MOLECULAR AND CELLULAR BIOLOGY, Aug. 1986, p. 3000–3004
0270-7306/86/083000-05$02.00/0

Vol. 6, No. 8

Human c-*ros*-1 Gene Homologous to the v-*ros* Sequence of UR2 Sarcoma Virus Encodes for a Transmembrane Receptorlike Molecule

HITOSHI MATSUSHIME, LU-HAI WANG,† AND MASABUMI SHIBUYA*

Department of Genetics, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108, Japan

Received 28 February 1986/Accepted 6 May 1986

We isolated a human gene (designated c-*ros*-1) homologous to the v-*ros* sequence of UR2 sarcoma virus. Ten exons, 1,414 base pairs spanning 26 kilobases, contained a tyrosine kinase domain, a transmembrane domain, and a part of an extracellular domain carrying an N glycosylation site which was not acquired by UR2 sarcoma virus. The predicted structure of c-*ros*-1 is unique among the *src* family and clearly distinct from the human insulin receptor.

UR2 sarcoma virus, a recent isolate of acutely transforming retrovirus of chickens (1), encodes for a fusion protein, $p68^{gag\text{-}ros}$, which has tyrosine-specific protein kinase activity (7). Nucleotide sequence analysis of the UR2 genome has revealed that the oncogene v-*ros* of UR2 carries a kinase domain homologous to those present in the oncogenes of the *src* family (12). The v-*ros* gene is considered to be derived from a cellular counterpart, proto-oncogene c-*ros* of chickens (15). Since the predicted chicken c-*ros* gene product, as well as the v-*ros* product, has a hydrophobic short stretch upstream of its kinase domain, it seems likely that the c-*ros* gene encodes for a transmembrane protein similar to the cell surface receptor for cell growth or differentiation factors (11, 12). Furthermore, recent reports have indicated that the deduced amino acid sequence of the kinase domain in the human insulin receptor (HIR) gene is highly homologous to the kinase domain in the v-*ros* sequence (6, 17). However, the phylogenetic conservation of the c-*ros* gene in mammalian species, including humans, and the relationship between the c-*ros* gene and the HIR gene have not yet been examined. In this study, we made an attempt to clarify these points.

High-molecular-weight genomic DNA was extracted from human placenta, mouse thymus, fish testis, *Drosophila melanogaster*, and yeast cells (*Saccharomyces cerevisiae*) and hybridized with a v-*ros*-specific probe (*Eco*RI-*Pvu*II 0.8-kilobase (kb) fragment; probe IV [see Fig. 2]). Under hybridization conditions of very relaxed stringency (20% formamide, 10% dextran sulfate, 1M NaCl; 37°C), we were able to detect clear bands in all DNAs examined, except for yeast DNA (Fig. 1). The c-*ros* sequence appeared to be conserved in vertebrate species from fish to mammals, including humans.

In human placenta DNA digested with the restriction endonuclease *Bam*HI, two discrete bands were detected at 15 and 10 kb (Fig. 1). Because these bands were observed only under very relaxed hybridization conditions, we molecularly cloned these *Bam*HI fragments, which may contain a portion(s) of the c-*ros* gene, before screening a large number of phages in a human genomic DNA library. With Charon 30 as a vector, four independent clones (two clones each from the 15- and 10-kb *Bam*HI fragments) were isolated by the method described by Benton and Davis (3) (Fig. 2). Partial DNA sequencing analysis by the dideoxy method (13) revealed that recombinant phage HYuros8 contained DNA sequences highly homologous to the v-*ros* sequence, whereas phage HYuros4 contained human sequences partially related to v-*ros* (data not shown). Both HYuros5 and HYuros1 were found to contain human sequences incidentally homologous to v-*ros*. Therefore, the human genes

