Oncogene (2003) 22, 5082–5090

www.nature.com/onc



# Data mining the p53 pathway in the Fugu genome: evidence for strong conservation of the apoptotic pathway

Morgane Le Bras[1,3], Karim Bensaad[1,4] and Thierry Soussi*[,1,2]

[1]*Laboratoire de Génotoxicologie des Tumeurs, Institut Curie, 26 rue d'Ulm, 75005 Paris, France;* [2]*Université P.M. Curie, 4 Place Jussieu, 75005 Paris, France*

**The p53 tumour suppressor gene belongs to a small family of related proteins that includes two other members, p63 and p73. Phylogenetic and functional studies suggest that p63 and p73 are ancient genes that have essential roles in normal development, whereas p53 seems to have evolved more recently to prevent cell transformation. In mammalian cells, a plethora of proteins have been found to specifically regulate p53 activity. The genome of the fish *Fugu rubripes* has been recently published. It is the second vertebrate genome for which the entire sequence is now available. Phylogenetic studies are essential in order to analyse and define signalling pathways important for cell cycle regulation. The presence or absence of a critical member in any pathway can shed light about the evolution of these pathways. The Fugu genome databank has been analysed for several members of the p53 network, including p53, p63 and p73. A good conservation of the network that regulates p53 stability and apoptosis has been found. We also discovered that some cofactors that cooperate with p53 for apoptosis are also well conserved and belong to multigene families not detected in the human genome.**
*Oncogene* (2003) **22,** 5082–5090. doi:10.1038/sj.onc.1206424

**Keywords:** p53 tumour suppressor gene; p53 gene family; apoptosis; phylogenetic studies; fugu rubripes; data mining

## Introduction

The p53 gene family, comprising p53, p63 and p73, is situated at the crossroads of a network of signalling pathways that are essential for apoptosis induced by genotoxic and nongenotoxic stresses (Vogelstein *et al*., 2000; Melino *et al*., 2002; Vousden and Lu, 2002). The p53 protein controls certain cell cycle checkpoints, whereas p63 and p73 have more specific functions in tissue differentiation (Levine, 1997; Yang *et al*., 2002). In normal unstressed cells, these upstream pathways predominantly include the binding by proteins such as MDM2 or JNK that promote p53 degradation via the ubiquitin/proteasome pathway (Alarcon-Vargas and Ronai, 2002). After genotoxic or nongenotoxic stresses, these pathways include signalling mainly performed by kinases and acetylases that regulate p53 stability and conformation, leading to activation of p53 transcriptional activity. A plethora of proteins have been found to bind various regions of p53 in order to regulate the specificity of its activity. Downstream signalling includes a large series of genes that are activated by the transactivating properties of p53. This occurs via specific DNA binding of the p53 protein to a p53 response element (p53 RE) that is found either in the promoter or in the intron of target genes (El-Deiry *et al*., 1992; Tokino and Nakamura, 2000). Regardless of the type of stress, the final outcome of p53 activation is either cell cycle arrest and DNA repair or apoptosis, but the mechanism leading to the choice between these two fates has not yet been elucidated (see Vousden and Lu (2002) for a discussion). Numerous studies have analysed the pattern of genes induced after p53 activation using global technologies such as SAGE, DNA array, suppression subtractive hybridization (SSH) or by cloning functional p53-binding sites (Polyak *et al*., 1996; Yu *et al*., 1999; Kostic and Shaw, 2000; Tokino and Nakamura, 2000; Zhao *et al*., 2000; Kannan *et al*., 2001). These studies emphasize the heterogeneity of the p53 response that is highly variable depending on the cell type, the nature and amount of DNA damage, the genetic background of the cells and the amount of p53 protein. Several criteria have been described to define a *bona fide* p53 response gene (Ko and Prives, 1996): (i) the existence of p53 RE specifically recognized by p53; (ii) the ability of these sites to act as a p53 RE activating transcription in a wild-type p53-dependent manner; (iii) the response of the element to p53 in the context of the endogenous promoter; and (iv) induction of the target gene after cellular stress such as DNA damage in cells containing wild type, but not a mutant form of p53. These criteria have recently been challenged, as several p53-induced genes lack p53 RE. It has also been shown that, in the PIG3 gene, p53 DNA binding does not need the canonical response element, but binds efficiently to a

*Correspondence: T Soussi, Institut Curie, 26, rue d'Ulm, 75248 Paris Cedex 05, France;
E-mail: Thierry.Soussi@curie.fr

[3]Current address: Université de Versailles Saint Quentin, CNRS FRE 2445-LGBC, 45 avenue des Etats-Unis, 78035 Versailles, France

[4]Current address: Beatson Institute for Cancer Research, Garscube Estate, Switchback Road, Bearsden, Glasgow G61. 1BD, Scotland
Received 16 December 2002; revised 8 January 2003; accepted 28 January 2003

polymorphic microsatellite sequence localized in the promoter (Contente *et al*., 2002).

Considerably less information is available about the p73 and p63 network. Recent data indicate that they are essential for the apoptotic function of p53 after DNA damage, a function that is supported by the observation that p73 and p63 can bind to the same DNA RE and therefore transactivate similar genes (Flores *et al*., 2002). On the other hand, the observation that neither p73 nor p63 are mutated in tumour cells suggests that they have an independent function (Melino *et al*., 2002). No specific p63 or p73 RE has been identified to date.

The genome of the fish *Fugu rubripes* has been recently published (Aparicio *et al*., 2002). It is the second vertebrate genome for which the entire sequence is now available. The Fugu genome is approximately 1/8 the size of the human genome because of (i) small intronic sequences, (ii) smaller intergenic distances and (iii) very little repetitive DNA. These features should allow rapid identification of important conserved regulatory regions. Numerous studies of gene families in Fugu have demonstrated that the Puffer fish genome contains approximately the same number of genes as mammalian genomes. The structure of genes is also conserved, as splice sites are located in positions identical to those found in man. Conservation of synteny should also allow the discovery of gene clusters that have been conserved through the course of evolution.

