

# Characterization of ROS1 cDNA from a human glioblastoma cell line

(tyrosine kinases/oncogenes/sevenless/receptors)

CARMEN BIRCHMEIER\*, KATHY O'NEILL†, MICHAEL RIGGS†, AND MICHAEL WIGLER†‡

\*Max-Delbrück-Laboratorium in der Max-Planck-Gesellschaft, Carl-von-Linné-Weg 10, 5000 Cologne 30, Federal Republic of Germany; and †Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, NY 11724

Contributed by Michael Wigler, March 29, 1990

**ABSTRACT** We have isolated and characterized a human ROS1 cDNA from the glioblastoma cell line SW-1088. The cDNA, 8.3 kilobases long, has the potential to encode a transmembrane tyrosine-specific protein kinase with a predicted molecular mass of 259 kDa. The putative extracellular domain of ROS1 is homologous to the extracellular domain of the sevenless gene product from *Drosophila*. No comparable similarities in the extracellular domains were found between ROS1 and other receptor-type tyrosine kinases. Together, ROS1 and sevenless gene products define a distinct subclass of transmembrane tyrosine kinases.

Oncogenes are defined as genetic elements that are able to induce malignant transformation. Many oncogenes are mutated or activated analogues of cellular genes that normally function in signal-transduction pathways. We have previously reported the isolation and characterization of the activated human *ROS1* gene, which we call *MCF3* (1). This oncogene, which was isolated by a transfection-tumorigenicity assay, encodes a transmembrane protein with a sequence typical of tyrosine kinases (2). *MCF3* was activated by a rearrangement in which all but eight amino acids of the *ROS1*-specific extracellular domain were replaced with sequences of unknown origin. Structurally, *MCF3* is very similar to the *erbB*, *fms*, *neu*, *trk*, and *kit* oncogenes (3–7). The normal cellular analogues of *erbB* and *fms* encode receptors for the epidermal growth factor and colony-stimulating factor, respectively (8, 9). Hence, we assume that *ROS1* also encodes a cellular receptor.

*ROS1* is not a ubiquitously expressed gene. In a survey of 40 different human tumor cell lines, *ROS1* was found to be expressed frequently in cell lines established from one particular type of human tumor, glioblastomas. *ROS1* transcripts were not found in a normal glial cell line or in adult brain tissue (10). Most glioblastoma cell lines express a *ROS1* transcript of identical length, 8.3 kilobases (kb), with the exception of one particular line, U-118 MG. This line expresses a 4-kb transcript and has a rearranged *ROS1* locus (10). The characterization of this mutant *ROS1* gene has been reported elsewhere (11).

We report here the isolation and sequence<sup>§</sup> of a *ROS1* cDNA from the SW-1088 glioblastoma cell line. This cell line expresses an 8.3-kb transcript. The cDNA can encode a protein of 259,000 daltons with a large extracellular domain, a transmembrane domain, and an intracellular domain with the characteristic sequence of a tyrosine protein kinase. It was previously noted that the products of the *Drosophila* gene *sevenless* and of *ROS1* have extensive homologies in their cytoplasmic domains (12). *Sevenless* is a gene required for normal eye development in the fruit fly and also encodes a transmembrane tyrosine-specific protein kinase (13, 14).

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.

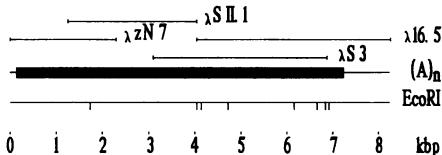


FIG. 1. Schematic representation of the *ROS1* cDNA from SW-1088 cells. Overlapping phage  $\lambda$  clones used for sequence determination and a diagram of the composite cDNA structure are shown. Untranslated sequence and translated sequence are indicated by a line and a box, respectively. The *EcoRI* restriction map and the nucleotide coordinates are indicated below.

We find that the extracellular domains of the *ROS1* and the *sevenless* gene products share similarities in size and sequence as well. The distribution of cysteine residues in the extracellular domains of these gene products do not fit the patterns of previously described classes of transmembrane protein kinases. Thus, *ROS1* and *Drosophila* *sevenless* gene encode a new structural class of transmembrane protein kinases.

## MATERIAL AND METHODS

**cDNA Library.** RNA from the human glioblastoma cell line SW-1088 was prepared by the guanidinium/CsCl method and purified on oligo(dT)-cellulose (15). Two cDNA libraries, one in phage  $\lambda$  gt10 and one in  $\lambda$  ZAP (Stratagene), were constructed by standard techniques (16). cDNA for the  $\lambda$  gt10 library was primed with oligo(dT). cDNA for the  $\lambda$  ZAP library was primed with a synthetic oligonucleotide of the sequence 5'-GGTTCACTAGCTGGCACCAGGGTAGTA-3', the antisense sequence of positions 2204–2230 of *ROS1* cDNA, and was cloned via *Not* I linkers into  $\lambda$  ZAP (17). cDNA fragments were used as probes to screen the libraries. Their coordinates were 6183–6649, 3160–4019, and 1207–1785. Phages containing *ROS1* cDNA were identified by plaque hybridization (15) and characterized by restriction mapping.

**Sequence and Analysis.** Nucleotide sequence determination was performed after subcloning into pUC118 (18) by using the dideoxynucleotide chain-termination method (19) and Sequenase (United States Biochemical). All of the coding sequence was determined in both orientations. The hydrophobic index was computed by the method of Kyte and Doolittle (20) with the PC/Gene program "SOAP" (IntelliGenetics). For the sequence comparison, the programs "COMPARE" and "DOT-PLOT" from the University of Wisconsin, Genetics Computer Group (UWGCG), were used (21).

<sup>‡</sup>To whom reprint requests should be addressed.

<sup>§</sup>The sequence reported in this paper has been deposited in the GenBank data base (accession no. M34353).

