THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 270, No. 24, Issue of June 16, pp. 14733–14741, 1995
Printed in U.S.A.

Characterization of the Mammalian YAP (Yes-associated Protein) Gene and Its Role in Defining a Novel Protein Module, the WW Domain*

(Received for publication, January 23, 1995, and in revised form, March 17, 1995)

Marius Sudol‡§, Peer Bork¶, Aaron Einbond‡, Kumar Kastury‖, Teresa Druck‖, Massimo Negrini‖, Kay Huebner‖, and David Lehman‡

From the ‡Laboratory of Molecular Oncology, The Rockefeller University, New York, New York 10021, the ¶European Molecular Biology Laboratory, D 69117 Heidelberg, Federal Republic of Germany, and the ‖Department of Microbiology and Immunology, Jefferson Cancer Institute, Jefferson Medical College, Philadelphia, Pennsylvania 19107

We report cDNA cloning and characterization of the human and mouse orthologs of the chicken YAP (Yes-associated protein) gene which encodes a novel protein that binds to the SH3 (Src homology 3) domain of the Yes proto-oncogene product. Sequence comparison between mouse, human, and chicken YAP proteins showed an inserted sequence in the mouse YAP that represented an imperfect repeat of an upstream sequence. Further analysis of this sequence revealed a putative protein module that is found in various structural, regulatory, and signaling molecules in yeast, nematode, and mammals including human dystrophin. Because one of the prominent features of this sequence motif is two tryptophans (W), we named it the WW domain (Bork, P., and Sudol, M. (1994) *Trends Biochem. Sci.* 19, 531–533). Since its delineation, more proteins have been shown to contain this domain, and we report here on the widespread distribution of the WW module and present a discussion of its possible function. We have also shown that the human YAP gene is well conserved among higher eukaryotes, but it may not be conserved in yeast. Its expression at the RNA level in adult human tissues is nearly ubiquitous, being relatively high in placenta, prostate, ovary, and testis, but is not detectable in peripheral blood leukocytes. Using fluorescence *in situ* hybridization on human metaphase chromosomes and by analyzing rodent-human hybrids by Southern blot hybridization and polymerase chain reaction amplification, we mapped the human YAP gene to chromosome band 11q13, a region to which the multiple endocrine neoplasia type 1 gene has been mapped.

One of the hallmarks of signal transduction processes is a specific physical interaction between proteins carried out by well demarcated and structured regions of the proteins, which are called domains (1–4). Our research has focused on molecular steps by which non-receptor-type protein-tyrosine kinases of the Src family signal in normal and transformed cells (4). In recent years, much of the attention has been concentrated on amino-terminal domains of protein-tyrosine kinases. At least three distinct structural domains, termed SH2, SH3 (SH for Src homology) and PH (for pleckstrin homology) are present in the non-receptor-type protein-tyrosine kinases and are also found in a wide variety of proteins implicated in signal transduction processes (5–8). The SH2 domains are known to interact specifically with phosphotyrosine-containing proteins, and the resulting complexes are involved in signal transduction events initiated by protein-tyrosine kinases (1, 9). The SH2 domain of Src protein-tyrosine kinases is not only involved in substrate recognition but is also necessary for the regulation of kinase activity through maintenance of a repressed conformation of the protein-tyrosine kinases (10, 11). The SH3 domains mediate noncovalent protein-protein interactions essential for cellular and intercellular signaling (12–21). For Src and other members of the family, it is presumed that binding of specific proteins to their SH3 domains may result in the modulation of their enzymatic activity and thus could be a part of the signaling mechanism of cellular and oncogenic forms of the Src family protein-tyrosine kinases (22–29). The PH domain was first defined as two repeats in pleckstrin, the major substrate for serine/threonine phosphorylation by protein kinase C in platelets (7, 8). In contrast to SH3 and SH2, the PH domain seems to bind a nonproteinaceous ligand, phosphatidylinositol 4,5-bisphosphate, implicating this domain in membrane-protein interaction (30).

Recently, we have identified, cloned, and characterized the cDNA for a novel chicken protein that binds to the SH3 (Src homology 3) domain of the Yes proto-oncogene product (31). The protein has a molecular mass of 65 kilodaltons (kDa) and is phosphorylated *in vivo* on serine. We named it YAP65 for Yes-associated protein of 65 kDa. Within the YAP65 (YAP for short) sequence, we identified a proline-rich motif that is involved in binding YAP to Yes kinase. YAP was also shown to bind to other signaling molecules that contain SH3 domains including Nck, Crk, and Src. In order to analyze the function of YAP in transgenic animals, we have cloned mammalian orthologs of chicken YAP. Two interesting findings resulted from our studies. First, the sequence comparison between mouse, human, and chicken YAPs showed an inserted sequence in mouse YAP representing an imperfect repeat of an upstream sequence. Unexpectedly, this sequence showed significant similarity with various regulatory and signaling molecules, and we have proposed that it may form a novel domain involved in protein-protein interaction (32). In this report, we show the

* This work was supported by National Institutes of Health Grant CA51083 (to K. H.), by grants from the Annenberg Foundation and Hebrew Technical Institute (to A. E.), by National Cancer Institute Grants CA45757 and CA01605, by Council for Tobacco Research-U. S. A. Inc. Grant 3035, by Human Frontier Science Program grant, and by the Klingenstein Award in the Neurosciences (to M. S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) X80507 (human YAP65) and X80508 (mouse YAP65).

§ To whom correspondence and reprint requests should be addressed: Laboratory of Molecular Oncology, Box 169, The Rockefeller University, 1230 York Ave., New York, NY 10021-6399. Tel.: 212-327-8811; Fax: 212-327-7582; E-mail: sudol@rockvax.rockefeller.edu.

widespread occurrence of this domain and point out new signaling molecules that contain the WW module. Second, the human YAP (hYAP) gene was localized to a short interval on chromosome 11q13 that harbors a gene for multiple endocrine neoplasia type 1. We have also shown that the hYAP gene is well conserved among higher eukaryotes and is expressed in most tissues.

EXPERIMENTAL PROCEDURES

cDNA Cloning and Sequencing—A chicken YAP cDNA corresponding to the coding region (31) was used as a probe to screen a λpCEV15 cDNA library derived from M426 human lung embryonic fibroblast cells (a gift from Dr. Stuart Aaronson, Ref. 33) and a 16-day mouse embryo cDNA library in λ EXlox™ (purchased from Novagen, Madison, WI). The low stringency conditions of hybridization were as follows: 5 × SSPE, 10 × Denhardt's, 2% SDS, 0.2 mg/ml salmon sperm DNA, and 10^6 cpm/ml ^{32}P-labeled cDNA at 65 °C overnight. The filters were washed twice for 20 min at room temperature with 2 × SSC, 0.05% SDS, and twice at 60 °C for 20 min with 0.1 × SSC, 0.1% SDS. Both libraries contained phages with a plasmid portion that carried the insert. The plasmids with inserts were easily rescued from the λ genome following published protocols (33, 34). The apparently complete sequence of the hYAP cDNA was contained in one recombinant plasmid pCEV15-hYAP6 with a *SalI-SalI* insert of 5 kb[1] pairs. The complete sequence of mouse YAP (mYAP) cDNA was contained in two overlapping clones, pEXlox-mYAP6 (2.3-kb *Eco*RI-*Hin*dIII insert) and pEXlox-mYAP20 (*Eco*RI-*Hin*dIII insert). Both strands of the cDNA clones were analyzed by direct sequence analysis using the Sanger method (35).

Southern and Northern Blot Analysis—Southern blot of genomic DNA from nine eukaryotic species was performed using the same conditions as for cDNA library screening (36). DNA sources were as follows: human, rhesus monkey, Sprague-Dawley rat, BALB/c mouse, dog, cow, rabbit, chicken, and *Saccharomyces cerevisiae*. Except for yeast and human DNAs, all other genomic DNAs were isolated from kidney tissue. Human DNA was isolated from placental tissue. DNA was digested with *Eco*RI, run on a 0.7% agarose gel, transferred to a charge-modified nylon membrane by blotting, and fixed by UV irradiation. The cDNA insert of the hYAP5 plasmid or human β-actin cDNA control probe were radioactively labeled to a specific activity of approximately 2 × 10^8 cpm/μg and were used as a probe for Southern (hYAP probe) and for Northern analysis (hYAP probe was used first, and, after stripping, the probe for β-actin was used). Poly(A)$^+$ RNAs were isolated from 16 different human tissues from healthy donors of both sexes (Clontech Laboratory Inc.). The RNAs (2 μg/lane) were run on a denaturing formaldehyde-1.2% agarose gel, transferred to a charge-modified nylon membrane by blotting, and fixed by UV irradiation. The hybridization conditions were: 5 × SSPE, 10 × Denhardt's solution, 100 μg/ml freshly denatured, sheared salmon sperm DNA, 50% formamide, and 2% SDS at 42 °C overnight (36). The blots were washed for 30 min at room temperature in 2 × SSC, 0.05% SDS and for 1 h at 50 °C in 0.1 × SSC, 0.1% SDS. Removal of the hYAP probe from the blot for subsequent hybridization with the human β-actin probe was achieved by incubating the blot for 10 min in sterile H_2O containing 0.5% SDS that was heated to 90 °C.

