The EMBO Journal vol.3 no.9 pp.2179–2183, 1984

Analysis of the gene coding for the murine cellular tumour antigen p53

Brigitta Bienz, Rina Zakut-Houri, David Givol and Moshe Oren

Department of Chemical Immunology, The Weizmann Institute of Science, Rehovot 76-100, Israel

Communicated by Y. Aloni

A genomic clone containing the functional gene for the murine p53 cellular tumour antigen was isolated and structurally characterised. The gene contains at least 11 exons and 10 introns, the first intron possessing a length of 6.1 kb. Attempts to determine the exact 5′ end of p53 mRNA were inconclusive, probably due to the presence of a remarkable stem and loop structure ($\Delta G° \approx -56$ kcal/mol) in the 5′ region of the gene. Suggestive similarities were found to exist between p53 and the protein product of the *myc* oncogene.

Key words: genomic clone/p53 tumour antigen/stem and loop structure/transformation protein

Introduction

The p53 cellular tumour antigen is a protein found in elevated levels in a variety of transformed cells (reviewed in Klein, 1982). Recently, a series of murine p53-specific cDNA clones have been isolated and structurally characterised (Oren and Levine, 1983; Oren *et al.*, 1983; Zakut-Houri *et al.*, 1983). The nucleotide sequence of these clones predicted the primary structure of the protein, whose size was found to be ~43 500 daltons. Analysis of genomic DNA by Southern blotting revealed that the mouse possesses two p53-specific genes. In previous work, one of these was shown to be a processed pseudogene (Zakut-Houri *et al.*, 1983), implying that the other member of the family encodes the protein. Employing somatic cell hybrids, the functional p53 gene was mapped to mouse chromosome 11, while the pseudogene was shown to reside in chromosome 14 (Czosnek *et al.* 1984; Rotter *et al.*, 1984). Analysis of polyadenylated RNA has demonstrated that the mature p53 mRNA possesses a size of ~2.0 kb (Oren *et al.*, 1983).

We now describe the detailed analysis of the functional gene, isolated from normal BALB/c liver DNA. The gene contains at least 11 exons and probably possesses a stable stem and loop structure in its 5′ region. Analysis of the predicted protein sequence points at the possible existence of differences between p53 from normal and some transformed tissues, as well as of suggestive similarities between p53 and the *myc* protein.

Results

Sequence analysis of the functional p53 gene

The overall organisation of the functional murine p53 gene was previously determined by heteroduplex analysis of clone Ch53-16, derived from Abelson murine leukaemia virus (MuLV) transformed C57 L/J cells (Zakut-Houri *et al.*, 1983). This information, as well as the comparison of restriction endonuclease maps between the genomic and cDNA clones, formed the basis for the selection of the sequencing strategy, as shown in Figure 1. The sequencing was performed on the genomic clone Ch53-7, derived from normal BALB/c liver DNA. The p53 gene is composed of 11 distinct exons which span a DNA region of ~12 kb, as compared with a mature p53 mRNA size of 2.0 kb (Oren *et al.*, 1983). All the exons, parts of the introns and the immediate 5′- and 3′-flanking regions of the gene were sequenced. The DNA sequence of the gene, the predicted protein sequence and the locations of the introns are shown in Figure 2.

Comparison of the sequence of the gene (Figure 2) with the previously pubished cDNA (Zakut-Houri *et al.*, 1983) reveals several differences some of which are due to inaccuracies in the published cDNA sequence, resolved upon resequencing of the corresponding cDNA regions (B. Bienz, unpublished data). Three differences, however, were found to represent true alterations between the genomic and the cDNA clones. These altered nucleotides are at positions 404, 503 and 702, as indicated in Figure 2. Each of these substitutions implies an amino acid difference between the p53 protein sequences predicted from the cDNA and the genomic DNA. Two of them are non-conservative changes: from glycine to glutamic acid (nucleotide 503) and from isoleucine to methionine (nucleotide 702), whereas the third is a conservative one: from alanine to valine (nucleotide 404). In addition to the differences between the genomic and the cDNA clones there are also two sequence alterations in the 5′-flanking region of the gene as compared with the published sequence. They most probably reflect strain-specific changes between the DNA of C57 L/J and BALB/c mice. A few additional differences can also be observed between the 3′ end of the gene and that of the p53 pseudogene, probably due to evolutionary diversification following the chromosomal integration of the processed pseudogene.

