Identification, basic characterization and evolutionary analysis of differentially spliced mRNA isoforms of human YAP1 gene

Christian J. Gaffney, Tsutomu Oka, Virginia Mazack, Dror Hilman, Uri Gat, Tomoki Muramatsu, Johji Inazawa, Alicia Golden, David J. Carey, Amjad Farooqe, Gerard Tromp, Marius Sudol

The YAP1 gene encodes a potent new oncogene and stem cell factor. However, in some cancers, the YAP1 gene plays a role of tumor suppressor. At present, the gene and its products are intensely studied and its cDNAs are used as transgenes in cellular and animal models. Here, we report 4 new potential mRNA splicing isoforms of the YAP1 gene, bringing the total number of isoforms to 8. We detected all 8 YAP1 isoforms in a panel of human tissues and evaluated the expression of the longest isoform of YAP1 (YAP1-2a) using Real Time PCR. All YAP1 isoforms are barely detectable in human leukocytes compared to fair levels of expression found in other human tissues. We analyzed the structure of the genomic region that gave rise to alternatively spliced YAP1 transcripts in different metazoans. We found that YAP1 isoforms, which utilize exon 6 emerged in evolution with the appearance of amniotes. Interestingly, 6 YAP1 isoforms, which contain the exon 5 extension, exon 6 or both would have their leucine zipper region disrupted in the predicted protein product, compared to the intact leucine zipper found in two YAP1 (x) isoforms. This observation has direct functional ramifications for YAP1 signaling. We also propose a normalized nomenclature for the mRNA splice variants of the YAP1 gene, which should aid in the characterization of signaling differences among the potential protein products of the YAP1 gene.

© 2012 Elsevier B.V. All rights reserved.
Over the past six years, the number of reports on the YAP1 gene has grown exponentially. Several YAP1 cDNA clones, which we and others have isolated, showed that apart from the presence or absence of the second WW domain (Sudol et al., 1995), YAP1 sequences varied in terms of inserted sequences at a position located within the TAD region (transcriptional activation domain) (Fig. 1) (e.g., Hao et al., 2008; Komuro et al., 2003). Since many studies have reported the use of a single cDNA of YAP1 in functional over-expression studies, but the YAP1 cDNAs used differed from each other either by the number of WW domains or by the variations within the TAD domain (e.g., Hao et al., 2008; Komuro et al., 2003; Muramatsu et al., 2011; Oka et al., 2008), we decided to define the spectrum of isoforms of the human YAP1 gene and propose a standardized nomenclature for their reference. In our study, we identified four additional, potential isoforms of human YAP1, which differ from the four isoforms reported recently (Muramatsu et al., 2011), proposing the existence of eight isoforms of YAP1 in total. We also investigated the exon–intron structure of the human YAP1 gene in detail and quantitatively measured the relative concentration against the following genes: $\alpha$-tubulin, $\beta$-actin, G3PDH and phospholipase A2 and used as templates for PCR. cDNA for YAP1-1 and YAP1-2 isoform sequences were confirmed by direct sequence analysis at Genewiz, (South Plainfield, NJ). 2.2. YAP1-1 and YAP1-2 isoform analysis

The first strand human pancreas cDNA preparation was used as a template for amplification of YAP1 isoforms. The amplified product was ligated into the pCR 2.1-TOPO vector and transformed into chemically competent E. coli (TOPO TA Cloning, Invitrogen, Carlsbad, CA), with an incubation time of 30 min at room temperature. The pCR 2.1 TOPO library was then transformed into ‘One Shot Top Ten’ competent cells via the One Shot Chemical Transformation Protocol, as outlined by the manufacturer, and plated on beta-galactose indicator plates. White colonies were selected from plates and plasmid DNA was prepared from each by use of the Qiaprep Spin Miniprep Kit (Qiagen, Maryland). Clones were then digested with EcoR1 (Promega, Madison, WI) and resolved on a 3% agarose gel.