		CCGCATTCAAGCTTCAAGCATTCAAAGGCTAAATGAAAAGGCTAAGTATTATTCAAAAGGCAAGT	71
72	ATCCTAATATAGCAAACAAACAAAGCAAATCATCAGCTACTCTCCAATTGAACTGATGAAGCCAAAATAATTCTATAGCAAATGGAGAAAATTAGCCGCATCTAAAATCTGCCATTGGTGAAGTG	206	
207	ATGAAAGAACATTACTGTCTTATCGAAGCTGTCAATTGCAACTCTGGCTGCCATTGGATTCTGTGCGAGTGACAGTTAAATCTGCTCAAAGTGTGTGTAACATACTCTGGGCCAGGCCT	341	
1	MetLysAsnIleTyrCysLeuIleProLysLeuValasnPhenAlaThrLeuGlyCysLeuTrpIleSerValVaIvaIg1nCysThrValLeuAsnSerCysLeuLysSerCysValThrLeuGlyIg1nLeu	45	
342	GACCTTGGCACACCATAATCTGAGTGAACCGTGTATCCAAGGATGTCACTTTGGAACTCTGAGTACAGAAAACCTGTCTTTAAAGTGTGGGAGTCGTTGAGGTTGGCTGTAGCAGCGCGGAAAGGTGCA	476	
46	AspLeuGlyThrProH1sAsnLeuSerGluProCysIleG1nGlyCysH1sPheTrpAsnSerValAspG1nLysAsnCysAlaLeuLysCysArgGluSerCysGluValG1yCysSerSerAlaG1nLysAla	90	
477	TATGAAGAGGAAGTACTGGAAAATGCAGACCTACCAACTGCTCCTTGCTTCTCCATTGGAAAGGCCACAATATGACATTGAACTCTGCAAACCTCTCTGGAGTAAATACATCATTGAGGAAAT	611	
91	TyrGluGluValLeuIuaasnAlaAspLeuProThrAlaProPheAlaSerSerIleGlySerH1sAsnMetThrLeuUrgTrpLysSerAlaAsnPhenSerG1yValLysTyrIleG1nTrpLysTyr	135	
612	GCACAACTCTGGGAAGCTGGACTTATCTAAAGCTGTGTCAGGCCCTACTGTGGTCAAGGCCCTGACCCCCCTACTGAGTACATTTCCAGGTTGGATCTTCACAGCGCAGTCGAGCTACTCC	746	
136	A1aG1nLeuLeuGlySerTrpThrTyrThrLysVal1SerArgProSerTyrValValLysProLeuH1sProPheThrGluTyrIlePheArgValValTrpIlePheThrAlaG1nLeuG1nLeuTyrSer	180	
747	CCTCCAAGTCCCAGTACAGGACTCATCCTCATGGATTCTCTGAAACTGACCTTTGTTAGGAATTATTGAGAGCTCAAGTCCCAGACTGTGAGAAGTCAGCTGGGATCACCTCAATTCCAGGGACCTATT	881	
181	ProProSerProSerTyrArgThrH1sProH1sGlyValProG1uThrAlaProLeuIleArgAsnIleGluSerSerProAspThrValG1uVal1SerTrpAspProG1nPhenProG1yGlyProIle	225	
882	TTGGGTTATAACTTAAGGCTGATCAGCAAAATCAAATTAGATGCAAGGGACACAGAGAACCGAGTTCCAGTTTACTTACCAAAATACTATCACAGGTTTCTATTGCGACAGTAATGAAGTTGGT	1016	
226	LeuG1yTyrAsnLeuArgLeuIleSerLysAsnG1nLysLeuAspAlaG1yThrG1nArgThrSerPheG1nPhenTyrSerThrLeuProAsnThrIleTyrArgPheSerIleAlaA1aValAsnG1uValG1y	270	
1017	GAGGGTCAGAAGCAGAACTCTAGTATTACCACTCTCAGCAGTTCAACAAAGGAAACAGTGGCTTCTTATCCAGAAAACCTCTCTAAAGAAAGAGATCTTAAACATTTAGTAGATGAAGCACATTG	1151	
271	GluGlyProG1uAlaG1uLysSerSerIleThrSerSerAlaValG1nG1nG1uG1nTrpLeuPheLeuSerArgLysThrSerLeuArgLysArgSerLeuLysHisLeuValAspG1uA1aH1sCys	315	
1152	CTTCGGTTGGATGCTATATACCATATAATTACAGGAATATCTGTTGATGTCACCCAGCAGAAATTGTTTATTCCTCTGAAGGAACCTCTCATATGGCGAAGAAGGCTGCCAACATGTCATGATGATCTGACCTGAGA	1286	
316	LeuArgLeuAspAlaIleTyrH1sAsnIleThrG1yIleSerVal1AspVal1H1sG1nIleVal1TyrPheSerGluG1yThrLeuIleTrpAlaLysLysAlaA1aAsnMetSerAspVal1SerAspLeuArg	360	
1287	ATTTTTACAGAGGTTCAAGGTTAATTCTCTATCTCCATAGATTGGCTTATCAAGAATGTTTCTCATCATGGATGAACTGCTGATGTTAGAGAACCTGCTCAACACATCGAGGAAAATTA	1421	
361	IlePheTyrArgG1ySerG1yLeuIleSerSerIleSerIleAspTrpLeuTyrG1nArgMetTyrPheIleMetAspG1uLeuVal1CysVal1CysAspLeuG1uAsnCysSerAsnIleG1uG1uIleThrPro	405	
1422	CCCTCTATTAGTCGACCTCAAAATATTGTCGATTCTACATACTGGATGTCCTTACCTCTGAGAGATGGCATTATAGCAGACCCCTCTGTCACCATCTGGCCGGTGTGAGAACGCTGTGCTGATTGTC	1556	
406	ProSerIleSerAlaProG1nLysIleVal1AlaAspSerTyrAsnG1yTyrVal1PheTyrLeuLeuArgAspG1yIleTyrArgAlaAspLeuProVal1ProSerG1yArgCysAlaG1uAlaVal1ArgIleVal1	450	
1557	GAGAGTTGCACTGAAAGACTTTGCAATCAAGCACAAGCCAAGCAGCATTTACTCTCAATGACACTGCCAACGTCATCTGTCACATTTCTGGATGCTCTGCTCCCATCTCATCTACCTGCATCCCC	1691	
451	GluSerCysThrLeuLysAspPheAlaIleLysProG1nAlaLysArgIleIleTyrPheAsnAspThrAlaG1nVal1PheMetSerThrPheLeuAspG1ySerAlaSerH1sLeuIleLeuProArgIlePro	495	
1692	TTTGTGATGAAAAGTTGCTTGTGAAAACAATGACTTTCTGTCACAGATGGCAAGGTCTTICCAACAGGATGCTTGTCTTAAATGAAATTCTCGTGGGATGTGACCTGAGTCACATAGAAGATT	1826	
496	PheAlaAspVal1LysSerPheAlaCysG1uAsnAsnAspPheLeuVal1ThrAspG1yLysVal1IlePheG1nLysAspAlaLeuSerPheAsnG1uPheIleVal1G1yCysAspLeuSerH1sIleG1uG1uPhe	540	
1827	GGGTTTGGTAACTTGGCTCATCTTGGCTCATCTCCAGTCGACCCCTCTGCCAGGCGCCGCCAGGAGCTTCCGGTCTGTTGGCTCTCACAGGCTCTGTCATGGAAGCCTCTGCCCTGGCCATAGGA	1961	
541	GlyPheG1yAsnLeuVal1IlePheGlySerSerG1nLeuH1sProLeuProG1yArgProG1nG1uLeuSerVal1LeuPheG1ySerH1sG1nA1aLeuVal1G1nTrpLysProProAlaLeuAlaIleG1y	585	
1962	GCCAATGTCATCTGATCAGTGTATATTGAACTCTTGAATTAGGCCCTCTGCCCTGGCAGACTGGACCTATGAGGTTGAAAGATATCACCACAGGCTCTGAGTCACACTCATATTCTCTGAACATAAGT	2096	
586	A1aAsnVal1IleLeuIleSerAspIleIleG1uLeuPheG1uLeuG1yProSerAlaTrpG1nAsnTrpThrTyrG1uVal1LysVal1SerThrG1nAspProG1uVal1ThrH1sIlePheLeuAsnIleSer	630	
2097	GGAACCATGTCGAATGTCACCTGAGCTGAGAGTCATGAAATACAAGGTTCTGAGAGCAAGTCTCACAAGGAGGCCCTGGTCAGAGGCCCTCATGGGCTACTACCTGGTCAGCAGTAGTGAACCA	2231	