The sequence around the splice sites and the precise lengths of the introns are shown in Figure 3. In general there is good agreement with the consensus sequence (Breathnach and Chambon, 1981). The sizes of the larger introns were calculated from detailed restriction enconuclease mapping. It is noteworthy that the first intron, immediately followed by the initiation codon ATG, has a remarkable length of 6.1 kb.

Characterisation of the 5′ region of the p53 gene

To determine the 5′ end of the functional p53 gene, S1 nuclease mapping was carried out. For this purpose, RNA from IB-9 cells was annealed with the 5′ end-labeled *Ban*I-*Sac*I fragment of the genomic DNA (nucleotides −63 to −457, Figure 2). This fragment was chosen because it includes the 5′ end of the longest available cDNA clone pp53-176 (Zakut-Houri *et al.*, 1983), located at position −157. Unexpectedly, when S1 nuclease digestion was performed under standard conditions (37°C to 45°C, Berk and Sharp, 1977) no prominent protected fragment could be seen (data not shown). However, when the digestion was perform-

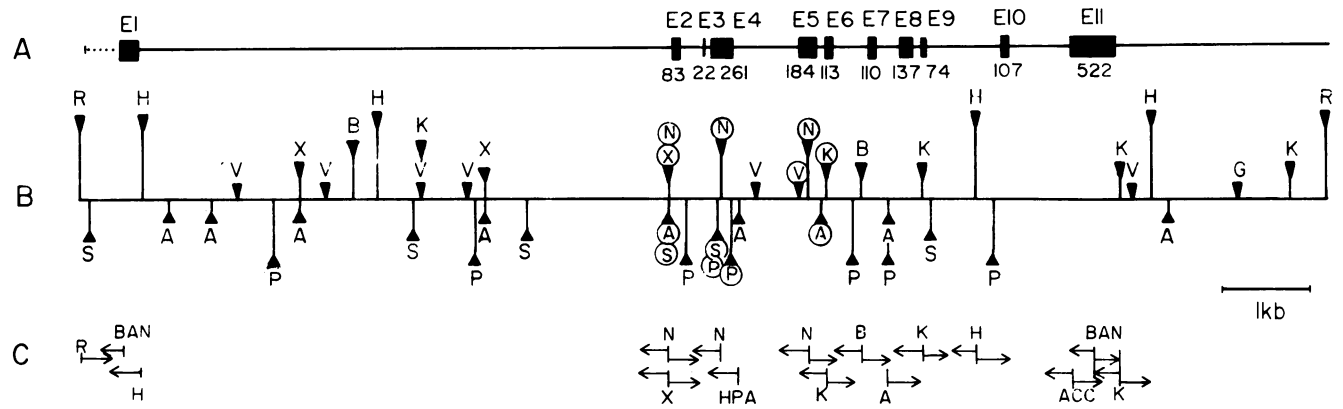

Fig. 1. (**A**) Overall structure of the genomic clone Ch53-7. Exons are shown as boxed regions, introns as lines. The numbers below the boxes represent the exact sizes (in bp) of the corresponding exons. (**B**) Restriction endonuclease map of the genomic clone Ch53-7. Sites in the genomic DNA which are retained in the cDNA are circled. A, *Aval*; B, *Bam*HI; G, *Bgl*II; H, *Hind*III; K, *Kpn*I; N, *Nco*I; P, *Pst*I; R, *Eco*RI; S, *Sac*I; V, *Pvu*II; X, *Xho*I. (**C**) Sequencing strategy used for the genomic clone Ch53-7. ACC, *Acc*I; BAN, *Ban*I; HPA, *Hpa*II. DNA fragments were end-labeled at the indicated site and recut with an appropriate second restriction enzyme.

ed at 24°C or at 0°C, distinct DNA fragments could be discerned (Figure 4). The length of these protected fragments depended on the S1 nuclease digestion temperature. At 24°C, DNA was protected mainly up to nucleotide position −112, resulting in a major fragment which was ~50 nucleotides long. At 0°C, predominant DNA protection was up to nucleotide position −156, giving rise to a major fragment of ~94 nucleotides. At both temperatures an additional minor DNA fragment was detected, which resulted from an S1 nuclease protection of the DNA up to nucleotide position −216. This minor fragment possesses the approximate length of 154 nucleotides. These findings suggest that the 5′ region of p53 mRNA possesses unique structural features which dramatically affect the results obtained by S1 nuclease analysis. The nature of this structure and its possible implications are discussed below.

