Oncogene (2003) 22, 5325–5337
© 2003 Nature Publishing Group All rights reserved 0950-9232/03 $25.00
www.nature.com/onc

Regulation of constitutive expression of mouse PTEN by the 5′-untranslated region

Baoguang Han[1], Zizheng Dong[1], Yang Liu[1], Qun Chen[1], Katsuyuki Hashimoto[2] and Jian-Ting Zhang*,[1]

[1]*Department of Pharmacology and Toxicology, I.U. Cancer Center and Walther Oncology Center/Walther Cancer Institute, Indiana University School of Medicine, Indianapolis, IN 46202, USA;* [2]*National Institute of Infectious Diseases, Tokyo, Japan*

PTEN tumor suppressor serves as a major negative regulator of survival signaling mediated by PI3 kinase/AKT/protein kinase B pathway, and is inactivated in various human tumors. Elucidation of mechanisms responsible for PTEN expression is important for providing insight into strategies to control the loss of PTEN expression in human cancers. Although recent studies suggested that p53 and Egr-1 can modulate induced PTEN expression, the mechanism responsible for ubiquitous constitutive expression of PTEN remains elusive. PTEN mRNA contains a highly conserved and GC-rich 5′-untranslated region (5′-UTR). Recently, it has been shown that the long 5′-UTR sequences of several growth-regulated mRNAs contain promoters that can generate mRNAs with shorter 5′-UTRs. In this paper, we tested whether the 5′-UTR sequence of mouse PTEN contains a promoter that is responsible for constitutive expression of PTEN. We found that the long 5′-UTR sequence of mouse PTEN severely inhibits translation of PTEN and a heterologous gene firefly luciferase. Deletion of the most 5′-UTR sequence would enhance translation efficiency 100-fold. We also showed that the 5′-UTR sequence of mouse PTEN does not have an internal ribosome entry site (IRES) that can mediate cap-independent initiation of translation. Instead, we found that the 5′-UTR sequence of mouse PTEN contains a strong promoter that drives the production of a transcript with shorter 5′-UTRs, which can be translated with higher efficiency. This promoter was mapped to the region between −551 and −220 bases upstream of the translation start codon. Cotransfection analysis using *Drosophila* SL2 cells showed that Sp1 is one of the major transcription factors that can constitutively activate this promoter. Two endogenous PTEN transcripts with 5′-UTRs of 193 and 109 bases were found in DU145 and H226 cell lines. Based on these observations, we conclude that the PTEN expression may be regulated at both transcriptional and translational levels, and that the 5′-UTR sequence of PTEN contains a promoter that is responsible for constitutive PTEN expression.

Oncogene (2003) **22**, 5325–5337. doi:10.1038/sj.onc.1206783

Keywords: PTEN; IRES; promoter; 5′-UTR

Introduction

PTEN/MMAC1/TEP1 (referred as PTEN in the remaining text of this paper) is a tumor suppressor gene that maps to the 10q23.3 and has been shown to be deleted or mutated in many human tumors including glioblastomas, endometrial neoplasms, hematological malignancies, and prostate and breast cancers (Ali *et al.*, 1999; Cantley and Neel, 1999; Dahia, 2000; Di Cristofano and Pandolfi, 2000). In addition, germline mutations in PTEN cause Cowden syndrome (CS), characterized by multiple hamartomas and a high proclivity for developing cancer (Cantley and Neel, 1999). Although PTEN sequence is highly homologous with dual-specificity protein phosphatase, its major substrate is phosphatidylinositol triphosphate (PIP_3). It has been suggested that PTEN is a major negative regulator of PIP_3 and negatively regulates the survival signaling mediated by the PI3 kinase/AKT/protein kinase B pathway (Maehama and Dixon, 1999; Leslie and Downes, 2002). Loss of PTEN function or expression results in an increased concentration of PIP_3 and AKT hyperactivation, which leads to the protection of cells from various apoptotic stimuli (Stambolic *et al.*, 1998). In contrast, overproduction of PTEN induces growth suppression via cell cycle arrest and/or induction of apoptosis, and inhibits cell adhesion and migration (Dahia, 2000). Recently, PTEN has been demonstrated to serve as a negative regulator for proliferation of human neural stem cells and mammary epithelial cells (Backman *et al.*, 2001; Groszer *et al.*, 2001; Kwon *et al.*, 2001; Li *et al.*, 2002).

With the critical role in antagonizing PI3 kinase pathways, PTEN is anticipated to be the target of complex control mechanisms (Leslie and Downes, 2002). However, very little is known about the regulation of PTEN expression. PTEN expression is ubiquitous and constitutive, while it can also be altered by

*Correspondence: Jian-Ting Zhang; Department of Pharmacology and Toxicology, IUCC, Indiana University School of Medicine, 1044 W. Walnut Street, R4-166, Indianapolis, IN 46202, USA;
E-mail: jianzhan@iupui.edu
Received 15 November 2002; revised 13 May 2003; accepted 13 May 2003

various biological stimuli. It was reported that PTEN expression in human keratinocytes decreased when the cells were treated with TGF-β (Li and Sun, 1997) and PTEN expression increased dramatically when myeloid leukemia cells were induced to differentiate into either granulocytic or monocytic cells (Hisatake *et al.*, 2001). PTEN expression is also induced during neuronal differentiation (Ross *et al.*, 2001). It has been reported that PTEN expression can be induced by p53 and Egr-1, and functional p53 and Egr-1 binding sites were identified between bases −1190 and −1157, and between −947 and −939 upstream of the translation start codon (Stambolic *et al.*, 2001; Virolle *et al.*, 2001), respectively. This observation suggests that p53 and Egr-1 may be able to modulate the PI3 kinase/AKT/Protein kinase B pathway via regulating PTEN expression (Stambolic *et al.*, 2001). However, mechanisms responsible for constitutive expression of PTEN remain to be identified.

Recently, we proposed that promoters may exist in the long 5′-UTR sequences of genes related to cell growth (Han and Zhang, 2002). In this study, we attempted to address this hypothesis by studying the regulation mechanism of PTEN expression. Examination of the 5′ untranslated region (5′-UTR) of PTEN messenger RNA (mRNA) revealed several interesting features related to translational regulation. First, the 5′-UTR of PTEN mRNA is about 1 kb inferred from cDNA sequences of both mouse (GeneBank Accession number NM_008960) and human (GeneBank Accession number NM_000314) PTEN and confirmed by reverse transcription–polymerase chain reaction (RT–PCR) (Stambolic *et al.*, 2001). This length of 5′-UTR is significantly longer than the average length of 100–300 bases for eucaryotic 5′-UTRs. Second, the 5′-UTR of both mouse and human PTEN mRNA contains four upstream translation start codons. Third, the 5′-UTR of PTEN mRNA is G/C-rich with a predicted complex secondary structure. Lastly, the 5′-UTRs of both human and mouse PTEN mRNA are well conserved with an overall homology of >90%. These features of PTEN mRNA strongly suggest that the 5′-UTR of PTEN may inhibit efficient translation initiation by the conventional cap-dependent ribosome scanning mechanism. It is possible that PTEN mRNAs use other mechanisms such as internal ribosome entry site (IRES) to initiate translation (for reviews see Pestova *et al.* (2001) and Sachs (2000)). Alternatively, the long 5′-UTR of PTEN may contain a promoter or splice site to generate mRNA species with significantly shorter 5′-UTRs (Kozak, 2001; Schneider and Kozak, 2001), which can be translated efficiently (Han and Zhang, 2002).

In this study, we tested these possibilities and found that mRNAs with long 5′-UTRs of PTEN cannot be translated by the cap-dependent mechanism, whereas the mRNAs with shorter 5′-UTRs can. The long 5′-UTR of PTEN does not contain an IRES element to mediate IRES-dependent translation initiation. Instead, we demonstrated that the 5′-UTR of PTEN contains a strong promoter that produces mRNA with shorter 5′-UTRs. Using 5′ RACE and RNase protection assay, we found two endogenous PTEN transcripts with short 5′-UTRs of 109 and 193 bases, respectively, in human cancer cell lines. These mRNA species with shorter 5′-UTRs may be responsible for constitutive production of PTEN proteins. Interestingly, the promoter region in the 5′-UTR overlaps with the hypermethylated region identified in gastric tumor (Kang *et al.*, 2002) and endometrial carcinoma (Salvesen *et al.*, 2001), implicating that hypermethylation of this promoter may be important for the loss of PTEN expression observed in a number of tumors.

Results

The Long 5′-UTR of mouse PTEN mRNA inhibits translation in vitro

The 5′-UTR of mouse PTEN mRNA is 948-nucleotide long and contains 72% G and C residues. Analysis of the 5′-UTR using Zuker's mfold program (version 3.0) (Mathews *et al.*, 1999) showed that this region has potential to form various secondary structures and all of the structures predicted are extremely stable with a typical free energy as high as −432 kcal/mol, which is well beyond the energy shown to inhibit ribosome scanning. In addition, this region is also punctuated with four translation start codon AUGs that encode two short upstream open reading frames (uORFs) with 36 and 45 amino acids, respectively (two of the AUGs are located at the end of the first uORF in the same frame). Both uORFs end before −551 bases upstream of the physiological translation start codon. Generally, uAUGs or uORFs inhibit translation of the main ORF by interfering with the scanning process of the 40S ribosome complex. To analyse the effect of the 5′-UTR sequence of mouse PTEN on translation, we utilized an *in vitro* transcription and translation system. Mouse PTEN cDNA with (pG-WPTEN) and without (pG-DPTEN) the 5′-UTR sequence (Figure 1a) were cloned into pGEM-4Z under the control of an SP6 promoter. Run-off transcripts were generated by SP6 RNA polymerase, purified and quantified (Figure 1b), and an equal amount of RNAs were used to program cell-free translation in rabbit reticulocyte lysate (RRL). As shown in Figure 1c, no full-length PTEN protein was translated from transcripts containing the long 5′-UTR (WPTEN) (lane 1), whereas a significant amount of PTEN protein (DPTEN) was generated in the absence of the 5′-UTR sequence (lane 2). Clearly, the presence of the long 5′-UTR sequence drastically inhibits the translation of PTEN in RRL.

To analyse whether the protein-encoding sequence (ORF) influences the translational inhibition effect of the 5′-UTR, we cloned the 5′-UTR sequence of PTEN immediately upstream of an open reading frame encoding firefly luciferase (LUC) in a pGEM-4Z-based reporter vector, pG-LUC, resulting in the plasmid pG-5′LUC (Figure 1a). This construct was then used to program *in vitro* transcription (Figure 1b) and cell-free translation (Figure 1d) as described above. Again, the translation programmed by the pG-LUC transcript

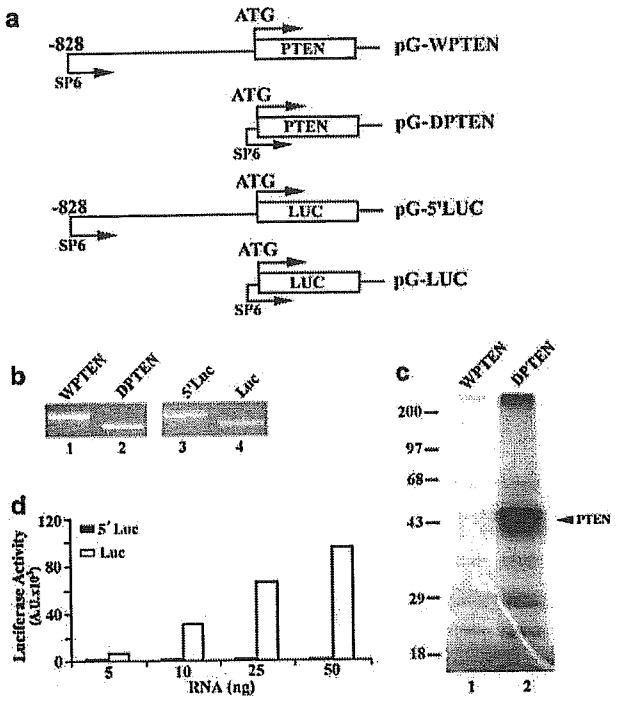

Figure 1 5′-UTR of PTEN inhibits translation of its downstream sequence in RRL. (**a**) Schematic diagram of *in vitro* transcription constructs. The transcription start site using SP6 RNA polymerase and the translation start codon AUG are indicated by arrows. (**b**) *In vitro* transcription. The capped RNA transcripts were generated by *in vitro* transcription using linearized plasmids and SP6 RNA polymerase, and 500 ng of each RNA transcript was separated using 1% agarose gel. (**c**) *In vitro* translation of mouse PTEN. The *in vitro* transcripts of mouse PTEN with (WPTEN) or without (DPTEN) the 5′-UTR sequence were used to program translation in RRL in the presence of [^{35}S]methionine. Protein product was separated by SDS–PAGE and visualized by autoradiography. (**d**) *In vitro* translation of firefly luciferase. 5, 10, 25, and 50 ng of *in vitro* transcripts of firefly luciferase with (5′Luc, solid bars) or without (Luc, open bars) the 5′-UTR sequence of mouse PTEN were used to program translation in RRL, followed by measuring luciferase activity by enzymatic assays

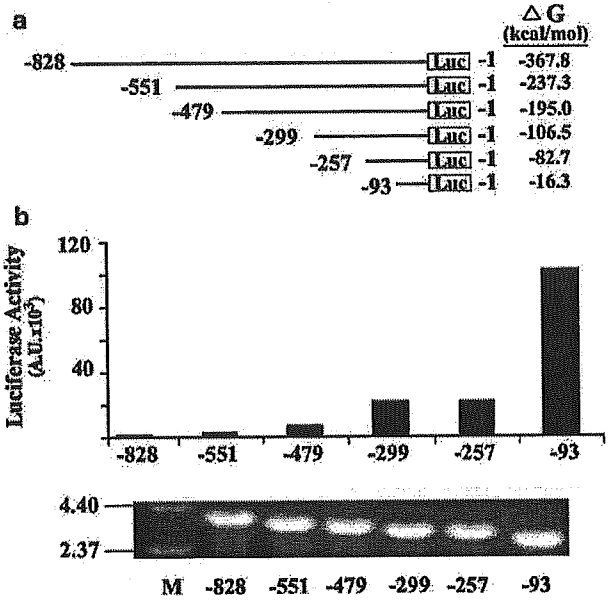

Figure 2 Effect of deletions of the 5′-UTR sequence of mouse PTEN on the translation of firefly luciferase in RRL. (**a**) Schematic diagram of sequential deletions in the 5′-UTR sequence of mouse PTEN. The free energy for the predicted secondary structure of the 5′-UTR sequence is shown on the right. The positions of the 5′ end of each deletion are indicated on the left. These mutant 5′-UTR sequences were engineered into pG-LUC for *in vitro* expression. (**b**) *In vitro* transcription and translation. The deletion mutants were used to generate *in vitro* transcripts (bottom), which were then used to program cell-free translation in RRL. LUC activities of the translated products were measured by enzymatic assays as described in Materials and methods

generated a significantly increasing amount of firefly LUC activity proportional to the amount of transcripts used, while no LUC activity was detected from the translation programmed by pG-5′LUC transcripts. Thus, the 5′-UTR sequence of mouse PTEN inhibits translation of transcripts encoding a heterologous protein and is independent of the protein-encoding sequence.