Chromosomal Localization—For Southern blot hybridization, the hYAP cDNA insert (hYAP6 clone) was isolated and radiolabeled by random priming to a specific activity of 10^8 cpm/0.1 μg, and 10^8 cpm was used for each filter hybridization; for FISH, the entire hYAP cDNA was labeled with biotin by nick translation (36, 38).

Hybrid DNAs were from previously described rodent-human hybrid cell lines (37–39) or from the NIGMS Human Genetic Mutant Cell Repository (Coriell Institute, Camden, NJ). Hybrids retaining partial chromosomes 11 and 6 have also been described (38, 39). Hybrid DNAs were tested for the presence of YAP specific human *Sst*I and *Pst*I restriction fragments detected by radiolabeled hYAP probe using standard Southern hybridization methods. In addition, oligonucleotide primers were prepared for amplification of a 208-bp fragment of the hYAP 3′-untranslated region (UTR) representing nucleotides 2135 through 2341. The forward primer, an 18-mer starting at position 2135 of the cDNA sequence, was 5′GGAAATGGCCACTGCAGA3′, and the reverse primer, a 20-mer starting at position 2323, was 5′CCCTAAGCTAAAGCTAATCT3′. These primers were used to amplify the hYAP 3′-UTR fragment from mouse, hamster, human, and hybrid DNAs under the following cycling conditions: 94 °C, 5 min for denaturation; 30 cycles of 94 °C for 30 s and 72 °C for 30 s; and a final cycle at 72 °C for 5 min.

Chromosomal Fluorescence in Situ Hybridization (FISH)—The procedure used in this study has been described in detail (38). Probes were prepared by nick translation using biotin-labeled 11-dUTP (BioNick Kit, Life Technologies, Inc.). Hybridization of biotin-labeled probes was detected with fluorescein isothiocyanate-conjugated avidin. Metaphase chromosomes were identified by Hoechst-33528 staining and UV irradiation (365 nm), followed by 4′,6-diamidino-2-phenylindole staining to produce the banding pattern. The fluorescent signal was observed with filter block I3 (BP450–490/LP515; Leitz Orthoplan) on the background of red chromosomes stained with propidium iodide. Q-banding was observed with filter block A (BP340–380/LP430).

Computer-aided Analysis of Protein Sequences—Analyses of sequence homology and secondary structures of the polypeptides were performed as described previously (40) using the following computer programs: BLASTP (41) for initial data base searches, PROFILES (42) and PATTERNS (43) for selective identification of the current set of the WW domains, PHD (44) for predicting secondary structures, and MoST (45) for calculating a probability of matching the alignment by chance.

RESULTS

Cloning of Human and Mouse YAPs—Using a cDNA fragment encoding the chicken YAP as a probe, we screened λ phage plaques of a human lung embryonic fibroblast cDNA library. Of 13 positive clones, two (hYAP5 and hYAP6) with the longest inserts (approximately 3 and 5 kb long, respectively) were analyzed further. Initial analysis of the DNA sequence showed that hYAP5 cDNA is included within the hYAP6 clone. The result of direct sequence analysis of both strands of the hYAP6 cDNA is shown in Fig. 1. The longest open reading frame predicted a protein product of 493 amino acids with significant sequence similarity to the chicken YAP (Fig. 2).

In parallel experiments, we isolated a mouse ortholog of YAP using the same chicken YAP cDNA as a probe. We screened a mouse embryo (16 day) cDNA library in the λ EXlox vector. Of 7 positive clones, 2 (mYAP6 and mYAP20) were shown to contain long inserts (approximately 2.3 and 3.6 kb long, respectively); the clones overlapped giving rise to a 4-kb-long cDNA sequence terminating with a poly(A) stretch. As for the hYAP, the longest open reading frame predicted a protein product with significant sequence similarity to the chicken YAP. However, an additional sequence of 38 amino acids was present in the middle of the sequence (Fig. 2). Visual inspection of the insert sequence suggested that it is an imperfect duplication of a sequence found upstream (see *underlined sequences* in Figs. 1 and 2). We have subjected this sequence to more detailed analysis and found that the motif shares significant sequence and structural similarities with sequences found in various regulatory and signaling proteins (32). Alignment of the chicken YAP, mYAP, and hYAP also revealed long stretches of amino acid sequences that were perfectly conserved (Fig. 2). Interestingly, the proline-rich sequence (Fig. 2, indicated with a *number sign*), implicated in binding chicken YAP to the SH3 domain of Yes, is 100% conserved among the three sequences.

Evolutionary Conservation of YAP—A high degree of sequence similarity between hYAP, mYAP, and chicken YAP was confirmed by Southern blot analysis of the genomic DNAs digested with *Eco*RI enzyme (Fig. 3). Genomic DNA from other higher eukaryotes also showed hybridization with the hYAP radioactive probe. However, no specific signal was detected in yeast *S. cerevisiae*.

Expression of YAP Transcript—A major transcript of approximately 5 kb was detected by Northern blot in various human tissues. An additional band migrating below 2.4 kb was de-

[1] The abbreviations used are: kb, kilobase(s); bp, base pair(s); BCL1, B cell leukemia/lymphoma 1 human locus; CCND1, cyclin D1 human locus; FISH, fluorescence *in situ* hybridization; YAP, Yes-associated protein of 65 kilodaltons from chicken, also named YAP65; hYAP, human Yes-associated protein; mYAP, mouse Yes-associated protein; MAP, microtubule-associated protein; MEN1, multiple endocrine neoplasia type 1; PH, pleckstrin homology; SH, Src homology; UTR, untranslated region; W, tryptophan; www, world wide web.