Discussion

This work describes the detailed structural analysis of the functional murine p53 gene. The gene is composed of a relatively high number of exons (at least 11), 10 of which contribute to the translated region of the mRNA. A careful comparison between the p53 protein sequence predicted from the previously described cDNA clones and that predicted from the genomic clone (Figure 2) reveals three amino acid substitutions: alanine to valine (nucleotide 404), glycine to glutamic acid (nucleotide 503) and isoleucine to methionine (nucleotide 702). These differences could merely represent cloning artefacts or could reflect true sequence differences. In the latter case it is unclear whether such alterations may affect in any way the activity of the protein.

So far, practically nothing is known about the function of p53 and its relationship, if any, to other cellular proteins. However, since p53 is considered a transformation associated protein, it was of interest to see whether it possesses any features in common with known oncogenes. Based on the currently available information, the best candidate appeared to be the product of the *myc* oncogene, since both proteins accumulate in the nuclei of transformed cells (Donner *et al.*, 1982; Rotter *et al.*, 1983), both bind DNA (Donner *et al.*, 1982; D. Lane, personal communication), both are cell cycle correlated (Reich and Levine, 1984; Kelly *et al.*, 1983; Campisi *et al.*, 1984) and both appear to be induced at an early stage following mitogenic stimulation of resting cells (Milner and Milner, 1981; Mercer *et al.*, 1982, 1984; Kelly *et al.*, 1983). A possible relationship between p53 and the *myc* protein was already suggested by Weiss (1982); we therefore compared the primary structure of p53 with that of the murine *myc* protein (Bernard *et al.*, 1983). Although no statistically significant homology could be demonstrated at the level of the whole molecules, several suggestive similarities could be detected. Thus both proteins possess a relatively acidic N-terminal domain and a basic C-terminal domain. Furthermore, as shown in Figure 5, there is also an apparently similar pattern in the positioning of basic residues in the C-terminal domain between the two proteins, as well as a weak homology between a 5′-proximal region of p53 and a stretch of acidic residues in the *myc* protein. A structural relationship has recently been demonstrated between the products of the *myc, myb* and E1A genes (Ralston and Bishop, 1983). All of these, like p53, are proteins located in the nuclei of transformed cells (Abrams *et al.*, 1981; Donner *et al.*, 1982; Jochemsen *et al.*, 1982; Ralston and Bishop, 1983; Rotter *et al.*, 1983). The *myc* and E1A proteins also appear to serve similar functions in the process of malignant transformation (Land *et al.* 1983; Ruley, 1983). In this respect, it is noteworthy that the best conservation between the E1A and *myc* products is exhibited within the same acidic region that also shares structural features with p53 (Figure 5). We have recently been able to show that a possible functional homology may in fact exist between p53 and the *myc* product, as revealed in experiments employing the transformation of primary rat embryo fibroblasts (Land *et al.*, 1983; Ruley, 1983; D. Eliahu *et al.*, in preparation). It is therefore tempting to speculate that some or all of the features pointed out in Figure 5 may indeed be pertinent to the mode of action of p53.

Two approaches, S1 nuclease analysis and primer extension experiments, were taken to determine the precise 5′ end of the gene. Neither provided a definite answer to this question. Discrete, easily detectable protected fragments were obtained only when S1 nuclease digestion was performed at temperatures lower than those used routinely. The position of the major S1 nuclease cut at 24°C coincides with that of the major stop point observed in primer extension experiments (Zakut-Houri *et al.*, 1984). Nevertheless, it probably does not represent the 5′ end of the p53 mRNA, since the cDNA clone