To further analyse the translational inhibition effect of the 5′-UTR sequence of PTEN, we generated several sequential deletion mutants from the 5′ end of the 5′-UTR in pG-5′LUC constructs (Figure 2a). These deletion mutants contain putative secondary structures with different free energies from −367 to −16 kcal/mol predicted by Zuker's mfold program (Mathews *et al.*, 1999). As shown in Figure 2b, translation of firefly LUC was very inefficient for transcripts with 5′-UTRs that have 479 nucleotides or longer. Further deletion resulted in an increase in translation. However, significant translation occurred only to the transcript with 93 bases of the 5′-UTR sequence with a predicted energy of −16 kcal/mol. It is ~100-fold more translatable than the full-length transcript. These results are consistent with a model that the stable secondary structure in the 5′-UTR sequence of PTEN mRNA inhibits the ribosome scanning process for cap-dependent translation initiation.

5′-UTR sequence of PTEN enhances expression of the second cistron in dicistronic test

The above results clearly demonstrate that the 5′-UTR of PTEN mRNA can effectively suppress ribosome scanning *in vitro*. However, PTEN protein has been reported to be constitutively expressed and its expression can be induced to a high level by biological stimulation. Thus, if PTEN mRNA with the full-length 5′-UTR is translated, it may use an alternative mechanism such as IRES-mediated translation initiation. To test this hypothesis, the 5′-UTR (−828 to −1 bases) of PTEN was cloned into the intergenic region of a dicistronic vector pRF (Stoneley *et al.*, 2000) to obtain pR-PTEN-F (Figure 3a). The internal ribosome entry site (IRES) sequence of human rhinovirus (HRV) was engineered in the same way and was used as a positive control. The pRF-based constructs contain an SV40 promoter to direct the transcription of dicistronic RNA encoding *Renilla* LUC as the first cistron and firefly LUC as the second cistron. Translation of the first

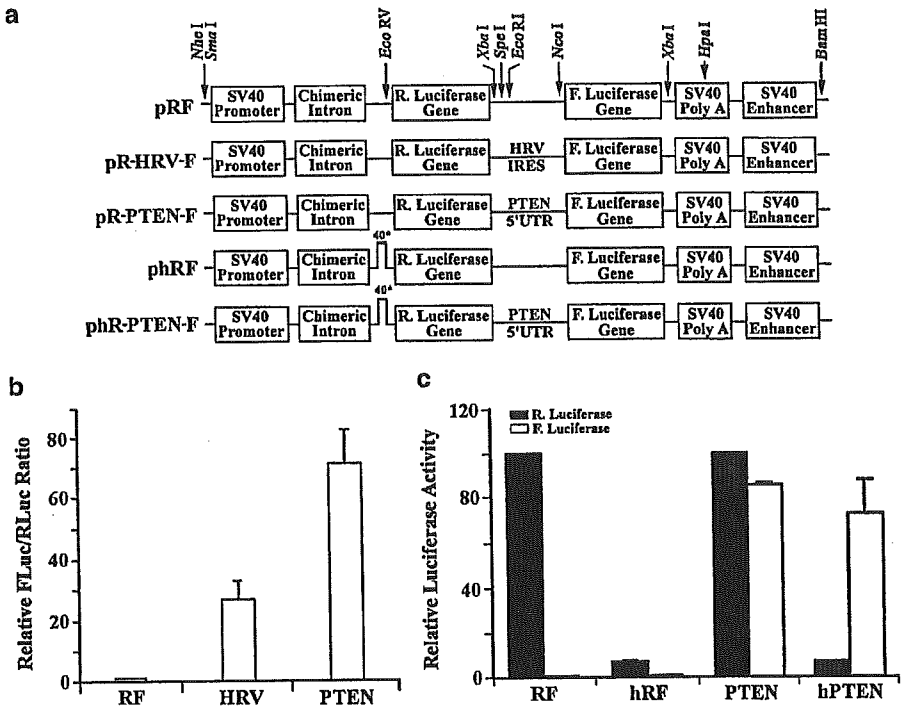

Figure 3 Dicistronic DNA test of the 5′-UTR sequence of mouse PTEN. (**a**) Schematic diagram of dicistronic constructs. The 5′-UTR sequence of mouse PTEN and the IRES element of HRV are cloned into the intergenic region. The inverted U-shaped structure in phRF and phR-PTEN-F represents the inserted hairpin loop with a free energy of −40 kcal/mol. The locations of several relevant restriction enzyme sites are shown by arrows. (**b**) Relative LUC activities generated by the dicistronic constructs in HeLa cells. HeLa cells were transfected with pRF, pR-HRV-F, and pR-PTEN-F constructs. At 24 h following transfection, cells were harvested and the *Renilla* and firefly LUC activities were measured and the relative ratios were calculated and normalized to that of the vector-transfected cells (RF). (**c**) Effect of hairpin structure on the expression of LUCs. HeLa cells were transfected with phRF and phR-PTEN-F, which has a hairpin structure immediately upstream of the ATG codon of the *Renilla* LUC gene. At 24 h following transfection, cells were harvested and *Renilla* and firefly LUC activities were measured in comparison with that generated by pRF and pR-PTEN-F, respectively. Both *Renilla* (solid bars) and firefly LUC (open bars) activities were normalized to pRF control. All the data were from three independent experiments

cistron (*Renilla* LUC) serves as an indicator of cap-dependent translation, while translation of the second cistron (firefly LUC) reflects the IRES activity-associated with the inserted intergenic sequence. This approach is considered as a 'gold standard' for characterizing cellular IRES (Sachs, 2000). These dicistronic constructs were transfected into HeLa cells and both *Renilla* and firefly LUC activities were measured. As shown in Figure 3b, the 5′-UTR sequence of PTEN stimulated the expression of firefly LUC by ~70-fold over negative vector control and about 2–3-fold over positive pR-HRV-F control. Similar results were also obtained using another cell line H1299 (data not shown). These results suggest that the 5′-UTR sequence of PTEN may contain an IRES element or it may enhance read-through or jumping from the first cistron.

To determine whether the effect of the 5′-UTR sequence of PTEN on the translation of the second cistron is due to enhanced read-through or jumping from the first cistron, we inserted a synthetic hairpin in front of the first cistron and created phRF and phR-PTEN-F (Figure 3a). The hairpin structure has a free energy of −40 kcal/mol and is expected to significantly inhibit the cap-dependent translation initiation of the first cistron. Therefore, it will decrease potential ribosome read-through or jumping but will not affect the translation from IRES-mediated initiation of the second cistron. As shown in Figure 3c, the insertion of the hairpin resulted in ~90% decrease of the *Renilla* LUC activity (first cistron) in both phRF and phR-PTEN-F as compared with pRF and pR-PTEN-F, respectively, suggesting that the hairpin structure inhibits the cap-dependent translation. However, the level of firefly LUC activity (second cistron) produced by phR-PTEN-F was not significantly affected when compared with pR-PTEN-F. This result suggests that the enhanced expression of the second cistron by the 5′-UTR sequence of PTEN was independent of the cap-dependent translation initiation of the first cistron. Thus, it is not due to enhanced read-through or jumping from the first cistron.

To define the boundaries in the 5′-UTR sequence of PTEN for enhancing expression of the second cistron, a series of deletion mutants of the 5′-UTR of PTEN were engineered for a dicistronic assay in HeLa cells (Figure 4a). As shown in Figure 4b, deletion from both ends of the 5′-UTR resulted in a gradual decrease in stimulating firefly LUC expression. A construct pR-hPTEN-F that

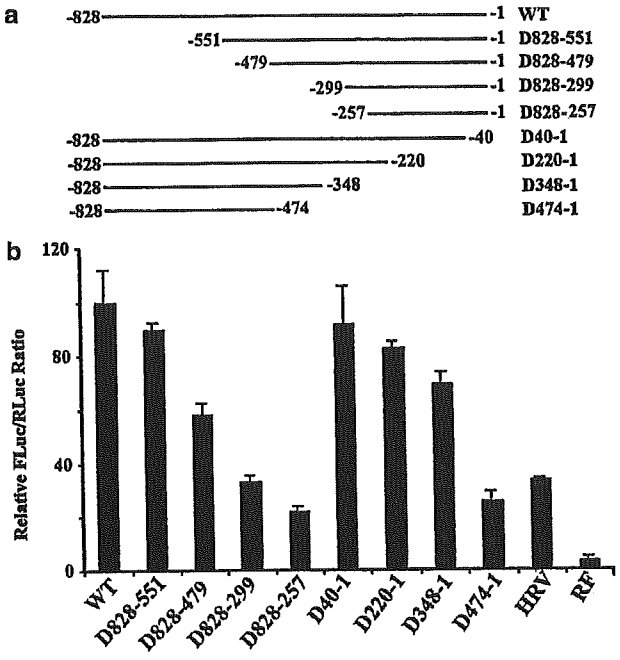

Figure 4 Deletion mapping of the 5′-UTR sequence of mouse PTEN in dicistronic DNA test. (**a**) Schematic diagram of sequential deletions of the 5′-UTR sequence of mouse PTEN. The positions of the 5′ and 3′ ends of each deletion are indicated on the left and right, respectively. These mutant 5′-UTR sequences were engineered into the dicistronic vector pRF at the intergenic region. (**b**) Relative luciferase activities generated from constructs containing the wild-type and mutant 5′-UTR sequences of mouse PTEN. HeLa cells were transfected with the constructs shown in (**a**), and 24 h following transfection, *Renilla* and firefly LUC activities were measured and the ratio of firefly to *Renilla* LUC was calculated and normalized to the wild-type (WT) control. The empty vector pRF and the plasmid containing HRV IRES were also used as negative and positive controls, respectively. The data were from three independent experiments

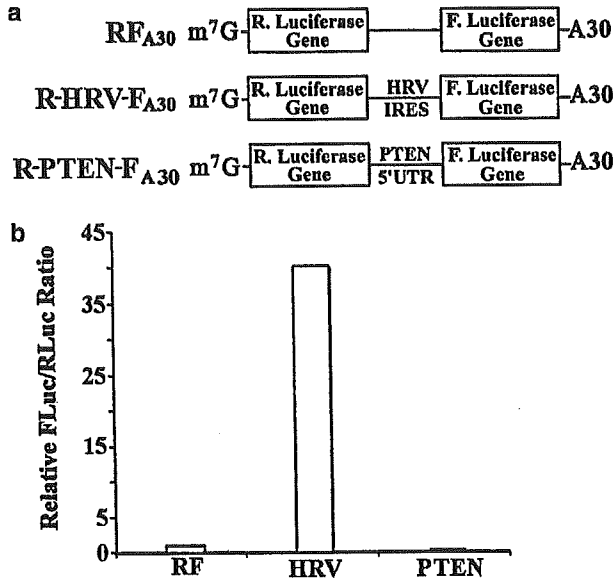

Figure 5 Translation of dicistronic mRNA in HeLa cells. (**a**) Schematic diagram of the dicistronic mRNA used for translation in HeLa cells. *In vitro* transcripts with 5′ cap (m^7G) and 3′ poly A tail (A_{30}) were synthesized using T7 RNA polymerase from linearized vector (RF_{A30}) and constructs containing the IRES of HRV (R-HRV-F_{A30}) and the 5′-UTR of PTEN (−551 to −1) (R-PTEN-F_{A30}). (**b**) Relative LUC activity from dicistronic mRNAs in HeLa cells. HeLa cells were transfected with the dicistronic mRNAs, and 8 h following transfection, *Renilla* and firefly LUC activities were measured and the relative ratios were calculated and normalized to that of the vector-transfected cells (RF_{A30})

encodes the −492 to −1 region of the 5′-UTR sequence of human PTEN enhanced the expression of the second cistron by 50-fold (data not shown), equivalent to that of the D828-479 construct of mouse PTEN (Figure 4b). Similar results with these constructs have also been observed by using another cell line H1299 (data not shown). Since the deletion mutant D828-551 and D220-1 retains the most enhancing activity, the 230-nucleotide central region of the 5′-UTR (−551 to −220) may contain important elements for stimulating firefly LUC expression. It is noteworthy that deletion mutants that retain about 200 nucleotides of the 5′-UTR sequence at the 3′-end (D828-257) or at the 5′-end (D474-1) still have the enhancing activity comparable to HRV IRES, suggesting that the enhancing activity also exists at both 5′- and 3′-ends of the 5′-UTR sequence of PTEN.

