

Complete Sequence Analysis of a Gene (OS-9) Ubiquitously Expressed in Human Tissues and Amplified in Sarcomas

Yan A. Su, Carolyn M. Hutter, Jeffrey M. Trent, and Paul S. Meltzer¹

Laboratory of Cancer Genetics, National Center for Human Genome Research, National Institutes of Health, Bethesda, Maryland

Amplification and overexpression of genes involved in cellular growth control occur frequently in human tumors. Using a chromosome microdissection-based hybrid-selection strategy, we recently identified two novel genes (OS-9 and OS-4) within 12q13–15, a region frequently amplified in human cancers. We now report further characterization of the full-length OS-9 cDNA sequence. This cDNA sequence consists of 2785 bp from which an open reading frame (ORF) with 667 amino-acid residues was deduced. The predicted polypeptide was water soluble and acidic. We also demonstrate that the OS-9 gene encoded a 2.8-kb mRNA transcribed in all 16 human tissues examined, suggesting that OS-9 is ubiquitously expressed in human tissues. OS-9 was co-amplified with *CDK4* in three of five sarcoma tissues. Homology analysis of the amino-acid sequence reveals significant similarities between OS-9 and two ORFs deduced from genomic sequences in *Caenorhabditis elegans* and *Saccharomyces cerevisiae*. The region of similarity extended over 200 residues (approximately one-third of each ORF), and eight cysteines were conserved in all three ORFs. These observations suggest that this region comprises a functional domain present in a novel evolutionarily conserved gene family defined by OS-9.

© 1996 Wiley-Liss, Inc.*

Key words: Neoplasm, genetics, chromosome 12q

INTRODUCTION

In human cancers, gene amplification is a common mechanism by which increased dosage of a gene leads to its overexpression. Amplification of cellular oncogenes has been observed in tumor cell lines and primary tumor tissues, suggesting that overexpression of these genes provides tumor cells with a selective growth advantage in vitro and in vivo.

The human osteosarcoma cell line OsA-C1 [1] contains a homogeneously staining region (hsr) composed of sequences derived from 12q13–15. Several amplified genes have been mapped within this hsr, including those for a zinc-finger protein (*GLI*) [1], a member of the transmembrane 4 superfamily (*SAS*) [2,3], cyclin-dependent kinase 4 (*CDK4*) [4], a transcription factor (*CHOP*) [5], and a modulator of p53 (*MDM2*) [6,7]. Two additional novel genes (OS-9 and OS-4) were recently identified from this hsr [8]. Because the OS-9 gene was shown to be amplified in additional sarcoma specimens, we further characterized its cDNA. The 2785-bp full-length cDNA encodes an open reading frame (ORF) of 667 amino-acid residues. Northern blot analysis demonstrated OS-9 expression in all human tissues examined. Analysis of the amino-acid sequence revealed a significant similarity with two ORFs deduced from genomic sequences of *Caenorhabditis elegans* [9] and *Saccharomyces cerevisiae* [10,11].

MATERIALS AND METHODS

Library Screening and Generation of Plasmid DNA

The generation of the cDNA library (in bacteriophage λ ZAPII) highly enriched for transcribed sequences from the 12q hsr has been previously described [8]. The library was plated at a density of 628 plaque-forming units per 150-mm petri dish in *Escherichia coli* XL1-blue cells. Plaque lifts and hybridizations were performed by standard methods [12]. A total of 30 plaques was picked after hybridization with the OS-9 probe and used to excise pBluescript (Stratagene, La Jolla, CA) plasmids containing OS-9 cDNA inserts according to the manufacturer's protocol.

5' cDNA Synthesis and Cloning

Synthesis and amplification of the 5' end of the OS-9 gene were performed according to published methods [13,14] with the following modifications. OsA-C1 was cultured as previously described [8]. Total RNA was isolated by standard methods [15]. Poly(A)⁺ RNA was isolated from the total RNA with

¹Corresponding author: NCHGR/NIH, Bldg. 49, Room 4A10, 49 Convent Drive, Bethesda, MD 20892-4470.

Abbreviations: ORF, open reading frame; hsr, homogeneously staining region; CDK4, cyclin-dependent kinase 4; INT-1, integration site; utr, untranslated region; EST, expressed tagged sequence.