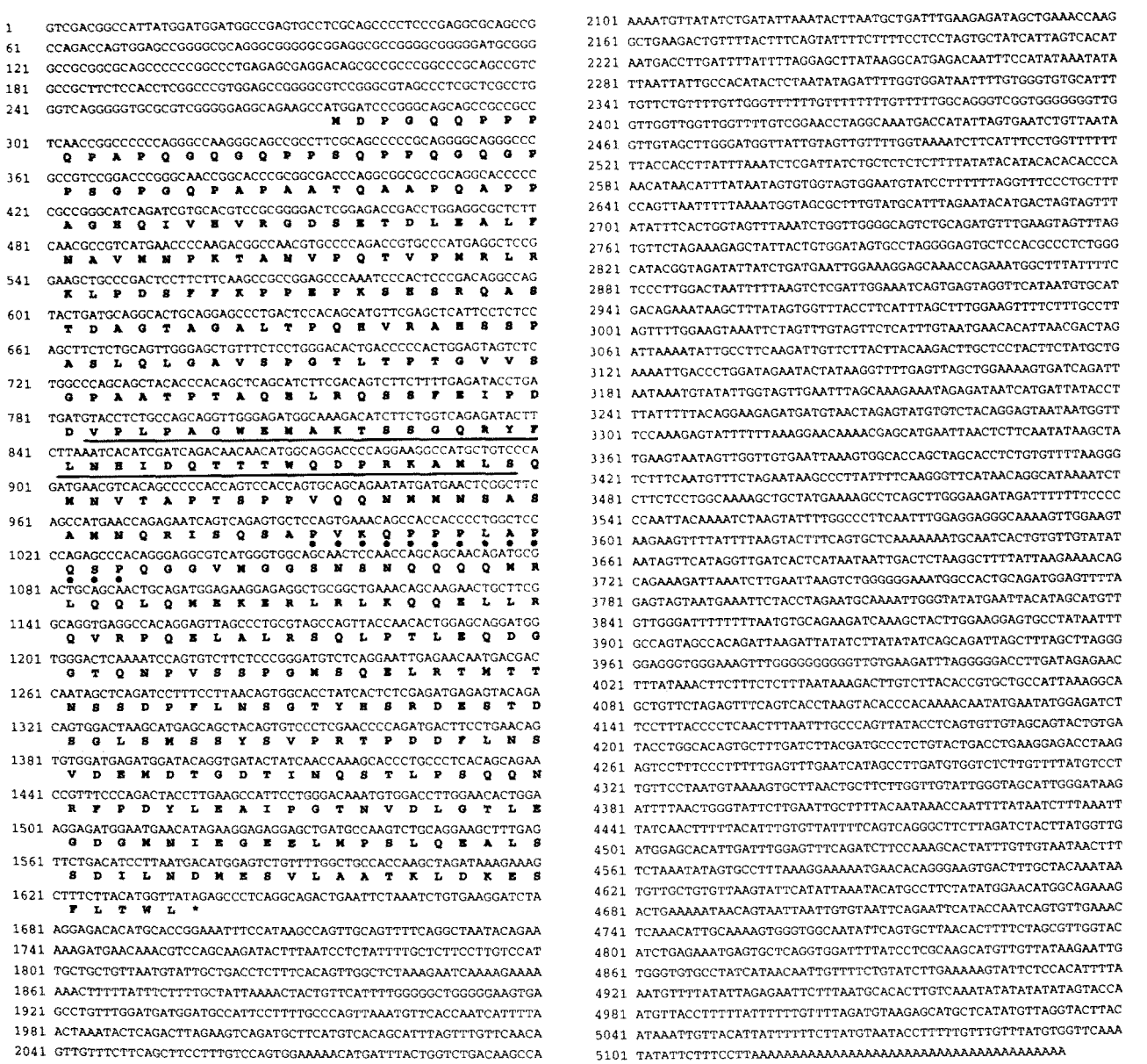

FIG. 1. **Nucleotide and deduced amino acid sequences of human YAP.** The 5154-base pair human YAP cDNA encodes 493 amino acids and is terminated at nucleotide 1638 marked by an *asterisk*. A putative protein module, termed the WW domain, is *underlined*. A proline-rich sequence implicated in binding between YAP and various SH3 domains is indicated with *black dots*.

tected in some of the tissues (see Fig. 4, *lanes K, M,* and *O* for example). The expression of hYAP mRNA is relatively high in placenta, prostate, testis, ovary, and small intestine (Fig. 4, *lanes C, K, L, M,* and *N*). Relatively lower levels of the message were found in the brain, liver, and spleen (Fig. 4, *lanes B, E,* and *I*). We could not detect hYAP mRNA in the preparation of human peripheral blood leukocytes even on overexposure of the blot (Fig. 4, *lane P*).

Chromosomal Localization—The hYAP cDNA detected two loci, one on chromosome 11 (11q13) and another on chromosome 6 (6q23-qter). When human DNA was digested with *Sst*I restriction enzyme and probed with radioactive hYAP cDNA, two strongly hybridizing bands, one of 16 kb pairs and another migrating above 23 kb pairs, were detected (not shown). In addition, we also observed less strongly hybridizing bands. In the same analysis, rodent DNA digested with *Sst*I and probed with hYAP cDNA showed fainter bands distinguishable from the hYAP specific fragments (not shown). When DNAs from a panel of rodent-human hybrids, each carrying a few human chromosomes, were tested for the presence of the hYAP locus, it was observed that the two strongly hybridizing bands segregated independently and thus were on different chromosomes (not shown). The results of the analysis of the rodent-human hybrid panel are summarized in Fig. 5*A*. These data illustrate that one hYAP specific locus maps to chromosome 11q and the other to chromosome 6. The less strongly hybridizing bands did not seem to segregate with either of the two major bands. The locus on chromosome 11q was the most intensely hybridizing band and was thus presumed to represent the cognate hYAP gene.

In order to demonstrate that the 11q locus was indeed the locus of the cognate gene, oligonucleotide primers flanking an approximately 200-bp region of the 3′-UTR (cDNA positions 2135–2341, inclusive of primers) were synthesized and used in polymerase chain reaction amplification with mouse, hamster, human, and hybrid DNAs as templates. Amplification products were detected after electrophoresis on 1.5% agarose gels containing ethidium bromide. No product was amplified from ro-

```
          1                                                  50
HYAP  MDPGQQPPPQ PAPQGQGQPP SQPPQGQGPP SGPGQPAPAA TQAAPQAPPA
MYAP  MEPAQQPPPQ PAPQGPA.PP SVSPAGTP.. .......... ..AAPPAPPA
YAP   MDPGQPQPQQ P.PQAAQPPA PQQAAPQPPG AGSGAPGGAA QPPGAGPPPA

          51                                                 100
HYAP  GHQIVHVRGD SETDLEALFN AVMNPKTANV PQTVPMRLRK LPDSFFKPPE
MYAP  GHQVVHVRGD SETDLEALFN AVMNPKTANV PQTVPMRLRK LPDSFFKPPE
YAP   GHQIVHVRGD SETDLEALFN AVMNPKGANV PHTLPMRLRK LPDSFFKPPE

          101                                                150
HYAP  PKSHSRQAST DAGTAGALTP QHVRAHSSPA SLQLGAVSPG TLTPTGVVSG
MYAP  PKSHSRQAST DAGTAGALTP QHVRAHSSPA SLQLGAVSPG TLTASGVVSG
YAP   PKAHSRQAST DAGTAGALTP QHVRAHSSPA SLQLGAVSPG TLTPSGVVTG

          151                                                200
HYAP  PAATPTAQHL RQSSFEIPDD VPLPAGWEMA KTSSGQRYFL NHIDQTTTWQ
MYAP  PAAAPAAQHL RQSSFEIPDD VPLPAGWEMA KTSSGQRYFL NHNDQTTTWQ
YAP   P.GAPSSQHL RQSSFEIPDD VPLPPGWEMA KTPSGQRYFL NHIDQTTTWQ

          201                                                250
HYAP  DPRKAMLSQM NVTAPTSPPV QQNMMNSAS. .......... ..........
MYAP  DPRKAMLSQL NVPAPASPAV PQTLMNSASG PLPDGWEQAM TQDGEVYYIN
YAP   DPRKAMLSQM NVTAPTSPPV QQNLMNSAS. .......... ..........

          251                                                300
HYAP  .......... .......AMN QRISQSAPVK QPPPLAPQSP QGGVMGGSNS
MYAP  HKNKTTSWLD PRLDPRFAMN QRITQSAPVK QPPPLAPQSP QGGVLGGGSS
YAP   .......... .......AMN QRISQSAPVK QPPPLAPQSP QGGVMGGSSS
                                ###############

          301                                                350
HYAP  NQQQQMRLQQ LQMEKERLRL KQQELLRQVR PQELALRSQL PTLEQDGGTQ
MYAP  NQQQQIQLQQ LQMEKERLRL KQQELFR... .QELALRSQL PTLEQDGGTP
YAP   NQQQQMRLQQ LQMEKERLRL KHQELLR... .QELALRSQL PTMEQDGGSQ

          351                                                400
HYAP  NPVSSPGMSQ ELRTMTTNSS DPFLNSGTYH SRDESTDSGL SMSSYSVPRT
MYAP  NAVSSPGMSQ ELRTMTTNSS DPFLNSGTYH SRDESTDSGL SMSSYSIPRT
YAP   NPVSSPGMSQ ELRTMTTNSS DPFLNSGTYH SRDESTDSGL SMSSYSVPRT

          401                                                450
HYAP  PDDFLNSVDE MDTGDTINQS TLPSQQNRFP DYLEAIPGTN VDLGTLEGDG
MYAP  PDDFLNSVDE MDTGDTISQS TLPSQQSRFP DYLEALPGTN VDLGTLEGDA
YAP   PDDFLNSVDE MDTGDSISQS NIPSHQNRFP DYLEAIPGTN VDLGTLEGDG

          451                                           493
HYAP  MNIEGEELMP SLQEALSSDI LNDMESVLAA TKLDKESFLTWL*
MYAP  MNIEGEELMP SLQEALSSEI L.DVESVLAA TKLDKESFLTWL*
YAP   MNIEGEELMP SLQEALSSDI LNDMESVLAA TKPDKESFLTWL*
```

FIG. 2. **Alignment of the human (*HYAP*), mouse (*MYAP*) and chicken (*YAP*) YAP amino acid sequences.** Positions that differ in at least one amino acid are indicated in *bold*. Spaces in the alignment were introduced arbitrarily and are indicated with *dots*. The sequences corresponding to the putative WW domain are *underlined*. Note that in mYAP a second WW domain is present. Proline-rich sequences implicated in binding between YAP and various SH3 domains are conserved and indicated with a *number sign*.

dent DNAs or from DNA hybrids containing human chromosome 6 without chromosome 11, hybrid Nu9 for example. The expected 200-bp fragment was observed after amplification from DNA templates derived from human or hybrids retaining chromosome 11 but not 6, hybrids 734 and 7298 for example (data not shown). These data suggest that the chromosome 6 locus represents a YAP-related locus rather than a pseudogene.

