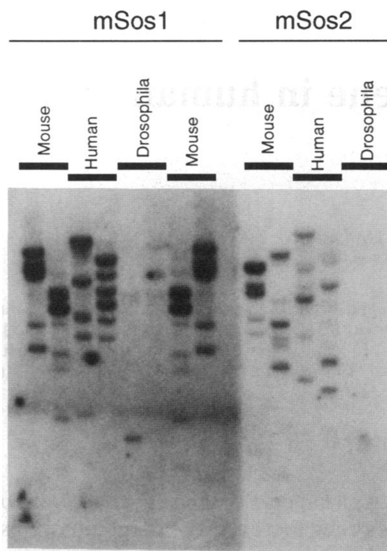

FIG. 4. Genomic Southern blot analysis of m*Sos-1* and m*Sos-2* genes. Human, mouse, and *D. melanogaster* DNA, restricted with either *Eco*RI (left lanes) or *Hin*dIII (right lanes) and subjected to Southern analysis, was hybridized separately with fragments corresponding to most of the coding regions of the two genes. Strong hybridization to human DNA was apparent with both genes after washing under conditions of high stringency. Relatively little cross hybridization was apparent between m*Sos-1* and m*Sos-2* when used to probe mouse DNA under these conditions. Therefore, the large number of bands apparent in the human DNA lanes probably reflects hybridization of each cDNA probe to several exons of genes corresponding to m*Sos-1* and m*Sos-2*, respectively, rather than hybridization to a larger number of related genes.

findings have suggested a model in which PTKs regulate the activity of ras proteins by modulating the activity of *Sos* (5). However, direct regulation of *Sos* by PTKs remains to be demonstrated biochemically. Several components of this proposed signaling pathway have been identified in the mouse previously, including the epidermal growth factor receptor and a number of ras and ras-related proteins. The identification of murine homologues of *Sos* provides further evidence of a conservation of this signaling pathway in vertebrates.

The biochemical activities of the Sos and mSos proteins remains to be determined. However, the postulated requirement for a factor that stimulates GTP addition on ras proteins (12) and the detection of such activity in protein extracts (13–16) or indirectly in permeabilized cells (35) all indicate the presence of a factor(s) that provides a counterbalance to the activity of the *ras*GAP proteins in mammalian cells. The work described here suggests the presence of at least two such proteins in the mouse and human. The availability of these genes provides a valuable tool for further analyzing biochemical interactions between tyrosine kinases and ras proteins in vertebrates.

We thank J. Allen and G. Begley for the gift of RNA from certain cell lines and members of the Walter and Eliza Hall Institute computing facility for assistance with DNA sequence analysis. We thank S. Khoury and J. Haralambidis for their synthesis of oligonucleotides and M. Kafali for technical assistance. This work was supported by funds from an Institute block grant provided by the National Health and Medical Research Council, Australia.