Dynabeads Oligo (dT)₂₅ (DynaL, Great Neck, New York) following the manufacturer's instructions. The 5' end of the OS-9 cDNA was synthesized with avian myeloma virus reverse transcriptase (CLONTECH, Palo Alto, CA). Poly(G) was added to the 5' end of the cDNA with terminal DNA transferase (GIBCO BRL, Gaithersburg, MD). Both OS-9-specific and 5'-end primers for the polymerase chain reaction amplification of the cDNA contain the sequence (CUA)₄ at their 5' ends, so the polymerase chain reaction products were directly cloned into the UDG cloning vector pAMP10 (GIBCO BRL).

DNA Sequencing and Sequence Analysis

The plasmids containing OS-9 cDNA inserts were sequenced by automated fluorescence sequencing. The sequence analysis of eight plasmid inserts derived from our cDNA library generated a consensus sequence of approximately 2 kb. Using the known sequence, we synthesized and cloned the 5' end of the OS-9 cDNA. Five plasmids containing 5'-end cDNA inserts were sequenced and analyzed. Both strands of all of the above-mentioned plasmids were sequenced. Thus, the nucleotide sequence of the full-length OS-9 cDNA was determined from the consensus sequence of both strands of 13 plasmids containing overlapping cDNA inserts. Multiple sequence alignments and consensus determinations were performed with DNASTAR software (version 1.58; DNASTAR Inc., Madison, WI). The current sequence databases were searched by using the BLASTN, BLASTP, and FASTA programs [16,17].

Northern and Southern Hybridization

The OS-9, human β -actin, *CDK4*, and integration site-1 (*INT-1*) probes were described previously [8,18] and labeled by random priming [19]. Northern and Southern hybridizations were performed by standard methods [20]. Filters were washed to a stringency of $0.1 \times$ standard saline citrate at 42°C. The relative intensity of the bands in radiograph was measured by densitometry and analyzed by using IP Lab Spectrum software (Signal Analytics Co., Vienna, VA).

RESULTS

OS-9 cDNA Sequence

Figure 1 shows the full-length OS-9 cDNA sequence and the predicted sequence of the longest ORF. The 2785-bp nucleotide sequence of the cDNA was determined from the consensus sequence of both strands of a total of 13 plasmids containing overlapping cDNA inserts. The full-length cDNA consisted of three segments: an 85-bp 5' untranslated region (UTR), nt 1–85; a 2001-bp coding region, nt 86–2086; and a 698-bp 3' UTR, nt 2087–2785. This nucleotide sequence encoded an ORF of 667 amino-acid residues. The deduced polypeptide consisted of three parts: an N-terminal hydrophobic sequence (residues

1–31), a cysteine-containing hydrophilic sequence (residues 32–337), and an acidic region with multiple predicted hydrophilic helices (residues 338–667). Eight cysteine residues and a possible nuclear targeting signal [21,22] within the ORF are also indicated in Figure 1. The calculated molecular weight of the polypeptide was 75 556 Da, and the estimated isoelectric point was 4.60. The cDNA contains a 48-bp nucleotide sequence with 16 GAD trinucleotide repeats (where D = G, A, or T), which encode an acidic domain composed of glutamic acid (E) and aspartic acid (D) (residues 414–429). The 3' UTR was 698 bp long with a polyadenylation signal 13 bp upstream from the polyadenylated site.

Homology Analysis

Comparison of the OS-9 cDNA sequence with the GenBank database (version 93.0) using the BLASTN program [16,17] revealed identical matches with 15 partial cDNA sequences (expressed, ESTs) derived from human fetal brain, spleen, liver, sequence tags and adult liver, ovary, placenta, and a colorectal cancer. The GenBank accession numbers of these ESTs are listed in Figure 1. No significant identity with any known gene was revealed by using the current GenBank and European Molecular Biology Laboratory databases, except for the partial OS-9 sequence, which we previously submitted. However, comparison of the OS-9 ORF with these databases by using the BLASTP program revealed a significant similarity with an ORF (F48E8.4) deduced from cosmid CELF48E8 mapped to chromosome III of *C. elegans* [9] and a somewhat weaker similarity with an ORF (YD9609.11) predicted from cosmid 9609 on chromosome IV of *S. cerevisiae* [10,11]. BLASTP aligned a 211 amino-acid segment (residues 62–267 (in brackets in Figure 1)) of OS-9 with F48E8.4 ($P = 1.8 \times 10^{-15}$) and YD9609.11 ($P = 1.4 \times 10^{-5}$). Within this homologous region (Figure 2), the identities between OS-9 and F48E8.4, OS-9 and YD9609.11, and F48E8.4 and YD9609.11 were 27.0% (57/211), 19.6% (45/230), and 22.5% (51/227), respectively, with identity plus similarity being 68.7% (145/211), 53.0% (122/230), and 60.0% (137/227), respectively. The most notable feature of this alignment is the conservation of eight cysteine residues in all three ORFs. In addition, the homologous region of all three proteins was preceded by a hydrophobic leader peptide. Furthermore, all three ORFs were predicted to be acidic proteins ($pI = 4.31, 4.60, \text{ and } 5.12$ for YD9609.11, OS-9, and F48E8.4, respectively).