To further define the chromosome positions of these loci, small panels of DNAs from hybrids carrying partial chromosomes 11 or 6 were also tested for the presence of the YAP loci with the results summarized in Fig. 5*B*. Because the hYAP cognate locus was present in hybrid 7298 but absent in hybrid CE4, the gene maps between the centromere and the CCND1/BCL1 locus on chromosome 11, whereas the hYAP-related locus maps to 6p21 to 6qter.

To confirm and refine the above localizations, fluorescence *in situ* hybridization (FISH) with the hYAP cDNA probe was performed on normal human metaphase chromosomes. Using FISH, we detected 51 signals at chromosome 11q13 on 27 metaphases and only 12 signals on the q-terminal one-third of chromosome 6. The FISH results are summarized to the left of the chromosome idiograms shown in Fig. 5*B*.

Since the hYAP gene was mapped to 11q13, centromeric to the BCL1 major breakpoint region, possibly within the chromosomal region which is amplified in a significant fraction of human mammary carcinomas, a panel of 17 mammary carcinoma cell line DNAs was tested for evidence of amplification of the hYAP gene. Four of these DNAs had shown amplification of the CCND1/BCL1 gene (from 3- to 10-fold), but none showed evidence of an amplified hYAP gene (data not shown). Thus, the hYAP gene is most likely centromeric to the chromosome region commonly amplified at 11q13 in mammary carcinomas.

Protein Domain—The presence of an extra sequence in the murine YAP as compared to the human and chicken orthologs focused our attention on this motif and led us to propose it as a new protein module, the WW domain (32). The domain appears to contain β-strands grouped around four conserved aromatic positions (Fig. 6). Two of these positions are most frequently occupied by tryptophans, hence the name, the WW domain. Other important features of the domain are a high content of

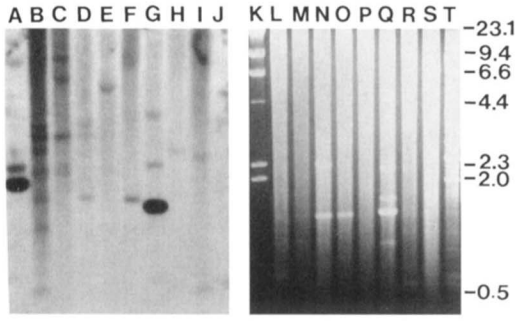

FIG. 3. **Southern blot analysis of genomic DNA from nine eukaryotic species.** The genomic DNA (4 μg) was digested with *Eco*RI, resolved on 0.7% agarose gel, transferred to a charge-modified nylon membrane by blotting, and fixed by UV irradiation. The DNA corresponding to the entire coding region of the hYAP cDNA was used as a probe. *Left panel (A–J)* represents results of hybridization with the hYAP cDNA probe, and the *right panel (K–T)* shows results of staining the agarose gel with ethidium bromide to check for even DNA loading and clear satellite bands. *Lanes A* and *K* contain λ *Hin*dIII DNA markers with sizes indicated in kilobases on the *right side* of the *right panel*. *Lanes B* and *L* contain human; *C* and *M*, monkey; *D* and *N*, rat; *E* and *O*, mouse; *F* and *P*, dog; *G* and *Q*, cow; *H* and *R*, rabbit; *I* and *S*, chicken; and *J* and *T*, yeast DNA. The exposure time was 4 days.

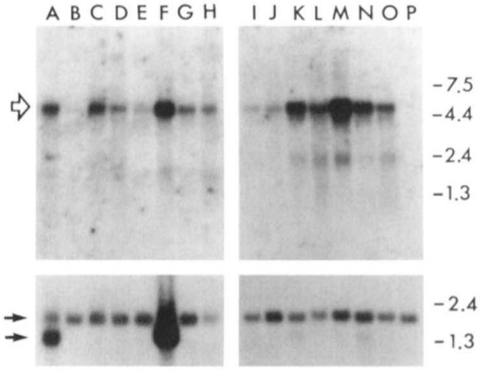

FIG. 4. **Northern blot analysis of poly(A)$^+$ RNA from 16 different human tissues.** Poly(A)$^+$ RNAs (2 μg each) from adult human tissues were run on a denaturing formaldehyde-1.2% agarose gel, transferred to a charge-modified nylon membrane by blotting, and fixed by UV irradiation. The radiolabeled cDNA corresponding to the entire coding region of the hYAP (insert of the hYAP6 clone) was used as a probe (*upper panel*). For normalization and to ensure the intactness of the RNA, the blot was hybridized with a radiolabeled cDNA encoding human β-actin (*lower panel*). *Lane A*, heart; *B*, brain; *C*, placenta; *D*, lung; *E*, liver; *F*, skeletal muscle; *G*, kidney; *H*, pancreas; *I*, spleen; *J*, thymus; *K*, prostate; *L*, testis; *M*, ovary; *N*, small intestine; *O*, colon; and *P*, peripheral blood leukocytes. An *open arrow* indicates hYAP mRNA. *Two arrows* indicate β-actin mRNAs. Note that heart and skeletal muscle and to lesser degree prostate and small intestine contain an extra form of β-actin mRNA that is of 1.6–1.8 kb. The exposure times were 3 days for hYAP and 2 h for β-actin.

polar amino acids and the presence of prolines distributed preferentially toward both termini of the linear sequence. One of the carboxyl-terminal prolines at the seventh amino acid from the end is invariably conserved (Fig. 6). The length of the WW domain was set at 38 residues, which corresponds to the length of the second WW motif (the insert) identified in the mouse ortholog of YAP (Fig. 2). Interestingly, the sequence similarity among WW domains ends rather abruptly beyond the 38 amino acids. If indeed the 38 amino acids of the WW motif compose a structured domain, the size would be relatively small compared with the SH2, SH3, or PH domain. There are two other features of the WW motif which also suggest a protein module. As with the SH3 domain, the WW sequence occurs frequently in multiple repeats within the same molecule: from two repeats in mouse YAP, for example, to four repeats in the human homolog of Nedd-4 (Fig. 7). In addition, most of the proteins that contain the WW motif(s) also contain other functional domains, either catalytic or structural (*e.g.* Rsp5 and 38D4, see "Discussion").

Examination of the primary sequence of the WW domains indicated that WW domains of YAP, Nedd-4 (Ref. 46), and Rsp5 (Ref. 47) show more similarity to each other than to WW domains of other proteins. It is likely that these domains share certain functional features; for example, they could interact with similar ligands or localize the proteins to similar cellular compartments. When the repeats of the WW domains within the same protein are examined, the second or third WW domain does not necessarily show as high a sequence similarity to the first WW domain as one would expect from a recent evolutionary duplication event, but does show a high similarity to WW domains of other proteins. For example, the second domain in mYAP is more similar to one of the WW domains of the yeast Rsp5 gene product than to the first domain in mYAP (Fig. 6). This suggests that multiple WW domains within the same molecule may not be redundant but could have evolved to carry out subtly diverged functions.

The domain turns out to be even more widespread than initially reported (32). As many as 11 new sequences with WW modules have been recently deposited in the gene banks (Fig. 6). Anticipating the number of WW sequences to grow rapidly, we have provided updated information on the WW domain via world wide web (www) (http://www.embl- heidelber.de/~bork/ww1.html) since December 1994. Both the alignment and a diagram with the modular structure of proteins containing the WW domain(s) (Figs. 6 and 7) are available via www network and will be updated by us (P. B. and M. S.) frequently.

The new list of proteins with the WW motifs confirmed our initial conclusion that like the SH2, SH3, and PH domains, the WW domain occurs in a variety of structural and signaling molecules with no apparent common functions. Three new proteins that contain the WW domain could provide a clue to the role of this module in major signaling pathways. One of the human proteins, named ORF1 (D29640), contains a WW domain just upstream from the carboxyl-terminal sequence that shows similarity to yeast Ras GTPase activator protein. A nematode (*Caenorhabditis elegans*) protein named 38D4 [Z46241] harbors a WW domain at the amino terminus, followed by a PH domain in the middle and a carboxyl-terminal sequence that is conserved in the breakpoint cluster region, *n*-chimaerin, and p85 subunit of the phosphoinositol 3-kinase (48, 49). A gene product named Msb1, with one WW domain, was isolated in a genetic screen in yeast and was implicated in the MAP kinase pathway.[2] Taken together, these data suggest an involvement of the WW domain in the Ras and/or MAP kinase signaling pathways.