Phylogenetic studies are essential in order to analyse and define signalling pathways important for cell cycle regulation (Wahl and Carr, 2001). The presence or absence of a critical member in any pathway can shed light about the evolution of these pathways. Comparison of the yeast and human genomes has led to the discovery of the conservation of numerous pathways, many of which are important in human disease such as DNA repair or cellular checkpoint in cancer (Steinmetz *et al*., 2002). On the other hand, the unicellular structure of yeast prevents the analysis of pathways specific to multicellular organisms, such as differentiation or organogenesis for example. The p53 pathway described above also appears to be lacking in yeast despite the presence of the upstream pathways such as Mec1/Rad3 (ATR) or Tel1 (ATM). A p53 gene homologue has been recently described in *Drosophila melanogaster* and *Caenorhabditis elegans*, but this ancestor protein more closely resembles the p63 gene from which the p53 gene was derived at a period which remains to be defined (Brodsky *et al*., 2000; Jin *et al*., 2000; Ollmann *et al*., 2000; Derry *et al*., 2001; Schumacher *et al*., 2001).

We have analysed the Fugu genome databank for several members of the p53 network, including p53, p63 and p73. There is a very good conservation of the network that regulates p53 stability and apoptosis. We also discovered that some cofactors that cooperate with p53 for apoptosis are also well conserved and belong to multigene families not detected in the human genome.

## Results and discussion

### *The p53 family*

The easy identification of Fugu p53 is not surprising, as rainbow trout p53 cDNA was already identified more than 10 years ago and the p53 cDNAs have been described more recently for several other fish species (Caron de Fromentel *et al*., 1992; Krause *et al*., 1997; Cachot *et al*., 1998). Exon 7 is missing because of the lack of sequencing of 200 bp in the gene. In mammals and *Xenopus laevis*, exon 1 is noncoding and divergent among the various species. In the present analysis, as no expressed sequences were available, exon 1, if present, was not detected. Use of the 5′ UTR of various fish p53 as probes for BLAST analysis was also unsuccessful. The genomic structure of the Fugu gene is identical to that of the human gene with a fully conserved intron/exon organization (Figure 1). As expected for the Fugu genome, the coding region of the p53 gene (exons 2–11) is very compact in the fish (2.5 kb) compared to its mammalian counterpart (9 kb). The Fugu p53 protein does not display any particular features compared to other fishes or mammalian p53 protein (Figure 1A, supplementary material). The p53 sequences for 10 fish species are currently available. The phylogenetic tree deduced from the alignment of these proteins has been constructed (Figure 1B, supplementary material). All these fishes belong to the class of Actinopterygii (ray-fined fishes). The phylogenetic tree shows two groups corresponding to the Neoteleostei subclass (medaka, flounder, fugu and tetraodon) and the Ostariophysi subclass (catfish, zebrafish and barbus), respectively.

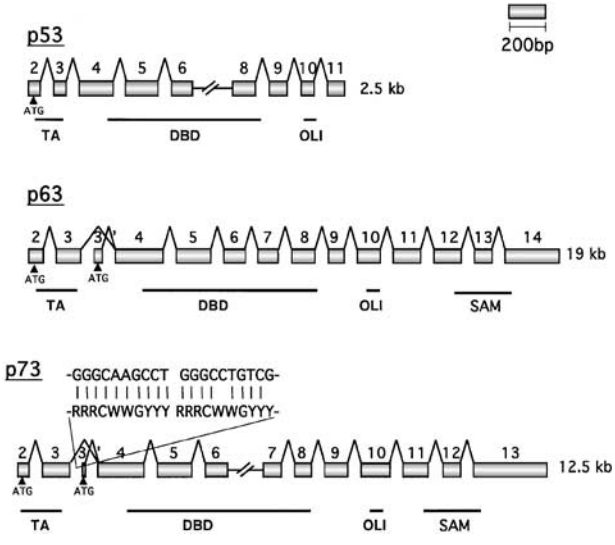


**Figure 1** Genomic organization of the p53, p63 and p73 gene in Fugu. Intron sizes have not been drawn to scale. For each gene, only coding exons have been indicated. For p53 and p73, sequence uncertainty is responsible for certain gaps on this map. For the p73 gene, the sequence of the p53RE found in intron 3 is shown above the consensus defined for this sequence. TA, transactivation domain; DBD, DNA-binding domain; OLI, oligomerization domain; SAM, sterile α motif

Although the trout is an Euteleosti like medaka, flounder, Fugu and tetraodon, it has a rapidly divergent lineage. Unfortunately, no p53 from coelancanthiforma or dipnoi that are closer from the terapod lineage are available.

The structure of the p63 and p73 genes is shown in Figure 1. For both genes, the first noncoding exon, if any, was not found. For the p73 gene, there is a degree of uncertainty around intron 6 and the beginning of exon 7 because of the missing sequence information. There is marked compaction of these two genes: 20 kb for p73 (100 kb in humans) and 13 kb for p63 (100 kb in humans). This feature can be very useful, as it should allow cloning and expression of these full-length genes with their own promoter in order to provide a deeper insight into the regulation of their expression. The exon/intron organization is well conserved compared to their human counterpart. Although the p53 gene codes for a major RNA species translated into one protein, mammalian p73 and p63 genes code for multiple isoforms with variations at the amino- and carboxy-termini. The carboxy-terminus isoforms are generated by numerous alternative splicings of exons 11, 12 and 13 that code for the sterile $\alpha$ motif (SAM) domain of p63 and p73. The amino-terminus isoforms are generated by either alternative splicing or the use of an internal promoter localized in intron 3. Although the biological function of the carboxy-terminus isoforms has not been clearly defined, the amino-terminus isoforms (collectively named $\Delta$N) lack the transactivation domain, but retain the DNA-binding domain and could act as dominant-negative regulators of their full-length counterparts via hetero-oligomerization. The amino-terminus of the $\Delta$N p63 and $\Delta$N p73 generated by the use of this second promoter is encoded by an alternative exon (3') that is not used in the full-length protein. In both p73 and p63 genes, we have identified exon 3' and the sequence of the Fugu $\Delta$N p73 and $\Delta$N p63 proteins. It has been recently shown that this internal promoter localized in intron 3 of the human p73 gene contains a p53 RE allowing feedback regulation of p73, and perhaps p53 and p63 by $\Delta$N p73 (Grob *et al*., 2001; Kartasheva *et al*., 2002; Nakagawa *et al*., 2002). In humans, the consensus sequence for p53 RE is a repeat of the decamer motif RRRCWWGYYY separated by 0–13 of random sequence (El-Deiry *et al*., 1992). This degenerate nature of the p53 RE might be critical for the regulatory control of the various genes that are activated by p53, and could also be important for the distinction of a hierarchy of transactivation among the various p53 family members. It is generally assumed that this consensus is conserved in the mouse, but no systematic studies have been performed to verify this point. The main argument suggesting that this consensus could be conserved in various species is derived from the finding that *X. laevis* p53 binds efficiently to this consensus sequence in a similar way to human p53 (Hardy-Bessard *et al*., 1998). *X. laevis* p53 also efficiently transactivates human p53 target genes, such as p21WAF1 or bax. This p53 RE is also efficiently recognized by p73 and p63. Recent chromatin immunoprecipitation (CHIP) experiments indicate that p53, p63 and p73 can bind to similar promoter regions *in vivo*, but it remains to be demonstrated whether or not this is a direct binding (Flores *et al*., 2002). A putative p53 RE has been found in intron 3 of Fugu p73 (Figure 1 and Table 1). Of the 20 residues of this sequence, 18 match the consensus sequence. No similar sites have been observed either in the whole Fugu p53 or p63 gene or elsewhere in the p73 gene. Similar p53 RE has also been identified in other regulatory regions of Fugu genes known to be regulated by p53 (see below). Although experimental evidence is necessary to ensure that this promoter is really regulated *in vivo* by p53 (and p73), the strong homology of the p53 RE suggests that it is a *bona fide* target for this protein. The finding of a p53RE at this position in the p73 gene indicates that this regulation of p73 expression by any member of the p53 family is important for their various signalling pathways.