Northern Hybridization

To evaluate the expression status of OS-9, membranes containing poly(A)⁺ RNA from multiple adult tissues (CLONTECH) were hybridized with the OS-9 probe and a control probe (human β -actin) that was used to normalize RNA loading. The result demonstrated that OS-9 was transcribed in all 16 human tissues (Figure 3). It was also noted by densitometry

TTGCACTCTCCACACCCCTTTCTTTTCGTCGCTCTTCGCTATTATTTCTCCCGCGTCTCTCTGCATAAGAAGGGGAACGAAAG ATG GCG GCG 94
M A A
 GAA ACG CTG CTG TCC AGT TTG TTA GGA CTG CTG CTT CTG GGA CTC CTG TTA CCC GCA AGT CTG ACC GGC GGT GTC 169
E T L L S S L L G L L L G L L L P A S L T G G V
 GGG AGC CTG AAC CTG GAG GAG CTG AGT GAG ATG CGT TAT GGG ATC GAG ATC CTG CCG TTG CCT GTC ATG GGA GGG 244
G S L N L E E L S E M R Y G I E I L P L F V M G G
 CAG AGC CAA TCT TCG GAC GTG GTG ATT GTC TCC TCT AAG TAC AAA CAG CGC TAT GAG TGT CGC CTG CCA GCT GGA 319
Q S Q S S D V V I V S S K Y K Q R Y E C R L F A G
 GCT ATT CAC TTC CAG CGT GAA AGG GAG GAG GAA ACA CCT GCT TAC CAA GGG CCT GGG ATC CCT GAG TTG TTG AGC 394
A I H F Q R E R E E T P A Y Q G P G I P E L L S
 CCA ATG AGA GAT GCT CCC TGC TTG CTG AAG ACA AAG GAC TGG TGG ACA TAT GAA TTC TGT TAT GGA CGC CAC ATC 469
P M R D A P C L L K T K D W W T Y E F C Y G R H I
 CAG CAA TAC CAC ATG GAA GAT TCA GAG ATC AAA GGT GAA GTC CTC TAT CTC GGC TAC TAC CAA TCA GCC TTC GAC 544
Q Q Y H M E D T S E I K G V L Y L G Y Y Q S A A F D
 TGG GAT GAT GAA ACA GCC AAG GCC TCC AAG CAG CAT CGT CTT AAA CGC TAC CAC AGC CAG ACC TAT GGC AAT GGG 619
W D D E T A K A S K Q H R L K R Y H S Q C T Y G N G
 TCC AAG TGC GAC CTT AAT GGG AGG CCC CGG GAG GCC GAG GTT CGG TTC CTC TGT GAC GAG GGT GCA GGT ATC TCT 694
S K C D L N G R P R E A E V R F L C D E G A G I S
 GGG GAC TAC ATC GAT CGC GTG GAC GAG CCC TTG TCC TGC TCT TAT GTG CTG ACC ATT CGC ACT CCT CGG CTC TGC 719
G D Y I D R V D GAG CCC TTG TCC TGC TCT TAT GTG CTG ACC ATT CGC ACT CCT CGG CTC TGC
 CCC CAC CCT CTC CTC CGG CCC CCA CCC AGT GCT GCA CCA CAG GCC ATC CTC TGT CAC CCT TCC CTA CAG CCT GAG 844
P H P L L R P P S A A P Q A I L C H P S L Q P E
 GAG TAC ATG GCC TAC GTT CAG AGG CAA GCC GAC TCA AAG CAG TAT GGA GAT AAA ATC ATA GAG GAG CTG CAA GAT 919
E Y M A Y V Q R Q A D S K Q Y G D K I I E E L Q D
 CTA GGC CCA CAA GTG TGG AGT GAG ACC AAG TCT GGG GTG GCA CCC CAA AAG ATG GCA GGT GCG AGC CCG ACC AAG 994
L G S P Q V W S E T K S G V A P C A K M A G A S P T K
 GAT GAC AGT AAG GAC TCA GAT TTC TGG AAG ATG CTT AAT GAG CCA GAG GAC CAG GCC CCA GGA GGG GAG GAG GTG 1069
D D S K D S D F W K M L N E P E D Q A P G G E E V
 CCG GCT GAG GAG CAG GAC CCA AGC CCT P GAG GCA GCA GAT TCA GCT TCT GGT GCT CCC AAT GAT TTT CAG AAC AAC 1144