DISCUSSION

Our results describe the molecular cloning, expression, and chromosomal localization of the human YAP gene, which encodes a protein implicated in binding to the SH3 domain of the Yes tyrosine kinase. We have also described cDNA cloning of the mouse YAP homolog, whose sequence provided a clue for the identification of a novel protein module, designated the WW domain (32), which is present in various regulatory, signaling, and structural molecules. The following aspects of this work deserve further comment: (i) the expression profile of YAP mRNA in human adult tissues, (ii) the high degree of sequence conservation among YAPs from higher eukaryotes, (iii) the chromosomal localization of the human YAP gene, and (iv) the widespread occurrence of the WW domain and its possible role as a module mediating protein-protein interaction.

[2] K. Matsumoto, personal communication.

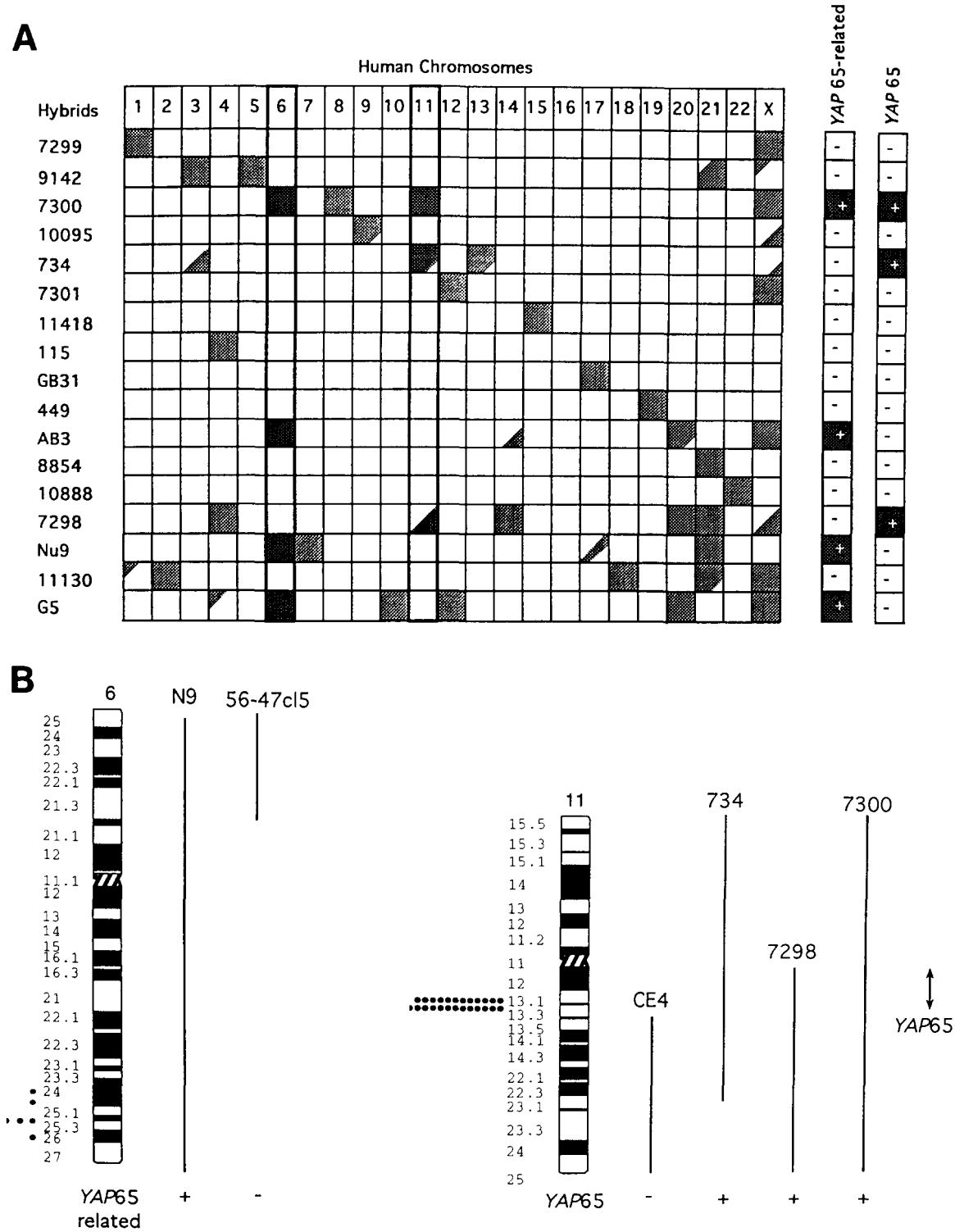

FIG. 5. **Chromosomal localization of the hYAP gene.** *A*, presence of the hYAP loci in a panel of 17 rodent-human hybrids. DNA (10 μg) from various rodent-human hybrids was cleaved with restriction enzyme *SstI*, electrophoresed, transferred to nylon filter, and hybridized to radiolabeled hYAP cDNA probe. ■ indicates that the hybrid named in the *left column* contains the chromosome indicated in the *upper row*; ◪ indicates the presence of the long arm of the chromosome (or part of the long arm represented by a smaller fraction of *stippling*); ◩ indicates the presence of the short arm (or partial short arm) of the chromosome; and □ indicates the absence of the chromosome listed above the column. The column for chromosomes 6 and 11 are *boldly outlined* and *stippled* to highlight correlation of the presence of these chromosomes (or region of the chromosomes) with the presence of the hYAP loci. The patterns of retention of the loci in the panel are shown to the *right* of the figure where the presence of a locus in a hybrid is indicated by a *stippled box with a plus sign*, and the absence of a locus is indicated by an *open box enclosing a minus sign*. *B*, regional chromosomal localization of hYAP loci. *Chromosome 6*, the portion of chromosome 6 present in specific hybrids is represented by the *solid line* to the right of the chromosome 6 idiogram. Hybrids were tested by filter hybridization as described under "Experimental Procedures." The presence or absence of the hYAP-related locus is indicated below the lines representing individual hybrids. The hYAP-related locus was present only in hybrids which retained chromosome region 6p21–6qter in common. Results of fluorescence *in situ* hybridization (FISH) to normal human metaphases is illustrated to the *left* of the chromosome 6 idiogram where each *filled circle* represents two fluorescent signals. *Chromosome 11*, hybrids carrying partial fragments of chromosome 11 are illustrated to the *right* of the chromosome 11 idiogram with results of filter hybridization