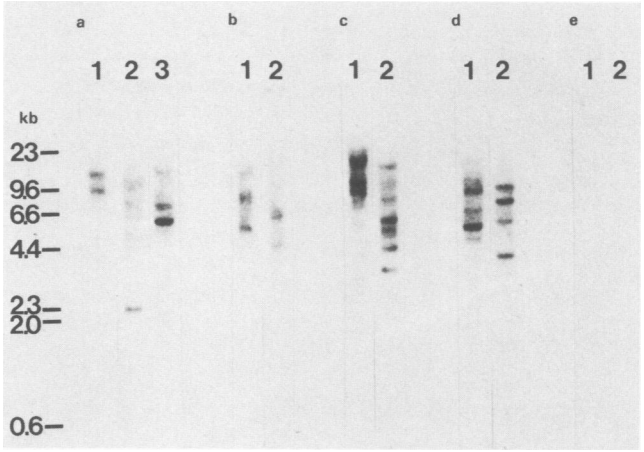


FIG. 1. Hybridization of a v-*ros*-specific probe to genomic DNAs. Cellular DNAs (10 μg in panels a to c, 3 μg in panel d, and 2 μg in panel e) were digested with various endonucleases, electrophoresed on 0.8% agarose gels, and transferred to nitrocellulose filters (16). These filters were hybridized with a v-*ros*-specific probe (probe IV in Fig. 2) under hybridization conditions of low stringency (see text). Panels a to c were exposed to X-ray film at −70°C for 3 days; panels d and e were exposed for 6 days with an intensifying screen. Molecular weight markers were lambda DNAs digested with *Hin*dIII. Genomic DNAs used in Southern blot analyses were from human placenta (a), mouse thymus (b), fish testis (c), *D. melanogaster* (d), and *S. cerevisiae* (e) cells. These DNAs were digested with the following endonucleases. (a) Lanes: 1, *Bam*HI; 2, *Eco*RI; 3, *Hin*dIII. (b) Lanes: 1, *Bam*HI; 2, *Eco*RI. (c) Lanes: 1, *Bam*HI; 2, *Hin*dIII. (d) Lanes: 1, *Eco*RI; 2, *Hin*dIII. (e) Lanes: 1, *Bam*HI; 2, *Eco*RI.

* Corresponding author.
† Present address: The Rockefeller University, New York, NY 10021.

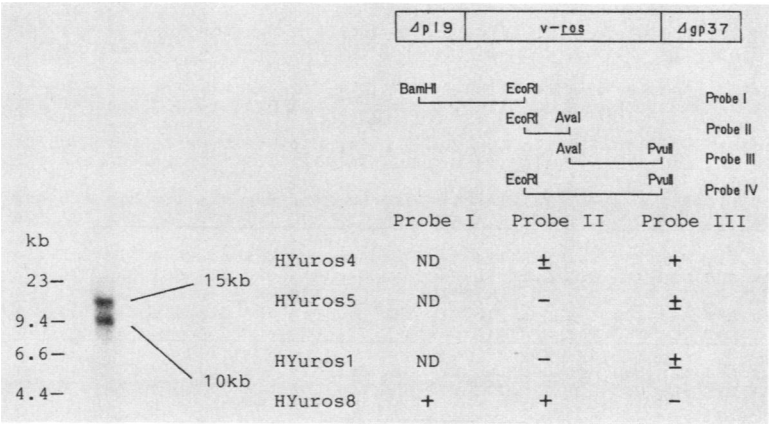
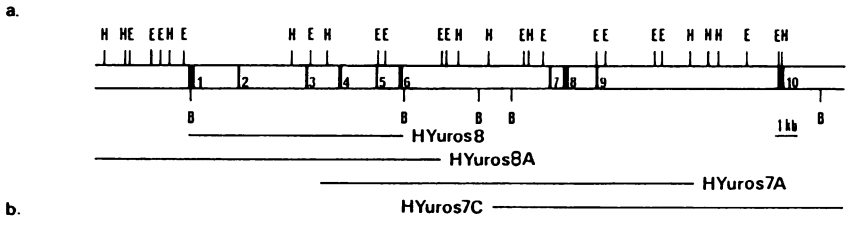