The sequences of the Fugu p73 and p63 proteins are 63 and 75% homologous to their human counterparts (for the full-length protein). The homology of $\Delta$N proteins is higher, as the main divergence lies in the amino-terminus that is missing in these proteins. Like their mammalian counterpart, Fugu p73 and p63 contain a SAM domain in their carboxy-terminus. No noticeable specific feature was observed in these proteins.

### *The mdm2 family*

MDM2 is an essential regulator of p53 stability (Daujat *et al*., 2001; Woods and Vousden, 2001). Owing to its ubiquitin ligase activity, it induces p53 degradation via the proteosomal pathway. Mdm2 is a transcriptional target of p53 because of the presence of p53 RE localized in intron 2 of the mdm2 gene. The mdm2 cDNA from zebrafish and *X. laevis* have already been identified, but the genomic organization of these two genes is unknown (Marechal *et al*., 1997; Thisse *et al*., 2000). In the present analysis, the coding exon of the Fugu mdm2 gene has been completely identified (Figure 2). The genomic organization is similar to that of the human gene with 10 coding exons distributed in a 5 kb region (25 kb in humans). The Fugu MDM2 protein (F-mdm2) shares all the functional domains identified in other mdm2 proteins (Figure 2, supplemen-

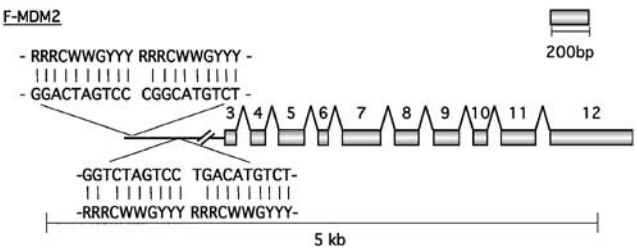


**Figure 2** Genomic organization of the F-mdm2 gene. Intron sizes have not been drawn to scale. Only coding exons are indicated. The sequence of the two p53RE found in intron 2 are shown above the consensus defined for this sequence

tary material). Several reviews have analysed the phylogenetic conservation of the mdm2 gene and protein (Piette *et al*., 1997; Momand *et al*., 2000). Three conserved regions have been identified: CRI contains the p53-binding sites, CRII contains a cysteine residue essential for binding Zn, and CRIII contains the RING finger essential for the ubiquitin ligase activity of mdm2. All these features are conserved in the Fugu protein (Figure 2, supplementary material). Analysis of intron 2 of the fdm2 gene has identified two p53 RE with a very strong homology with the consensus sequence (Figure 2 and Table 1), indicating a strong selection pressure for this regulation to ensure that p53 is always maintained at a minimum level in unstressed cells. A first mdm2 family member has been identified and has been named mdmx (Shvarts *et al*., 1996). Although it binds efficiently to p53 and inhibits its transactivating properties, mdmx does not lead to p53 degradation. The importance of the mdmx/p53 interaction has been demonstrated by the observation that homozygous loss of mdmx results in midgestational embryo lethality, a phenotype that is completely rescued by the absence of p53. Mice homozygous for both mdmx and p53 null mutations are viable and appear to be developmentally normal (Parant *et al*., 2001). Although mdmx, like mdm2, also regulates p73 transcriptional activity, there is still considerable uncertainty about the biological function of mdmx. Compared to the human and mouse genomic organization, the F-mdmx gene is very condensed (3.5 versus 20 kb in the mouse). It contains eight coding exons compared to 11 in the mouse due to fusion of exons 6, 7 and 8 in the fish. One-third of the protein is encoded by the last exon containing CRII and CRIII. These domains are well conserved in F-mdmx protein (Figure 2, supplementary material).