Protein/Species	Position	Sequences of the WW Domains	Accession #
Yap/Human	171	VPLPAGWEMAKTSS.GQRYFLNHIDQTTTWQDPRKAMLS	X80507
Yap/Chick	169	VPLPPGWEMAKTPS.GQRYFLNHIDQTTTWQDPRKAMLS	X76483
Yap/Mouse-1	151	VPLPAGWEMAKTSS.GQRYFLNHNDQTTTWQDPRKAMLS	X80508
Yap/Mouse-2	218	GPLPDGWEQAMTQD.GEVYYINHKNKTTSWLDPRLDPRF	X80508
Nedd4/Mouse-1	?	SPLPPGWEERQDVL.GRTYYVNHESRRTQWKRPSPDDDL	D10714
Nedd4/Mouse-2	?	SGLPPGWEEKQDDR.GRSYYVDHNSKTTTWSKPTMQDDP	D10714
Nedd4/Mouse-3	?	GPLPPGWEERTHTD.GRVFFINHNIKKTQWEDPRLQNVA	D10714
Nedd4/Human-1	218	SPLPPGWEERQDIL.GRTYYVNHESRRTQWKRPTPQDNL	D42055
Nedd4/Human-2	375	SGLPPGWEEKQDER.GRSYYVDHNSRTTTWTKPTVQATV	D42055
Nedd4/Human-3	448	GFLPKGWEVRHAPN.GRPFFIDHNTKTTTWEDPRLKIPA	D42055
Nedd4/Human-4	500	GPLPPGWEERTHTD.GRIFYINHNIKRTQWEDPRLENVA	D42055
Rsp5/Yeast-1	228	GRLPPGWERRTDNF.GRTYYVDHNTRTTTWKRPTLDQTE	L11119
Rsp5/Yeast-2	331	GELPSGWEQRFTPE.GRAYFVDHNTRTTTWVDPRRQQYI	L11119
Rsp5/Yeast-3	387	GPLPSGWEMRLTNT.ARVYFVDHNTKTTTWDDPRLPSSL	L11119
56G7/Caeel-1	229	TPPESHWKTYLDAK.KRKFYVNHVTKETRWTKPDTLNNN	Z46793
56G7/Caeel-2	372	QPLPSGWECITMNN..RTVFLNHANKETSFYDPRIRRFE	Z46793
Dmd/Human	3052	TSVQGPWERAISPN.KVPYYINHETQTTCWDHPKMTELY	P11532
Dmd/Ray	253	TSVQGPWERAISPN.KVPYYINHQTQTTCWDHPKMTELY	M37645
Utro/Human	2813	TSVQLPWQRSISHN.KVPYYINHQTQTTCWDHPKMTELF	X69086
Ykb2/Yeast-1	1	...MSIWKEAKDAS.GRIYYYNTLTKKSTWEKPKELISQ	P33203
Ykb2/Yeast-2	39	LLRENGWKAAKTAD.GKVYYYNPTTRETSWTIPAFEKKV	P33203
Yo61/Caeel-1	49	PSVESDWSVHTNEK.GTPYYHNRVTKQTSWIKPDVLKTP	P34600
Yo61/Caeel-2	94	QPQQGQWKEFMSDD.GKPYYYNTLTKKTQWVKPDGEEIT	P34600
Amoe/Acaca	?	KMSVDGWKQYFTAE.GNAYYYNEVSGETSWDPPSSLQSH	M60954
FE65/Rat	12	SDLPAGWMRVQDTS.GTYYWHIP.TGTTQWEPPGRASPS	X60468
Ess1/Yeast	29	TGLPTPWTARYSKSKKREYFFNPETKHSQWEEPEGTNKD	P22696
Msb1/Human	249	IVLPPNWKTARDPE.GKIYYYHVITRQTQWDPPTWESPG	not yet
ORF1/Human	679	GDNNSKWVKHWVKG.GYYYYHNLETQEGGWDEPPNFVQN	D29640
38D4/Caeel	97	RDLLNGWFEYETDV.GRTFFFNKETGKSQWIPPRFIRTP	Z46241
K015/Caeel	?	QNPDDAWNEFNAPD.GRKYYFNSITQENTWEKPKALIDQ	D34959
Yfx1/Yeast	9	PQVPSGWKAVFDDEYQTWYYVDLSTNSSQWEPPRGTTWP	Z46255
Db10/Tobac	8	PTLPKPWKGLVDGTTGFIYFWNPETNDTQYERPVPSSHA	D16247
Consensus Line:		LPtGWE ttt Gt YYhNH TtTTtW tPt t	
Secondary Structure:		llllll eee llll eeeee llll lllllllll	

FIG. 6. **Alignment of selected WW domains.** Computer programs used in the analysis of the protein sequences and in predicting their secondary structures were as described under "Experimental Procedures." The consensus line displays conserved features (*capitals*, conserved amino acids; *h*, hydrophobic; *t*, turn-like or polar). Two amino acids, **W** and **P**, are 100% conserved in all WW domains listed and are shown in *bold*. *Dots* indicate spaces introduced in order to optimize the alignment. *Question marks* denote lack of the full open reading frame for a given protein; therefore, the precise location of the WW domain within a protein was not possible. (For more information on the entries, see Ref. 32 and "Results" for the availability of constantly updated information on the WW alignment through the www network.)

The expression profile of YAP mRNA in adult human tissues is broad with relatively high levels of the message in placenta, prostate, testis, ovary, and small intestine (Fig. 4). We did not detect YAP mRNA in peripheral blood leukocytes, and, in contrast to results obtained with chicken YAP, relatively low levels of the mRNA were detected in adult human brain (Fig. 4, *lane B*). Two factors could account for the quantitative difference observed in the brain tissues. First, in chickens we used cerebellum and telencephalon for our studies and not the entire brain, as was done in the mRNA preparation from the human source. Second, the age of the chicken brain was 2 weeks, whereas the sample of the human brain mRNA was from individuals of various ages and sex. The size of hYAP mRNA was estimated at approximately 5 kb, which is in good agreement with the size of its cDNA (5.1 kb, see Fig. 1).

The YAP sequence is highly conserved among higher eukaryotes, as shown by sequence comparison among human, mouse, and chicken YAP, as well as by "Zoo-blot" analysis. Our data from the comparative Southern blot analysis on genomic DNA from yeast (*S. cerevisiae*) showed hybridization of hYAP cDNA with distinct DNA bands on the blot. However, these bands coincided with the so-called satellite DNA bands and therefore probably represented signals of nonspecific hybridization, rather than hybridization with a YAP homolog or a YAP-related gene present in yeast.

The human YAP gene maps to chromosome 11q13 centromeric to the CCND1/BCL1 locus and could thus be near the locus for the multiple endocrine neoplasia type 1 familial gene (MEN1) (50–53). MEN1 is an autosomal dominant disorder characterized by a high frequency of peptic ulcer disease and primary endocrine abnormalities involving the pituitary, parathyroid gland, and pancreas. Schimke (54) postulated that the MEN1 mutation may involve derepression of a "primitive" gene, possibly a proto-oncogene, coding for a protein that promotes the growth of endocrine glands. A high resolution radiation hybrid map of the proximal long arm of chromosome 11, containing the MEN1 and BCL1 gene loci, pointed to the *sea* proto-oncogene as one of the potential candidates for the MEN1

to the hYAP cDNA shown below the lines representing hybrids; the hYAP cognate locus is present only in hybrids which retain 11cen → 11q13 in common. Hybrid CE4 retains a der 14 (14pter → 14q32::11q13 → 11qter) from a B cell leukemia with a break in the BCL1 major breakpoint region and is negative for the hYAP locus. Thus, the hYAP gene is centromeric to the CCND1/BCL1 locus. Results of FISH on normal human metaphases is illustrated to the *left* of the chromosome 11 idiogram where each *filled circle* represents two fluorescent signals.

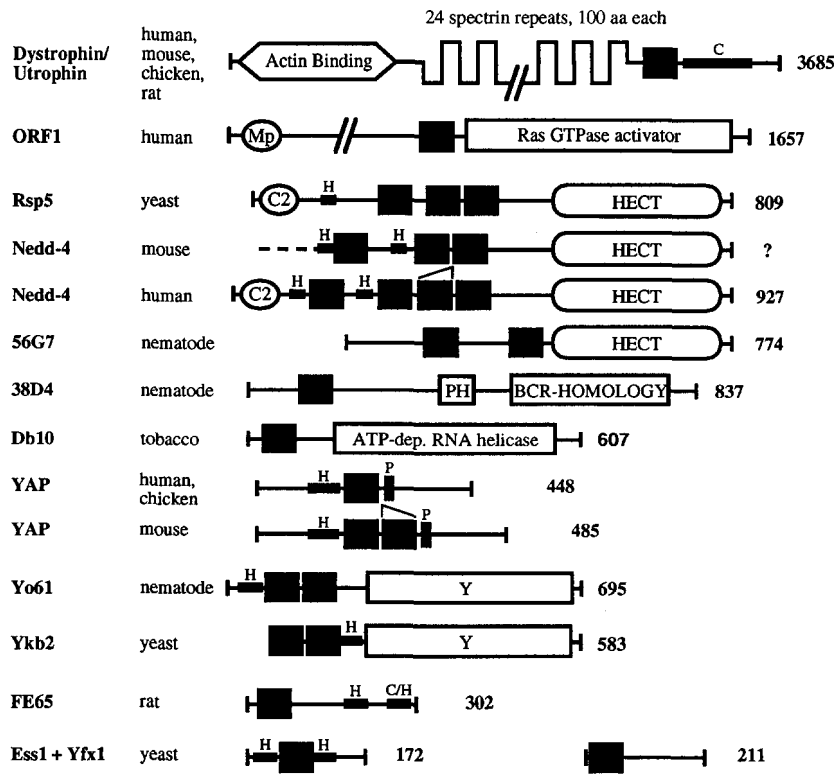

FIG. 7. **Modular structure of proteins containing the WW domains.** The WW domain is indicated by a *black box*. The C2 domain is shared with some forms of protein kinase C, synaptotagmins, and *C. elegans* Unc-13 protein; the *HECT* domain is found at the carboxyl termini of Rsp5, E6-AP, UreB1, Nedd4, a hypothetical yeast protein (Ykbo), and 56G7 protein from *C. elegans*, and it encodes ubiquitin-protein ligase activity (58–60); other catalytic domains shown are: Ras-GTPase activator-like domain; breakpoint cluster region homology domain that is shared with Grb-1, *n*-chimaerin, and other signaling molecules (48, 49); and an ATP-dependent RNA helicase; *PH*, pleckstrin homology domain; actin-binding domain shares similarity with actinin and spectrin; the *Y* domain is common to Yo61 and Ykb2 and its function is not known; *Mp* domain shows similarity to a fly muscle protein mp20; *C/H* indicates a region rich in cysteine and histidine; *P* denotes a proline-rich region in YAP implicated in binding to the SH3 domain of Yes, Src, and Nck (31). *Dashed lines* indicate that only partial sequence data were available. For the purpose of clarity, some of the proteins and domains were not drawn to scale.

locus (55, 56). The hYAP DNA probe should be a valuable marker to refine the chromosomal map around the MEN1 locus and perhaps to identify the MEN1 gene. The localization of hYAP to chromosome 11q13 also allows prediction of the location of the mouse ortholog on chromosome 19 or 7 (Ref. 57).