P A E E Q D P S A A D S A S G A P N D E D E D
 GTG CAG GTC AAA GTC ATT CGA AGC CCT GCG GAT TTG ATT CGA TTC ATA GAG GAG CTG AAA GGT GGA ACA AAA AAG 1219
V Q V K V I R S P A D L I R F I E E L K G G T K K
 GGG AAG CCA AAT ATA GGC CAA GAG CAG CCT GTG GAT GAT GCT GCA GAA GTC CCT CAG AGG GAA CCA GAG AAG GAA 1294
G K P N I G Q E Q P V D A A E V P Q R E P E K E
 AGG GGT GAT CCA GAA CGG CAG AGA GAG ATG GAA GAA GAG GAG GAT GAG GAT GAG GAT GAG GAT 1369
R G D P F E E E E E D E D E D E D E D E D
GAA CGG CAG TTA CTG GGA GAA TTT GAG AAG GAA CTG GAA GGG ATC CTG CTT CCG TCA GAC CGA GAC CGG CTC CGT 1444
E R Q L L G E F E K E L E G I L L P S D R D R L R
 TCG GAG GTG AAG GCT GGC ATG GAG CGG GAA CTG GAG AAC ATC ATC CAG GAG ACA GAG AAA GAG CTG GAC CCA GAT 1519
S E V K A G M E N I I Q E T E K E L D P D
 GGG CTG AAG AAG GAG TCA GAG CGG GAT CGG GCA ATG CTG GCT CTC ACA TCC ACT CTC AAC AAA CTC ATC AAA AGA 1594
G L K K E S E R D R A M L A L T S T L N K L I K R
 CTG GAG GAA AAA CAG AGT CCA GAG CTG GTG AAG AAG CAC AAG AAA AAG AGG GTT GTC CCC AAA AAG CCT CCC CCA 1669
L E E R Q Q P E L V K K H K R K R
 TCA CCC CAA CCT ACA GAG GAG GAT CCT P GAG CAC AGA GTC CGG GTC CGG GTC ACC AAG CTC CGT CTC GGA GGC CCT 1744
S P Q P T E H E D P GAG CAC AGA GTC CGG GTC CGG GTC ACC AAG CTC CGT CTC GGA GGC CCT
 AAT CAG GAT CTG ACT GTC CTC GAG ATG AAA CGG GAA AAC CCA CAG CTG AAA CAA ATC GAG GGG CTG GTG AAG GAG 1819
N Q D L T V L E M K R E N P Q L K Q I E G L V K E
 CTG CTG GAG AGG GAG GGA CTC ACA GCT GCA GGG AAA ATT GAG ATC AAA ATT GTC CGC CCA TGG GCT GAA GGG ACT 1894
L L E R E G L T A A G K I E I K I V R P W A E G T
 GAA GAG GGT GCA CGT TGG CTG ACT GAT GAG GAC ACG AGA AAC CTC AAG GAG ATC TTC TTC AAT ATC TTG GTG CCG 1969
E E G A R W L T D E D T R N L K E I F F N I L V P
 GGA GCT GAA GAG GCC CAG AAG GAA CGC CAG CGG CAG AAA GAG CTG GAG AGC AAT TAC CGC CGG GTG TGG GGC TCT 2044
G A E E A Q K E R Q R Q K E L E S N Y R R V W G S
 CCA GGT GGG GAG GGC ACA GGG GAC CTG GAC GAA TTT GAC TTC TGA GACCAACACTACACTTGACCCCTTCACGGAATCCAGACTCT 2129
P G G E G T G D L D E F D F *
 TCCTGGAGTGCTTGCCTCTCCACCTCCACCTGGAACCCCTGAGGGCCAAACAGCAGAGTGAGCTGAGCTGTGGACCTCTCGGGCAACTCTGT 2229
 GGGTGTGGGGCCCTGGGTGAATGCTGCTGCCCTGTGGCAGCCACCTTGAGACCTCACCGGGCCTGTGATTTGCTCTCCGAACTCTCACTCAATC 2329
 CTCCTCTCTCTCTCTGCTTTCTCTGTTATGTCCTCCCTAATGATAGGATATTCCTGCTGCTACCTGGAGATTTCAGTAGGATCTTTTGGAGTGGAGGT 2429
 GGGTAGAGAGCAAGGAGGGCAGGACACTTACAGGCACTGAGCAAGCAGGCCCCACCTGCGCTTGTAGTGTGTTGGAGTCTTTTACCTCTCTCTAT 2529
 TGAATGCTCTGGGATTTCT 2629
 TGGTAGAATGTGCTTATCATCTCCAGCACAATCCAGCGAAGAGGTGTGAAGCACCACCATGTTCTTGAACAATCAGGTTCTCTAATAAACAACCTG 2729
 ACCATCA_n 2785