5′-UTR of PTEN does not display an IRES activity in dicistronic mRNA assay

The above results suggest that the 5′-UTR of PTEN may (1) contain an IRES activity that enhances the translation of firefly LUC from the dicistronic mRNA by internal initiation (for reviews see Pestova *et al.* (2001) and Sachs (2000)), (2) contain a promoter that directs transcription of the firefly LUC gene (Han and Zhang, 2002), and/or (3) contain a splicing acceptor site, that creates a splicing variant with only the second cistron of the firefly LUC gene. To distinguish between these possibilities, we generated dicistronic RNAs *in vitro* from the dicistronic constructs and used them to program translation both in HeLa cells and in RRL. RNA transfection allows a direct analysis of whether the 5′-UTR sequence of PTEN in the intergenic region of a dicistronic mRNA can enhance the translation of the second cistron without transcriptional interference. For purposes of RNA transfection, plasmids pSP-RF_{A30}, pSP-R-PTEN-F_{A30}, and pSP-R-HRV-F_{A30} were engineered and used for producing dicistronic transcripts containing m^7GpppG cap and polyadenylated tail *in vitro* (Figure 5a). The *in vitro* transcripts were introduced into HeLa cells by lipofectin encapsulation. At 8 h following transfection, cell lysates were prepared for LUC activity measurement. As expected, the firefly LUC of vector RNA was very poorly translated and its activity (arbitrary units) only represented about 0.16% of *Renilla* LUC (data not shown). It increased to about 6.7% with the dicistronic RNAs containing HRV IRES. Therefore, HRV IRES significantly stimulated the translation of firefly LUC about 40-fold over vector control (Figure 5b). However, no stimulation of firefly LUC expression was observed with the 5′-UTR of

PTEN. The expression of the second cistron firefly LUC in the presence of the 5′-UTR of PTEN was about 0.003% of *Renilla* LUC activity, significantly less than that observed with vector control. Again, this observation confirms the conclusion that the presence of the 5′-UTR sequence of PTEN significantly prevents ribosome scanning initiated from the 5′ end. Similar results were observed using the dicistronic RNA transcripts to program cell-free translation in RRL supplemented with HeLa extract (data not shown). Thus, it appears that the 5′-UTR sequence of PTEN does not contain an IRES element to mediate internal ribosome entry.

5′-UTR sequence of PTEN contains a ubiquitously functional promoter

The above results prompted us to explore whether the 5′-UTR sequence of PTEN contains a promoter. For this purpose, we used a promoterless dicistronic assay as described previously (Han and Zhang, 2002) by simply removing the unique SV40 promoter together with the intron sequence from the pRF-based dicistronic constructs. These promoterless dicistronic constructs (Figure 6a) were then transfected into HeLa cells for determination of both *Renilla* and firefly LUC activities (Figure 6b). As we have shown previously (Han and Zhang, 2002), both the *Renilla* and firefly LUC activities were minimal, but detectable for pRF(-P) vector control. Only a twofold increase in firefly LUC activity was observed with pR-HRV-F(-P) construct. This small increase was in dramatic contrast with the 30-fold increase associated with the pR-HRV-F construct (Figure 3b). Thus, the enhanced expression of firefly LUC from pR-HRV-F construct was not due to production of monocistronic transcripts by a promoter present in the HRV sequence. However, the pR-PTEN-F(-P) construct generated more than 160-fold firefly LUC activity over vector control. Thus, the significant stimulation of the firefly LUC expression by the 5′-UTR sequence of PTEN was likely due to the presence of a strong promoter in this region.

The existence of a constitutively active promoter in the 5′-UTR sequence of PTEN is unexpected and may help explain the relatively constant and high levels of PTEN mRNA and protein in most primary cells (Stambolic *et al*., 2001). To determine whether this promoter is ubiquitously used, we analysed its activity in various human cell lines including HEK293 (transformed primary embryonic kidney cells), H1299 (large-cell lung carcinoma cells), K562 (erythroleukemia cells), and DU145 (prostate cancer cells). As shown in Figure 6c, the 5′-UTR sequence displays significant promoter activities in all the cell lines tested, at a level similar to that observed in HeLa cells. Thus, this promoter in the 5′-UTR of PTEN is ubiquitously and constitutively active.

To determine directly whether the transcript derived from the promoter in the 5′-UTR sequence of PTEN exists in cells, poly(A) mRNAs were isolated for Northern blot analysis 24 h following transfection of pRF, pR-PTEN-F, pR-PTEN-F(-P), and pRF(-R). The pRF(-R) construct, which lacks the *Renilla* LUC gene (Han and Zhang, 2002), was used as a monocistronic control. This vector is expected to produce transcript encoding only firefly LUC using the SV40 promoter. The pRF vector was used as a dicistronic control, which produces only dicistronic mRNA using SV40 promoter. As shown in Figure 7, dicistronic transcript from control pRF (lane 1, indicated by an asterisk) and the monocistronic transcript from control pRF(-R) (lane 4, indicated by an arrowhead) were detected as expected. The dicistronic transcript from pR-PTEN-F (lane 2, indicated by an asterisk) is bigger than that from pRF (lane 1), consistent with the presence of the 5′-UTR sequence of PTEN in the intergenic region. The transcript produced from pR-PTEN-F(-P) has a size slightly larger than that from pRF(-R) (compare lanes 3 and 4), suggesting that it is a monocistronic mRNA produced by the promoter in the 5′-UTR of PTEN. The same monocistronic transcript was also observed with pR-PTEN-F (lane 2, indicated by an arrowhead). These observations confirm that the 5′-UTR sequence of PTEN has a strong promoter for transcription of the second cistron firefly LUC. It is noteworthy that the monocistronic RNA bands from both pR-PTEN-F and pR-PTEN-F(-P) on Northern blot (lanes 2 and 3, indicated by an arrowhead) are much wider than that

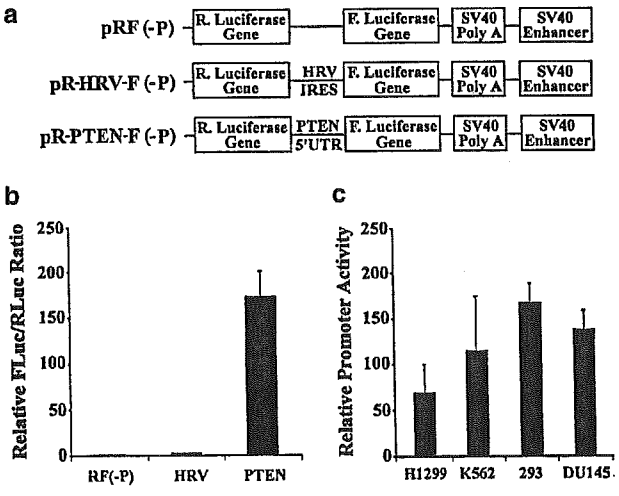

Figure 6 5′-UTR sequence of mouse PTEN contains a ubiquitously active promoter. (**a**) Schematic diagram of promoterless dicistronic construct of pRF(-P), pR-HRV-F(-P), and pR-PTEN-F(-P). The sequences of SV40 promoter and chimeric intron were removed from the parental dicistronic constructs shown in Figure 3. (**b**) Relative LUC activities generated from the promoterless dicistronic constructs. The promoterless constructs were transfected into HeLa cells, and 24 h following transfection, cells were harvested for determination of *Renilla* and firefly luciferase activity. The relative ratios between firefly and *Renilla* LUC activities were calculated and normalized to that of the vector-transfected cells (pRF (-P)). (**c**) Promoter activity of the 5′-UTR sequence in various human cell lines. Human H1299, K562, HEK293, and DU145 cells were transfected with promoterless constructs and cell lysates were prepared 24 h following transfection for LUC measurement as described for HeLa cells in panel b. Relative promoter activity was calculated as the ratio of firefly LUC activity of pR-PTEN-F(-P) to the vector control pRF(-P). The data were from three independent experiments

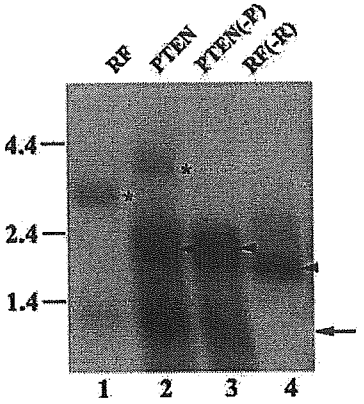

Figure 7 Northern blot analysis of RNA products generated by promoter of mouse PTEN. PolyA mRNAs were isolated following transfection with pRF (lane 1), pR-PTEN-F (lane 2), pR-PTEN-F(-P) (lane 3), and pRF(-R) (lane 4) and used for Northern blot analysis as described in Materials and methods. The asterisks and arrowheads indicate the dicistronic and monocistronic RNA transcripts, respectively. The arrow shows unknown RNAs hybridized with the probe which have been consistently observed in previous studies (Coldwell *et al*., 2000; Han and Zhang, 2002)

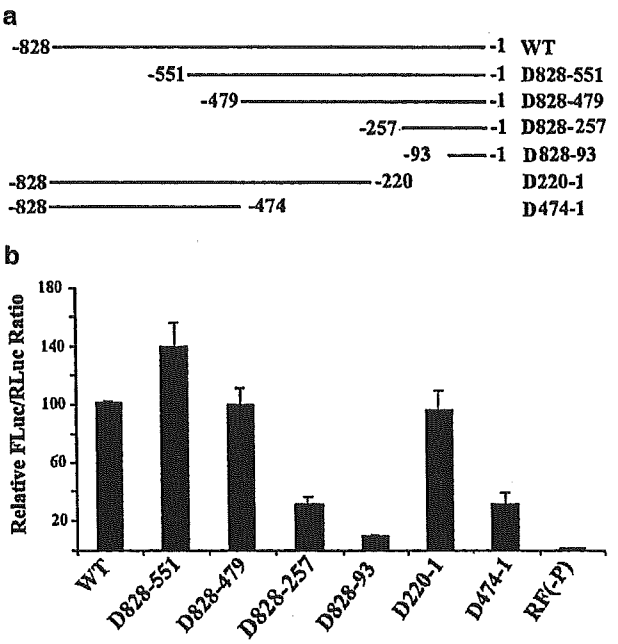

Figure 8 Deletion mapping of the promoter activity in the 5′-UTR sequence of mouse PTEN. (**a**) Schematic diagram of the deletions in the 5′-UTR of mouse PTEN. The positions of 5′ and 3′ ends of each deletion are indicated on the left and right, respectively. These mutant 5′-UTRs were engineered into the promoterless dicistronic vector pRF(-P) at the intergenic region. (**b**) Relative LUC activity generated from the wild-type and mutant 5′-UTR sequences of mouse PTEN. HeLa cells were transfected with the constructs shown in (**a**), and 24 h following transfection, *Renilla* and firefly LUC activities were measured and the ratio of firefly to *Renilla* LUC was calculated and normalized to the wild-type (WT) control. The data were from three independent experiments

from pRF(-R) (lane 4), suggesting that these transcripts may be derived from multiple transcription start sites. Indeed, RT–PCR analysis showed the existence of transcripts starting from a very upstream region of the 5′-UTR (data not shown).

Transactivation of the promoter in the 5′-UTR sequence of PTEN by Sp1

To identify the boundaries of the DNA region that are responsible for promoter activity, deletion mutants were generated from either 5′ or 3′ ends of the 5′-UTR of PTEN (Figure 8a). These deletion mutants were engineered into the promoterless dicistronic vector and then used to determine their ability to direct firefly LUC expression. As shown in Figure 8b, deletion from either end of the 5′-UTR resulted in a gradual decrease in stimulating firefly LUC expression. The full stimulation activity resides in the region of −551 to −220 since deletion mutants D828-551 and D220-1 still contain full-promoter activity. In addition, the deletion mutants containing only the 5′ (D474-1) or 3′ (D828-257) region of the 5′-UTR sequence also have significant promoter activity. Widespread promoter activity within the 5′-UTR sequence of PTEN may be responsible for multiple transcriptional initiation sites observed with Northern blots. These observations are consistent with the results obtained from conventional dicistronic DNA constructs (see Figure 4).