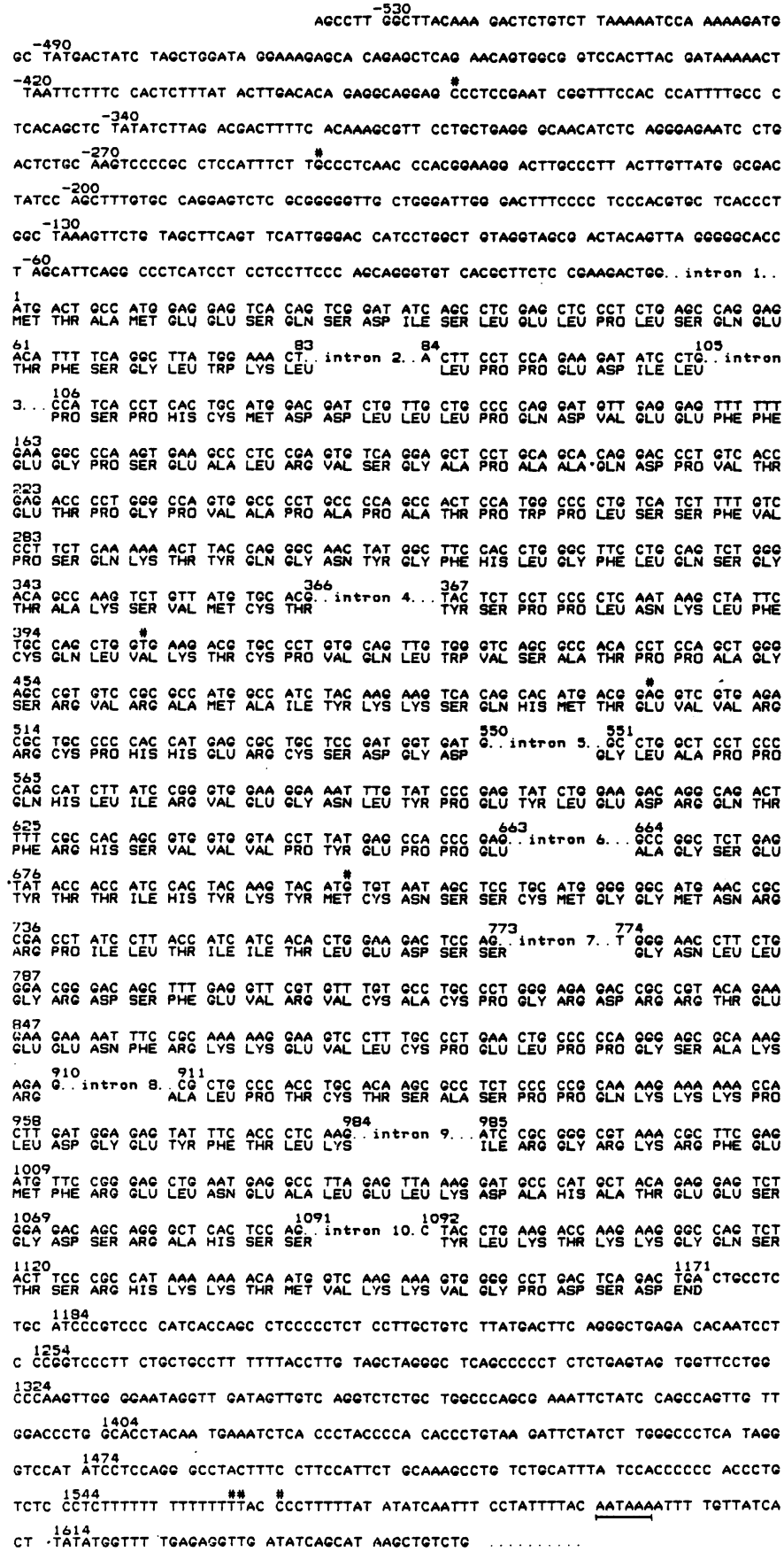

Fig. 2. DNA sequence of the genomic clone Ch53-7. The DNA sequence is given in parallel with the predicted amino acid sequence of the protein p53. Numbers refer to the positions of the nucleotides starting from the first nucleotide of the initiation codon ATG. Interruptions in the sequence mark the positions of the introns. The polyadenylation signal is underlined. Nucleotide substitutions within the genomic DNA as compared with the previously published DNA sequence are indicated by # above the relevant nucleotides.