The function of the WW domain in YAP and in other proteins remains to be determined. The occurrence of the domain in yeast proteins provides a powerful genetic system that could be employed to analyze the function of the WW motif *in vivo*. The *rsp5* gene was identified as a suppressor of mutations in the *spt3* gene, which encodes a transcription factor that interacts with TATA-binding protein (Ref. 47).[3] It is unlikely that Rsp5 and Spt3 proteins interact, since Rsp5 mutations suppress a deletion of Spt3. It is more likely that this interaction is indirect because the *rsp5* gene was recently isolated by researchers in several other laboratories studying various aspects of cytoplasmic signaling in yeast.[4] One of the explanations for this apparent diversity of roles in signaling could come from biochemical studies of Rsp5, showing that its catalytic domain (designated *HECT*, Fig. 7) can form a high energy thioester bond with ubiquitin, therefore proving that the Rsp5 protein is indeed a ubiquitin-protein ligase (58–60). It is likely that *nedd-4*, the gene whose expression is modulated during early development of the central nervous system, encodes the similar enzymatic activity as Rsp5 (60). Since ubiquitination is directly related to protein metabolism, and the WW domains in Rsp5 and Nedd-4 could be considered as molecular adhesive to anchor the ligase to the appropriate targets, one could speculate that a ligand for the WW domain, in general, is of a proteinaceous nature and that this domain represents a module mediating protein-protein interaction. Two types of preliminary data support our hypothesis. (i) We have recently identified and cloned two cDNAs for low molecular weight proteins that bind specifically to the WW domain of hYAP.[5] The analyses of the partial DNA sequences of the putative ligands suggest two novel gene products. We anticipate that as in the case of 3BP1 and 3BP2 proteins, and the SH3 domain of Abl (61), the isolation of two independent sequences by a functional assay will help us to define the region that interacts directly with the WW domain. It is hoped that the two sequences will share a significant sequence similarity over this region. (ii) We have recently determined preliminary NMR spectra of the WW domain of hYAP; the results suggest a structured domain.[6]

In speculating about the possible function of the WW domain in the specific context of a given protein, in addition to Rsp5 and Nedd-4, we would like to briefly mention three other examples: dystrophin, Msb1, and YAP.

The human dystrophin gene encodes a large molecule that is classified as a cytoskeletal protein, based on its similarity to α-actinin and β-spectrin (62). Mutations of this gene cause the degenerative diseases, Duchenne and Becker muscular dystrophies (62). The WW domain of dystrophin is located in the carboxyl-terminal part of this large protein, close to the site which connects the molecule with membranes through the β-dystroglycan receptor (63). Although the WW domain is extremely well conserved among dystrophins from different vertebrates (100% between human and chicken, for example) (64) and the preliminary data we obtained recently on the existence of protein ligands and structure for the WW domain are pointing to this part of the molecule as a putative "signaling module," there is no direct evidence to support our daring proposal of the WW domain as a signaling site in dystrophin (32). The current survey of available point mutations and short deletions identified in the dystrophin (65) from muscular dystrophy patients does not show any genetic lesions that would map to the conserved residues within the WW domain, although it is clear that any stop-codon mutation that maps before the WW domain and abrogates the carboxyl terminus of dystrophin causes the more severe, Duchenne type of dystrophy. With the availability of an animal model for muscular dystrophy (mdx mice), one

[3] F. Winston, personal communication.
[4] F. Winston and A. Hopper, personal communication.
[5] H. Chen and M. Sudol, unpublished data.
[6] A. Ulrich, M. Hyvonen, H. Chen, M. Sudol, H. Oschkinat, and M. Saraste, unpublished data.

could gather genetic evidence as to the role of the dystrophin WW module directly, through transgenic approaches.

The *msb1* gene encodes a 424-amino-acid long protein that contains one WW domain. *msb1* stands for "mammalian suppressor of *bck1*" because this gene was isolated as a suppressor of the *bck1* mutation using a cDNA library prepared from human cells.[2] Yeast *bck1* gene encodes a kinase (MAP kinase kinase kinase), which functions downstream of Pkc1 (yeast protein kinase C) and upstream of Mkk1/Mkk2 (MAP kinase kinase) (66). Surprisingly, the *msb1* gene was also found to be able to suppress the *mpk1* (yeast MAP kinase) mutation.[2] Therefore, it is possible that *msb1* may function downstream of *mpk1*. This genetic system provides a powerful tool to assay the role of the WW domain in the protein kinase C and MAP kinase pathways in yeast and to extend these studies to a mammalian model.

The role of the WW domain in YAP is not known, although one could speculate that this module connects the Src family tyrosine kinases with serine kinases (31). The lack of an apparent catalytic domain in YAP, the tandem location of the WW domain(s) and the SH3 binding proline-rich motif, plus YAP expression in most tissues are reminiscent properties of the adaptor-like molecules, Grb-2 or Crk. We have recently isolated a chicken isoform of YAP with two WW domains (see www update), providing suggestive evidence that mouse YAP is an isoform generated by alternative splicing.[7] This recent result suggests the presence of two isoforms of YAP in human as well as in mouse. Since two putative ligand proteins for "the first" WW domain of hYAP were identified, we hope that these reagents will advance functional and structural studies of this new module.

Acknowledgments—Special thanks are due to Hidesaburo Hanafusa for his enthusiastic encouragement, support, and constructive advice concerning our work on the ligand for the WW domain; Drs. Fred Winston, Anita Hopper, Peter Pavlik, Peter Howley, Teresa Zoladek, and Kunihiro Matsumoto for sharing with us unpublished results on the *rsp5* and *msb1* genes; Drs. Hartmut Oschkinat, Anne Ulrich, Marko Hyvonen, and Matti Saraste for collaboration on the structure of the WW domain and for permission to mention their preliminary data; Dr. Stuart Aaronson for the gift of the human cDNA library; Henry Chen and Dan Stieglitz for stimulating discussions and help with the preparation of some parts of the manuscript; Drs. Hidesaburo Hanafusa, Tadashi Yamamoto, Beatrice Knudsen, Peter Howley, Fred Winston, Andrea Musacchio, and Matti Saraste for valuable comments on the first version of the manuscript; Drs. Georges Calothy, Cecile Bougeret, Sahng-June Kwak, and Paul Fehlner and Terry Solomon for insightful discussions on the potential medical implications of our work on the WW motif in human dystrophin; and Dr. Bonnie Kaiser for enthusiastic support.