Analysis of the potential transcription factor binding sites in the −551 to −220 region of the 5′-UTR sequence using MatInspector (Quandt *et al*., 1995) showed several consensus sites for transcription factors such as Sp1, Ets-1, Egr-1, Ap-1, Ap-2, and Ap-4 (Figure 9a). To determine directly whether the above transcription factors could functionally modulate promoter activity in the 5′-UTR sequence of PTEN, *Drosophila* SL2 cells, which are deficient in Sp1-, Sp3-, and Ets-related proteins (Courey and Tjian, 1988; Dennig *et al*., 1996), were used. The reason for using insect instead of mammalian cells is that Sp1-, Sp3-, and Ets-related factors are expressed in virtually all mammalian cells, which could affect the interpretation of this experiment. We introduced the pR-PTEN-F(-P) construct along with *Drosophila* expression plasmids pPacSp1 or pPacUEts-1 into *Drosophila* SL2 cells. As shown in Figure 9b, pPacSp1 stimulated the promoter activity in the 5′-UTR sequence of PTEN by 15-fold, whereas it stimulated vector control only by threefold. Ets-1 did not significantly stimulate the promoter activity in the 5′-UTR sequence of PTEN. These results suggest that Sp1 is a possible transcription factor that regulates the constitutive expression of PTEN by acting at the promoter within the 5′-UTR of PTEN. Since there are several Sp1 binding sites in the 5′-UTR sequence, it remains to be determined which Sp1 element is responsible for the promoter activity.

Detection of endogenous transcripts originating from the 5′-UTR promoter in human cancer cell lines

Previous research with human PTEN using Northern blot detected heterogeneous transcripts of a variety of

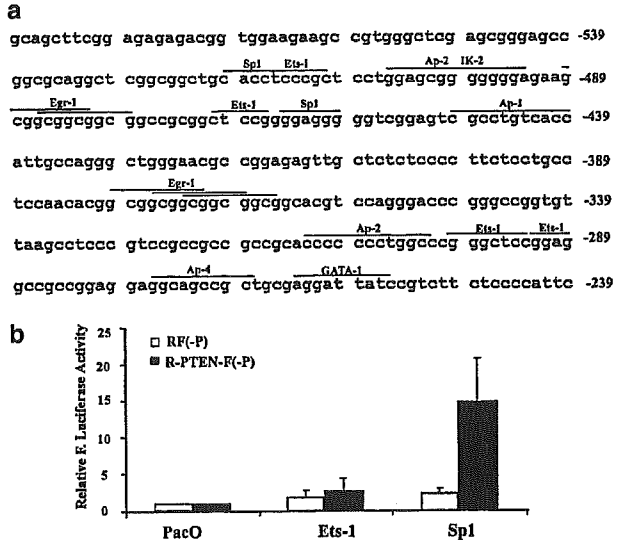

Figure 9 Sp1 protein activates the promoter in the 5′-UTR sequence of mouse PTEN in *Drosophila* SL2 cells. (**a**) Partial DNA sequence of the 5′-UTR sequence of mouse PTEN. The sequence corresponding to −588 to −239 upstream of translation start codon is shown. The consensus sequences for transcription factor-binding sites predicted by MatInspector are indicated by lines above the sequence with the name of the factors. (**b**) Relative LUC activities in SL2 cells. SL2 cells were transfected with pR-PTEN-F(-P) (solid bars) or empty vector pRF(-P) (open bars) together with *Drosophila* expression vector (PacO), or plasmids expressing transcription factors Sp1 and Ets-1. At 48 h following transfection, the cells were harvested and firefly LUC activities were measured and normalized to that of vector-transfected cells. The data were from three independent experiments

sizes between 2 and 5 kb (Whang *et al.*, 1998). This heterogeneity might partly be due to the existence of different promoters, such as the ones discussed above, and different transcription start sites. To test this hypothesis and to determine whether endogenous PTEN transcripts with a shorter 5′-UTR sequence exist, we performed 5′ rapid amplification of cDNA ends (RACE) and an RNase protection assay using RNAs isolated from the human prostate cancer cell line DU145 and human lung cancer cell line H226, which are known to express PTEN (Whang *et al.*, 1998; Soria *et al.*, 2002). As shown in Figure 10a, multiple products with various lengths ranging from about 150 to 1100 bp were produced from 5′ RACE using gene-specific primer 10-bases upstream of the translation start codon ATG (Figure 10c). This observation indicates that PTEN transcripts in DU145 cells have different 5′-UTRs and likely multiple transcription start sites exist, consistent with our findings shown in Figures 7 and 8. The two small products (indicated by asterisks) were then excised from the gel, cloned, and sequenced. The sequencing data (Figure 10a) suggest that the two small transcripts are products from transcription start sites at −109 and −193 bases upstream of the translation start codon, respectively. This observation is consistent with our promoter mapping, which showed that the 5′-UTR promoter resides in the region between −551 and −220 (Figure 8).

As an independent approach to demonstrate the existence of endogenous PTEN transcripts with shorter 5′-UTRs, we performed an RNase protection assay. Using a probe covering the region of −483 to −60 in the 5′-UTR sequence of human PTEN (Figure 10c), we generated three protected products with estimated sizes of 360, 135, and 59 bases, respectively, with RNAs isolated from both DU145 and H226 cell lines (Figure 10b). However, no protected RNAs were found with yeast tRNA control. While the products of 135 and 59 base bands are likely derived from transcripts with 5′-UTRs of 109 and 193 bases as shown by 5′ RACE, respectively, the product of 360 bases is possibly produced from the transcripts with a longer 5′-UTR. Together, these results demonstrate that endogenous PTEN transcripts with shorter 5′-UTRs exist in human cell lines and they are likely responsible for efficient translation to produce proteins.

Discussion

Whereas efficiently translated mRNAs, such as cellular transcripts that encode globins and some viral mRNAs, have rather short 5′UTRs (<100 bases), many mRNAs encoding oncogene products or factors related to cell growth have long 5′-UTRs (Kozak, 1987, 1991). In many cases, the long 5′-UTR presents a hindrance to cap-dependent translation initiation (Willis, 1999). How such mRNAs are translated remains to be answered. As one of the alternative mechanisms, IRES-mediated translation initiation has recently been proposed and demonstrated for some of these mRNAs. With this mechanism, ribosome is directly recruited to the translation start codon, thereby overcoming the translational inhibition by the long and structured 5′-UTR sequence (Willis, 1999; Sachs, 2000; Pestova *et al.*, 2001). However, the existence of such a mechanism is recently questioned and argued (Kozak, 2001; Schneider and Kozak, 2001). Furthermore, detailed analysis of the reported IRES element in the 5′-UTR sequence of eIF4G (Gan & Rhoads, 1996; Gan *et al.*, 1998) demonstrated that it is in fact a promoter (Han and Zhang, 2002). It was also found that the long 5′-UTRs of human Sno and mouse Bad contain promoters, suggesting that promoters in the 5′-UTR sequences are more prevalent than anticipated. These observations led us to hypothesize that promoters in the 5′-UTR sequence provide an alternative way of producing a translatable mRNA with shorter 5′-UTRs (Han and Zhang, 2002).

The long and structured 5′-UTR sequence of PTEN mRNA is highly conserved between mouse and human. Conservation of the 5′-UTR sequence is also observed for other proto-oncogenes or factors such as VEGF and PDGF (Bernstein *et al.*, 1997; Huez *et al.*, 1998). The conservation of the 5′-UTR sequences suggests that they may play an important role in regulating gene expression. In this report, we showed that the 5′-UTR sequence of PTEN inhibits translation of its downstream

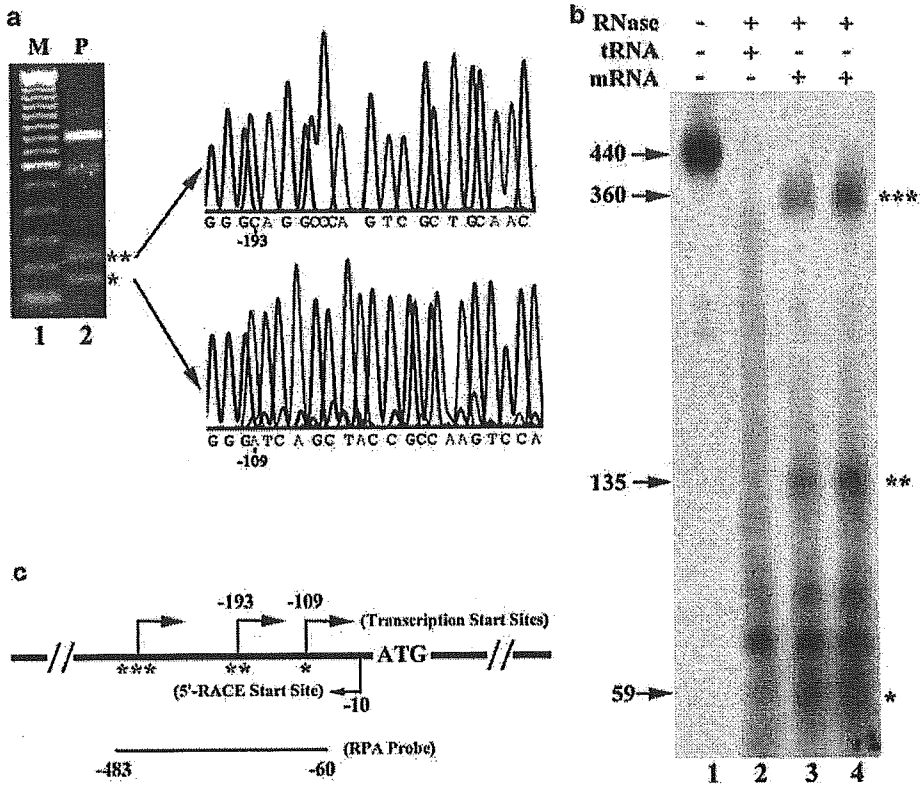

Figure 10 Endogenous PTEN transcripts with shorter 5′-UTRs exist in human cancer cell lines. (**a**) Mapping 5′-UTRs of endogenous PTEN transcripts by 5′ RACE. Total RNAs (2 μg) from DU145 cell line were used for 5′ RACE as described in Materials and methods. The final DNA products were separated by agarose gel electrophoresis (lane 2). The 5′ end sequencing results of the two small products were shown on the right. (**b**) Mapping 5′-UTRs of endogenous PTEN transcripts by an RNase protection assay. The RNase protection assay with 10 μg of RNAs from DU145 (lane 3) and H226 (lane 4) cells, 10 μg of yeast tRNA (lane 2), and a probe covering the region of −483 and −60 in the 5′ UTR sequence of human PTEN was performed as described in Materials and methods. The three distinctively protected products with estimated sizes of 360, 135, and 59 bases are indicated. Lane 1 shows the mobility and the estimated size of the probe used. (**c**) Schematic diagram showing relative locations of transcription start sites, the primer used in the 5′ RACE, and the probe used for RPA. The asterisks show the possible corresponding positions in the PTEN sequence for the products generated from both 5′ RACE (panel a) and RNase protection assay (panel b)

sequence in an *in vitro* system. A transcript with a 5′-UTR of only 93 bases is 100-fold more translatable than the transcript with a long 5′-UTR of 828 bases. The inhibitory effect of the long 5′-UTR is similar to that of the 5′-UTR of PDGF, which inhibits translation about 40-fold (Rao *et al*., 1988).

Inhibition of translation by 5′-UTR led us to explore other mechanisms such as IRES that may be used for PTEN translation. The 5′-UTR sequence of PTEN enhanced the expression of the second cistron by about 70-fold in the traditional dicistronic DNA assays and translation of the second cistron persisted when a hairpin structure was used to block translation of the first cistron (Figure 3). However, no such enhancing activity was observed in the dicistronic RNA assay, indicating that the 5′-UTR of PTEN may not have IRES. This conclusion is supported by our promoterless dicistronic test, which revealed that the enhancing effect was due to the existence of promoter activities in the 5′-UTR of PTEN. In addition, Northern blot analysis clearly detected the monocistronic transcript derived from the promoter in the 5′-UTR sequence of PTEN (Figure 7). By construction of a variety of deletion mutants from both 5′ and 3′ end of the 5′-UTR, we mapped the promoter activity to the region −551 to −220. Our experiments using *Drosophila* SL2 cells showed that the promoter in the 5′-UTR sequence of PTEN responds to Sp1, a ubiquitous transcription factor that is not only important for constitutive expression of many house-keeping genes but also critical for induced expression of many growth-related genes (Black *et al*., 2001). Finally, using the RNase protection assay and 5′ RACE, we identified two endogenous PTEN transcripts with short 5′-UTRs of 109 and 193 bases in both human cancer cell lines H226 and DU145, which were known to express human PTEN (Whang *et al*., 1998; Soria *et al*., 2002).

In the past, promoters in the 5′-UTR sequences have primarily been ignored and lack attention. Promoter studies normally begin with primer extension aiming at identification of the transcription start site furthest away from the translation initiation start codon. Subsequent analysis usually focuses on the promoter regions that are responsible for the generation of the transcript with the

longest 5′-UTR. Few of the past studies considered the impact of translational regulation and the fact that mRNAs with long 5′-UTRs generated by the promoter they studied may not be translatable. For the same reason, the reporter activity detected in the presence of both the upstream promoter and the long 5′-UTR sequence might be due to the contribution of another promoter in the 5′-UTR sequence that produces a transcript with a shorter 5′-UTR. Consequently, many promoters in 5′-UTRs were never discovered. Such a strategy may introduce confusion to the researchers in the translation field who are working to understand how mRNAs with long 5′-UTRs can be translated.