631	GlyThrMetLeuAsnVal1ProG1uLeuG1nSerAlaMetLysVal1SerVal1ArgAlaSerSerProLysArgProG1yProTrpSerG1uProSerVal1G1yThrThrLeuVal1ProAla1SerG1uPro	675	
2232	CCATTATCATGGCTGTGAAAAGAGATGGCTTGGAGTAACACATTAAATAGCTTGGCCAGGAGAGTCTTCTGATATAGGAAATGTCAGACATGGATTGGTATAACAAACAGCCTCTACTACAGT	2366	
676	ProPheIleMetAlaVal1LysG1uAspG1yLeuTrpSerLysProLeuAsnSerPheG1yProG1yGluPheLeuSerSerAspIleG1yAsnVal1SerAspMetAspTrpTyrAsnAsnSerLeuTyrTyrSer	720	
2367	GACACGAAAGGCCGACCTTGTGGCTGCTGAATGGGACCGGATATCTCAGAGAATTACACCTACAGCAGGAGCAGGGCTTACGTTGGTGGGATCTGACTGGACACACTCTATTCTAGTGGAA	2501	
721	AspThrLysG1yAspVal1PheVal1TrpLeuAsnG1yIleSerG1uAsnTyrH1sLeuProSerIleAlaG1yAlaG1yAlaLeuAlaPheG1uTrpLeuG1yIhsPheLeuTyrTrpAlaG1yLys	765	
2502	ACATATGTCATCAAAGGAGCTGTCGTTGAGGGACACAGACATGGTACCTGGGATGACATGGTGAATGACATGGTGGGATGTTCTGAGTGGGCTCTGTATTGGTTGGGTCAGACACTCTATTCTAGTGGAA	2636	
766	ThrTyrVal1IleG1nArgG1nSerVal1LeuThrG1yH1sThrAspIleVal1ThrH1sVal1LysLeuLeuVal1AsnAspMetVal1Val1AspSerVal1G1yTyrLeuTyrTrpThrLeuTyrSerVal1G1u	810	
2637	AGCACAGACAAATGGGAAAGTCTCTGACTACAGACACAGCCTGGTTCTGGGAAAGGTAATTGCTCTAAGCTTACGTTGAGCTGATGGGCTCTGTATTGGTTGGGTCAGACAGTCATGTT	2771	
811	SerThrArgLeuAsnG1yGluSerSerLeuVal1LeuG1nThrG1nProTrpPheSerG1yLysLysVal1IleAlaLeuThrLeuAspLeuSerAspG1yLeuLeuTyrTrpAlaG1yLys	855	
2772	CACCTGTACACAGCTTCTGGGAGCAGAGCACTGGGGATACCCACATCAGAGAATTTCGAGCTGAGGACTCTCTGAAATTCTCCAGAATGCACTGACTATAGTGGTCGGCTTCTGGATCAATGGC	2906	
856	H1sLeuTyrThrAlaVal1LeuArgG1yG1nSerThrG1yAspThrThrIleThrG1uPheAlaIlaTrpSerThrSerG1uIleSerG1nAsnAlaLeuMetTyrTyrSerG1yArgLeuPheTrpIleAsnG1y	900	
2907	TTAGGATTATCACAACTCAAGAAATAGGTCAAGAACCCAGTGTCTGTTGGAAACCCAGCAGATTAACTGAGTCAACATTATTCAAGCATCCCTAGGCCAGGAAACTTTCTCTACCCCTAAAG	3041	
901	PheArgIleIleThrG1nG1uIleG1yG1nLysThrSerVal1SerVal1LeuG1uProAlaArgPheAsnG1nPhenThrIleIleG1nThrSerLeuLysProLeuProG1yAsnPhenSerPheThrProLys	945	
3042	GTTATTCCAGATTCTGTCAGAGCTCTTCTAGGATTGAGGAAATGCTCAAGGTTTCAAATCTGAGCTTCAAGGTTTCAAATCTGAGGTTGAGGCTTCTGAGCTGAGGTTGAGACTGGGCTACT	3176	
946	V1IleProAspSerVal1G1nG1uLysSerPheArgIleG1uG1yAsnAlaSerSerPheG1nIleLeuTrpAsnG1yProProAlaVal1AspSerVal1G1yTyrLeuTyrTrpThrLeuTyrSerVal1G1u	990	
3177	AAGTCTGGCTAGTGAACACACTCTTACCTGTTACTGTGAGGACTGGACCTTATGCTTAAATCTCTGACTCTCTTACCTACTGGGAAAGGCCAAACACATCTGTCACTTCGA	3311	
991	LysPheLeuAlaSerG1uG1nH1sSerLeuProVal1PheThrVal1G1uG1yLeuG1uProTyrAlaLeuPheAsnLeuSerVal1IleAlaLeuThrLeuAspLeuSerAspG1yLeuLeuTyrTrpLysThrSerLeuLeuArg	1035	
3312	GCACCTGAAACAGCTTCCATCAGCACCAAGAGAACCCAGAATATTATACCAAGTGGAAAATGCTGCAACAGAATGAAGTGGTGGGAAATTAGGTGGAACAAACCTAACGATGAAAATGGGTGTTAAC	3446	
1036	AlaProG1uThrVal1ProSerAlaProG1uAsnProArgIlePheIleLeuProSerG1yLysCysCysAsnLysAsnG1uVaIvaIvaIvaIg1uPheArgTrpAsnLysProLysH1sG1uAsnG1yVal1LeuThr	1080	
3447	AAATTGAAATTTCTACAAATATCCAATCAAAGTATTACAAACAAACATGTAAGAGCTGGATTGCTGTCATGTCACTCCCTAGTGTGATGTCATGCTTCACTGAGTCAGGCTCAGATGCTTATTGCT	3581	
1081	LysPheG1uIlePheTyrAsnIleSerAsnG1nSerIleThrAsnLysThrCysG1uAspTrpIleAlaVal1AsnVal1ThrProSerVal1MetSerPheG1nLeuG1yMetSerPheProArgCysPheIleAla	1125	
3582	TTCCAGGTTAGGCCCTTACATCTAACGGGCCAGGACCATGCTGAGCTGAAAGCTACAAACATCAGAAAATCACCCATTCTCCTACATCTCTTCTGGAACAAAGATAGTTTTAGATAGTGGAT	3716	
1126	PheG1uVal1ArgAlaPheThrSerLysG1yProG1yProTyrAlaAspVal1Val1LysSerThrSerG1uIleAsnProPheProHisLeuIleThrLeuLeuG1yAsnLysIleVal1PheLeuAspMetAsp	1170	
3717	CAAATCAGTTGTTGGACGTTTCAAGGAGAGTATCACAGTATTACAGTACTGGGTTACACAGCTGATAATGAGATGGGATATTAGTGTGAGGAGGACTCACTCTCTGCACTGCAACATGCTCTAGCT	3851	
1171	G1nAsnG1nVal1Val1TrpThrPheSerAlaG1uArgVal1IleSerAlaVal1CysTyrThrAlaAspAsnG1uMetG1yTyrTyrAlaG1uG1yAspSerLeuPheLeuH1sAsnArgSerSer	1215	
3852	GAGCTTCTCAAGAGTTCAGCTGGTTTGTGATATCACAGTATTACAGTACTGGGTTACAGGACACCTCTACTTGCACAGAATGCAAGTATTGAGTGTGAGTGTGACACAAAGGTC	3986	
1216	GluLeuPheG1nAspSerLeuVal1PheAspIleLeuThrVal1IleLeuAspTrpIleSerArgH1sLeuTyrPheAlaLeuLysG1uSerG1nAsnG1yMetG1nVal1PheAspVal1AspLeuG1uH1sLysVal1	1260	
3987	AAATATCCCAGAGGAGATTCAACATAGGAATTCAACAAATATTCTTCTGTTATCTCTTAAAGTCGCTGTTGAGGAGCAGAGTGGCTACCAAGATGCTTACAGTATTGAGTGTGAGTGTGACACAAAGGTC	4121	
1261	LysTyrProArgG1uVal1LysIleH1sAsnArgAsnSerThrIleIleSerPheSerVal1ItyProLeuLeuSerArgLeuPheTyrTrpG1uVal1SerAsnPhenPheG1yTyrTyrSerIleIle	1305	
4122	AGTCACACTTGGCACCGAAATTCTGCAACCCACAGCTACAAACCAACAAAACAAAGGAATCAATGTCATGTTGAGTGTGACTGAATTGAGTGTGAGTGTGACACAAAGGCTATTGATACCTCTAACCTAGAAGAACCA	4256	
1306	SerH1sThrLeuH1sArgIleLeuG1nProAlaIleThrAsnG1nTg1nAsnLysArgAsnG1nCysSerCysAsnVal1ItrhG1uPheG1uLeuSerG1yAlaMetAlaIleAspThrSerAsnLeuG1uLysPro	1350	