FIG. 2. Molecular cloning of human c-*ros* genes. Four independent clones were isolated. HYuros4 and HYuros5 were derived from phages recombined with 15-kb *Bam*HI fragments; the other two clones were derived from 10-kb *Bam*HI fragments. These four clones were hybridized with various v-*ros* probes under hybridization conditions of low stringency. Symbols: +, hybridization; ±, weak hybridization; −, no detected hybridization. ND, Not determined. Probes used are illustrated at the top of the figure. Molecular size markers (in kilobases) are indicated to the left of the gel.

found in HYuros8 and HYuros4 were designated human c-*ros*-1 and human c-*ros*-2, respectively, and the human c-*ros*-1 gene was further characterized in this study. The details of the analysis of the human c-*ros*-2 gene will be described elsewhere.

We constructed a human genomic DNA library by the method described by Maniatis et al. (9), and gene walking of overlapping c-*ros*-1 DNA sequences of about 60 kb was carried out. By restriction endonuclease mapping and by DNA sequencing of DNA fragments which hybridized with various v-*ros* probes, seven exons (exons 4 to 10) were found to encode for the entire kinase domain of this gene (Fig. 3a and 4). However, an approximately 240-base-pair region at the 5′ end of the v-*ros* gene, including a possible transmembrane domain, was not detected by cross-hybridization between the human 60-kb DNA sequence and the v-*ros* sequence. We expected that the chicken c-*ros* DNA fragment might be useful for isolating the transmembrane domain of the human c-*ros*-1 gene more efficiently than could be done with the v-*ros* sequence as a probe. The ^{32}P-labeled 5′ region of chicken c-*ros* DNA (5.2-kb *Eco*RI fragment [11]) was hybridized with the human DNA fragment upstream of the kinase domain (exon 1 in Fig. 3a), and then the nucleotide sequence of this region was determined. Although we did not obtain an exon(s) for a transmembrane domain in this region, to our surprise we found a new exon surrounded by a consensus splice acceptor site and a donor site (Fig. 3b); the predicted amino acid sequence had a potential site of N-linked glycosylation (Asn-Gly-Ser, amino acid residues 19 to 21; Fig. 4). In chicken c-*ros* DNA, an exon highly homologous to this human sequence at the level of predicted amino acids was also observed (Fig. 5). Since the amino acid sequence in this new exon (amino acid residues 1 to 63) was not present in the



b.

Exon number	Splice acceptor	Splice donor
1.	ATTTATTATTTTTTGCAG AAA-	-GAG GTATGTTAC
2.	TCTGATTTATATTATTAG ATG-	-TTG GTAAGTATA
3.	TCATTTTTTCCTTTATAG TCT-	-ACA GTATGTAGC
4.	AATGTTGCTATTTTACAG TAC-	-AAG GTAATGTGA
5.	TGTTTTTATGTTGTATAG ACT-	-GAG GTAGCTCTG
6.	TGTCTTTCCACCTTTCAG CAA-	-ACG GTAGGCAGT
7.	TTCTTTCATTTCTTACAG TTT-	-CAG GTACAATTC
8.	ATATAACTTCTGCCACAG GGA-	-ATG GTAAGTTTA
9.	ACCCTCTACTATTCTTAG GTC-	-TCT GTAAGTTAA
10.	CTTTCCCCATATATTTAG GTG-	(-AAG GTAAGTTTG)/-TGA Opal

FIG. 3. Restriction map and gene organization of the human c-*ros*-1 gene. (a) Four overlapping clones spanning 32-kb of cellular DNA are indicated. HYuros8 was a recombinant phage with Charon 30 *Bam*HI arms; the other clones, HYuros8A, HYuros7A, and HYuros7C, had Charon 4A *Eco*RI arms. The restriction map was determined by double or triple digestion with the following various endonucleases: B, *Bam*HI; E, *Eco*RI; and H, *Hin*dIII. The positions of exons in the human c-*ros*-1 gene were determined by DNA sequencing. Black and white boxes indicate exons and introns, respectively. The numbers to the right of the black boxes indicate exon numbers. (b) Possible splice junctions in the human c-*ros*-1 gene.