2.3. Expression of YAP1-2 $\delta$ isoform in human tissues

YAP1-2- $\delta$ TaqMan primers (Applied Biosystems, Foster City, CA) were as follows, Forward, 6FAMAAGAAAGGACCCCTTTGGCTAGA CCCAAMGBNFQ and Reverse, VICGCACTGCTGCGCC TACCTMGBN FQ. The forward primer was located within the second WW domain of YAP, whereas the reverse primer corresponded to the sequences in exons 5A, 5B (also labeled as 5 extension) and 6, with the DNA corresponding to the coding region for QVRPQAMR (see Fig. 2B).
Quantitative real-time PCR was performed using the Applied Biosystems 7500 Fast Real-Time PCR System. Our TaqMan gene expression assay was custom designed to measure the expression level of the YAP1 isoform in 16 human tissue types. Comparative Ct quantitation experiments were run under standard cycling conditions using 10 µl reactions containing 3 µl of cDNA. GAPDH was used as the endogenous control. All reactions were run in triplicate.

2.4. Evolution of YAP1 gene

YAP1 and its paralog TAZ (WWTR1) were queried in the ENTREZ GENE database using the web-interface. For non-bilaterians, sequences were first verified with blastp against Refseq database. The sequences of the exons relevant to the different human isoforms were visualized on the genomic locus of different metazoans using the ENTREZ GENE graphic interface. Exon positions were manually assigned and corrected when needed.

2.5. Sequence and structural analysis of leucine zippers of YAP1 isoforms

The amino acid sequence alignment of the putative leucine zippers of various YAP1 isoforms was performed using ClustalW (Thompson et al., 1994). Structural models of putative leucine zippers of YAP1-1α isoform were built using the MODELLER software based on homology modeling (Marti-Renom et al., 2000). Briefly, the solution structure of the leucine zippers of Jun transcription factor (PDB# 1JUN) was used as a template. It should be noted that the putative leucine zippers of YAP1-1α share greater than 25% sequence identity with those from Jun. Importantly, the putative leucine zippers of YAP1-1α contain a total of five highly-conserved signature leucine residues, designated herein L1–L5 in one chain and L1’–L5’ in the other chain, at every seventh position in a manner akin to the leucine zippers of Jun. A total of 100 atomic models were calculated and the structure with the lowest energy, as judged by the MODELLER Objective Function, was selected for rendering using RIBBONS (Carson, 1991).
3. Results

3.1. Identification of YAP1 splice isoforms

Forward and reverse primers of human YAP1 were designed to identify not only both YAP1-1 and YAP1-2 isoforms, the latter of which differs from the former by the presence of an additional WW domain (Sudol et al., 1995), but also to identify suspected splicing isoforms at amino acid position 290 for YAP1-1 and 328 for YAP1-2. Surprisingly, gel analysis of RT-PCR products from the amplification reaction of human pancreas cDNA, which was performed under stringent conditions, revealed eight products, which we provisionally named as YAP1-1α, β, γ and YAP1-2 α, β, γ, δ (Fig. 1). In a pilot experiment, the pancreas cDNA gave us the best yields and quality of the resolved bands, therefore we continued to use this cDNA as template for PCR amplifications, cloning and sequence analysis. The novelty of our data is that four additional isoforms of YAP, namely YAP1-1α and γ plus YAP1-2β and δ, were detected in addition to four other isoforms that were recently reported in normal esophagus and in an esophageal cancer cell line (Muramatsu et al., 2011). The PCR amplified DNA fragments were subcloned into pCR2.1 TOPO vector and submitted to direct DNA sequence analysis. The DNA sequences and the predicted amino acid sequences of all four new YAP1 isoforms plus one isoform that was reported previously (Muramatsu et al., 2011) but was not deposited in the sequence data banks were submitted to the EMBL-EBI Nucleotide Sequence Database and assigned the following accession numbers: HE864159, HE864160, HE864161, HE864162, and HE864163.

3.2. Genomic structure of YAP1 locus

The structure of the human YAP1 gene comprises 9 exons (Fig. 3). Interestingly, the first WW domain resides within exons 2 and 3, which are flanked by the longest intron. The second WW domain is located within a single exon, exon 4. Exon 5 contains a short extension, which we named 5B*. More precisely, an alternative splice donor site 12 nucleotides into intron 5 generates some alternative transcripts containing the amino acids VRPQ. Note two 5' splice donors in intron 5, (5A, CAGGTGAGG) (5B*, CAGGTTAGA). The 5A splice donor is fairly canonical, only the last G appears less often than T. In the 5B* splice donor, the first 5 nucleotides are conserved, including the obligatory GT, the 6th nucleotide is a T (least preferred), the 7th and 8th nucleotides are conserved, and the 9th nucleotide is the least preferred base. From the presence of least preferred bases in two sites, the 5B* could be a weaker 5' splicing site than 5A.