FIG. 2. (*Figure continues on the opposite page.*)

4257	TTGATATACTTGCACAAAGCACAAGAGACTGGCAATGGATCTGGAGGCTGTCAGTTGGAGAGTTATCACAGTACCTGCTATGCTCGAGAAAAACCCCTGTTAGCTTAACGTGGATGGAGATCTTATA	4391
1351	LeuIleTyrPheAlaLysAlaGluIleTrpAlaMetAspLeuGluGlyCysGlnCysTrpArgValIleThrValProAlaMetLeuAlaGlyLysThrLeuValSerLeuThrValAspGlyAspLeuIle	1395
4392	TACTGGATCATCACAGCAAAGGACAGCACACAGATTTCAGGCAAAGAAAAGGAAATGGGCCATCGTCTCCAGGTGAAGGCCCTAAGGAGTAGGCATATCTGGCTTACAGTTCAGTTATGCGCTTACGCCCTTCCA	4526
1396	TyrTrpIleIleThrAlaLysAspSerThrGlnIleTyrGlnAlaLysLysGlyAlaIleValSerGlnValLysAlaLeuArgSerArgHisIleLeuAlaTyrSerSerValMetGlnProPhePro	1440
4527	GATAAACGGCTTCCTGCTCTAGCTCACAGACACTGAGAACCAACTATACTTAATGCCACTAACACTAGCCTCACAACTAGCATTTACCTCGCCAGAACAAACCTCACATGGTATGGCATACCAGCCCTACTCCA	4661
1441	AspLysAlaPheLeuSerLeuAlaSerAspThrValGluProThrIleLeuAsnAlaThrAsnThrSerLeuThrIleArgLeuProLeuAlaLysThrAsnLeuThrTrpTyrGlyIleThrSerProThrPro	1485
4662	ACATACCTGGTTTATTATGAGCAGGAAAACAGCTGACTGAAATATAGAATTCTGGAGACAGTAGCTTATTGAGAATTCTACACCATTTCAACATACATGATAACAGATA	4796
1486	ThrTyrLeuValTyrAlaGluValAsnAspArgLysAsnSerSerAspLeuLysTyrArgIleLeuGluPheGlnAspSerIleAlaLeuIleGluAspLeuGlnProPheSerThrTyrMetIleGlnIle	1530
4797	GCTGAAAAAAATTATTTCAGATCTTGGAACATTACCAACAGGAAAAGGATTTGGGGAAAAGCTAAAGGAGTACCGAGGAGCTGAGCTCATTAATACACTGTCGGCTCAGACACCAGCCCTATT	4931
1531	AlaValLysAsnTyrTyrSerAspProLeuGluHisLeuProProGlyLysGluIleTrpGlyLysThrLysAsnGlyValProGluAlaValGluIleLeuAlaAsnThrThrValArgSerAspThrSerLeuIle	1575
4932	ATATCTGGAGAGAACATCTCACAGCCAAATGGACCTAAAGAATCAGTCAGCTTACGGCAATCTCACACCTGGCCCTAATTCGAAACTCTTAAGACAAAGTGAATTCTCAAAATGGAGGCTACTCTC	5066
1576	IleSerTrpArgGluSerHisLysProAsnGlyProLysGluSerValArgTyrGlnLeuAlaIleSerHisLeuAlaLeuIleProGluThrProLeuArgInSerGluPheProAsnGlyArgLeuThrLeu	1620
5067	CTTGTACTAGACTGCTGGTGGAAATATTATGTTAAAGGTTCTGCCACTTGAGGAAATGTTGTTACAGAGACTCATCTGTCACTGTTAACACACAGGAAACCTTATTCTG	5201
1621	LeuValThrArgLeuSerGlyAsnIleTyrValLeuLysValLeuAlaCysHisSerGluGluMetTrpCysThrGluSerHisProValThrValGluMetPheAsnThrProGluLysProTyrSerLeu	1665
5202	GTTCAGAGAACACTGTTGCAATTAAATGGAAGGCTCATTGAATGTTAACCTCATCAGATTGGGTTGAGCTACAGAAGTGAATACATGAGTTTACCATGTTAACACTCATGAGCCAAAGGCTCT	5336
1666	ValProGluAsnThrSerLeuGlnPheAsnTrpLysAlaProLeuAsnValAsnLeuIleArgPheTrpValGluLeuGluLysTrpLysTyrAsnGluPheThrYHisValIleThrSerCysSerGlnGlyPro	1710
5337	GCTTATGTCGTAATATCACAAATCTACAAACCTTAACTTCATATAATGTCAGAGTAGTGGGGTTAAAGACGGGAAAGTACGACCTCACTTCCAGAAAGCTTAAAGACAAAGCTGGAGTCCAAATAAA	5471
1711	AlaTyrValCysAsnIleThrAsnLeuGlnProTyrThrSerTyrAsnValArgValValValTyrLysThrGlyGluAsnSerThrSerLeuProGluSerPheLysThrLysAlaGlyValProAsnLys	1755
5472	CCAGGCATTCCAAATTAATGAGGGAGTAAATCAACAGTGGGAAAGCTGAGGATAATGGAGTGAAGATTACACTATATCCTGAGATAAGAACAGCACTTCAAATAATTACAGAACAGAAAT	5606
1756	ProGlyIleProLysLeuGluGlySerLysAsnSerIleGlnTrpGluLysAlaGluAspAsnGlyCysArgIleThrTyrTyrIleLeuIleArgLysSerThrSerAsnAsnLeuGlnAsnGlnAsn	1800
5607	TTAAGGTGGAAGATGACATTAATGGATCTGAGTAGTGGTTGCACATGGAAAGTCCAAAACCTGAAAGGAATATTCTAGTTAGCTAGTACGCTCAAATAATCTAGGGTTGGTGAATATAGTGGAACTAGT	5741
1801	LeyArgTrpLysMetThrPheAsnGlySerCysSerSerValCysThrTrpLysSerLysAsnLeuLysGlyIlePheGlnPheArgValValAlaAsnAsnLeuGluGlyIleTyrSerGlyIleSer	1845