It has recently been shown that the AKT signalling pathway promotes phosphorylation and translocation of mdm2 in the nucleus to downregulate the p53 protein. This occurs after a mitogenic signal to ensure that p53 activity does not impair the cell cycle. This pathway is negatively regulated by PTEN phosphatase that dephosphorylates phosphatidylinositol (3,4,5) triphosphate (Pdtlns(3,4,5)P3), the second messenger produced by phosphatidylinositol 3 kinase that activates AKT. PTEN is a *bona fide* tumour suppressor gene whose germline and somatic inactivating mutations have been

**Table 1** Summary of the analysis of various genes involved in the p53 pathway

| *Gene*[1] | *Presence in Fugu* | *Identity (%)*[2] | *Scaffold*[3] | *SINFRUG*[4] | *p53 RE*[5] |
|---|---|---|---|---|---|
| p53 | Yes | 40% | 126 | 055229 | NR |
| p63 | Yes | 75% | 1141 | 053290 | NR |
| p73 | Yes | 63% | 67 | 064827 | **GGGCAAGCCT GGGC**CT**GTC**G (10/8) |
| ASPP1 | F-ASPP1a | 64% | 706 | 065875/065903 | NR |
| | F-ASPP1b[6] | 52% | 2940 | 061039 | NR |
| ASPP2 | F-ASPP2a | 64% | 1907 | 073718/073/731 | NR |
| | F-ASPP2b | 46% | 5697 | 056627/056629 | NR |
| | F-sASPP | 47%[7] | 700 | 064889/064903 | NR |
| JMY | F-JMY1 | 55% | 452 | 067493 | NR |
| | F-JMY2 | 57% | 442 | 074791 | NR |
| | F-JMY3 | 40% | 1675 | 075705 | NR |
| Apaf-1 | Yes | | 2551 | 070113 | **GAACTTGTCG AGACAGGTAC** (9/8)<br>**GAGCAGGTTC CAGCTGGCCT** (9/8)<br>**AAACTTGAGC** (N6) C**AGCTGGCCT** (8/8)[8] |
| Bax | NF | | | | |
| Cyclin G | Yes | 60% | 166 | 063043 | **GGACATGTCC** C**GACATGAC**G (10/7) |
| FAS/CD95 | NF | | | | |
| Mdm2 | Yes | 40% | 4374 | 079196 | **GGACTAGTCC** C**GGCATGTCT** (10/9)<br>**GGTCTAGTCC TGACATGTCT** (9/9) |
| mdmx | Yes | 45% | 1086 | 058743 | |
| Noxa | NF | | | | |
| p19ARF | NF | | | | |
| p21WAF | NF | | | | |
| p53AIP | NF | | | | |
| p53DINP1 | Yes | 33%[9] | 1545 | 024653 | **AGACCTGGTC AGTCAGGACT** (8/7)<br>**AGACCTGGTC** (N5) **GGACTTGTC**G (8/9)[10] |
| p53R2 | Yes | 72% | 1276 | 052604 | **GTACAAGC**AG **GCCCAGGCCT** (7/7) |
| PIDD | Yes | 46% | 4351 | 052731 | **AAGCATGCCC** (N10) **ATCCCAGTTT** (10/7) |
| PIG3 | Yes | 60% | 573 | 063138 | **GAGCTAGTTT** (N10) **AGTCATGGTT** (10/8) |
| PTEN | F-PTEN1 | 89% | 105 | 052295 | **AGCCGAGGCA GAGCAGGCTT** (6/9) |
| | F-PTEN2 | | 8240 | 084311 | NF |
| PUMA | NF | | | | |

[1]Several genes such as BTG2, although present in the fugu genome, could not be analysed due to too much uncertainty and gaps in the sequence; [2]Identity versus the human protein; [3]Position of the gene in the database localized at (http://www.ensembl.org/Fugu_rubripes/); [4]Gene name (http://www.ensembl.org/Fugu_rubripes/); [5]Bold residues and numbers in brackets indicate homology with the consensus sequence; [6]The protein sequence is partial, missing about 120 residues in the amino-terminus; [7]Identity with the mouse protein as the full-length sequence of human protein is not available; [8]Sites 2 and 3 overlap; [9]The protein sequence is partial, missing about some residues in the carboxy-terminus; [10]Sites 1 and 2 overlap. NF: Not found; NR: Not relevant

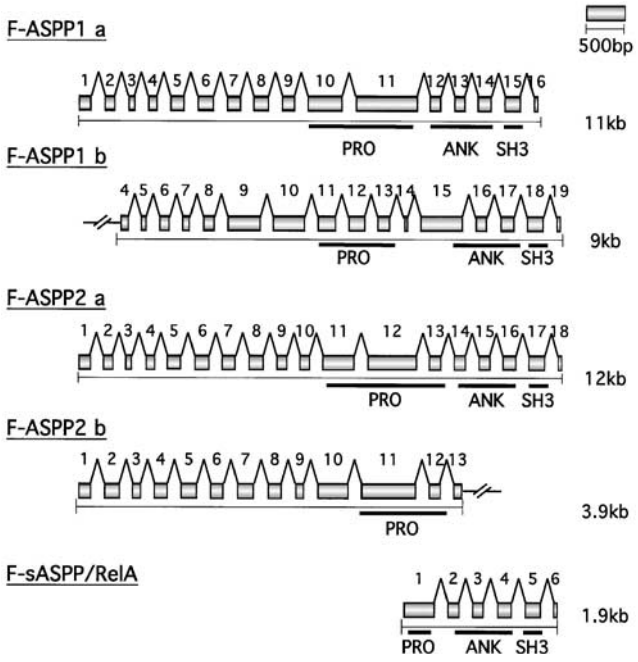


**Figure 3** Genomic organization of the five ASPP genes. Intron sizes have not been drawn to scale. For each gene, only coding exons are indicated. The protein domain, ankyrin (ANK), proline-rich (PRO) and Sarc homology (SH3) are shown under their corresponding coding exons. The sequences of F-ASPP1b and F-ASPP2b are not complete because of sequence gaps in the bank

found in hereditary cancer (Cowden disease) and brain cancer, respectively. The discovery that PTEN is a p53 target gene supports the concept of major crosstalk between the AKT and p53 signalling pathways (Stambolic *et al*., 2001). Following DNA damage, p53 can downregulate MDM2 nuclear translocation via PTEN activation to short-cut mitogenic activation (Mayo and Donner, 2002). The Fugu PTEN gene has already been described (Yu *et al*., 1999). The gene is highly compressed compared to the human homologue (7.5 kb versus 100 kb). In the present study, the Fugu PTEN protein is one of the most highly conserved proteins (89% identity). During screening of the Fugu databank, we isolated a second PTEN-related gene. The sequence of this gene is not complete, as there is a considerable sequence uncertainty in this region (Figure 3, supplementary material). The five coding exons detected encode for a protein that is 80% identical to the other Fugu PTEN protein (Figure 3, supplementary material). In humans, a PTEN pseudogene, corresponding to a processed sequence, has been detected in chromosome 9. In the present observation, this second PTEN homologue is encoded by a different exon with a similar organization to that of the first gene. The lack of information about the total gene organization precludes any further analysis to determine whether or not this gene is active. A p53 RE has been detected in the first intron of the Fugu gene (Table 1). This finding suggests that the crosstalk between the PTEN/AKT and MDM2 pathways is an ancient feature established before the divergence between fish and mammals.