In the case of the PTEN promoter study, Stambolic *et al.* (2001), using a conventional approach, identified that the transcription start site for transcripts with the longest 5′-UTR was located between −951 and −925 bases upstream of the translation start codon. However, our studies showed that a transcript with more than 500 bases of 5′-UTR could not be translated efficiently and it does not appear to contain IRES. The question thus remains as to how the mRNA of PTEN with a 5′-UTR of 925–950 bases is translated. Use of the promoter in the 5′-UTR sequence to produce mRNAs with shorter 5′-UTRs would provide a solution to this problem. In fact, of the 100 EST clones found by searching the GeneBank using the 5′-UTR of mouse PTEN as a query sequence, 50 clones may have a 5′-UTR less than 500 bases. Our RACE and RPA study (Figure 10), promoter mapping (Figure 8), and Northern blot analysis (Figure 7) suggested the existence of transcripts with various lengths of 5′-UTR. Furthermore, using RPA and 5′ RACE, we identified two endogenous transcripts with 5′-UTRs of 109 and 193 bases, respectively, that are likely originated from the 5′-UTR promoter. However, this possibility was not considered in the Stambolic *et al.* (2001) study and a large part (427 bp immediately upstream of the translation start codon) from the 5′-UTR sequence of PTEN was deleted in their constructs. By doing so, their promoter activity is only fourfold over the vector control as compared to about 100-fold shown in our study (Figure 6). They, therefore, failed to explain the p53-independent, relatively high and constant expression of PTEN mRNA and protein in a variety of cells.

In another study (Virolle *et al.*, 2001), the transcription start site at −1013 bases upstream of the translation start codon was considered simply based on the longest cDNA sequence of human PTEN. Nevertheless, the whole 5′-UTR sequence was used, which led to the identification of the Egr-1 factor binding site located downstream of the postulated transcription start site. However, the existence of promoters in the 5′-UTR sequence, which may direct alternative transcription compatible with translation, was not considered. Although not discussed in their study, their results are consistent with ours in the present study. Furthermore, they also showed that the promoter activity in the 5′-UTR is about half of that generated in the presence of both 5′-UTR and the upstream promoter. Again, these data were not discussed and they still considered in their figures that the transcription start site is located at the upstream. This type of routine practice in promoter studies will likely generate incorrect conclusions on the regulation of gene expression, specifically for PTEN because it is known that the mRNAs of PTEN with long 5′-UTRs cannot be translated.

Our observations also confirm that the translation of PTEN mRNAs is regulated by the secondary structure of their 5′-UTRs. Only mRNAs with shorter 5′-UTRs that have less free energy in secondary structure can be significantly translated. Retrospectively, it is noteworthy that human PTEN transcripts of heterogeneous sizes between 2 and 5 kb are observed on Northern blot (Whang *et al.*, 1998). Our RPA and 5′ RACE results (Figure 10) demonstrated that 5′-UTRs of endogenous PTEN mRNAs in human cancer cell lines are heterogeneous at least in their length with the 5′-UTRs. Interestingly, in the human PTEN promoter study (Virolle *et al.*, 2001), the Δ3′ transcript with a short 5′-UTR (~258 bases) has a much higher basal and Egr-1-stimulated expression than the transcript with a longer 5′-UTR (~1031). Although these authors were focusing on the positive and negative elements responsible for transcriptional regulation of PTEN expression, it cannot be ruled out that the transcripts with a short 5′-UTR are in fact translated with higher efficiency than the transcripts with a long 5′-UTR. Furthermore, the relative abundance of PTEN mRNA does not correlate well with its protein level in advanced prostate cancer tissues as compared with normal tissues (Whang *et al.*, 1998), suggesting that translational control may play an important role in regulating PTEN expression.

Since the promoter in the 5′-UTR, which was mapped to the region between −551 and −220 bases upstream of the translation start codon, contributes to production of translation-compatible transcripts, and functions in a variety of cell lines tested (Figure 6), it is tempting to suggest that the promoter in the 5′-UTR is a major promoter that controls ubiquitous and constitutive PTEN protein production. Upstream promoters or enhancers may function through regulating the production of transcripts from the promoter in the 5′-UTR. Switching off the 5′-UTR promoter will, therefore, lead to reduced or complete loss of PTEN expression. Like many other tumor suppressor genes, expression of PTEN has been postulated to be regulated by epigenetic mechanisms. PTEN expression can be lost or greatly reduced in some tumors without any mutation in the coding sequence of the PTEN gene in endometrial, breast, prostate, ovarian, non-small cell lung cancer, and melanocytic tumors (reviewed in Mutter, 2001; Leslie and Downes, 2002; Soria *et al.*, 2002). This is supported by an observation that PTEN expression in the prostate cancer xenograff LuCap-35 and lung cancer cell line H1299 was recovered upon treatment with the demethylating agent 5-azadeoxycytidine (Whang *et al.*, 1998; Soria *et al.*, 2002). However, the detailed mechanism underlying epigenetic regulation of PTEN expression is yet to be identified. In this respect, it is interesting to note that the 5′-UTR-encoding sequence of both human and mouse PTEN contains two CpG islands, which may

be potential methylation targets as predicted using an online program (http://www.ebi.ac.uk/emboss) (data not shown). Indeed, the presence of methylation in the 5′-UTR sequences (−405 to −104) was reported to be frequent in both endometrial and gastric carcinomas (Salvesen *et al.*, 2001; Kang *et al.*, 2002). This region with high frequency of methylation overlaps with the region containing the promoter mapped in this study (−551 to −220). Methylation of this region may shut off 5′-UTR promoter activity and constitutive expression of PTEN and, therefore, inactivate the expression of PTEN in some tumors. We are currently testing this hypothesis by determining the effect of methylation of the 5′-UTR sequence on promoter activity.

Since the promoter in the 5′-UTR sequence of PTEN may serve as a major promoter controlling constitutive expression of PTEN, detailed promoter mapping may lead to discovery of *cis*-elements and transcription factors that govern constitutive PTEN expression. Analysis of putative transcription factor binding sites showed that the promoter in the 5′-UTR sequence contains several binding sites for transcription factors such as Sp1, Egr-1, Ets-1, Ap-1, Ap-2, and Ap-4. We have shown that Sp1 is involved in the activation of this promoter. Interestingly, the Sp1 site located at −82/−77 plays a critical role in the p53/p53 site (at 2281/2262)-mediated synergistic transactivation of the p21 promoter (Koutsodontis *et al.*, 2001). Similarly, conditional formation of transcriptional Sp1–p53 regulatory complexes has been reported recently in a number of other promoters (Borellini and Glazer, 1993; Gualberto and Baldwin, 1995; Ohlsson *et al.*, 1998; Torgeman *et al.*, 2001). Sp1/p53 synergism may serve as a general mechanism for transcriptional activation of p53 target genes. Therefore, it is tempting to hypothesize that the possible synergism between the p53 site located in −1190 to −1157 of human PTEN promoter and the multiple Sp1 sites located in the 5′-UTR region also function on PTEN promoter. On the other hand, multiple consensus Egr-1 sites (GCGGCGGCG) identified in human PTEN promoter (Virolle *et al.*, 2001) were also found in the 5′-UTR sequence of mouse PTEN (Figure 9a). Whether these Egr-1 sites are functional and whether the Sp1 and Egr-1 transcription factors interplay on the promoter in the 5′-UTR sequence of PTEN remain to be addressed in future studies.

Materials and methods

Materials

Restriction enzymes, m^{7}GpppG cap analogue, and Pfu polymerase were purchased from New England Biolabs, Amersham/Pharmacia Biotech, and Stratagene, respectively. The SL2 cell line was from ATCC. Sp6 RNA polymerases, RNasin, RNase-free DNase, RRL, LUC assay 'Stop & Glo' kit, and pSP64 PolyA plasmid were from Promega. RNeasy Mini Kit and Oligotex mRNA Mini Kit were from Qiagen. Rediprime II Random Prime Labeling System, α[^{32}P]dCTP and α[^{32}P]CTP were from Amersham Biosciences. The Sephadex G-25 Quick Spin Columns (TE) for radiolabeled DNA and RNA purification was from Roche Diagnostics. MAGNA nylon transfer membrane was from Osmonics Inc. Zero Blunt PCR Blunt PCR Cloning kit, Schneider's *Drosophila* culture medium, Lipofectamine plus and Lipofectin transfection reagent, and the 5′-RACE system for rapid amplification of cDNA ends were purchased from Invitrogen. Oligonucleotides were synthesized by Sigma-Genosys. IMAGE EST clones were obtained from either ATCC or Research Genetics. MAXIscript *in vitro* transcription kit and RPA III ribonuclease protection assay kit were products of Ambion. TaKaRa LA *Taq* polymerase with GC buffer was purchased from TaKaRa Bio Inc.

Construction of plasmids

The cDNA clone mncb-0146 (GeneBank Accession #: AU035162) was from the Sugano mouse brain mncb library (Suzuki *et al.*, 2000). Double-strand DNA sequencing showed that this clone encodes the full-length mouse PTEN protein and confirmed that the 5′-UTR sequence (−828 to −1) was identical to that shown in GeneBank (Accession # *NM_008960*). The cDNA, following sequencing, was cloned into pGEM-4Z (Promega) at *Eco*RI and *Xba*I sites, resulting in the plasmid pG-WPTEN (Figure 1a). In the deletion construct pG-DPTEN, the 5′-UTR sequence from −828 to −8 was removed by replacing the *Eco*RI/*Bgl*II fragment representing −828 to +318 with a PCR fragment representing −8 to +318. The plasmid pG-LUC was generated by cloning *Eco*RI–*Xba*I fragment of pRF (Stoneley *et al.*, 2000), corresponding to the firefly LUC-encoding region into pGEM-4Z. To generate pG-5′LUC, the −828 to −1 of the 5′-UTR sequence of mouse PTEN was amplified by PCR and inserted into pG-LUC between the *Eco*RI and *Nco*I sites. The 5′-UTR deletion mutants of mouse PTEN (−551/−1, −479/−1, −299/−1) were generated by cloning *Xho*I-*Nco*I, *Not*I-*Nco*I, *Sma*I-*Nco*I fragments into pG-LUC, respectively. The −257/−1 and −93/−1 deletion mutants were generated by PCR.

Dicistronic constructs were generated based on pRF. The 5′-UTR sequence of mouse PTEN (−828 to −1) from pG-LUC was cloned into pRF, resulting in pR-PTEN-F. In addition, a 492 bp of the 5′-UTR sequence of human PTEN cDNA (−492 to −1) was amplified from human IMAGE 2157760 (GeneBank Accession #: AI480306) and cloned into pRF, resulting in pR-hPTEN-F. Dicistronic constructs with a hairpin structure were generated by inserting into *Eco*RV site of pRF and pR-PTEN-F a double-stranded oligonucleotide with sense sequence of 5′-ATCAAAAGCGCAGGTCGCGACCGCG-CATGCGCGGTCGCGACCTGCGCTAAAGAT-3′. The dicistronic constructs containing the truncated 5′-UTR of PTEN (−551/−1, −479/−1, −299/−1, −257/−1) were generated by cloning these cDNA fragments from the pG-LUC-based constructs (see above) into pRF. The dicistronic constructs containing other truncations (−828/−220, −828/−348, and −828/−474) were obtained by deletions from pR-PTEN-F between *Bst*XI and *Nco*I, *Sma*I and *Nco*I, and between *Not*I and *Nco*I. The −828/−40 construct was generated by cloning a PCR fragment into pRF. The promoterless dicistronic constructs were generated as previously described (Han and Zhang, 2002).

Constructs containing poly(A) for *in vitro* transcription were engineered using the vector pSP64 PolyA (Promega) that has 30 bp (dA-dT) sequence. The *Eco*RV–*Xba*I fragment of the pRF vector containing the *Renilla* LUC gene was first cloned into pSP64 PolyA vector at the *Xba*I and blunted *Hind*III sites to generate the plasmid pSP-R$_{A30}$. The *Xba*I fragment of pR-HRV-F containing the IRES of HRV and the firefly LUC gene

was then isolated and cloned into pSP-R_{A30} at the *Xba*I site to generate pSP-R-HRV-F_{A30}. The pSP-RF_{A30} plasmid was obtained by removing the IRES sequence of HRV from the pSP-R-HRV-F_{A30} by digestion with *Spe*I and *Nco*I. To engineer pSP-R-PTEN-F_{A30}, the *Xho*I–*Nco*I fragment containing −551 to −1 sequence from pR-PTEN-F was used to replace the HRV IRES fragment in pR-HRV-F_{A30} construct. All the above constructs were confirmed by double-strand DNA sequencing.

In vitro *transcription and translation*

In vitro transcription and translation were performed as previously described (Zhang and Ling, 1991). The pG-WPTEN-, pG-DPTEN-, and pG-LUC-based plasmids were linerized by *Xba*I and pRF_{A30}-based plasmids were linearized by *Eco*RI. The capped transcripts were synthesized in the presence of 1 mM m^7GpppG and purified using a Qiagen RNeasy Mini kit. Capped RNA transcripts (5–50 ng) were used to program cell-free translation in RRL in a final volume of 10 μl containing 6.5 μl RRL. Translation product was either visualized by autoradiography or by measurement of LUC activities.