5742	GAGAAATTATTATTAGTGGAGATGATTGGGATACCGAGAACAAAGTTCTACTTACTATTAGTGGAAATTCTGGTTACAATCCCAACTGACCTTGTCTGGCATAGAAGATTAAAGAATCAAAA	5876
1846	GluAsnIleIleLeuValGlyAspAspPheTrpIleProGluThrSerTyrIleLeuThrIleValIleGlyIlePheLeuValValThrIleProLeuThrPheValIleProGluHisArgLeuLysAsnGlnLys	1890
5877	AGTGCAGGAAAGGGGTGACAGTGCCTTAAACGAGAACAGGTTGGCTGAGGCTGGCAGGGCTGGCAGCCGGAGTAGGCTCTGGCTATGCAAAATACACTACTCTCCAAACCCAAGGAGATTGAA	6011
1891	SerAlaLysGluGlyValThrValLeuIleAsnGluAspLysGluLeuAlaGluLeuArgGlyLeuAlaAlaGlyValGlyLeuAlaAsnAlaCysTyrAlaIleHisThrLeuProThrGlnGluIleGlu	1935
6012	AATCTTCTGCCTTCCCCTGGGAAACACTGACTCTGCCTCTGGCTGGGAGCTTTGGAGATGTTGAAAGGACAGCAGTGGACATCTTGGAGTTGGAGAATCTGGAGGATGAA	6146
1936	AsnLeuProAlaPheProArgGluLysLeuThrLeuArgLeuLeuGlySerGlyAlaPheGlyGluValTyrGluValAlaPheIleLeuGlyValGlyIleLysValAlaValLys	1980
6147	ACTTTGAGAAGGGTCCACAGCAGGAGAATGAGATTCTGAGGAGGACATCTGATGAGCAATTAAATCATCCCAACATTCTGAGCAGCTGGAGTTGGCTCTGCTGAATGAGAACCCCAATACATTATC	6281
1981	ThrLeuLysGlySerThrAspGlnGluLysIleGluPheLeuLysGluAlaAlaHisLeuMetSerLysPheAsnHisProAsnIleLeuLysGlnLeuGlyValCysLeuLeuAsnGlnProGlnTyrIleIle	2025
6282	CTGGAACTGATGGAGGAGGAGACCTTCTACTTATTCGCTAAAGCCGGATGGCACGTTGTTACCTCCCTGCTGACCTTGTGAGCTGTGTTAGATATTCAAAAGGCTGTGCTACTTG	6416
2026	LeuGluLeuMetGluGlyAspLeuLeuThrTyrLeuArgLysAlaArgMetAlaIleThrPheTyrGlyProLeuLeuThrLeuValAspLeuValAspIleSerLysGlyCysValIleTyrLeu	2070
6417	GAACGGATGCATTTCATTACAGGGATCTGGCAGCTGCAAATTGCTTCTCGTGAAGACTATACAGTCACGGAGTAGTGAAGATTGGAGACTTGGACTCGCCAGAGACATCTAAAAATGATTACTAT	6551
2071	GluArgMetHisPheIleHisArgAspLeuAlaAlaArgAsnCysLeuValSerValLysAspTyrThrSerProArgIleValIleGlyAspPheGlyLeuAlaArgAspIleTyrLysAsnAspTyrTyr	2115
6552	AGAAAAGAGGGGAAGGCCCTGCTCCAGGTTGGTGGAGCTTCAGGCTCCAGGAAAGTTGATGGATGGAATCTTCAACTCACTGATGATGGTCTTTGGAGATTCTGATTTGGAGATTTAACTCTGGCTCATCAG	6686
2116	ArgLysArgGlyGluGlyLeuLeuProValArgTrpMetAlaProGluSerLeuMetAspGlyIlePheThrThrGlnSerAspValTrpSerPheGlyIleLeuIleTrpGluIleLeuThrGlyHisGln	2160
6687	CCTTATCCAGCTATTCCACCTTGATGTTAAACTATGTCGAAACAGGAGGAGACTGGGACCAAGAAATTGTCCTGATGATCTGTTGGAAATTATGACCCAGTGTGGCTGGCTCAAGAACCCGACCAAGA	6821
2161	ProTyrProAlaHisSerAsnLeuAspValLeuAsnTyrValGlyGlyArgLeuGluProProArgAsnCysProAspLeuTrpAsnLeuMetThrGlnCysTrpAlaGlnGluProAspGlnArg	2205
6822	CCTACTTTCATAGAATTCAAGACCAACTTCAGTTACAGAAATTCTTAAATAGCATTATCAGTGCAGAGATGAGCAGAACACAGTGGAGTCATAAATGAAAGCTTGAAGGAGATGCGCGATGTG	6956
2206	ProThrPheHisArgIleGlnAsnGlnLeuIlePheArgAsnPhePheLeuAsnSerIleTyrGlnCysArgAspGluAlaAsnAsnSerGlyIleAsnGluSerPheGlyGluAspGlyAspVal	2250
6957	ATTGTTGAACTTACAGATGACATTGCAAGTGTGTTAACTGGAAACGAAGAACCCGAGAAGGGTTAACATATGGTACTTGCCTACAGAATGTGCCAAGGTTGAAGAAAAGTCTGAGGGCTCTAGGCTCCCCAG	7091
2251	IleCysLeuAsnSerAspAspIleMetProValIleLeuMetGluIleThrLysAsnArgGluGlyLeuAsnTyrMetValIleAlaIleThrGluCysGlyGluGluLysSerGluGlyProLeuGlySerGln	2295
7092	GAATCTGAATCTTGTGCTGAGGAAAGAGAGAACAGAACACATGAGCAGAACAGATTCTGCCAAGAAAACAAGTGTGCTTACTGCCCTCTGGCAAGGCTGAAACTATGCTGCTACTCAGCT	7226
2296	GluSerGluSerCysGlyLeuArgLysGluGluLysGluProHisAlaAspLysAspPheCysGlnGluLysGlnValAlaTyrCysProSerGlyLysProGluGlyLeuAsnTyrAlaCysLeuThrHisSer	2340
7227	GGATATGGAGATGGGCTGATTAATAGCGTTGTTGGAAATAGAGAGTTGAGATAAACACTCTCATCGTAGTTACTGAAAGAAAACACTCTGCTAGATGATAATGTCATGGGGTCTATAACTCCAAATAAA	7361
2341	GlyTyrGlyAspGlySerAspEnd	2347
7362	CAATGCAACGTTCC	