*Apoptosis stimulating protein p53 (ASPP) family*

In humans, the ASPP genes (ASPP1 and ASPP2) have been recently shown to be involved in p53-induced apoptosis (Samuels-Lev *et al*., 2001). ASPP2 was originally identified as 53BP2, a protein that binds to the DNA-binding region of the p53 protein via an ankyrin motif (Gorina and Pavletich, 1996; Iwabuchi *et al*., 1994). It was subsequently shown that 53BP2 is the carboxy-terminus of a larger protein called ASPP2. A similar protein, ASPP1, has also been identified and the full-length cDNA clones of the two proteins have been isolated (Samuels-Lev *et al*., 2001). These two proteins share a similar structure, a proline-rich region in the centre of the protein followed by four ankyrin domains and an SH3 region in the carboxy-terminus. They are 48% identical with stronger conservation at the amino-terminus and carboxy-terminus. These two proteins have been shown to stimulate the apoptotic function of p53 via specific activation of p53 transactivating activity of apoptotic genes (bax, PIG3). No activation of the genes involved in growth arrest, such as p21, has been detected. Human ASPP1 and ASPP2 genes are localized in chromosomes 4 and 1 and are composed of 17 and 18 exons, respectively. These genes are conserved in the mouse, but no information is available for other species. The specific function of ASPP1 versus ASPP2 is unknown. Using either mouse ASPP1 and ASPP2 as probes, we have identified four ASPP genes in the Fugu genome: 2 ASPP1 (F-ASPP1a and b) and 2 ASPP2 genes (F-ASPP2a and b). A fifth gene, encoding a shorter version of ASPP (F-sASPP), has also been discovered (Figure 3). The identity between the two Fugu ASPP2 proteins is 50% with a high degree of conservation in the amino- and carboxy-terminus of the two proteins. These two FUGU proteins are 46 and 64% identical to the human ASPP2 protein and also present a greater conservation of the extremity of the protein (Figure 4, supplementary material). The carboxy-terminus contains the four ankyrin repeats followed by the SH3 domain. The situation is similar for ASPP1 except that one 5′ exon is missing for one of the two genes because of the uncertainty of the sequence in the 5′ end of the gene. F-ASSP1a and F-ASPP1b proteins are 55% identical, and are 64 and 52% identical to the human protein, respectively. Conservation of the various domains is also similar to ASPP2. Careful analysis of the human genome reveals only one ASPP1 and one ASPP2 gene. The high degree of homology between human ASPP1 and ASPP2 and their similar function suggest that they are derived from a common ancestor. This type of duplication occurred before separation with fish, as Fugu also contains the two genes. The finding of four ASPP genes in the Fugu is more surprising, as there is no evidence to suggest that part of the Fugu genome has undergone duplication. The fifth protein that is homologous to ASPP has been named F-sASPP for short Fugu ASPP, as it is homologous to the carboxy-terminus of ASPP (Figure 4, supplementary material). This protein contains the proline-rich region followed by the four ankyrin repeats

and the SH3 domain. A protein homologous to F-sASPP called RAI for RelA associated inhibitor (Yang *et al*., 1999) has also been found in the human genome. It was identified by a two-hybrid screen to be a partner and inhibitor of the p65 subunit of NF-$\kappa$B. It would be interesting to determine whether this protein has any function related to the ASPP/p53 network in mammalian.

*The JMY gene*

Junction mediated regulatory protein (JMY) was identified in the mouse by a two-hybrid screen using the p300-binding protein as bait (Shikama *et al*., 1999). The mouse JMY protein possesses a cluster of potential phosphorylation sites for cyclin-dependent kinases (CDK) in the amino-terminus and a carboxy-terminus that contains a long stretch of proline residues. This 100 kDa protein associates *in vivo* with p300 via its central region. It regulates p53-dependent transcription and augments p53-dependent apoptosis. After DNA damage, a ternary complex containing p300, p53 and JMY can be detected. The presence of a direct association between p53 and JMY has not been described. In humans, we have detected two proteins homologous to mouse JMY. One is localized on chromosome 5 and encodes for a 634-residue protein, while a second 809-residue protein is encoded by a gene localized in chromosome 15. The short protein is 36% identical to the carboxy-terminus of the longer protein. The mouse JMY is 90% identical to the short JMY protein, confirming the localization of JMY gene in chromosome 5 by Shikama *et al*. (1999). The function of the protein encoded by chromosome 15 is unknown. Using the mouse JMY sequence as a probe, three JMY-related genes were detected in Fugu (Figure 2, supplementary material). These genes encode three large proteins (865 residues for F-JMY1, 742 for F-JMY2 and 724 for F-JMY3). The homology between these three proteins is shown in Table 1. F-JMY3 is more closely related to the human protein localized on chromosome 15 (45% identity), whereas the other two proteins are more closely related to the short protein expressed by chromosome 5 (52 and 53% identity). The amino-terminus of F-JMY1 and F-JMY2 also present significant homology with the amino-terminus of the large human protein (Figure 5, supplementary material). These p300 cofactors therefore appear to belong to a multigene family. It would be interesting to analyse the specific activating function of each of these cofactors. This study would be fairly complex, as it has been shown that the mouse JMY gene undergoes alternative splicing encoding various forms that could have opposite effects. In the present study, p300 was not analysed in Fugu, as it is already known to be well conserved between vertebrates and invertebrates (Goodman and Smolik, 2000).