ACTGG : GTAAGTAATT ....$^{intron1}_{6100bp}$.... CTCTCTACAG : ATGAC

AAACT : GTGAGTGGAT ....$^{intron2}_{270bp}$.... TGTTTTCCAG : ACTTC

TCCTG : GTAAGGCCCA ....$^{intron3}_{90bp}$.... CCATCCACAG : CCATC

GCACG : GTGAGTGGGC ....$^{intron4}_{780bp}$.... TCTCTTCCAG : TACTC

TGATG : GTAAGCCCTC ....$^{intron5}_{80bp}$.... TTGCTCTTAG : GCCTG

CCGAG : GTCTGTAATT ....$^{intron6}_{390bp}$.... TCTTCCCCAG : GCCGG

TCCAG : GTAGGAAGGC ....$^{intron7}_{320bp}$.... TCCCGGATAG : TGGGA

GAGAG : GTACGCAGGC ....$^{intron8}_{80bp}$.... CCTTCCACAG : CGCTG

TCAAG : GTACCAAGGC ....$^{intron9}_{830bp}$.... TCTGCTGCAG : ATCCG

TCCAG : GTAAGTGGCC ....$^{intron10}_{700bp}$.... CTTCCTATAG : CTACC

$^{A}_{C}AG : GT^{A}_{G}AGT$ ....*consensus*... $^{C\,C\,C\,C\,C\,C}_{T\,T\,T\,T\,T\,T}N^{C}_{T}AG : G^{G}_{T}$

Fig. 3. Nucleotide sequence around the exon-intron junctions. Nucleotides deviating from the consensus sequence (shown at the bottom of the Figure) are marked with dots. The full sizes of the introns are given with an accuracy of ±5%

pp53-176 comprises nucleotides located upstream to this position. All these findings, namely the lack of a major S1 nuclease-protected fragment at 37°C, unexpected protection at lower temperatures and the failure to extend the cDNA all the way to the 5′ end of the p53 mRNA, are hard to explain unless the 5′ region of this RNA possesses some unusual structural feature. Indeed, close examination of the DNA sequence in this region reveals that a stem and loop structure can be easily formed, as shown in Figure 6. The predicted change in the standard free energy ($\Delta G°$) of such a hairpin structure is approximately −56 kcal/mol, when formed in the corresponding RNA (Tinocco *et al.*, 1983). When S1 nuclease digestion is performed at 24°C it is conceivable that the double-stranded RNA stem displaces the DNA from the DNA-RNA hybrid, resulting in an S1 nuclease cut (−112) downstream to the end of this stem. This effect can also account for the premature stop (−114) observed in primer extension experiments. At 0°C, it is possible that the displaced DNA also forms a stem and loop structure resulting in a preferential S1 nuclease cut (−156) in the middle of the single-stranded DNA loop. The existence of such a peculiar structure might affect the conformation of both the DNA and the transcribed RNA, depending on the temperature and the ionic strength. Such a structure, if stable under physiological conditions, may influence the stability of p53 mRNA or the efficiency at which it is translated (Kozak, 1980; Saito *et al.*, 1983).

The minor longer fragment observed at both temperatures could, in principle, represent an authentic RNA-protected DNA fragment. The end point of the DNA protection at nucleotide −216 cannot represent an intron-exon junction, since there is no splice acceptor sequence around this position. Employing the CAT assay system (Gorman *et al.*, 1982) we have recently obtained evidence suggesting the existence of a functional promoter near the 5′ end of the genomic insert,