FIG. 2. Nucleotide sequence of ROS1 cDNA from SW-1088 cells and deduced amino acid sequence. The putative signal sequence is underlined. The start of the sequence that is also present in *MCF3* is indicated by an arrow. The putative transmembrane domain is underlined, and the tyrosine kinase domain is boxed. The positions where the ROS1 cDNA of SW-1088 differs from the other ROS1 coding published sequences are marked by arrowheads. Nucleotide and amino acid coordinates are indicated in the margin.

## RESULTS

The ROS1 cDNA was isolated in several steps from two libraries that were prepared from poly(A)<sup>+</sup> RNA of the glioblastoma cell line SW-1088. An oligo(dT)-primed library was first screened by hybridization with a *ROS1*-specific probe derived from the previously isolated cDNA of the *MCF3* gene (2). In the subsequent steps, we used 5' sequences from new cDNA isolates as probes and finally screened a second library prepared with an internal *ROS1*-specific primer (for details see *Material and Methods*). Four overlapping cDNA clones, which together span 8.3 kilobase pairs (kbp), were chosen for further sequence analysis (Fig. 1).

The sequence of the composite ROS1 cDNA (Fig. 2) has one large open reading frame, which starts at position 207 and

ends at position 7247 with two consecutive termination codons. Approximately 1 kbp of 3' untranslated sequence follows, which was not fully sequenced. We have assigned the ATG at positions 207–209 as the initiating codon because upstream termination codons exist in all three reading frames and because the nucleotide sequence flanking this ATG fulfills Kozak's criteria for an authentic initiation codon (22). Moreover, the downstream stretch of 36 amino acids has all the features of a signal sequence (23), which includes a hydrophobic stretch of 21 amino acids, clearly identifiable on a hydropathic profile (Fig. 3). By analogy with other signal-peptide cleavage sites (24), the amino terminus of the mature *ROS1* gene product would be Cys-37 (Fig. 2). The predicted molecular weight of the *ROS1* gene product, without post-