Figure 1. The 2785-bp nucleotide sequence of the OS-9 cDNA with the deduced longest ORF, 667 amino-acid residues (GenBank accession number U41635). The amino-acid sequence is indicated under the nucleotide sequence. A hydrophobic "signal" sequence at the N terminus of the ORF is underlined. The polypeptide segment (211 residues) in brackets is homologous to an ORF (F48E.4) from *C. elegans* [9] and an ORF (YD9609.11) from *S. cerevisiae* [10,11]. Eight cysteine (C) residues conserved in these ORFs are shown in bold. The 48 nt of trinucleotide repeat GADs

(D = G, A, or T) encoding an acidic domain is underlined. The possible nuclear targeting signal is in the shaded box. The asterisk indicates the stop codon. The polyadenylation signal AAUAAA is boxed. "n" represents 49 A residues sequenced. Partial nucleotide sequences within the OS-9 cDNA match perfectly with 15 ESTs derived from the Human Genome Project. The GenBank accession numbers of these ESTs are T52905, T86580, T48324, T57573, T52904, T07270, T54460, T57526, T86757, M85690, T69075, T72427, T48323, R31165, and T24754.

YD9609.11	(61) IVN-MGDDL- EC FIQNASTQLNDVLEDSNEHSNSEKTALLTKTLNQGK--TIFDKLNERCIFYQAG-FWIY EC PG
F48E8.4	(78) FVTSKGGQ KFAC -----S--LPDVEDVKDKPKSSKNP---KIYGDALA--ASF--YVDKCVKLRGNHWSYIL CRG
OS-9	(62) IVSSKYKQRY EC -----R--LP-AGAIHFQREEREETP---AYQGPGIPELLSP-MRDAPCL-LKTKDWWTY EC YG
Consensus	.V.skg.q..e C -----s--Lpdv.....sektp---k..g.g.--.sf-----C..l....wWiYe.C.G
YD9609.11	IEFVQFHGRVNTKTGEIVNRDESLVYRLGPKKANVE--EREFELLYDDVG-YISEIIGSGD IC DTV G --AERMVEIQYV
F48E8.4	QTIEQVHG-----E-PGQEGYVKNILGLFDGSLK--MPSYQESTDDRL-LFVEETYS GTF CDLE EE YREPRMTSVRYE
OS-9	RHIQQYHM-----E-DSEIKGEVLVLYGYSQAFDWDDETAKASKQHRLKRYHSQTYNGSK CD LNG--RPREA EV RFL
Consensus	..i.Q.Hg-----E-.....v..LG.....--.....s.ddrl.-y.setygsG..CDl.g--.pRm.evry.
YD9609.11	CGGSNSGPST-IQWVRETKICVYEAQVTIPELCNL-ELLAKNEDQKNAS PIL CRMP--AKSKI-GSNSIDLITKY (277)
F48E8.4	CDAQLSTNEVYISSVVEVKPCQYLMIVKVGTLCRYEFLPASQ SNT KK--IGCQ-PFLRKEDVRQLLERQLEEKQ (281)
OS-9	CDEGAGISGDYIDRVDEPLSCSYVLTI RT PRLC PH PLLRPPPSAAPQA--ILCH-PSLQPEEY MA VQRQADSKQ (267)
Consensus	Cd...s....yI..V.E.k.C.Y...v..p.LC..pellp.....--IlC.-P.l.ke.....rql..Kq

Figure 2. Amino-acid sequence alignment of OS-9 with an ORF (F48E8.4) from *C. elegans* [9] and an ORF (YD9609.11) from *S. cerevisiae* [10,11]. The dashes between the residues indicate gaps to optimize the alignments. The vertical lines indicate identical residues. The double and single dots between the sequences represent high and low similarities of the residues, respectively. In the consensus sequence, the capital letters indicate identical residues in all three proteins, lowercase letters

represent residues conserved in two of the three proteins, and the dots indicate residues different in all three proteins. The eight conserved cysteine residues are indicated in bold. The numbers in parentheses indicate the positions of the amino-acid residues in the corresponding proteins. The protein alignment was created with the Lipman-Pearson program in the DNASTAR software, Version 1.58.