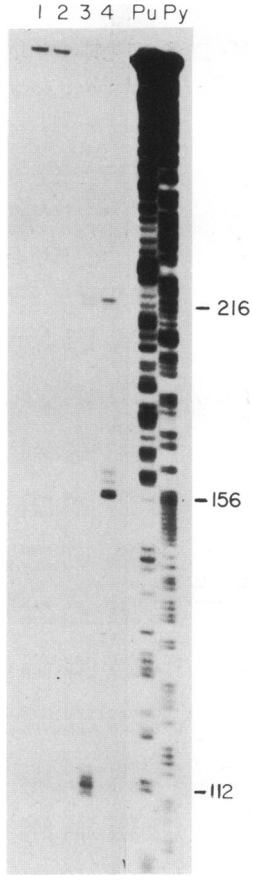

Fig. 4. S1 nuclease analysis in the 5′ region of the genomic clone Ch53-7. The *BanI-SacI* fragment derived from the 5′ region of Ch53-7 DNA was subjected to S1 nuclease analysis as described in Materials and methods. The radiolabeled DNA fragment was reannealed in the absence (**lanes 1** and **2**) or presence (**lanes 3** and **4**) of IB-9 RNA. S1 nuclease digestion was performed at 24°C (**lanes 1** and **3**) or 0°C (**lanes 2** and **4**). The S1 nuclease digestion products are shown in parallel with a sequence ladder generated by purine-specific (Pu) or pyrimidine-specific (Py) chemical degradation of the initial *BanI-SacI* fragment. Numbers on the right denote the locations of the main S1 nuclease cuts, deduced from the positions of the respective bands relative to the sequencing ladder (see Figure 2).

```
                I                    II                  III
p53   QDVEEFFEGPSE...KKKPLDGEYFTLKIRGRKRFEMFR...KTKKGQSTSRHKKTMVKKVGPDSD
       * **  *  *  ***          *  **   *  *        ****  **
myc   SDSEEEEQEDEEE...KRRTHNVLERQRRNELKRSFFALR...KLTSEKDLLRKRREQLKHKLEQLRNSGA
       * *** **
E1A   SDSEDEQDENGM
```

Fig. 5. Comparison between parts of the sequences of p53 and the products of the *myc* gene and the Ad12 E1A region. Shown in the Figure are the regions encompassing residues (from left to right) 48−59 (I), 316−339 (II) and 367−390 (III) of p53 and 250−261 (I), 355−378 (II) and 412−439 (III) of the mouse c-*myc* (Bernard *et al.*, 1983). Asterisks indicate sites at which an acidic (I) or a basic (II and III) amino acid is present in both proteins. For the region I the corresponding amino acid sequence of Ad12 E1A is given as well (residues 117−128). Arrows mark the C-terminal residue of p53 and *myc*.

between the *Eco*RI and *Hind*III sites (see Figure 1; B. Bienz unpublished results). Therefore the 5′ end of the p53 mRNA might indeed be located at around nucleotide −216. This interpretation is challenged by the absence of a properly positioned TATA box in this region, although the presence of a TATA box upstream to eucaryotic genes is not an absolute prerequisite (Weissman, 1980). If p53 mRNA indeed starts around position −216 it includes an additional initiation

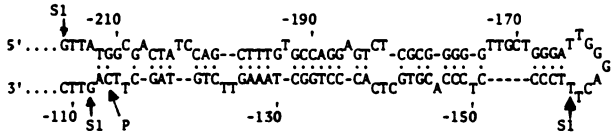

Fig. 6. Putative secondary structure of the region between nucleotides −216 and −108. The major cut points of the S1 nuclease (see Figure 4) are indicated (S1), as well as the major reverse transcriptase stop site observed in primer extension studies (P). Strong hydrogen-bonds are marked with dots. Gaps (indicated by dashes) are included to maximize the complementarity.

codon AUG (positions −213 to −211), preceding the one used for the synthesis of p53. Utilisation of this AUG could give rise to a peptide of only 31 amino acids, due to the presence of a termination codon at positions −120 to −118.

Materials and methods

Extraction and hybridisation of nucleic acids

The extraction of plasmid and bacteriophage DNA and of polyadenylated RNA from IB-9 tumours were described previously (Oren *et al.*, 1983).

Isolation of p53-specific genomic clones

BALB/c liver DNA was digested to completion with *Eco*RI and fractionated over a sucrose gradient (Maniatis *et al.*, 1978). Fractions enriched for the 16-kb fragment harbouring the functional p53 gene were detected by blotting and hybridisation to nick-translated pp53-176 DNA. A genomic library of such enriched DNA was constructed in bacteriophage Charon 4A and screened by plaque hybridisation (Benton and Davis, 1977) with the probe described above, which led to the isolation of clone Ch53-7.

Subcloning, restriction endonuclease mapping and DNA sequencing

Phage Ch53-7 DNA was digested with *Eco*RI and *Hind*III and ligated into *Hind*III- or *Hind*III plus *Eco*RI-digested, phosphatase-treated pBR322 DNA. The five resulting types of subclones were identified by comparing their insert lengths with the corresponding fragments of digested Ch53-7 DNA. Restriction enzyme fragments of phage and plasmid DNA were analysed by agarose gel electrophoresis, blotting and hybridisation to several nick-translated p53 cDNA fragments (Southern, 1975). The DNA sequence was determined by the chemical degradation method employing DNA fragments which were labeled at their 3′ ends by reverse transcriptase (Maxam and Gilbert, 1980).