Exon 4, (which encodes the second WW domain), the exon 5 extension, 5B*, (which codes for amino acids VRPQ), and exon 6, (which codes for amino acids AMRNINPSTANSPKCQ), are alternatively spliced to generate the 8 primary transcripts of YAP1 gene (Figs. 1–3).

3.3. Detection of 8 transcripts in human organs using non-quantitative PCR

We next sought to determine the relative expression profiles of all 8 YAP1 isoforms across numerous tissues. Commercial preparations of cDNA generated from polyA+ mRNA from human organs, which were normalized for concentration against 4 genes (see Materials and methods) were used as templates for PCR. We assayed cDNAs from colon, small intestine, leukocytes, spleen, kidney, liver, skeletal muscle, brain, placenta, lung, heart, ovary and testes. With the exception of leukocytes, all other tissues or organs expressed all known isoforms of YAP1 (Supp. Figs. 2 and 3). We also noted that in most of the organs the two shortest isoforms, namely YAP1-1α and β were expressed at significantly lower levels, compared to other isoforms (Supp. Fig. 2).

3.4. Expression of the YAP1-2 δ isoform using TaqMan primers and quantitative PCR

To provide an example of the quantitative expression profile of one of the isoforms of YAP1, we selected YAP1-2δ isoform because of the possibility of designing PCR primers that would not amplify

---

**Fig. 3.** Structure of the YAP1 gene and the 8 alternatively spliced transcript isoforms. **Top Panel,** structure of the gene and its primary transcript drawn to scale. **Middle panel,** structure of primary transcript with introns of arbitrary length, but exons drawn close to scale with respect to one another. **Bottom Panel,** structure of eight mature (spliced) RNA transcripts. Dark green indicates transcribed regions and un-translated regions of mature transcripts; dark blue, no identified protein domain; red, WW domain; yellow with orange border, beginning of the TAD domain; light blue with black border, alternatively spliced exon 6 sequence, AMRNINPSTANSPKCQ.
any other isoforms (Fig. 2B). The forward primer was placed within the second WW domain, whereas the reverse primer was at the very junction of spliced transcripts that included exon 5B* extension and exon 6 (Fig. 2B and Materials and methods). Indeed, the amplification reaction resulted in a single DNA fragment of the correct size (not shown).

From the TaqMan-RT-PCR results, we calculated a fold change using as a reference the lowest expression level found in leukocytes (Fig. 4). Ovary and placenta expressed the highest level of the YAP1-2β isoform, whereas skeletal muscle and brain expressed lower levels of the isoform, but still more than 2.5 fold higher level than leukocytes (Fig. 4). Equal amounts of cDNAs from human organs, which were normalized for the expression of 4 housekeeping genes (see Materials and methods) were used in the TaqMan based Real Time PCR, run in triplicate. Although the normalization process is reassuring, we do assume that the cDNA synthesis efficiency was similar for all the RNA templates isolated from different organs.

3.5. Evolution of the splicing isoforms of YAPI gene

We decided to analyze the evolutionary changes in the structure of the genomic region that includes the exons 5 and 6 coding for the carboxy-terminal isoforms of YAPI. In the basal metazoan, sea anemone, (Nemastrotella vectensis) the YAPI gene is encoded by two exons only (Fig. 5). In this animal, the exon 2 harbors the sequence equivalent to the human exon 5, ending with a codon for the residue Q, which is followed by the sequence equivalent to the start of the human exon 7, initiating with the codon for residue E. In the basal chordate, sea squirt, (Ciona intestinalis) the exon equivalent to human exon 5 ends in a similar sequence, with Q as the last residue, and is followed by the exon 7 equivalent that starts at a more carboxy-terminal sequence of YAPI protein.