*Other p53 target genes*

Although many genes are regulated by p53, the purpose of this paper is not to provide an exhaustive analysis of the presence of all these p53 target genes, but rather to analyse some of the more representative p53 target genes. The most popular p53 target is the cyclin-dependent kinase inhibitor (CDKI) p21WAF1, which is involved in cell growth arrest after DNA damage. P21 belongs to the kinase inhibitor family (KIP) that includes p21WAF1, p27Kip1 and p57Kip2. The second family (INK4) includes p15INK4B, p16INK4A, p18 and p19. Analysis of the Fugu genome for the presence of CDKI p21WAF1 led to the discovery of the partial sequence of five CDKI from the KIP family (data not shown). The cyclin-binding region and the CDK-binding site have been conserved in the five proteins. None of these five proteins present a particular homology that could be related to p21, such as the PIP domain involved in the binding to PCNA (Warbrick, 1998). No equivalent of p21 has yet been detected in other nonmammalian species. Only two CDKI, p27Xil and p28Kix1, have been found in *X. laevis*. It has been suggested that the appearance of CDKI could be very recent compared to CDK, found in all species from yeast to human (Lee *et al*., 2002). In mammals, the INK4 locus encodes two different proteins by the use of alternative splicing and protein translation of two different reading frames (Sherr and Weber, 2000). p16INK4A is a CDKI involved in the negative regulation of the cell cycle during the G1 phase. The second protein, named p19ARF, binds to mdm2 after oncogene activation leading to delocalization of mdm2 and activation of p53. It has been previously shown that the Fugu INK4 locus only encodes the p15/16INK4A protein and no ARF could be encoded by the same gene (Gilley and Fried, 2001). It was then suggested that either no ARF was present in Fugu or it was encoded by a gene that is independent of the INK4 gene that could have recombinated after divergence with fish to give rise to the INK4/ARF locus (Gilley and Fried, 2001). Analysis of the total Fugu genome for the presence of the ARF protein did not provide any positive results, suggesting that this gene appeared very recently.

We then analysed the Fugu genome for p53 target genes involved in either DNA repair or apoptosis (Table 1) and which are known to have *bona fide* p53 RE either in promoter or intronic sequences. No orthologue related to bax, Fas, p53AIP1, PUMA or NOXA was detected. Several genes with an amino-terminus region containing a cysteine-rich domain related to the extracellular domain of Fas were found, but none of them have a death domain. Similarly, for bax protein, several proteins with Bax homology domain (BH) were identified, but the sequence divergence precludes any relation with their human counterpart (data not shown). One of the most conserved apoptotic target genes identified to date is the PIG3 gene. The protein is 60% identical compared to its human counterpart (Figure 6, supplementary material). To date, this protein has only been identified in mouse and in human. Analysis of the gene promoter revealed a putative p53 RE composed of three decamers separated by 5 and 11 bp, respectively (Table 1). Although the expression of this gene and its product have been widely

investigated after various types of p53 induction, the function of this quinone oxidoreductase homologue has still not been elucidated, apart from the fact that it is involved in p53-mediated apoptosis. The high level of conservation of PIG3 suggests that it has an important function, either related or unrelated to p53. The apoptotic protein activator (Apaf-1) gene is also well conserved (60% identity between human and Fugu Apaf-1). The WD repeats found at the carboxy-terminus are also well conserved (Figure 7, supplementary material). They are involved in protein–protein interactions. Apaf-1 protein is essential for activation of caspase 9 by binding to cytochrome *c*. The Fugu gene has a complex genomic organization with at least 24 exons (data not shown). A p53 RE has been detected in the Apaf-1 promoter from Fugu (Table 1).

The other genes that are well conserved in the Fugu genome include cyclin G, PIDD, p53DINP1 and p53R2. Cyclin G was one of the first p53-regulated genes discovered, but its function remained unknown until recently (Okamoto and Beach, 1994). It has been shown that cyclin G binds to enzymatically active phosphatase 2A (PP2A) holoenzymes and Mdm2. This binding stimulates the ability of PP2A to dephosphorylate Mdm2, suggesting that cyclin G could modulate phosphorylation of Mdm2 and thereby regulate both Mdm2 and p53 (Okamoto *et al*., 2002). The Fugu cyclin G protein is 60% identical to its human counterpart (Figure 8, supplementary material). The putative cyclin G promoter also contains a p53 RE. PIDD is a new protein that has been shown to promote p53-induced apoptosis (Lin *et al*., 2000). The originality of this protein concerns its unusual structure with a death domain in the carboxy-terminus and several leucine-rich repeats in the amino-terminus. The Fugu protein contains the same features and is 46% identical to its human counterpart with more marked conservation in the amino- and carboxy-terminus, emphasizing the importance of these specific domains (Figure 9, supplementary material). No protein partner has been isolated to date. p53 RE has been localized in intron 1 of the gene. p53DINP1 is also a p53-induced gene whose product does not share any homology with any known proteins (Okamura *et al*., 2001). This protein stimulates phosphorylation of p53 at serine 46, which is necessary for apoptosis. The Fugu DINP1 protein is 33% homologous to the human protein. A p53 RE has been detected in the promoter region of this gene (Figure 10, supplementary material). Finally, the p53R2 gene, which encodes a protein similar to the ribonucleotide reductase small subunit, has also been identified in the Fugu genome (Figure 11, supplementary material). It has been suggested that this protein could be involved in the p53 checkpoint for repair of damaged DNA (Tanaka *et al*., 2000). In Fugu, this protein is 72% identical to its human counterpart. A p53 RE has been identified in the first exon of the gene. This is the first time that such a localization has been found and more experimental evidence is necessary to confirm its function.

In summary, conservation of the network of genes that acts upstream and downstream to p53, p73 and p63 is heterogeneously conserved. Although data mining can be a powerful approach for the dissection of signalling pathways, it is associated with a number of pitfalls and limitations. Current gene finding programmes require long open reading frames that could prevent identification of small proteins such as CDKI or Bcl2 family members. The sequence of the fugu genome has not been fully determined in the chromosomes and there are still a number of gaps that could prevent identification of several genes. We have not analysed the conservation of the various kinases that have been shown to activate p53, as many of them, such as ATM, ATR, CHK1 and 2, are known to be conserved from yeast to human. The mdm2 signalling pathway, including cyclin G, mdmx and, to a lesser extent, PTEN, is well conserved, emphasizing the importance of tight regulation of p53 breakdown. Apoptosis is an essential function of the various members of the p53 family. p53 acts both in the death receptor pathways and in the stress pathway (also called the mitochondrial pathway) via transactivation of a plethora of target genes. Whether all of these genes or only a subset are necessary in each p53-dependent apoptotic situation has yet to be elucidated. The finding that the majority of these genes contain a *bona fide* p53 RE is a strong argument suggesting that this site is an important feature in the regulation of p53 activity via specific DNA binding. This is the first time that p53 RE has been reported in nonmammalian species. Experimental evidence is necessary to demonstrate the biological activity of these sequences, but several lines of experimental evidence already suggest that mammalian and *X. laevis* p53 are interchangeable and behave in similar ways (Hardy-Bessard *et al*., 1998; Bensaad *et al*., 2001).