1. Cantley, L. C., Auger, K. R., Carpenter, C., Duckworth, B., Graziani, A., Kapeller, R. & Soltoff, S. (1991) *Cell* **64,** 281–302.
2. Smith, M. R., DeGudicibus, S. J. & Stacey, D. W. (1986) *Nature (London)* **320,** 540–543.
3. Feig, L. A. & Cooper, G. M. (1988) *Mol. Cell. Biol.* **8,** 3235–3243.
4. Arioan, R. V., Koga, M., Mendel, J. E., Ohshima, Y. & Sternberg, P. W. (1990) *Nature (London)* **348,** 693–699.
5. Simon, M. A., Bowtell, D. D. L., Dodson, G. S., Laverty, T. R. & Rubin, G. M. (1991) *Cell* **67,** 701–716.
6. Bonfini, L., Karlovich, C. A., Dasgupta, C. & Banerjee, U. (1992) *Science* **255,** 603–606.
7. Neuman-Silberberg, F. S., Schejter, E., Hoffmann, F. M. & Shilo, B. Z. (1984) *Cell* **37,** 1027–1033.
8. Rogge, R. D., Karlovich, C. A. & Banerjee, U. (1991) *Cell* **64,** 39–48.
9. Broek, D., Toda, T., Michaeli, T., Levin, L., Birchmeier, C., Zoller, M., Powers, S. & Wigler, M. (1987) *Cell* **48,** 789–799.
10. Robinson, L. C., Gibbs, J. B., Marshall, M. S., Sigal, I. S. & Tatchell, K. (1987) *Science* **235,** 1218–1221.
11. Jones, S., Vignais, M.-L. & Broach, J. R. (1991) *Mol. Cell. Biol.* **11,** 2641–2646.
12. Wittinghofer, A. & Pai, E. F. (1991) *Trends Biochem. Sci.* **16,** 382–387.
13. Wolfman, A. & Macara, I. G. (1990) *Science* **248,** 67–69.
14. Downward, J., Riehl, R., Wu, L. & Weinberg, R. A. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 5998–6002.
15. Huang, Y. K., Kung, H. F. & Kamata, T. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 8008–8012.
16. West, M., Kung, H. & Kamata, T. (1990) *FEBS Lett.* **259,** 245–248.
17. Frohman, M. A. & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8998–9002.
18. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
19. Higgins, D. G. & Sharp, P. M. (1988) *Gene* **73,** 237–244.
20. Gonda, T. J., Gough, N. M., Dunn, A. R. & de Blaquiere, J. (1985) *EMBO J.* **4,** 2003–2008.
21. Valencia, A., Chardin, P., Wittinghofer, A. & Sander, C. (1991) *Biochemistry* **30,** 4637–4648.
22. Mozer, B., Marlor, R., Parkhurst, S. & Corces, V. (1985) *Mol. Cell. Biol.* **5,** 885–889.
23. Lowe, D. G., Capon, D. J., Delwart, E., Sakaguchi, A. Y., Naylor, S. L. & Goeddel, D. V. (1987) *Cell* **48,** 137–146.
24. Hariharan, I. K., Carthew, R. W. & Rubin, G. M. (1991) *Cell* **67,** 717–722.
25. Chamonis, J. H., Kalekine, M., Gondre, B., Garreau, H., Boy-Marcotte, E. & Jacquet, M. (1986) *EMBO J.* **5,** 375–380.
26. Damak, F., Boy-Marcotte, E., Le-Roscouet, D., Guilbaud, R. & Jacquet, M. (1991) *Mol. Cell. Biol.* **11,** 202–212.
27. Hughes, D. A., Fukui, Y. & Yamamoto, M. (1990) *Nature (London)* **344,** 355–357.
28. Powers, S., Gonzales, E., Christensen, T., Cubert, J. & Broek, D. (1991) *Cell* **65,** 1225–1231.
29. Crechet, J. B., Poullet, P., Mistou, M.-Y., Parmeggiani, A., Chamonis, J., Boy-Marcotte, E., Damak, F. & Jacquet, M. (1990) *Science* **248,** 866–868.
30. Culvenor, J. G., Harris, A. W., Mandel, T. E., Whitelaw, A. & Ferber, E. (1981) *J. Immunol.* **126,** 1974–1977.
31. Kongsuwan, K. E., Webb, E., Housiaux, P. & Adams, J. M. (1988) *EMBO J.* **7,** 2131–2138.
32. Lozzio, C. B. & Lozzio, B. B. (1975) *Blood* **45,** 321–334.
33. Leon, J., Guerrero, I. & Pellicer, A. (1987) *Mol. Cell. Biol.* **7,** 1535–1540.
34. Furth, M. E., Aldrich, T. H. & Condon-Cardo, C. (1987) *Oncogene* **1,** 47–58.
35. Downward, J., Graves, J. D., Warne, P. H., Rayter, S. & Cantrell, D. A. (1990) *Nature (London)* **346,** 719–723.