An extension of exon 5, shown in green (Fig. 5) can be first found in the vertebrate, Zebrafish (Danio rerio) and in another fish, pufferfish (Tetraodon nigroviridis), in the form of a sequence coding for three additional amino acids (RPQ). The amphibian representative, the pipid frog (Xenopus tropicalis) demonstrates an extension of 4 amino acids (VRPQ), which are identical to those added to the human isoform β when compared to isoform α. Importantly, the addition of the codon for V (GTG) before the RPQ coding sequence creates a novel consensus site for an intron-donor start site, which gives rise to alternative splicing products, isoforms β and δ, which include the sequence we have termed “Exon 5-ext”, while in the other isoforms it is spliced out. We can assume that VRPQ-containing-isoforms can be found in most or all tetrapods.

The first ortholog of human exon 6, (marked in blue, Fig. 5) was identified in the amniote, an arboreal lizard (Anolis carolinensis). When compared to humans, this lizard has a very similar exon 6-like genomic sequence coding for 16 amino acids, and remarkably conserved exon-intron donor and acceptor sites. The genomic structure of all further amniotes up to man is also highly conserved, and we can thus expect the existence of these isoforms in them as well. The evolutionary appearance of exon 6 may be due to adjustments of amniotes to life on land or to the formation of amniote egg and its specific structures.

Interestingly, the YAPI paralog gene, TAZ (WWTR1), (Kanai et al., 2000) resembles the primordial state of YAPI in this region and shows no sign of either one of the evolutionary novel exons (Fig. 5).

In conclusion, we can see an elaboration of the YAPI gene structure along evolution, which led to the appearance of these isoforms. The basal form of YAPI has no introns 3’ to the WW domains and lacks both the exon 5-extension and the exon 6-like sequences. In early chordates, an intron was first introduced into that region, allowing further complexity. We first see a possible extension of the primordial exon 5, which may have occurred in the early vertebrate lineage after the formation of the TAZ gene. This was followed by a formation of an alternative intron start site, which allowed the existence of the extended isoform as well as the primordial form in the same organism. The next stage, allowing even more complexity, is first detected in amniotes with the “birth” of a new exon, in what was previously intron 5. This event doubled the number of YAPI isoforms that could be expressed.

It is worthy to note that exon 5-ext (VRPQ) is not identified in the NCBI annotations of genomic sequences. The lack of this exon extension is due to the low representation of this sequence in the consensus CDS protein set (CCDS) database (CCDS ID CCDS44716). CCDS has only one partial representation of this extension in GenBank, AK304485.1 (1). In contrast, there are more than 10 better quality EST’s that lack this extension. This can be seen in the evidence viewer for YAPI in ENTREZ GENE, geneID, 10413 (2).

3.6. Sequence and structural analysis of coiled coil region of YAPI isoforms

To further understand the differences between the various isoforms of YAPI identified here and how these differences may affect their biological function, we next performed detailed sequence and structural analysis (Fig. 6). Our amino acid sequence alignment suggests that the various isoforms of YAPI differ by the insertion of between 4 and 20 residues within the putative leucine zipper located within the TAD domain (Fig. 6a). It is noteworthy that the leucine zipper is a highly conserved protein module involved in mediating protein–protein interactions pertinent to a plethora of cellular activities (Landschulz et al., 1988). The leucine zipper is characterized by the presence of a signature leucine at every seventh position within the minimum four successive heptads of amino acid residues. Importantly, leucine zippers are thermodynamically unstable in isolation, but adopt continuous amphipathic α-helices in the context of a dimer, either through homo-association or hetero-association with other cellular partners, by virtue of their ability to wrap around each other into a parallel dimeric coiled coil with a slight left-handed twist (O’Shea et al., 1989a, 1989b, 1992). Ironically, while the YAPIα isoforms fulfill the above criteria of a stretch of amino acid sequence capable of generating natively folded leucine zippers (Fig. 6b) primarily through inter-digation of signature leucines within each chain, the insertion of amino acids (4–20