The absence of p53 target genes involved in the regulation of cell growth arrest such as p21 is puzzling but not unexpected as it was not found in *X. laevis* or other fishes. The finding that DNA damage can induce growth arrest in frog cells suggests that either another CDKI could be the target of p53 or that an alternative mechanism is involved (Bensaad *et al*., 2001). This observation together with the fact that the fly and the *C. elegans* ancestor of the p53 family are devoid of any growth arrest activity despite apoptotic activity suggest that the cell cycle arrest function of p53 is recent. In fly and *C. elegans*, only one member of the p53 family appears to be devoid of any growth arrest function despite apoptotic activity. Taken together, all these data suggest that growth arrest regulation is a recent property of the p53 protein. DNA array analysis has revealed thousands of genes regulated by p53 (Kannan *et al*., 2001; Kostic and Shaw, 2000; Polyak *et al*., 1996; Tokino and Nakamura, 2000; Yu *et al*., 1999; Zhao *et al*., 2000). The present analysis could be the first step of a more exhaustive study, in which systematic analysis of every gene that has been shown to be induced by p53 could be analysed with systematic comparison with the human (and mouse) genome. The systematic presence for a conserved p53 RE could be a very strong indicator to validate this analysis, as it would reinforce the concept that these genes are directly regulated by p53.

## Methods

Fugu genome searches were performed using Blast and two Fugu genome databases (Aparicio *et al*., 2002). The first database is located at the Sanger Institute (http://www.ensembl.org/Fugu_rubripes/) release 8.1.1 (07/2002) and the second is located at the Institute for Systems Biology (http://genome.jgi-psf.org/fugu6/fugu6.home.html) release 3.0 (08/2002). Protein alignment was determined with ClustaIX software. For gene analysis, splice junctions were screened using NetGene 2 (http://www.cbs.dtu.dk/services/NetGene2/) (Hebsgaard *et al*., 1996).

## Acknowledgements

MLB is supported by a fellowship from the Société Française du Cancer, Association pour la Recherche sur le Cancer and Fondation pour la recherche médicale, KB by a fellowship from the Ligue Nationale contre le Cancer (Comité National). Work in the laboratory of TS is supported by the Association pour la Recherche sur le Cancer, Ligue Nationale contre le Cancer (comité de Paris) and Ministère pour l'enseignement supérieur et la Recherche.

## Note added in proof

Since the submission of this manuscript, Bergamashi *et al*. (2003) have described the discovery of iASPP, a shorter version of ASPP, in the human genome. iASPP acts as a repressor of p53 apoptosis. Sequence comparison indicates that it corresponds to the F-sASPP described in the present manuscript. The conservation of this pathway suggests that control of p53 apoptosis has been strongly conserved through evolution.

The supplementary figures can be downloaded from URL: http://www.nature.com/ONC and figures can be downloaded from the author's web site http://p53.curie.fr/.

## References

Alarcon-Vargas D and Ronai Z. (2002). *Carcinogenesis*, **23,** 541–547.

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D and Brenner S. (2002). *Science*, **297,** 1301–1310.

Bensaad K, Rouillard D and Soussi T. (2001). *Oncogene*, **20,** 3766–3775.

Bergamaschi D, Samuels Y, O'Neil NJ, Trigiante G, Crook T, Hsieh JK, O'Connor DJ, Zhong S, Campargue I, Tomlinson ML, Kuwabara PE and Lu X. (2003). *Nat. Genet.*, **33**, 162–167.

Brodsky MH, Nordstrom W, Tsang G, Kwan E, Rubin GM and Abrams JM. (2000). *Cell*, **101,** 103–113.

Cachot J, Galgani F and Vincent F. (1998). *Comp. Biochem. Physiol. [C]*, **120,** 351–356.

Caron de Fromentel C, Pakdel F, Chapus A, Baney C, May P and Soussi T. (1992). *Gene*, **112,** 241–245.

Contente A, Dittmer A, Koch MC, Roth J and Dobbelstein M. (2002). *Nat Genet*, **30,** 315–320.

Daujat S, Neel H and Piette J. (2001). *Trends Genet*., **17,** 459–464.

Derry WB, Putzke AP and Rothman JH. (2001). *Science*, **294,** 591–595.

El-Deiry WS, Kern SE, Pientenpol JA, Kinzler KW and Vogelstein B. (1992). *Nat. Genet*., **1,** 45–49.

Flores ER, Tsai KY, Crowley D, Sengupta S, Yang A, McKeon F and Jacks T. (2002). *Nature*, **416,** 560–564.

Gilley J and Fried M. (2001). *Oncogene*, **20,** 7447–7452.

Goodman RH and Smolik S. (2000). *Genes Dev*., **14,** 1553–1577.

Gorina S and Pavletich NP. (1996). *Science*, **274,** 1001–1005.

Grob TJ, Novak U, Maisse C, Barcaroli D, Luthi AU, Pirnia F, Hugh B, Graber HU, De Laurenzi V, Fey MF, Melino G and Tobler A. (2001). *Cell Death Differ*., **8,** 1213–1223.

Hardy-Bessard AC, Garay E, Lacronique V, Legros Y, Demarquay C, Houque A, Portefaix JM, Granier C and Soussi T. (1998). *Oncogene*, **16,** 883–890.

Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P and Brunak S. (1996). *Nucleic Acids Res.*, **24,** 3439–3452.

Iwabuchi K, Bartel PL, Li B, Marraccino R and Fields S. (1994). *Proc. Natl. Acad. Sci. USA*, **91,** 6098–6102.

Jin S, Martinek S, Joo WS, Wortman JR, Mirkovic N, Sali A, Yandell MD, Pavletich NP, Young MW and Levine AJ. (2000). *Proc. Natl. Acad. Sci. USA*, **97,** 7301–7306.