Cell culture, DNA and RNA transfection

Human HeLa and HEK293 cells were cultured using DMEM, while H1299, K562, DU145, and H226 cells were maintained in RPMI1640, supplemented with 10% fetal bovine serum at 37°C with 5% CO_2. Schneider's *Drosophila* cell line 2 was maintained in Schneider's *Drosophila* medium supplemented with 10% fetal bovine serum at room temperature with atmospheric CO_2.

DNA transfection of human cell lines was performed with Lipofectamine plus reagents according to the manufacturer's protocol. In a 24-well plate, approximately 1×10^5 cells/well were plated and transfected with 0.4 μg DNA. Cells were harvested 24 h following transfection for LUC assay. Transfection of *Drosophila* SL2 cells was performed according to the protocol as previously described (Han *et al*., 2001).

RNA transfection was performed using the cationic liposome-mediated method as previously described (Han and Zhang, 2002). Briefly, approximately 2×10^5 cells/well were seeded onto six-well plates on the day before transfection. Cells were washed once with Opti-MEM I-reduced serum medium (GIBCO-BRL) and left in the incubator with some medium during preparation of the liposome–polynucleotide complexes. Opti-MEM I medium (1 ml) in a 12×75 mm polystyrene snap-cap tube was mixed with 12.5 μg of Lipofectin reagent and 5 μg capped mRNA. The liposome–RNA–medium mixture was immediately added to cells. At 8 h following transfection, cells were harvested and processed for LUC analysis.

Northern blot analysis

Subconfluent H1299 cells in 10-cm plates were transfected with 4 μg/plate constructs using Lipofectamine Plus. At 24 h following transfection, the total RNAs were extracted using an RNeasy Mini Kit. Residual plasmid DNA in the total RNA was digested with RNase-free DNase. The polyA mRNAs were then isolated from 250 μg of total RNAs using an Oligotex mRNA Mini Kit. One-fifth of the mRNAs were separated in 1% agarose gels in the presence of formaldehyde and MOPS buffer and blotted onto MAGNA nylon membranes. The blots were hybridized with a ^{32}P-labeled firefly LUC DNA probe (1656 bp), which was isolated by cleaving pRF with *Nco*I and *Xba*I and labeled using the Rediprime II Random Prime Labeling System.

5′ RACE and RNase protection assay

5′ RACE was performed based on the protocol provided by the supplier. Briefly, 2 μg total RNA from human prostate cancer cell line DU145 was used. The primer used for first-strand cDNA synthesis was GSP1A: 5′-TGGCGGTGTCA-TAATGT-3′, located in 238 bases downstream of the translation start codon ATG. The primers used for first-round PCR are 5′ RACE Abridged Anchor Primer from the supplier (36 bases), human PTEN-specific primer GSP2: 5′-CCATCCTCTTGATATCTCCTTTTG-3′, located 58 bases downstream of the translation start codon ATG. The primers used for the second and final PCR are Abridged Universal Amplification Primer (AUAP) from the supplier and PTEN-specific primer PTENEXT1: 5′-CCTGTGGCTGAAGAAAA-AGGAGGAGAGAGAT-3′, located at 10 bases upstream the translation start codon ATG. Takara *Taq* polymerase and GC buffer II are used for PCR amplification. The reaction conditions were 94°C for 2 min followed by 40 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1 min. PCR products were purified from gel, blunted with Klenow enzyme, and then cloned using the PCR Zero Blunt Cloning kit. Individual clones were sequenced for determination of the transcription start sites.

Ribonuclease Protection Assay (RPA) was performed using the RPAIII kit according to the supplier's instructions. Briefly, the RNA probe was produced by first cloning the region of −770 to −60 of human PTEN 5′-UTR sequence into PCR-Blunt vector at *Spe*I and *Xba*I sites. The resulting plasmid was linearized with *Not*I and transcribed using T7 RNA polymerase in the presence of 0.5 mM each of ATP, GTP, UTP, and 0.01 mM CTP supplemented with 3.12 μM a[^{32}P] CTP. The ^{32}P-labeled probe was digested with DNase and purified using a Sephadex G-25 Quick Spin Column. About 0.5×10^6 c.p.m. of probe was hybridized to 10 μg total RNA at 45°C overnight followed by digestion with RNase T1/A at 37°C. The reaction was then stopped and the protected RNAs precipitated before separation by electrophoresis on a 6% acrylamide/8 M urea gel for autoradiography.

Abbreviations

SL2, Schneider's *Drosophila* cell line 2; kb, kilobase(s); PCR, polymerase chain reaction; 5′-UTR, 5′-untranslated region; IRES, internal ribosome entry site; PAGE, polyacrylamide gel electrophoresis; mRNA, messenger RNA; RRL, rabbit reticulocyte lysate; HRV, human rhinovirus; LUC, luciferase; VEGF, vascular endothelial growth factor; PDGF, platelet-derived growth factor.

Acknowledgements

We are indebted to Dr AE Willis for pRF and pGL3RHRV plasmids, Drs Guntram Suske and Etty Benveniste for pPacO and pPacSp1 plasmids, and Dr Philip Marsden for pPac-UEts-1 plasmid. This work was supported in part by National Institutes of Health Grants CA64539 and GM59475, and by Department of Defense Grant DAMD17-02-1-0073. J-TZ is a recipient of a Career Investigator Award from the American Lung Association.

References

Ali IU, Schriml LM and Dean M. (1999). *J. Natl. Cancer Inst.*, **91**, 1922–1932.
Backman SA, Stambolic V, Suzuki A, Haight J, Elia A, Pretorius J, Tsao MS, Shannon P, Bolon B, Ivy GO and Mak TW. (2001). *Nat. Genet.*, **29**, 396–403.
Bernstein J, Sella O, Le SY and Elroy-Stein O. (1997). *J. Biol. Chem.*, **272**, 9356–9362.
Black AR, Black JD and Azizkhan-Clifford J. (2001). *J. Cell Physiol.*, **188**, 143–160.
Borellini F and Glazer RI. (1993). *J. Biol. Chem.*, **268**, 7923–7928.
Cantley LC and Neel BG. (1999). *Proc. Natl. Acad. Sci. USA*, **96**, 4240–4245.
Coldwell MJ, Mitchell SA, Stoneley M, MacFarlane M and Willis AE. (2000). *Oncogene*, **19**, 899–905.
Courey AJ and Tjian R. (1988). *Cell*, **55**, 887–898.
Dahia PL. (2000). *Endocr. Relat. Cancer*, **7**, 115–129.
Dennig J, Beato M and Suske G. (1996). *EMBO. J.*, **15**, 5659–5667.
Di Cristofano A and Pandolfi PP. (2000). *Cell*, **100**, 387–390.
Gan W, LaCelle M and Rhoads RE. (1998). *J. Biol. Chem.*, **273**, 5006–5012.
Gan W and Rhoads RE. (1996). *J. Biol. Chem.*, **271**, 623–626.
Groszer M, Erickson R, Scripture-Adams DD, Lesche R, Trumpp A, Zack JA, Kornblum, HI, Liu X and Wu H. (2001). *Science*, **294**, 2186–2189.
Gualberto A and Baldwin Jr AS. (1995). *J. Biol. Chem.*, **270**, 19680–19683.
Han B, Liu N, Yang X, Sun HB and Yang YC. (2001). *J. Biol. Chem.*, **276**, 7937–7942.
Han B and Zhang JT. (2002). *Mol. Cell. Biol.*, **22**, 7372–7384.
Hisatake J, O'Kelly J, Uskokovic MR, Tomoyasu S and Koeffler HP. (2001). *Blood*, **97**, 2427–2433.
Huez I, Creancier L, Audigier S, Gensac MC, Prats AC and Prats H. (1998). *Mol. Cell. Biol.*, **18**, 6178–6190.
Kang YH, Lee HS and Kim WH. (2002). *Lab. Invest.*, **82**, 285–291.
Koutsodontis G, Tentes I, Papakosta P, Moustakas A and Kardassis D. (2001). *J. Biol. Chem.*, **276**, 29116–29125.
Kozak M. (1987). *Nucleic Acids Res.*, **15**, 8125–8148.
Kozak M. (1991). *J. Cell. Biol.*, **115**, 887–903.
Kozak M. (2001). *Mol. Cell. Biol.*, **21**, 1899–1907.
Kwon CH, Zhu X, Zhang J, Knoop LL, Tharp R, Smeyne RJ, Eberhart CG, Burger PC and Baker SJ. (2001). *Nat. Genet.*, **29**, 404–411.
Leslie NR and Downes CP. (2002). *Cell Signal.*, **14**, 285–295.
Li DM and Sun H. (1997). *Cancer Res.*, **57**, 2124–2129.
Li G, Robinson GW, Lesche R, Martinez-Diaz H, Jiang Z, Rozengurt N, Wagner KU, Wu DC, Lane TF, Liu X, Hennighausen L and Wu H. (2002). *Development*, **129**, 4159–4170.
Maehama T and Dixon JE. (1999). *Trends Cell Biol.*, **9**, 125–128.
Mathews DH, Sabina J, Zuker M and Turner DH. (1999). *J. Mol. Biol.*, **288**, 911–940.
Mutter GL. (2001). *Am. J. Pathol.*, **158**, 1895–1898.
Ohlsson C, Kley N, Werner H and LeRoith D. (1998). *Endocrinology*, **139**, 1101–1107.
Pestova TV, Kolupaeva VG, Lomakin IB, Pilipenko EV, Shatsky IN, Agol VI and Hellen CU. (2001). *Proc. Natl. Acad. Sci. USA*, **98**, 7029–7036.
Quandt K, Frech K, Karas H, Wingender E and Werner T. (1995). *Nucleic Acids Res.*, **23**, 4878–4884.
Rao CD, Pech M, Robbins KC and Aaronson SA. (1988). *Mol. Cell. Biol.*, **8**, 284–292.
Ross AH, Lachyankar MB and Recht LD. (2001). *Neuroscientist*, **7**, 278–281.
Sachs AB. (2000). *Cell*, **101**, 243–245.
Salvesen HB, MacDonald N, Ryan A, Jacobs IJ, Lynch ED, Akslen LA and Das S. (2001). *Int. J. Cancer*, **91**, 22–26.
Schneider R and Kozak M. (2001). *Mol. Cell. Biol.*, **21**, 8238–8246.
Soria JC, Lee HY, Lee JI, Wang L, Issa JP, Kemp BL, Liu DD, Kurie JM, Mao L and Khuri FR. (2002). *Clin. Cancer Res.*, **8**, 1178–1184.
Stambolic V, MacPherson D, Sas D, Lin Y, Snow B, Jang Y, Benchimol S and Mak TW. (2001). *Mol. Cell*, **8**, 317–325.
Stambolic V, Suzuki A, de la Pompa JL, Brothers GM, Mirtsos C, Sasaki T, Ruland J, Penninger JM, Siderovski DP and Mak TW. (1998). *Cell*, **95**, 29–39.
Stoneley M, Subkhankulova T, Le Quesne JP, Coldwell MJ, Jopling CL, Belsham GJ and Willis AE. (2000). *Nucleic Acids Res.*, **28**, 687–694.
Suzuki Y, Ishihara D, Sasaki M, Nakagawa H, Hata H, Tsunoda T, Watanabe M, Komatsu T, Ota T, Isogai T, Suyama A and Sugano S. (2000). *Genomics*, **64**, 286–297.
Torgeman A, Mor-Vaknin N, Zelin E, Ben-Aroya Z, Lochelt M, Flugel RM and Aboud M. (2001). *Virology*, **281**, 10–20.
Virolle T, Adamson ED, Baron V, Birle D, Mercola D, Mustelin T and de Belle I. (2001). *Nat. Cell Biol.*, **3**, 1124–1128.
Whang YE, Wu X, Suzuki H, Reiter RE, Tran C, Vessella RL, Said JW, Isaacs WB and Sawyers CL. (1998). *Proc. Natl. Acad. Sci. USA*, **95**, 5246–5250.
Willis AE. (1999). *Int. J. Biochem. Cell Biol.*, **31**, 73–86.
Zhang JT and Ling V. (1991). *J. Biol. Chem.*, **266**, 18224–18232.

BMC Genomics

BioMed Central

Research article

Open Access

Molecular cloning, genomic characterization and over-expression of a novel gene, *XRRA1*, identified from human colorectal cancer cell HCT116^{Clone2_XRR} and macaque testis

Felix M Mesak[1,2], Naoki Osada[3,4], Katsuyuki Hashimoto[3], Qing Y Liu[5] and Cheng E Ng*[1,2]

Address: [1]Centre for Cancer Therapeutics, Ottawa Regional Cancer Centre, 503 Smyth Rd., Ottawa, ON, K1H 1C4, Canada, [2]Faculty of Medicine, University of Ottawa, 451 Smyth Rd., Ottawa, ON, K1H 8M5, Canada, [3]Division of Genetic Resources, National Institute of Infectious Disease, Toyama 1-23-1, Shinjuku-ku, Tokyo 162-8640, Japan, [4]Current address: Department of Ecology and Evolution, University of Chicago, 1101 E57th Str., Chicago, IL 60637, USA and [5]Institute for Biological Sciences, National Research Council of Canada, 1200 Montreal Rd, M-54, Ottawa, ON, K1A 0R6, Canada

Email: Felix M Mesak - Felix.Mesak@orcc.on.ca; Naoki Osada - nosada@uchicago.edu; Katsuyuki Hashimoto - khashi@nih.go.jp; Qing Y Liu - Qing_Yan.Liu@nrc.ca; Cheng E Ng* - Cheng.Ng@orcc.on.ca

* Corresponding author

Published: 09 August 2003

BMC Genomics 2003, 4:32

Received: 10 April 2003
Accepted: 09 August 2003

This article is available from: http://www.biomedcentral.com/1471-2164/4/32

© 2003 Mesak et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: As part of our investigation into the genetic basis of tumor cell radioresponse, we have isolated several clones with a wide range of responses to X-radiation (XR) from an unirradiated human colorectal tumor cell line, HCT116. Using human cDNA microarrays, we recently identified a novel gene that was down-regulated by two-fold in an XR-resistant cell clone, HCT116^{Clone2_XRR}. We have named this gene as X-ray radiation resistance associated 1 (*XRRA1*) (GenBank BK000541). Here, we present the first report on the molecular cloning, genomic characterization and over-expression of the *XRRA1* gene.