---


Fig. 4. The relative levels of YAPI-2β isoform expression in human tissues and organs. The data are represented in two forms. On the left, the distribution is summarized with a box-and-whisker plot: the median is indicated by the bold line, the range between the first and third quartiles is indicated by the box, the whiskers extend to 1.5 times the inter-quartile range (IQR), and circles indicate outliers (outside the 1.5 IQR). On the right, individual relative expression values are shown labeled with the source tissue. The expression levels were measured by TaqMan Real-Time PCR using human cDNAs normalized for expression to four house-keeping genes (see Materials and methods). Note that the expression of YAPI-2β isoform in leukocytes is significantly lower than that in other tissues.
residues) between the third and fourth heptad in YAP1β, YAP1γ and YAP1δ isoforms would likely disrupt the amino acid sequence requirement of leucine zippers, thereby leading to their thermodynamic destabilization. Accordingly, the inability of YAP1β, YAP1γ and YAP1δ isoforms to adopt leucine zippers may have serious consequences on protein–protein interactions required for their hetero-association with other cellular partners. In short, our sequence and structural analysis suggests that the YAP1β, YAP1γ and YAP1δ isoforms would not be able to participate in a wide variety of protein–protein interactions driven by the leucine zippers of YAP1α isoforms.

Interestingly in TAZ protein, the YAP1 paralog, in the region that corresponds to the leucine zipper of YAP1 neither a coiled coil region could be detected (Lupas et al., 1991; Lutenic et al., 2012) nor TAZ gene undergoes splicing in this region (Fig. 5).

4. Discussion

We identified 4 novel potential mRNA splicing isoforms of the human YAP1 gene, bringing the total number of isoforms to 8 (Komuro et al., 2003; Muramatsu et al., 2011; Sudol et al., 1995). We investigated the exon–intron structure of the human YAP1 gene and showed that differential use of exon 4, exon 5 extension and exon 6 gives rise to new isoforms of YAP1. We evaluated the relative levels of expression of the longest isoform of YAP1 (YAP1-2δ) in human tissues, using Real Time PCR, and showed that it is widely expressed, except in leukocytes. We also traced the evolutionary appearance of exon 6 to amniotes. Detailed analysis of the predicted protein region where the splicing of exon 5 extension and exon 6 occur, indicates that a leucine zipper would be disrupted in these isoforms compared to the unspliced α isoforms of YAP1.
YAP1. Finally, we proposed a simple nomenclature to distinguish the eight isoforms of YAP1 unequivocally.

The following aspects of this report deserve further comment: (i) the existence of the actual protein products for the 8 isoforms of YAP1; (ii) the known signaling differences among YAP1 isoforms; (iii) the relatively low expression of YAP1 isoforms in leukocytes; (iv) the disruption of the predicted leucine zipper in those YAP1 isoforms that contain exon 5 extension and exon 6; (v) the importance of the normalization of the nomenclature for YAP1 isoforms in the rapidly evolving field of Hippo pathway signaling.

We can easily distinguish YAP1-1 and YAP1-2 isoforms because they differ by 38 amino acids of the WW domain when YAP1 is precipitated with polyclonal antibodies. We cannot, however, resolve individual α-δ isoforms on a regular or gradient SDS-PAGE (e.g., Oka et al., 2008). Although various YAP1 cDNA clones have been generated and used by investigators in published reports, each encoding a distinct form of YAP1, the clones that we report here are novel. However, we cannot be sure that all 8 isoforms described here are expressed as proteins in cells. Moreover, various post-synthetic modifications, including serine (Sudol, 1994) and tyrosine phosphorylation (Levy et al., 2008) plus sumoylation (Lapi et al., 2008) have been shown for YAP1 protein and could add further to the heterogeneity of products resolved on SDS-PAGE gels, so that even the two dimensional gel chromatography may not be the right tool to identify the protein products of YAP1. Either mass spectrometry analysis or a generation of monoclonal antibodies against spliced-in peptide sequences could provide answers as to whether all 8 isoforms of YAP1 are expressed as proteins.

Most of the studies published on YAP1 employ one of the YAP1-2 isoforms that contain two WW domains. Expression clones of YAP1-1 with one WW domain are used less frequently. However, YAP1-1 clearly differs in signaling from YAP1-2. For example, our group has shown that YAP1-1 does not bind the p73 factor and cannot induce apoptosis when HEK293 cells are stressed by low serum conditions (Oka et al., 2008). In addition, YAP1-1 does not interact with angiomotin while YAP1-2 does, and angiomotin has the ability to inhibit the pro-apoptotic function of YAP1-2 by preventing its nuclear localization (Oka et al., 2012).