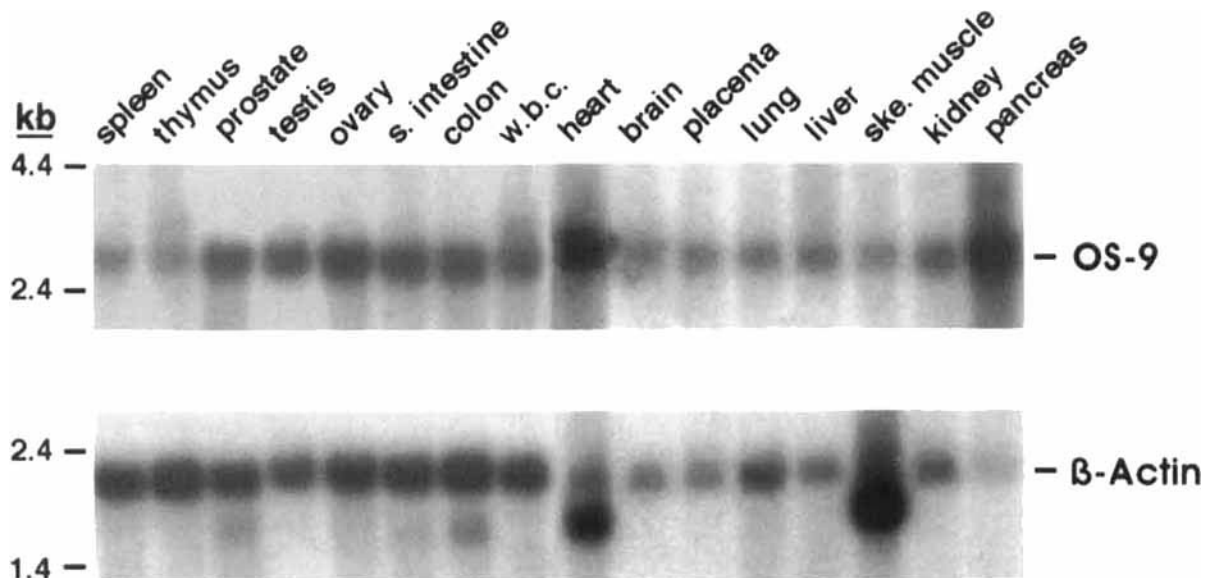


Figure 3. Northern analysis of the OS-9 gene. The OS-9 probe detected 2.8-kb transcripts in all 16 human tissues. The numbers indicate molecular markers (kb). The levels of OS-9 mRNA were twofold to threefold higher in heart and pancreas and twofold lower in spleen and thymus than in the rest of the

tissues. The human β -actin probe was used as a control for loading error. The weak and strong bands below β -actin in the lanes of prostate, colon, heart, and skeletal muscle are isoforms of β -actin mRNA. s, small; w.b.c., white blood cells; ske, skeletal.

that similar levels of *OS-9* mRNA were detected in most tissues except that the levels were twofold to threefold higher in heart and pancreas and twofold lower in spleen and thymus.

Southern Blot Analysis

To evaluate the *OS-9* amplification relative to the closely linked gene *CDK4*, which falls within 40 kb of *OS-9* [23], Southern blot analysis was performed for DNA samples from six different human sarcoma tissues including five previously noted to carry amplification of the *CDK4* gene [18]. The blot was hybridized sequentially with probes for *OS-9*, *CDK4*, and finally *INT-1* (a gene that maps proximal to the amplified sequence) as a control for loading error. The results (Figure 4) demonstrate that *OS-9* was amplified in three of five sarcoma tissues with *CDK4* amplification. The relative intensity of each band was measured by densitometry. After normalization relative to *INT-1*, in comparison to placental DNA, *CDK4* was amplified more than fivefold in MFH7, LIP2, and MFH20 and more than threefold in MFH6 and MFH21. *OS-9* was amplified to the same degree as *CDK4* in MFH20, LIP2, and MFH21 but was not amplified in MFH6 or MFH7. Neither probe detected amplification in MFH9.