Results: We found that *XRRA1* was expressed predominantly in testis of both human and macaque. cDNA microarray analysis showed three-fold higher expression of *XRRA1* in macaque testis relative to other tissues. We further cloned the macaque *XRRA1* cDNA (GenBank AB072776) and a human *XRRA1* splice variant from HCT116^{Clone2_XRR} (GenBank AY163836). *In silico* analysis revealed the full-length human *XRRA1*, mouse, rat and bovine *Xrra1* cDNAs. The *XRRA1* gene comprises 11 exons and spans 64 kb on chromosome 11q13.3. Human and macaque cDNAs share 96% homology. Human *XRRA1* cDNA is 1987 nt long and encodes a protein of 559 aa. XRRA1 protein is highly conserved in human, macaque, mouse, rat, pig, and bovine. GFP-XRRA1 fusion protein was detected in both the nucleus and cytoplasm of HCT116 clones and COS-7 cells. Interestingly, we found evidence that COS-7 cells which over-expressed XRRA1 lacked Ku86 (Ku80, XRCC5), a non-homologous end joining (NHEJ) DNA repair molecule, in the nucleus. RT-PCR analysis showed differential expression of *XRRA1* after XR in HCT116 clones manifesting significantly different XR responses. Further, we found that *XRRA1* was expressed in most tumor cell types. Surprisingly, mouse *Xrra1* was detected in mouse embryonic stem cells R1.

Conclusions: Both *XRRA1* cDNA and protein are highly conserved among mammals, suggesting that *XRRA1* may have similar functions. Our results also suggest that the genetic modulation of *XRRA1* may affect the XR responses of HCT116 clones and that *XRRA1* may have a role in the response of human tumor and normal cells to XR. *XRRA1* might be correlated with cancer development and might also be an early expressed gene.

Background

Treatment with X-radiation (XR) remains a major modality for eradicating cancer and it has been estimated that half of all cancer patients undergo radiotherapy at some point during their treatment [1]. At least partly because of the relevance of radiotherapy to cancer treatment, the genetic control of the expression of resistance, or conversely, sensitivity to XR has been investigated for several decades. It is now generally accepted that there are several components to the cellular response to XR stress: sensing of the DNA damage, cell signaling pathways, and repair of the DNA damage [2]. However, the molecular mechanisms of how cells detect, communicate, and cope with XR-induced DNA damage are not well characterized. There is strong evidence that ataxia telangiectasia mutated (ATM) protein act as damage sensors and that their activation initiates mobilization and/or activation of repair complexes [3]. In addition, the ATM kinase activity mediates the prompt induction of various signaling pathways with some of these events leading to either cellular apoptosis or cell cycle arrest [4]. There is also evidence that non-homologous end joining involving Ku86/Ku70 is likely the major DNA double strand breaks rejoining pathway following XR with a smaller role for homologous recombination in post replicated DNA [5].

Many of the studies on the genetic basis of cellular response to XR have benefited either from mutants that have increased ionizing radiation (IR) sensitivity compared to the parental cell lines or from the study of human genetic syndromes that are associated with increased IR sensitivity (i.e. Fanconi's anemia, Bloom's, and Werner's syndromes) [6,7]. By contrast, mutants or human genetic syndromes associated with increased IR resistance (as exemplified by Li Fraumeni) are rare [8]. Furthermore, Bennett et al [9] recently identified 107 new loci that are associated with response to γ-radiation in yeast. While over 50% of these newly identified putative yeast genes shared homology with human genes, it remains to be established that the human counterparts of these genes are actually associated with similar responses to γ-radiation.

In order to further investigate the genetic basis of tumor cell radioresponse, we have isolated several clones with a wide range of responses to X-radiation from an unirradiated human colorectal tumor cell line, HCT116. Interestingly, we found a cell clone, $HCT116^{Clone2\_XRR}$ that was resistant to both fractionated and single dose X-radiation when compared to the parental cells. In addition, we also isolated two other clones: (a) $HCT116^{CloneK\_XRS}$ that was more sensitive than the parental cells, and (b) $HCT116^{Clone10}$ that possessed similar X-radiation responses as the parental cells, to both of these types of X-radiation (Qutob, S.S., Ng, C.E., to be published elsewhere). Global gene analysis between cells $HCT116^{Clone2\_XRR}$ and $HCT116^{Clone10}$ using cDNA microarray identified an unknown GenBank expression sequence tag (EST) accession no. R40588, as one of the spots on the 19,200 human cDNA microarray (University Health Network, Ontario Cancer Institute, Toronto), that was down regulated by approximately two-fold in $HCT116^{Clone2\_XRR}$ in comparison with $HCT116^{Clone10}$ cells (Qutob, S.S., Ng, C.E., to be published elsewhere).

Here we describe the *in silico* analysis of R40588 that resulted in the identification of a novel gene, *XRRA1*. We cloned the *XRRA1* full-length cDNA from macaque testis and an *XRRA1* splice variant from $HCT116^{Clone2\_XRR}$ cells. We then extended our analysis to identify *XRRA1* in several mammals (i.e. mouse, rat, pig and bovine). This paper also reports the experiments to evaluate *XRRA1* expression in human and macaque normal tissue/organs and in various human cancer and normal cell lines, as well as over-expression of GFP-XRRA1 fusion protein in HCT116 clones and COS-7 cells. We did immunostaining of Ku86 in the presence of over-expressed GFP-XRRA1 24 hours after XR treatment at 4 and 10 Gy of COS-7 cells. This is the first report on the molecular cloning, genomic characterization and over-expression of the novel *XRRA1* gene.

Results

Genomic approaches revealed a novel XRRA1 *gene*

Using the EST sequence of R40588 as template, we performed ESTs "walking" either towards the 5' or 3' direction on both the sense and anti-sense strands that we refer to as ESTs-based ORF assembling (EBOA). By EBOA, we acquired a cDNA candidate that putatively contained start-, stop-codon and poly-A signal. ESTs that overlapped each other sequentially were as follows: GenBank accession nos R40588, BQ648278, BU187847, BE782795, BI819448, BQ925455, and BM563944. Out of this ORF candidate, we found 31 ESTs as primary sequences that could be constructed into one potential novel gene *in silico* (Table 1, Fig. 1). We have named this gene as X-ray radiation resistance associated 1 (*XRRA1*) (GenBank accession no. BK000541). BLAST [10] analysis demonstrated that it was a novel gene.

We used human genomic clones RP11-147I3 and CMB9-8M21 (GenBank accession nos AP001992 and AP000560) of chromosome 11q13.3 as template for GrailEXP [11] analysis. The analysis predicted 11 exons including a putative promoter region for the human (*Homo sapiens*, Hs) *XRRA1* gene (Fig. 2, Fig. 3A). These 11 exons spliced perfectly to create the mRNA for Hs *XRRA1*. The exon-intron junctions, comprising the 5' splice donor and 3' splice acceptor, follow Kozak consensus (Table 3) [12]. Cloning of the 5' region of *XRRA1* gene that

Table 1: ESTs that were used to build Hs *XRRA1* and Mm *Xrra1*.

Organism	EST	Nucleotides	Exons	Genomic Clone	Exons
Homo sapiens XRRA1 (1987 nt)	1. BM563944.1	1 – 822	1 – 7	AP001992.4	1 – 5
	2. BE780143.1	1 – 233	1 – 2	AP000560.4	3 – 11
	3. AW962535.1	1 – 175	1 – 2	AP001324.4	5 – 11
	4. BM921116.1	1 – 107	1		
	5. BQ925455.1	1 – 348, 513 – 1066	1 – 3, 4 – 7		
	6. AI651142.1 (+/-)	225 – 348, 513 – 726	2 – 3, 4 – 6		
	7. AW448937.1 (+/-)	234 – 348, 513 – 726	2 – 3, 4 – 6		
	8. AI651143.1 (+/-)	245 – 348, 513 – 726	2 – 3, 4 – 6		
	9. AW197451.1 (+/-)	270 – 348, 513 – 726	2 – 3, 4 – 6		
	10. AW895676.1 (+/-)	513–609	4 – 5		
	11. BE836017.1 (+/-)	513 – 609	4 – 5		
	12. BF810439.1	513 – 589	4 – 5		
	13. AI652186.1 (+/-)	526 – 726	5 – 6		
	14. BE313013.1	583 – 1190	5 – 8		
	15. BF891539.1 (+/-)	607 – 694	5 – 6		
	16. BI819118.1	897 – 1755	7 – 11		
	17. BQ319609.1	1271 – 1388	8 – 10		
	18. BU620818.1 (+/-)	1275 – 1976	9 – 11		
	19. BU620757.1 (+/-)	1312 – 1911	9 – 11		
	20. AI057634.1	1474 – 1981	10 – 11		
	21. BQ051594.1	1492 – 1979	10 – 11		
	22. BE782795.1	1500 – 1898, 1939–1987	11		
	23. AW189295.1	1502 – 1979	11		
	24. AI655334.1	1546 – 1976	11		
	25. AA854636.1	1556 – 1981	11		
	26. AI218588.1	1593 – 1979	11		
	27. AI457916.1	1594 – 1981	11		
	28. AW513519.1 (+/-)	1614 – 1758	11		
	29. AA961245.1 (+/-)	1680 – 1977	11		
	30. BU187847.1	1796 – 1987	11		
	31. AA953084.1 (+/-)	1845 – 1977	11		
Mus musculus Xrra1 (1903 nt)	1. BI990168.1	277 – 870	2 – 7	NW_000328	1 – 11
	2. AV046823.1	715 – 978	6 – 7		
	3. BE650319.1	857 – 1380	7 – 9		
	4. BU610007.1	1134 – 1459	8 – 10		
	5. AA144723.1	1144 – 1396	8 – 10		
	6. AI647383.1	1184 – 1372	8 – 9		
	7. BE956083.1 (+/-)	1381 – 1903	9 – 11		
	8. BQ840215.1	1398 – 1649	10 – 11		
	9. BE631695.1	1458 – 1604	10 – 11		
	10. BQ175761.1 (+/-)	1458 – 1903	10 – 11		
	11. BE956870.1 (+/-)	1471 – 1903	10 – 11		
	12. AI449753.1 (+/-)	1600 – 1901	11		
	13. AW125246.1 (+/-)	1704 – 1903	11		

comprised the first four exons from HCT116^{Clone2_XRR} and HCT116^{Clone10} revealed that Hs *XRRA1* had those exons. However, in addition to the one from HCT116^{Clone2_XRR}, we obtained one splice variant that lacked exon three and ended on exon four (GenBank accession no. AY163836) (Fig. 3A).