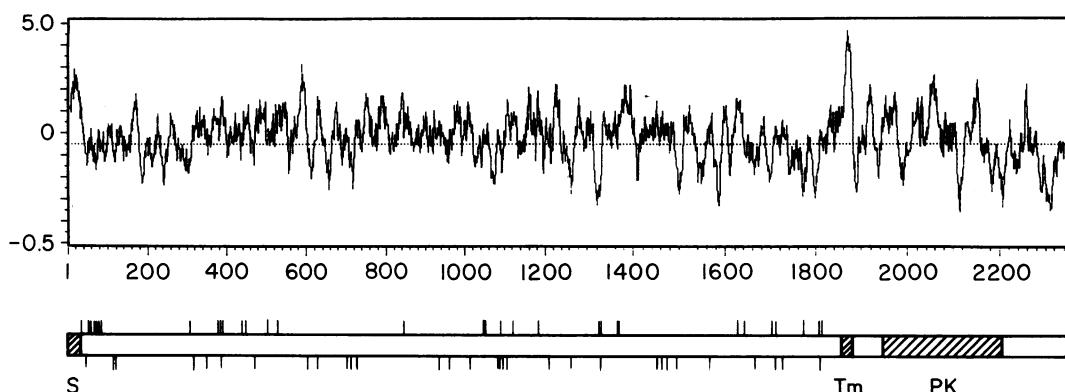


FIG. 3. Hydropathicity of the predicted *ROS1*-encoded protein. (Upper) The hydropathicity index (21) was determined with a window setting of 15 amino acids. Coordinates refer to amino acid positions of Fig. 2. (Lower) Hatched boxes represent the putative signal sequence (S), the transmembrane domain (Tm), and the tyrosine protein kinase (PK) domain. Vertical lines mark the positions of cysteine residues in the extracellular domain (above) and potential glycosylation sequences (below).

translational modification and after cleavage of the signal peptide, would thus be 259,000.

The amino acid sequence of the *ROS1* gene product from position 37 to position 1861 constitutes the putative extracellular domain and includes 31 potential N-linked glycosylation sites (Fig. 3). This sequence does not show similarities to the extracellular domains of previously described mammalian receptor-type tyrosine kinases. When compared to the epidermal growth factor or insulin receptor classes of tyrosine kinases (25), the *ROS1* protein is not rich in cysteine residues except for one cluster of 11 cysteines at the very amino terminus. The spacing of cysteine residues is also not similar to the spacing found in the ligand binding domain of the platelet-derived growth factor receptor, the prototype of the third class of transmembrane tyrosine kinases (25). The consecutive stretch of 21 hydrophobic amino acids (residues 1862–1882) constitutes the putative transmembrane domain, which is followed by four closely spaced arginine and lysine residues (Fig. 3). The carboxyl-terminal 464 amino acids constitute the cytoplasmic domain and include sequences typical for tyrosine-specific protein kinases.

Two partial human *ROS1* sequences have been reported previously. One, the cDNA sequence of the activated human *ROS1* gene, *MCF3*, encodes a truncated protein missing all but eight amino acids of the extracellular domain (2). It corresponds to nucleotides 5764–7375, which encode amino acids 1854–2347, of the sequence of the full-size *ROS1* cDNA. The other *ROS1* sequence was determined from a placental genomic DNA clone isolated by virtue of its homology to the chicken *v-ros* gene (26). It corresponds to nucleotides 5573–6941, which encode amino acids 1790–2245, of *ROS1* cDNA from SW-1088 cells. There are no differences between the coding sequences of the *MCF3* gene and the human placental *ROS1* gene in the region of overlap between the two. There are, however, five differences between these sequences and the sequence of the *ROS1* cDNA from SW-1088 cells (indicated in Fig. 2). The difference at position 6453 is silent. It changes the AGA codon (arginine) to CGA (Arg-2083). The differences at positions 6843, 6888, 6892, and 6991 change GAC (aspartic acid) to AAC (Asn-2213), AAG (lysine) to CAG (Gln-2228), TCC (serine) to TGC (Cys-2229), and GCT (alanine) to GTT (Val-2262), respectively. Apparently, the *ROS1* gene from the glioblastoma cell line SW-1088 has accumulated several mutations. The effect of these alterations on the physiological function of the gene product is presently unknown.

Previous investigators noted that the tyrosine kinase domains of the *ROS1* and the *Drosophila* sevenless gene products are closely related (12). In addition, we show here that both genes have also the potential to encode proteins of

similar overall structures—i.e., transmembrane tyrosine kinases with unusually large extracellular domains. Further comparison by DOTPLOT analysis (Fig. 4) demonstrates that the sequences of the *ROS1* and sevenless gene products can be aligned over more than 2200 amino acids and that homologies exist in the extracellular domains as well. The extent of similarity differs in distinct parts of the proteins. Extensive similarities are present in the intracellular domain, and patches of homologies are found over 900 amino acids located in the amino-terminal half of the extracellular domain. In addition, the sevenless gene product contains a cluster of cysteine residues (13, 14) that occurs at a position that is roughly comparable to the location of the cysteine cluster of the *ROS1* protein. The very amino-terminal 200 amino acids of the sevenless protein, which constitute a second potential transmembrane domain, have no equivalent in the *ROS1* protein. No comparable alignment could be found when the extracellular domains of the *ROS1* protein and other receptor-type tyrosine kinases (insulin, epidermal growth factor receptor, platelet-derived growth factor receptor, *c-kit*, and *c-fms*) were analyzed under the same stringency.

## DISCUSSION

We have described the sequence of an 8.3-kb transcript of the *ROS1* gene from the human glioblastoma cell line SW-1088.

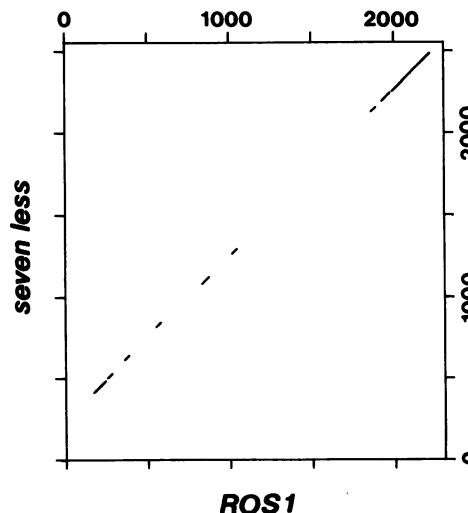


FIG. 4. DOTPLOT comparison of the amino acid sequences of *ROS1* and sevenless gene products. The predicted amino acid sequences of *ROS1* and sevenless gene products were analyzed for similarities at a window length of 30 and a stringency setting of 21 using the COMPARE/ALL option of the UWGCG package.