DISCUSSION

We previously identified the *OS-9* gene in the human osteosarcoma cell line OsA-Cl and demonstrated that it was amplified and overexpressed relative to human control tissues [8]. We now characterized the *OS-9* cDNA to begin assessing the impact of *OS-9* overexpression on tumor phenotype.

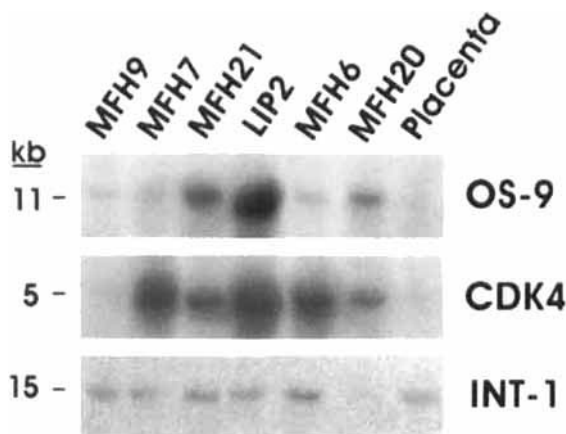


Figure 4. Southern detection of *OS-9* and *CDK4* amplification in human malignant fibrous histiocytoma (MFH) and liposarcoma (LIP). *OS-9* and *CDK4* were coamplified in three of five tissues (MFH21, LIP2, and MFH20) with *CDK4* amplification. Human placenta and a sarcoma without 12q13-15 amplification (MFH 9) are shown for comparison. The clinical characteristics of these tumor tissues have been described previously [18]. Ten micrograms of the indicated DNA was digested with *EcoRI*, size-fractionated on an agarose gel, and transferred to a nylon membrane for hybridization. The *INT-1* probe was used as a control for loading error.

The conclusion that the 2785-bp nucleotide sequence represents the full-length *OS-9* cDNA was supported by Northern analysis, which detected a message of approximately 2.8 kb in all 16 human tissues examined and in OsA-Cl [8]. Additionally, multiple plasmids containing the inserts from 5' cDNA synthesis had a common 5' end. The *OS-9* ORF predicted a polypeptide with 667 amino-acid residues and a calculated pI of 4.60. This polypeptide contained a short hydrophobic sequence followed by a long cysteine-containing hydrophilic sequence and then multiple predicted hydrophilic helices. Taken together, these data suggest that the *OS-9* polypeptide is a water-soluble acidic protein. The hydrophobic sequence (31 amino-acid residues) at the N terminus may act as a signal peptide for membrane insertion. However, the significance of this potential signal sequence as well as the possible nuclear targeting signal remains indeterminate.

The *OS-9* gene appears to encode a novel protein belonging to a previously uncharacterized evolutionarily conserved gene family in *C. elegans* and *S. cerevisiae*. Within this gene family, the similarity between human and *C. elegans* genes (68.7%) was greater than that between human and *S. cerevisiae* genes (53%), which apparently reflects the greater evolutionary distance between yeast and humans. Nevertheless, the exact alignment of the eight cysteine residues within the conserved regions in all three ORFs strongly suggests that these regions could form similar secondary and tertiary structures that may be essential to the function of the mature proteins. In addition, *OS-9* transcripts were detected in a variety of human normal tissues, suggesting that *OS-9* probably has an essential cellular function. The exact alignment of the *OS-9* cDNA sequence with the 15 ESTs derived from human fetal and adult tissues further verifies its ubiquitous expression.

We previously demonstrated *OS-9* co-amplification with *CDK4* in OsA-Cl [8]. We now demonstrated the amplification of *OS-9* in three of five human sarcoma tissues with *CDK4* amplification. Because the *CDK4* protein plays a key role in regulating the transition from the G₁ to S phases of the cell cycle by phosphorylating the retinoblastoma protein, *CDK4* is a likely primary target driving selection of amplification-bearing clones. The closely linked gene *OS-9* [23] may be amplified simply by a bystander effect. Alternatively, the co-amplification of *OS-9* and *CDK4* in four of six sarcomas suggests that *OS-9* has an oncogenic role in a subset of sarcomas. In either case, when amplification of a gene-dense region such as 12q13-15 occurs, it is important to consider the potential phenotypic effect of each gene, as many are frequently contained in the amplicon. The information presented here should facilitate the functional study of *OS-9* that will be required to evaluate the potential role of this gene in tumorigenesis. It is also noteworthy that the presence of homologous genes

in *C. elegans* and *S. cerevisiae* opens the possibility for functional studies of this novel gene family in these model organisms.