To test whether these exon-intron borders were actually unique to *XRRA1*, we searched for possible *XRRA1* homologues from other organisms. Thus far, we could only retrieve *Xrra1* containing genomic clones from mouse (*Mus musculus*, Mm) (GenBank accession no. NW_000328) and rat (*Rattus norvegicus*, Rn) (GenBank accession no. AC127923). Mm *Xrra1* gene is located on chromosome 7. GrailEXP analyses showed that Mm *Xrra1* gene also contained 11 exons. Alignment of mouse and rat genomic sequences resulted in information for rat exons, which also consist of 11 exons. Interestingly, all the exon-intron borders of Hs *XRRA1*, Mm and Rn *Xrra1* genes are well conserved (Table 3). We again employed

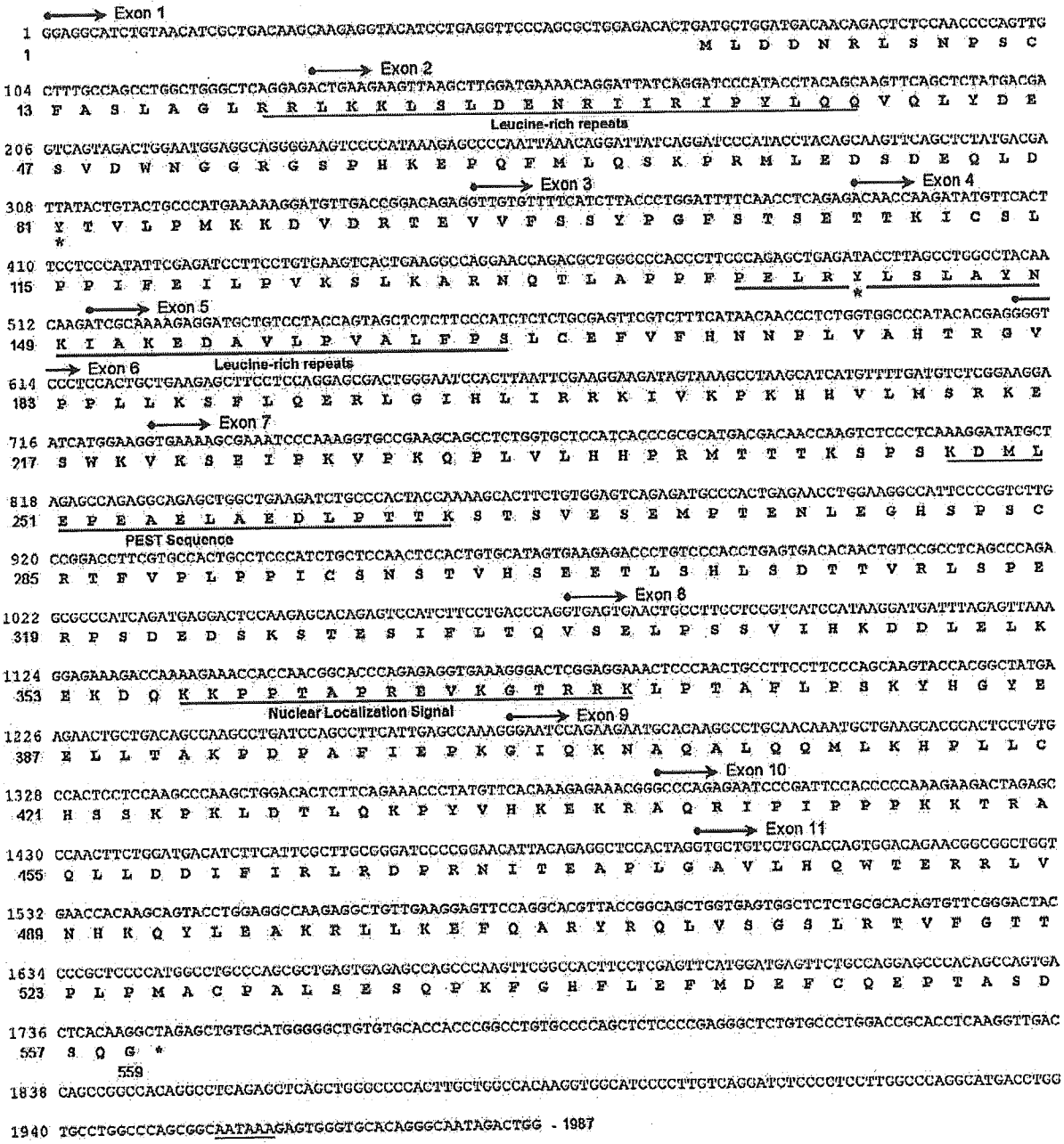

Figure 1
Human *XRRA1* cDNA sequences and its translation. The gene comprises 11 exons, and produces 559 aa protein. XRRA1 protein consists several motifs i.e. leucine rich repeats, PEST sequence, nuclear localization signal, and two sites of tyrosine phosphorylation (*). Poly A-signal, AATAA is underlined.

Figure 2
Promoter region of human *XRRA1* gene as predicted by GrailExp [11]. Two putative sites of TATA box were found at -506 and -400 from the start codon. Predicted transcription factor binding sites revealed several members of bZIP family such as C/EBPa, c-Jun, c-Fos, AP-1 and CREB.

EBOA to look for all possible ESTs of mouse and other organisms (Table 2). We used the longest human EST containing *XRRA1* cDNA sequence as template for the analysis. From the mouse dbEST, we obtained 13 ESTs as primary sequences, which assembled Mm *Xrra1* (GenBank accession no. BK000542). However, unlike the GrailEXP analysis, Mm *Xrra1* lacks exon one and part of exon two. Nevertheless, these two approaches verified the cDNA sequence for Mm *Xrra1*. We could not find any publicly available ESTs for rat. Other ESTs found to be homologues to *XRRA1* were from pig (*Sus scrofa*, Sc), horse (*Equus caballus*, Eq) and bovine (*Bos taurus*, Bt). We assembled Bovine (Bt) *Xrra1* (GenBank accession no. BK000644) from 5 ESTs. The Bt *Xrra1* gene is likely to have 11 exons with truncated exons 4 and 8 (Table 2, Fig. 3A).

We used cDNA libraries from macaque (*Macaca fascicularis*, Mf) brain and testis to obtain Mf *XRRA1*. Since we have created and sequenced these libraries, we were able to look for the Mf *XRRA1*. We found only three testis clones (QtsA-12093, 10089 and 20433) from 10,400 testis and 53,000 brain cDNAs that were homologous to Hs *XRRA1*. One of the clones, QtsA-20433 (GenBank

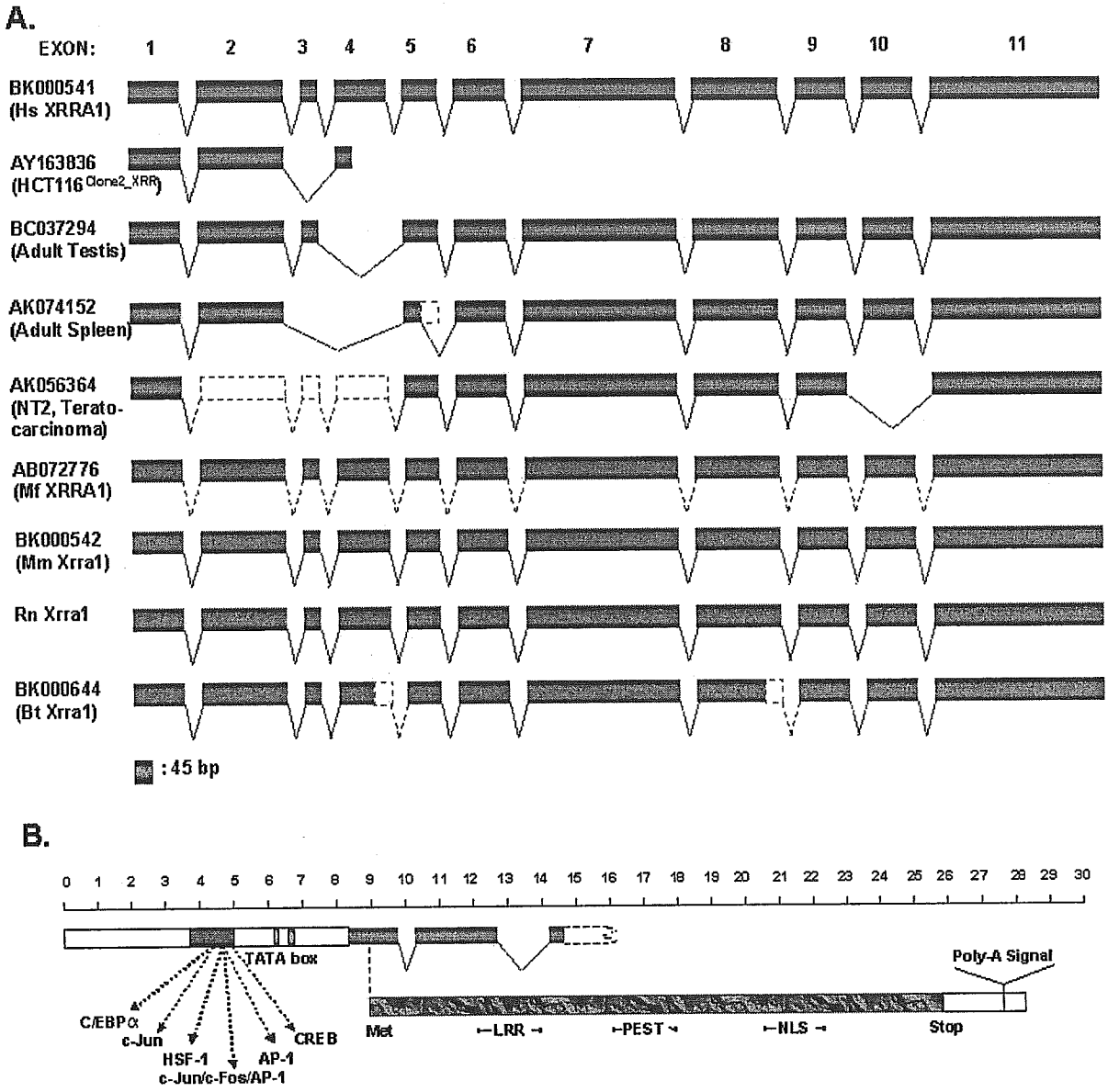

Figure 3
Representation of the *XRRA1* genomic structure. (A) The *XRRA1* gene spans 64 kb and comprises 11 exons. HCT116^{Clone2_XRR} cells contain 2 splice variants, one of which lacks exon 3 and is truncated. Other cDNA clones such as BC032794 lack exon 4, AK074152 lacks exon 3–4 and possible partial deletion of exon 5, and AK056364 lacks exon 2–4 and 10. The gene is located between genomic markers D11S916-D11S911 (80.1–84.2 cM) or at position 625.9 cR of NCBI RH Map. (B) The assembled Hs *XRRA1* comprises a putative promoter region, start/stop codons and poly-A signal, and predicted a protein with motifs such as LRR, PEST, and NLS. Scale bar is multiplied by 100.

Table 2: ESTs of Mm, Rn, Bt, Sc, and Eq *Xrra1* with corresponding nucleotides and exons to Hs *XRRA1*

Organism	EST	Corresponding Nucleotides of Hs *XRRA1*	Corresponding Exons of Hs *XRRA1*	Genomic Clone	Exons
Mus musculus Xrra1 (1903 nt)	BI990168.1	312 – 776	2 – 7	NW_000328	1 – 11
	AV046823.1	660 – 1011	6 – 7		
	BE650319.1	1017 – 1093, 1215–1414	7 – 8, 8 – 10		
	BU610007.1	1215 – 1383, 1400 – 1491	8 – 9, 10		
	AA144723.1	1215 – 1414	8 – 10		
	AI647383.1	1215 – 1383, 1400 – 1491	8 – 9, 10		
	BE956870.1 (+/-)	1437 – 1491, 1529 – 1599, 1668 – 1737, 1933–1966	10, 11		
	BQ840215.1	1436 – 1599	10 – 11		
	BE631695.1	1529 – 1599	11		
	BQ175761.1 (+/-)	1529 – 1599, 1668 – 1737, 1933 – 1966	11		
	BE956083.1 (+/-)	1529 – 1599, 1668 – 1737, 1933 – 1966	11		
	AI449753.1 (+/-)	1668 – 1737, 1933 – 1966	11		
	AW125246.1 (+/-)	1933 – 1966	11		
Rattus norvegicus *Xrra1* (1909 nt)				AC127923.2	1 – 11
Bos taurus Xrra1 (1759 nt)	BM106446.1	1 – 422	1 – 4		
	BI540915.1	542 – 703	5 – 6		
	BE722755.1	797 – 1248	7 – 8		
	BM254248.1	1308 – 1760	9 – 11		
	BE751905.1	1366 – 1761	9 – 11		
Sus scrofa Xrra1 (970 nt)	BF442510.1	407 – 708	4 – 6		
	BF442490.1	407 – 703	4 – 6		
	BF444411.1	407 – 514, 1208 – 1269	4, 8 – 9		
	BI341438.1	581 – 810, 877 – 978, 1006 – 1079	5 – 7, 7, 7 – 8		
Equus caballus Xrra1 (621 nt)	BM735121.1	1336 – 1575, 1658 – 1761, 1902 – 1970	9 – 11, 11, 11		

accession no. AB072776) had the full-coding sequence and showed strong identity to Hs *XRRA1*, sharing 96 and 95% identities in cDNA and protein sequences, respectively. SIM4 [13] and BLAST analysis showed that Mf *XRRA1* sequence dispersed properly into 11 exons on human chromosomal band 11q13 (Fig. 3A).

XRRA1 is conserved among human, macaque, mouse, rat, pig and bovine

Multiple alignments of both *XRRA1* cDNA and protein sequence from human, macaque, mouse, rat, pig, and bovine *Xrra1*s showed a conservative molecule. Percent identities at both the nucleotide and protein level among these mammals resulting from two by two alignments are listed in Table 4. As expected, phylogenetically, human *XRRA1* is closer to macaque *XRRA1* and both of them are closer to bovine and pig *Xrra1*s than murine and rat *Xrra1*s. We failed to find possible *XRRA1* homologues in other organisms ranging from microorganisms, plants, and animals. The latter included organisms whose genomes have been sequenced completely such as yeast, plasmodium, fruit fly, zebra fish and fugu.

The Hs *XRRA1* gene produces a novel 559 aa protein (Fig. 3B). The XRRA1 protein contains two possible sites of leucine-rich repeats (aa 20–41 and aa 138–164), a PEST sequence at aa 247–264, two tyrosine phosphorylation sites at aa 81 and aa 142, and bipartite nuclear targeting sequences are found at aa 357–373 and aa 485–501. Besides human, Mf XRRA1, Mm, Rn, and Bt Xrra1s retain the above-mentioned motifs.

XRRA1 responded differentially to XR in clones of varying radiation responses

The RT-PCR results confirmed earlier observation from our cDNA microarray analysis that unirradiated HCT116^{Clone2_XRR} cells had lower basal expression of *XRRA1* than unirradiated HCT116^{Clone10} cells (Fig. 4, Table 5). However, *XRRA1* expression in HCT116^{Clone2_XRR} cells increased almost immediately (~10 minutes) after XR and decreased thereafter. By