FIG. 4. Nucleotide sequence and predicted amino acid sequence of the human c-*ros*-1 gene. The putative transmembrane domain and kinase domain are underlined and boxed, respectively. A potential site of N-linked glycosylation is indicated by an asterisk. The horizontal arrows above the amino acid sequences indicate the junctions between exons and introns. The nucleotides above the dashed line indicate a possible splice donor site. Ala, Alanine; Cys, cysteine; Asp, aspartic acid; Glu, glutamic acid; Phe, phenylalanine; Gly, glycine; His, histidine; Ile, isoleucine; Lys, lysine; Leu, leucine; Met, methionine; Asn, asparagine; Pro, proline; Gln, glutamine; Arg, arginine; Ser, serine; Thr, threonine; Val, valine; Trp, tryptophan; Tyr, tyrosine.

v-*ros* sequence of UR2 sarcoma virus, it seemed most likely that this exon belonged to the extracellular domain of the c-*ros* gene not acquired in the viral genome.

By nucleotide sequencing of the entire cellular DNA of about 6 kb in length from exon 1 to exon 4 in Fig. 3a, two exons, exon 2 and exon 3, were detected. The predicted amino acid sequence of exon 2 showed an extremely hydrophobic stretch of 21 amino acids; that of exon 3 carried a sequence 63% homologous to the corresponding exon of the chicken c-*ros* gene (Fig. 4, Table 1). Although the nucleotide sequence of exon 2 greatly diverged from that of the chicken c-*ros* transmembrane domain and the peptide length was 2 amino acids shorter than that in chickens, we consider this hydrophobic stretch to be the transmembrane domain of the human c-*ros*-1 gene because of its partial homology to the hydrophobic amino acid sequence of the chicken c-*ros* gene and because its length is sufficient to pass through the lipid bilayer of the cell membrane. The extents of nucleotide homology of the regions of exon 2 and exon 3 with the corresponding regions of the chicken c-*ros* gene were 52 and 60%, respectively. This weak homology may explain the failure to detect cross-hybridization between these sequences and the v-*ros* sequence by Southern blot analysis. Such a high degree of heterogeneity in the nucleotide sequence outside the kinase domain of an oncogene between avian and human species has also been reported in the case of the c-*src* gene; exon 3 in the human c-*src* gene could not be detected by cross-hybridization using the v-*src* sequence as a probe (8).

Ten exons of the human c-*ros*-1 gene were identified within this 26-kb human DNA separated by 1- to 6-kb-long introns (Fig. 3a). The entire coding sequence and the predicted amino acid sequence of the human c-*ros*-1 gene are shown in Fig. 4. The kinase domain of the human c-*ros*-1 gene was found in the sequences from exon 4 to a part of exon 10 (Fig. 4). In these exons, the structure was highly

FIG. 5. Comparison of the amino acid sequences among the human c-*ros*-1 gene, the chicken c-*ros* gene, and the HIR gene. Symbols: colon (:), identical amino acid; +, conserved amino acid; −, deletion of amino acid.

homologous to chicken c-*ros* and v-*ros* genes not only in the DNA sequence but also in the predicted amino acid sequence, except for exon 10 (Table 1, Fig. 5). Furthermore, the predicted products of the human c-*ros*-1 gene and the chicken c-*ros* gene shared similar inserts of 2 to 5 amino acids (amino acid residues 252 to 254, 260 to 261, and 299 to 303) which were not present in any other members of the *src* family, and the splice junctions in these two genes were completely matched to each other (11). From these results, we conclude that the human c-*ros*-1 gene is a cellular DNA homolog of v-*ros* in the human genome. Recent studies on the structure of the HIR gene have shown that the kinase domain of the HIR gene deduced from the cDNA sequence is more homologous to that of v-*ros* than it is to those of other *src* family members. However, by comparison of the predicted amino acids of the human c-*ros*-1 gene and the HIR gene, these two molecules were demonstrated to be clearly different from each other; homology in the level of amino acids in the kinase domain was 48.5% (Fig. 5).