ACKNOWLEDGMENTS

The authors thank Dr. John Wooton for assistance in the computer analysis of the sequence. We also acknowledge Darryl Leja's help with graphic illustration.

Received October 9, 1995; revised December 5, 1995; accepted December 8, 1995.

REFERENCES

1. Roberts WM, Douglass EC, Peiper SC, Houghton PJ, Look AT. Amplification of the *gli* gene in childhood sarcomas. *Cancer Res* 49:5407-5413, 1989.
2. Meltzer PS, Jankowski SA, Dal cin P, Sandberg AA, Paz IB, Coccia MA. Identification and cloning of a novel amplified DNA sequence in human malignant fibrous histiocytoma derived from a region of chromosome 12 frequently rearranged in soft tissue tumors. *Cell Growth Differ* 2:495-501, 1991.
3. Jankowski SA, Mitchell DS, Smith SH, Trent JM, Meltzer PS. SAS, a gene amplified in human sarcomas, encodes a new member of the transmembrane 4 superfamily of proteins. *Oncogene* 9:1205-1211, 1994.
4. Khatib ZA, Matsushime H, Valentine M, Shapiro DN, Sherr CJ, Look AT. Coamplification of the CDK4 gene with MDM2 and *GLI* in human sarcomas. *Cancer Res* 53:5535-5541, 1993.
5. Aman P, Ron D, Mandahl N, et al. Rearrangement of the transcription factor gene *CHOP* in myxoid liposarcomas with t(12;16)(q13;p11). *Genes Chromosom Cancer* 5:278-285, 1992.
6. Oliner JD, Kinzler KW, Meltzer PS, George DL, Vogelstein B. Amplification of a gene encoding a p53-associated protein in human sarcomas. *Nature* 358:80-83, 1992.
7. Leach FS, Tokino T, Meltzer PS, et al. p53 Mutation and MDM2 amplification in human soft tissue sarcomas. *Cancer Res* 53:2231-2234, 1993.
8. Su YA, Trent JM, Guan XY, Meltzer PS. Direct isolation of genes encoded within a homogeneously staining region by chromosome microdissection. *Proc Natl Acad Sci USA* 91:9121-9125, 1994.
9. Wilson R, Ainscough R, Anderson K, et al. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature* 368:32-38, 1994.
10. Barrell B, Rajandream MA. GenBank accession number: Z49209, 1995.
11. Brandt P. GenBank accession number: X84162, 1995.
12. Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning: A Laboratory Manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, pp. 2.108-2.111.
13. Frohman MA, Dush MK, Martin GR. Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* 85:8998-9002, 1988.
14. Belyavski A, Vinogradova T, Rajewsky K. PCR-based cDNA library construction: General cDNA libraries at the level of a few cells. *Nucleic Acids Res* 17:2919-2932, 1989.
15. Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning: A Laboratory Manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, pp. 7.6-7.9.
16. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85:2444-2448, 1988.
17. Altschul S, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 215:403-410, 1990.
18. Smith SH, Weiss SW, Jankowski SA, Coccia MA, Meltzer PS. SAS amplification in soft tissue sarcomas. *Cancer Res* 52:3746-3749, 1992.
19. Feinberg AP, Vogelstein B. A technique for radio-labeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 132:6-13, 1983.
20. Brown T. Northern hybridization of RNA fractionated by agarose-formaldehyde gel electrophoresis. In: Ausubel FM, Brent R, Kingston RE, et al. (eds), *Current Protocols in Molecular Biology*, Green Publishing and John Wiley & Sons, New York, 1993, pp. 4.9.1-4.9.8.
21. Kalderon DB, Roberts L, Richardson WD, Smith AE. A short amino acid sequence able to specify nuclear location. *Cell* 39:499-509, 1984.
22. Robbins J, Dilworth SM, Laskey RA, Dingwall C. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: Identification of a class of bipartite nuclear targeting sequence. *Cell* 64:615-623, 1991.
23. Elkahoul AG, Meltzer PS, Guan XY, McNinch JS, Trent JM, de Jong PJ. Isolation of a cosmid sublibrary for a region of chromosome 12 frequently amplified in human cancers using a complex chromosome microdissection probe. *Genomics*, in press.