This transcript has the potential to encode a protein with an intracellular domain typical of tyrosine protein kinases, a transmembrane domain, a very large extracellular domain, and a putative amino-terminal signal peptide. Thus, *ROS1* resembles several protooncogenes that encode transmembrane tyrosine kinases that function as receptors. The extracellular domains of these proteins can be categorized by size and the distribution of cysteine residues. The amino acid sequence of the extracellular domain of the *ROS1* product resembles none of the known mammalian transmembrane tyrosine kinases. However, it does resemble the sequence of the extracellular domain of the *Drosophila* sevenless gene product in size, amino acid sequence, and distribution of cysteine residues. *ROS1* and sevenless genes thus encode members of a distinct subfamily of transmembrane tyrosine kinases. Other similarities in structure and function may therefore exist between these proteins.

An 8.3-kb transcript of the *ROS1* gene is found frequently in human glioblastoma cell lines but not in a primary glial cell line or in adult brain tissue. Rearrangement of the *ROS1* gene and expression of a truncated transcript has been observed in one particular glioblastoma cell line, U-118 MG (10). In the sequence from the glioblastoma cell line SW-1088, we found four amino acids changes carboxyl-terminal of the tyrosine kinase domain, a part of the protein implicated in regulation of enzyme activity. Point mutations in this region have been observed to activate the oncogenic potential of the *c-src* and the *c-erbB* genes (25, 27). Thus, even in the absence of chromosomal rearrangements, such mutations might alter the physiological function or activate the oncogenic potential of the *ROS1* gene product and contribute to the malignancy of glioblastomas. In primary human tumors of glial origin, frequent chromosomal abnormalities have been observed on chromosome 17 and chromosome 10 (28). *ROS1* is located on chromosome 6 (29). It should be interesting to examine whether a particular chromosomal abnormality or a specific stage in tumor progression correlates with expression of the *ROS1* gene in primary tumors. Since the sequence of the *ROS1* cDNA predicts a transmembrane protein accessible from the outside of the cell, the *ROS1* protein might provide a specific target for antibody-based diagnosis or therapy.

We thank Drs. A. Ullrich, J. Dangl, W. Birchmeier, and K. Ferguson for reading the manuscript and T. Etzold and K. Ferguson for help with computer analysis. We also thank U. Ringeisen and M. Ockler for their help with the art work and P. Bird for preparation of this manuscript. This work was supported by grants from the National Institutes of Health and the American Cancer Society, by the Pfizer Biomedical Research Award, and by Grant 0316 150a from the Bundesministerium fuer Wissenschaft und Technologie. M.W. is an American Cancer Society Research Professor.

1. Fasano, O., Birnbaum, D., Edlund, L., Fogh, J. & Wigler, M. (1984) *Mol. Cell. Biol.* **4**, 1695–1705.
2. Birchmeier, C., Birnbaum, D., Waitches, G., Fasano, O. & Wigler, M. (1986) *Mol. Cell. Biol.* **6**, 3109–3116.
3. Yamamoto, T., Nishida, T., Miyajima, N., Kawai, S., Ooi, T. & Toyoshima, K. (1983) *Cell* **35**, 71–78.
4. Hampe, A., Gobet, M., Sherr, C. & Galibert, F. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 85–89.
5. Bargmann, C., Hung, M. & Weinberg, R. (1986) *Nature (London)* **319**, 226–230.
6. Martin-Zanca, D., Hughes, S. & Barbacid, M. (1986) *Nature (London)* **319**, 743–748.
7. Yarden, Y., Kuang, W., Yang-Feng, T., Coussens, I., Munemitsu, S., Dull, T., Chen, E., Schlessinger, J., Francke, U. & Ullrich, A. (1987) *EMBO J.* **6**, 3341–3351.
8. Ullrich, A., Coussens, L., Hayflick, J., Dull, T., Gray, A., Tam, A., Lee, J., Yaden, Y., Libermann, T., Schlesinger, J., Downward, J., Mayes, E., Whittle, N., Waterfield, M. & Seuberg, P. (1984) *Nature (London)* **309**, 418–425.
9. Sherr, C., Rettenmier, C., Sacca, R., Roussel, M., Look, A. & Stanley, E. (1985) *Cell* **41**, 665–676.
10. Birchmeier, C., Sharma, S. & Wigler, M. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 9270–9274.
11. Sharma, S., Birchmeier, C., Nikawa, J., O'Neill, K., Rodgers, L. & Wigler, M. (1989) *Oncogene Res.* **5**, 91–100.
12. Hafen, E., Basler, K., Edstroem, J. & Rubin, G. (1987) *Science* **236**, 55–63.
13. Basler, K. & Hafen, E. (1988) *Cell* **54**, 299–311.
14. Bowtell, D., Simon, M. & Rubin, G. (1988) *Genes Dev.* **2**, 620–634.
15. Maniatis, T., Fritsch, E. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
16. Huynh, T., Young, R. & Davis, R. (1985) in *DNA Cloning, a Practical Approach*, ed. Glover, D. (IRL, Oxford), Vol. 1, pp. 49–78.
17. Colicelli, J., Birchmeier, C., Michaeli, T., O'Neill, K., Riggs, M. & Wigler, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 3599–3603.
18. Vieira, J. & Messing, J. (1987) *Methods Enzymol.* **153**, 3–11.
19. Sanger, F., Nicklen, S. & Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
20. Kyte, J. & Doolittle, R. (1982) *J. Mol. Biol.* **157**, 105–132.
21. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
22. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
23. Watson, M. (1984) *Nucleic Acids Res.* **12**, 5145–5146.
24. von Heijne, G. (1985) *J. Mol. Biol.* **184**, 99–105.
25. Yarden, Y. & Ullrich, A. (1988) *Annu. Rev. Biochem.* **57**, 443–478.
26. Matsushime, H., Wang, L. & Shibuya, M. (1986) *Mol. Cell. Biol.* **6**, 3000–3004.
27. Reynolds, A., Vila, J., Lansing, T., Potts, W., Weber, M. & Parson, T. (1987) *EMBO J.* **6**, 2359–2364.
28. James, D., Carlom, E., Nordenskjold, M., Collins, P. & Cavenee, W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2858–2862.
29. Rabin, M., Birnbaum, D., Young, D., Birchmeier, C., Wigler, M. & Ruddle, F. (1987) *Oncogene, Res.* **1**, 169–178.