In the overall structure, the human c-*ros*-1 gene carried an extracellular domain with a potential site of N-linked glycosylation, a hydrophobic 24-amino acid stretch, and a tyrosine kinase domain. These structural organizations are similar to those of the c-*erbB* (the gene for epidermal growth factor receptor), the c-*fms* (the gene for macrophage colony-stimulating-factor receptor), and the HIR genes (5, 6, 14, 17, 18). These results strongly suggest that the human c-*ros*-1 gene encodes for a transmembrane molecule which may function as a receptor for a cell growth or differentiation factor(s). Recently isolated transforming genes, *neu*/*erbB2* and *oncD*, appear to be derived from the same category of receptor-type proto-oncogene (2, 4, 10, 19).

The biological function of the c-*ros* gene product and the significance of the c-*ros* gene in tumorigenicity in animals remain to be elucidated. Expression of the c-*ros* gene in 7- to 14-day-old healthy chickens was strongly repressed in many tissues, but two to three copies per cell of c-*ros* RNA were detected in kidneys by liquid hybridization and Northern blotting methods (11, 15). These results might indicate that the c-*ros* gene has a function in a limited stage of development or in a particular cell population in some tissue such as the kidney. Molecularly cloned human c-*ros*-1 DNA may be very useful for examining the expression or abnormalities of this gene in normal or malignant tissues.

TABLE 1. Homology between the human c-*ros*-1 gene and the chicken c-*ros* gene in terms of nucleotide sequence and predicted amino acid sequence

Human c-*ros*-1 exon no.	No. of nucleotides identical to chicken c-*ros* (*n*)	% Homology	No. of amino acids identical to chicken c-*ros* (*n*)	% Homology
1	121[a] (191)[b]	63.4[a]	32[a] (63)[b]	50.8
2	44 (84)	52.4	9 (28)	32.1
3	81 (136)	59.6	29 (46)	63.0
4	121 (163)	74.2	42 (54)	77.8
5	55 (65)	84.6	19 (22)	86.4
6	102 (130)	78.5	34 (43)	79.1
7	68 (98)	69.4	24 (33)	72.7
8	156 (201)	77.6	57 (67)	85.1
9	96 (135)	71.1	29 (45)	64.4
10	100 (211)	47.4	25 (70)	35.7

[a] This possible extracellular domain in the chicken c-*ros* gene is not present in the c-*ros* sequence reported by Neckameyer et al. (11) and has been sequenced in this study.

[b] *n*, Total number of nucleotides or amino acids in each exon.

We thank Sadao Ishiwa for providing *D. melanogaster*. We are grateful to Takato O. Yoshida and Takeshi Odaka for helpful discussions and encouragement. The excellent technical assistance of Sachiko Yamaguchi is greatly appreciated. We also thank Yumiko Matsushime and Ryuko Hasui for preparing and typing the manuscript.

This work was supported by a grant-in-aid (60218008) for special project research on cancer-bioscience from the Ministry of Education, Science and Culture of Japan. L.H.W. was supported by Public Health Service grants CA18213 and CA29339 from the

National Cancer Institute. H.M. acknowledges a fellowship from the Japan Society for the Promotion of Science.