Kannan K, Amariglio N, Rechavi G, Jakob-Hirsch J, Kela I, Kaminski N, Getz G, Domany E and Givol D. (2001). *Oncogene*, **20,** 2225–2234.

Kartasheva NN, Contente A, Lenz-Stoppler C, Roth J and Dobbelstein M. (2002). *Oncogene*, **21,** 4715–4727.

Ko LJ and Prives C. (1996). *Gene Develop*., **10,** 1054–1072.

Kostic C and Shaw PH. (2000). *Oncogene*, **19,** 3978–3987.

Krause MK, Rhodes LD and Van Beneden RJ. (1997). *Gene*, **189,** 101–106.

Lee Y, Sultana R, Pertea G, Cho J, Karamycheva S, Tsai J, Parvizi B, Cheung F, Antonescu V, White J, Holt I, Liang F and Quackenbush J. (2002). *Genome Res.*, **12,** 493–502.

Levine AJ. (1997). *Cell*, **88,** 323–331.

Lin Y, Ma W and Benchimol S. (2000). *Nat. Genet*., **26,** 122–127.

Marechal V, Elenbaas B, Taneyhill L, Piette J, Mechali M, Nicolas JC, Levine AJ and Moreau J. (1997). *Oncogene*, **14,** 1427–1433.

Mayo L and Donner D. (2002). *Trends Biochem. Sci*., **27,** 462.

Melino G, De Laurenzi V and Vousden KH. (2002). *Nat. Rev. Cancer*, **2,** 605–615.

Momand J, Wu HH and Dasgupta G. (2000). *Gene*, **242,** 15–29.

Nakagawa T, Takahashi M, Ozaki T, Watanabe Ki K, Todo S, Mizuguchi H, Hayakawa T and Nakagawara A. (2002). *Mol. Cell. Biol*., **22,** 2575–2585.

Okamoto K and Beach D. (1994). *EMBO J*., **13,** 4816–4822.

Okamoto K, Li H, Jensen MR, Zhang T, Taya Y, Thorgeirsson SS and Prives C. (2002). *Mol. Cell*, **9,** 761–771.

Okamura S, Arakawa H, Tanaka T, Nakanishi H, Ng CC, Taya Y, Monden M and Nakamura Y. (2001). *Mol. Cell*, **8,** 85–94.

Ollmann M, Young LM, Di Como CJ, Karim F, Belvin M, Robertson S, Whittaker K, Demsky M, Fisher WW,

Buchman A, Duyk G, Friedman L, Prives C and Kopczynski C. (2000). *Cell*, **101,** 91–101.
Parant J, Chavez-Reyes A, Little NA, Yan W, Reinke V, Jochemsen AG and Lozano G. (2001). *Nat. Genet*., **29,** 92–95.
Piette J, Neel H and Marechal V. (1997). *Oncogene*, **15,** 1001–1010.
Polyak K, Waldman T, He TC, Kinzler KW and Vogelstein B. (1996). *Gene Develop*., **10,** 1945–1952.
Samuels-Lev Y, O'Connor DJ, Bergamaschi D, Trigiante G, Hsieh JK, Zhong S, Campargue I, Naumovski L, Crook T and Lu X. (2001). *Mol. Cell*, **8,** 781–794.
Schumacher B, Hofmann K, Boulton S and Gartner A. (2001). *Curr. Biol*., **11,** 1722–1727.
Sherr CJ and Weber JD. (2000). *Curr. Opin. Genet. Dev*., **10,** 94–99.
Shikama N, Lee CW, France S, Delavaine L, Lyon J, KrsticDemonacos M and LaThangue NB. (1999). *Mol. Cell*, **4,** 365–376.
Shvarts A, Steegenga WT, Riteco N, vanLaar T, Dekker P, Bazuine M, vanHam RCA, vanOordt WV, Hateboer G, vanderEb AJ and Jochemsen AG. (1996). *EMBO J*., **15,** 5349–5357.
Stambolic V, MacPherson D, Sas D, Lin Y, Snow B, Jang Y, Benchimol S and Mak TW. (2001). *Mol. Cell*, **8,** 317–325.
Steinmetz LM, Scharfe C, Deutschbauer AM, Mokranjac D, Herman ZS, Jones T, Chu AM, Giaever G, Prokisch H, Oefner PJ and Davis RW. (2002). *Nat. Genet*., **31,** 400–404.
Tanaka H, Arakawa H, Yamaguchi T, Shiraishi K, Fukuda S, Matsui K, Takei Y and Nakamura Y. (2000). *Nature*, **404,** 42–49.
Thisse C, Neel H, Thisse B, Daujat S and Piette J. (2000). *Differentiation*, **66,** 61–70.
Thompson JD, Higgins DG and Gibson TJ. (1994). *Nucleic Acids Res*., **22,** 4673–4680.
Tokino T and Nakamura Y. (2000). *Crit. Rev. Oncol. Hematol*., **33,** 1–6.
Vogelstein B, Lane D and Levine AJ. (2000). *Nature*, **408,** 307–310.
Vousden KH and Lu X. (2002). *Nat. Rev. Cancer*, **2,** 594–604.
Wahl GM and Carr AM. (2001). *Nat. Cell. Biol*., **3,** E277–E286.
Warbrick E. (1998). *BioEssays*, **20,** 195–199.
Woods DB and Vousden KH. (2001). *Exp. Cell Res*., **264,** 56–66.
Yang A, Kaghad M, Caput D and McKeon F. (2002). *Trends Genet*., **18,** 90–95.
Yang JP, Hori M, Sanda T and Okamoto T. (1999). *J. Biol. Chem*., **274,** 15662–15670.
Yu J, Zhang L, Hwang PM, Rago C, Kinzler KW and Vogelstein B. (1999). *Proc. Natl. Acad. Sci. USA*, **96,** 14517–14522.
Yu WP, Pallen CJ, Tay A, Jirik FR, Brenner S, Tan YH and Venkatesh B. (2001). *Oncogene*, **20,** 5554–5561.
Zhao R, Gish K, Murphy M, Yin Y, Notterman D, Hoffman WH, Tom E, Mack DH and Levine AJ. (2000). *Genes Dev*., **14,** 981–993.