We observed a relatively low level of YAP1 expression in leukocytes compared to other tissues. This finding could be interpreted as following: YAP1 and perhaps other components of the Hippo pathway do not need to be operative in free floating cells such as leukocytes. It is known that one of the main routes by which the Hippo pathway controls YAP1 activity is via clues from cell-to-cell junctions, which help determine the final size of organs and the body (Zhao et al., 2010). Leukocytes and white blood cells do not form junctions. Three items of information further support the finding that YAP1 does not play a major role in the biology of leukocytes. Analyzing polyA+ mRNA isolated from human peripheral blood leukocytes on a Northern blot, we could not detect the YAP1 transcript, which is widely expressed in other human tissues and is 5 kb long. (Sudol et al., 1995). In mice, whose hematopoetic cells lack two kinases, Mst1 and Mst2, which upstream of YAP1 in the canonical Hippo pathway, the number of mature T cells is significantly reduced (Mou et al., 2012). However, this defect in T cell proliferation seems unrelated to YAP1 (Mou et al., 2012). The ectopic expression of YAP1 in hematopoetic stem cells does not affect the quantity and function of these cells, neither during the steady state nor under stress conditions (Jansson and Larsson, 2012). This is in contrast to the effects of YAP1 on stem and progenitor cells in other organs.

Using individual mRNA splice isoforms of a gene for functional studies without specifying which isoform was used created controversy in the past, as exemplified by various reports using one of the four splice isoforms of ErbB4 receptor kinase (Sundvall et al., 2007; Veikkoalainen et al., 2011). The main message of our report is to propose a uniform nomenclature for the human YAP1 splice isoforms to avoid a similar scenario. This is important because the field of Hippo pathway signaling is progressing rapidly. A number of studies reported over-expression of individual and different isoforms of YAP1 in cells and animals (e.g., Hao et al., 2008; Komuro et al., 2003; Muramatsu et al., 2011; Oka et al., 2008; Schlegelmilch et al., 2011). Moreover, it is likely that exon 5 extension and exon 6 splicing variants of YAP1, which vary within the transcriptional activation domain (TAD), may select different repertoires of proteins in transcriptional complexes and affect the gene expression program in the YAP1 isoform- and tissue-specific manner. The disruption of the putative leucine zipper in exon 5 extension and exon 6 splicing variants of YAP1 may have functional consequences and we suggest that the two α isoforms of YAP1 will signal differently from β, γ and δ isoforms. Most likely the signaling differences between α and β, γ, δ isoforms of YAP1 will be subtle and a simple overexpression of these isoforms in cells followed by colony formation assay, for example, may not be the best way to characterize isoform-specific functions (Muramatsu et al., 2011).

Naming each of the isoforms of YAP1 either as suggested here (Figs. 2A and B) or by the number of amino acids in the coding region of YAP1 isoform (e.g., YAP1-450 aa = YAP1-1c; whereas YAP1-508 aa = YAP1-2b) will avoid miscommunications and should be helpful in the dissection of functional differences among various isoforms.

5. Conclusions

Four new mRNA splice isoforms of the human YAP1 gene were identified, making the total eight. Expression of all 8 isoforms of YAP1 is significantly lower in leukocytes compared to other tissues and organs. A new exon, exon 6 of the YAP1 gene, which is differentially spliced in various isoforms, appeared in evolution during the emergence of amniotes. All six isoforms of YAP1 with exon 5 extension, exon 6 or both, have their predicted leucine zipper regions disrupted. A new nomenclature for all 8 isoforms of YAP1 is proposed.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2012.08.025.

Acknowledgments

We thank our colleagues Kristin Berger, Kieran Harvey, Nikos Tapinos and Greg Yochum for valuable comments on the first version of the manuscript.

This work was supported by PA Breast Cancer Coalition Grants (#60707 and #920093) plus the Geisinger Clinic (to MS) and by funds from the National Institutes of Health (Grant# R01-GM083897) and the US Sylvester Braman Family Breast Cancer Institute to AF.

References


Sudol, M., 1994. Yes-associated protein (YAP65) is a proline-rich phosphoprotein that binds to the SH3 domain of the Yes proto-oncogene product. Oncogene 9, 2145–2152.