LITERATURE CITED

1. **Balduzzi, P. C., M. F. D. Notter, H. R. Morgan, and M. Shibuya.** 1981. Some biological properties of two new avian sarcoma viruses. J. Virol. **40**:268–275.
2. **Bargmann, C. I., M.-C. Hung, and R. A. Weinberg.** 1986. The *neu* oncogene encodes an epidermal growth factor receptor-related protein. Nature (London) **319**:226–230.
3. **Benton, W. D., and R. W. Davis.** 1977. Screening lambda gt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
4. **Coussens, L., T. L. Yang-Feng, Y.-C. Liao, E. Chen, A. Gray, J. McGrath, P. H. Seeburg, T. A. Libermann, J. Schlessinger, U. Francke, A. Levinson, and A. Ullrich.** 1985. Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with *neu* oncogene. Science **230**: 1132–1139.
5. **Downward, J., Y. Yarden, E. Mayes, G. Scrace, N. Totty, P. Stockwell, A. Ullich, J. Schlessinger, and M. D. Waterfield.** 1984. Close similarity of epidermal growth factor receptor and v-*erbB* on oncogene protein sequences. Nature (London) **307**:521–527.
6. **Ebina, Y., L. Ellins, K. Jarnagin, M. Edery, L. Graf, E. Clauser, J. Ou, F. Masiarz, Y. W. Kan, I. D. Goldfine, R. A. Roth, and W. J. Rutter.** 1985. The human insulin receptor cDNA: the structural basis for hormone-activated transmembrane signalling. Cell **40**:747–758.
7. **Feldman, R. A., L.-H. Wang, H. Hanafusa, and P. C. Balduzzi.** 1982. Avian sarcoma virus UR2 encodes a transforming protein which is associated with a unique protein kinase activity. J. Virol. **42**:228–236.
8. **Gibbs, C. P., A. Tanaka, S. K. Anderson, J. Radul, J. Baar, A. Ridgway, H.-J. Kung, and D. J. Fujita.** 1985. Isolation and structural mapping of a human c-*src* gene homologous to the transforming gene (v-*src*) of Rous sarcoma virus. J. Virol. **53**:19–24.
9. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual, p. 269–294. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. **Martin-Zanca, D., S. H. Hughes, and M. Barbacid.** 1986. A human oncogene formed by the fusion of truncated tropomyosin and protein tyrosine kinase sequences. Nature (London) **319**:743–748.
11. **Neckameyer, W. S., M. Shibuya, M.-T. Hsu, and L.-H. Wang.** 1986. Proto-oncogene c-*ros* codes for a molecule with structural features common to those of growth factor receptors and displays tissue-specific and developmentally regulated expression. Mol. Cell. Biol. **6**:1478–1486.
12. **Neckameyer, W. S., and L.-H. Wang.** 1985. Nucleotide sequence of avian sarcoma virus UR2 and comparison of its transforming gene with other members of the tyrosine protein kinase oncogene family. J. Virol. **53**:879–884.
13. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
14. **Sherr, C. J., C. W. Rettenmier, R. Sacca, M. F. Roussel, T. A. Look, and R. E. Stanly.** 1985. The c-*fms* proto-oncogene product is related to the receptor for the mononuclear phagocyte growth factor, CSF-1. Cell **41**:665–676.
15. **Shibuya, M., H. Hanafusa, and P. C. Balduzzi.** 1982. Cellular sequences related to three new *onc* genes of avian sarcoma virus (*fps*, *yes*, and *ros*) and their expression in normal and transformed cells. J. Virol. **42**:143–152.
16. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
17. **Ullich, A., J. R. Bell, E. Y. Chen, R. Herrera, L. M. Petruzzelli, T. J. Dull, A. Gray, L. Coussens, Y.-C. Liao, M. Tsubokawa, A. Mason, P. H. Seeburg, C. Grunfeld, O. M. Rosen, and J. Ramachandran.** 1985. Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. Nature (London) **313**:756–761.
18. **Ullich, A., L. Coussens, J. S. Hayflick, T. J. Dull, A. Gray, A. W. Tam, J. Lee, Y. Yarden, T. A. Libermann, J. Schlessinger, J. Downward, E. L. Mayes, N. Whittle, M. D. Waterfield, and P. H. Seeburg.** 1984. Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. Nature (London) **309**:418–425.
19. **Yamamoto, T., S. Ikawa, T. Akiyama, K. Semba, N. Nomura, N. Miyajima, T. Saito, and K. Toyoshima.** 1986. Similarity of protein encoded by the human c-*erb-B*-2 gene to epidermal growth factor receptor. Nature (London) **319**:230–234.