REFERENCES

1. Pawson, T., and Schlessinger, J. (1993) *Curr. Biol.* **3,** 434–442
2. Birge, R. B., and Hanafusa, H. (1993) *Science* **262,** 1522–1524
3. Bork, P. (1991) *FEBS Lett.* **286,** 47–54
4. Sudol, M. (1993) *The Molecular Basis of Human Cancer* (Neel, B., and Kumar, R., eds) pp. 203–224, Futura Publishing Co., Inc., Armonk, NY
5. Musacchio, A., Gibson, T., Lehto, V. P., and Saraste, M. (1992) *FEBS Lett.* **307,** 55–61
6. Mayer, B., and Baltimore, D. (1993) *Trends Cell Biol.* **3,** 8–13
7. Haslam, R. J., Kolde, H. B., and Hemmings, B. A. (1993) *Nature* **363,** 309–310
8. Mayer, B. J., Ren, R., Clark, K. L., and Baltimore, D. (1993) *Cell* **73,** 629–630
9. Pawson, T., and Gish, G. D. (1992) *Cell* **71,** 359–362
10. Kanner, S. B., Reynolds, A. B., Wang, H. C. R, Vines, R. R., and Parsons, J. T. (1991) *EMBO J.* **10,** 1689–1698
11. Roussel, R. R., Brodeur, S. R., Shalloway, D., and Laudano, A. P. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 10696–10700
12. Clark, S. G., Stern, M. J., and Horvitz, H. R. (1992) *Nature* **356,** 340–344
13. Musacchio, A., Noble, M., Pauptit, R., Wierenga, R., and Saraste, M. (1992) *Nature* **359,** 851–855
14. Yu, H., Rosen, M. K., Shin, T. B., Seidel-Dugan, C., Brugge, J. S., and Schreiber, S. L. (1992) *Science* **258,** 1665–1668
15. Booker, G. W., Gout, I., Downing, A. K., Driscoll, P. C., Boyd, J., Waterfield, M. D., and Campbell, I. D. (1993) *Cell* **73,** 813–822
16. Kohda, D., Hatanaka, H., Odaka, M., Mandiyan, V., Ullrich, A., Schlessinger, J., and Inagaki, F. (1993) *Cell* **72,** 953–960
17. Koyama, S., Yu, H., Dalgarno, D. C., Shin, T. B., Zydowsky, L. D., and Schreiber, S. L. (1993) *Cell* **72,** 945–952
18. Li, N., Batzer, A., Daly, R., Yajnik, V., Skolnik, E., Chardin, P., Bar-Sagi, D., Margolis, B., and Schlessinger, J. (1993) *Nature* **363,** 85–88
19. Noble, M. E., Musacchio, A., Saraste, M., Courtneidge, S. A., and Wierenga, R. K. (1993) *EMBO J.* **12,** 2617–2624
20. Rozakis-Adcock, M., Fernley, R., Wade, J., Pawson, T., and Bowtell, D. (1993) *Nature* **363,** 83–85
21. Musacchio, A., Wilmanns, M., and Saraste, M. (1994) *Prog. Biophys. Mol. Biol.* **61,** 283–297
22. Kato, J. Y., Takeya, T., Grandori, C., Iba, H., Levy, J. B., and Hanafusa, H. (1986) *Mol. Cell. Biol.* **6,** 4155–4160
23. Potts, W. M., Reynolds, A. B., Lansing, T. J., and Parsons, J. T. (1988) *Oncogene Res.* **3,** 343–355
24. Nemeth, S. P., Fox, L. C., DeMarco, M., and Brugge, J. S. (1989) *Mol. Cell. Biol.* **9,** 1109–1119
25. Hirai, H., and Varmus, H. E. (1990) *Mol. Cell. Biol.* **10,** 1307–1318
26. Seidel-Dugan, C., Meyer, B. E., Thomas, S. M., and Brugge, J. S., (1992) *Mol. Cell. Biol.* **12,** 1835–1845
27. Weng, Z., Taylor, J. A., Turner, C. E., Brugge, J. S., and Seidel-Dugan, C. (1993) *J. Biol. Chem.* **268,** 14956–14963
28. Liu, Y., Marengere, E. M., Koch, C. A., and Pawson, T. (1993) *Mol. Cell. Biol.* **13,** 5225–5232
29. Weng, Z., Taylor, J. A., Turner, C. E., Brugge, J. S., and Seidel-Dugan, C. (1993) *J. Biol. Chem.* **268,** 14956–14963
30. Harian, J. E., Hajduk, P. J., Yoon, H. S., and Fesik, S. W. (1994) *Nature* **371,** 168–170
31. Sudol, M. (1994) *Oncogene* **9,** 2145–2152
32. Bork, P., and Sudol, M. (1994) *Trends Biochem. Sci.* **19,** 531–533
33. Miki, T., Matsui, T., Heidaran, M. A., and Aaronson, S. A. (1989) *Gene (Amst.)* **83,** 137–146
34. Palazzolo, M. J., Hamilton, B. A., Ding, D., Martin, C. A., Raghavan, K. V., Mierendorf, R. C., Mead, D. A., Meyerowitz, E. M., and Lipschitz, H. D. (1990) *Gene (Amst.)* **88,** 25–36
35. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
36. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1987) *Current Protocols in Molecular Biology*, John Wiley and Sons, New York
37. Huebner, K., Druck, T., Croce, C. M., and Thiesen, H.-J. (1991) *Am. J. Hum. Genet* **48,** 726–740
38. Lou, Z., Kastury, K., Crilley, P., Lasota, J., Druck, T., Croce, C. M., and Huebner, K. (1993) *Genes Chrom. Cancer* **7,** 15–27
39. Nagarajan, L., Louie, E., Tsujimoto, Y., Balduzzi, P. C., Huebner, K., and Croce, C. M. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 6568–6572
40. Koonin, E. V., Bork, P., and Sander, C. (1994) *EMBO J.* **13,** 493–503
41. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410
42. Gribskov, M., McLachlan, A. D., and Eisenberg, D. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 4355–4358
43. Rhode, K., and Bork, P. (1993) *Comput. Appl. Biosci.* **9,** 183–189
44. Rost, B., and Sander, C. (1994) *Proteins* **19,** 55–72
45. Tatusov, R. L., Altschul, S. F., and Koonin, E. V. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 12091–12095
46. Kumar, S., Tomooka, Y., and Noda, M. (1992) *Biochem. Biophys. Res. Commun.* **185,** 1155–1161
47. Eisenmann, M. D., Arndt, K. M., Ricupero, S. L., Rooney, J. W., and Winston, F. (1992) *Genes & Dev.* **6,** 1319–1332
48. Boguski, M. S., and McCormick, F. (1993) *Nature* **366,** 643–654
49. Hall, C., Monfries, C., Smith, P., Lim, H. H., Kozma, R., Sohail, A., Vanniasingham, V., Leung, T., and Lim, L. (1990) *J. Mol. Biol.* **211,** 11–16
50. Schinzel, A., Frezal, J., and McKusick, V. A. (1993) *Genome Prior. Rep.* **1,** 658–699
51. Bystrom, C., Larsson, C., Blomberg, C., Sandelin, K., Falkmer, U., Skogseid, B., Oberg, K., Werner, S., and Nordenskjold, M. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 1968–72
52. Larsson, C., Skogseid, B., Oberg, K., Nakamura, Y., and Nordenskjold, M. (1988) *Nature* **332,** 85–87
53. Nakamura, Y., Iarsson, C., Julier, C., Bystrom, C., Skogseid, B., Wells, S., Oberg, K., Carlson, M., Taggart, T., O'Connell, P., Leppert, M., Lalouel, J. M., Nordenskjold, M., and White, R. (1989) *Am. J. Hum. Genet.* **44,** 751–755
54. Schimke, R. N. (1986) *New Engl. J. Med.* **314,** 1315–1316
55. Richard, C. W., Whithers, D. A., Meeker, T. C., Maurer, S., Evans, G. A., Myers, R. M., and Cox, D. R. (1991) *Am. J. Hum. Genet.* **49,** 1189–1196
56. Fujimori, M., Wells, S. A., and Nakamura, Y. (1992) *Am. J. Hum. Genet.* **50,** 399–403
57. O'Brien, S. J., Peters, J., Searle, A., Womack, J., and Marshall-Gram, J. (1993) *Genome Prior. Rep.* **1,** 758–809
58. Huibregtse, J. M., Scheffner, M., and Howley, P. (1995) *Cold Spring Harbor Symp. Quant. Biol.* **59,** in press
59. Scheffner, M., Huibregtse, J. M., Vierstra, R. D., and Howley, P. (1993) *Cell* **75,** 495–505
60. Huibregtse, J. M., Scheffner, M., Beaudenon, S., and Howley, P. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 2563–2567
61. Cicchetti, P., Mayer, B. J., Thiel, G., and Baltimore, D. (1992) *Science* **267,** 803–806
62. Ahn, A. H., and Kunkel, L. M., (1993) *Nature Genet.* **3,** 283–291
63. Tinsely, M. J., Blake, D., Zuellig, R. A., and Davis, K. E. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 8307–8313
64. Lemaire, C., Heilig, R., and Mandel, J. L. (1988) *EMBO J.* **7,** 4157–4162
65. Roberts, R. G., Gardner, R. J., and Bobrow, M. (1994) *Human Mutat.* **4,** 1–11
66. Irie, K., Takase, M., Lee, K. S., Levin, D. E., Araki, H., Matsumoto, K., and Oshima, Y. (1993) *Mol. Cell. Biol.* **13,** 3076–3083

[7] M. Sudol, unpublished data.