S1 nuclease analysis

Plasmid DNA of the 0.7-kb *Eco*RI-*Hind*III subclone was purified over a Biogel A-50m column (Biorad), digested with *Ban*I and *Sac*I and treated with calf intestinal phosphatase. The 5′ end of the *Ban*I site was labeled with T4 polynucleotide kinase to an approximate specific activity of 2×10^6 c.p.m./μg DNA. The resulting 395 nucleotide long fragment was isolated from a 5% polyacrylamide gel. Approximately 10^5 c.p.m. were used for hybridisation with 100 μg total IB-9 RNA. Denaturation was at 80°C for 15 min followed by annealing at 49°C for 14 h in 30 μl hybridisation buffer as described by Berk and Sharp (1977). S1 nuclease digestion was carried out at either 0°C for 2 h with 1000 units of S1 nuclease (Miles), or at 24°C for 1 h with 1000 units of S1 nuclease, or at 37°C or 45°C for 30 min with 500 units of S1 nuclease. Products were analysed on an 8% polyacrylamide gel containing urea in parallel with a sequencing ladder (A+G and C+T reactions) of the initial *Ban*I-*Sac*I fragment.

Acknowledgements

The expert technical assistance of Orit Pinhasi is gratefully acknowledged. We would like to thank Dr. V. Rotter and D. Wolf for a gift of Ch53-16, S. Katzav for a gift of IB-9 tumours. This work was supported by grants from the Leukemia Research Foundation, Inc., The BatSheva de Rotschild Fund for the Advancement of Science and Technology and the Leo and Julia Forchheimer Center for Molecular Genetics. M.O. is a scientist in cancer research supported by The Rose and George Blumenthal Fellowship of the Israel Cancer Research Fund.

References

Abrams,H.D, Rohrschneider,L.R. and Eisenman,R.N. (1981) *Cell,* **29**, 427-439.

Benton,DW. and Davis,R.W. (1977) *Science (Wash.),* **196**, 180-182.

Berk,A.J. and Sharp,P.A. (1977) *Cell,* **12**, 721-732.

Bernard,O, Cory,S., Gerondakis,S,. Webb,E. and Adams,J.M. (1983) *EMBO J.,* **2**, 2375-2383.

Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.,* **50**, 349-383.

Campisi,J., Gray,H.S., Pardee,A.B., Dean,M. and Sonenshein,G.E. (1984) *Cell,* **36**, 241-247.

Czosnek,H.H., Bienz,B, Givol,D, Zakut-Houri,R., Pravtcheva,D.D., Ruddle,F.H. and Oren,M. (1984) *Mol. Cell. Biol.,* in press.

Donner,P., Greiser-Wilke,I. and Moelling,K. (1982) *Nature,* **296**, 262-266.

Gorman,C.M., Moffat,L.F. and Howard,B.H. (1982) *Mol. Cell. Biol.,* **2**, 1044-1051.

Jochemsen,H., Daniels,G.S.G., Herthoghs,J.J.L., Schreier,P.I., Van den Elsen,P.J. and Van der Ebb,A.J. (1982) *Virology,* **122**, 15-28.

Kelly,K., Cochran,B.H., Stiles,C.D. and Leder,P. (1983) *Cell,* **35**, 603-610.

Klein,G., ed. (1982) *Advances in Viral Oncology,* Vol., **2**, published by Raven Press, NY.

Kozak,M. (1980) *Cell,* **19**, 79-90.

Land,H., Parada,L.F. and Weinberg,R.A. (1983) *Nature,* **304**, 596-602.

Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connell,C., Quon,D., Sim,D.K. and Efstratiadis,A. (1978) *Cell,* **15**, 687-701.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.,* **65**, 490-560.

Mercer,W.E., Nelson,D., De Leo,A.B., Old,L.J. and Baserga,R. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 6309-6312.

Mercer,W.E., Avignolo,C. and Baserga,R. (1984) *Mol. Cell. Biol.,* **4**, 276-281.

Milner,J. and Milner,S. (1981) *Virology,* **112**, 785-788.

Oren,M. and Levine,A.J. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 56-59.

Oren,M, Bienz,B., Givol,D., Rechavi,G. and Zakut,R. (1983) *EMBO J.,* **2**, 1633-1639.

Ralston,R. and Bishop,J.M. (1983) *Nature,* **306**, 803-806.

Reich,N.C. and Levine,A.J. (1984) *Nature,* **308**, 199-201.

Reich,N.C., Oren,M. and Levine,A.J. (1983) *Mol. Cell. Biol.,* **3**, 2143-2150.

Rotter,V., Abutbul,H. and Ben-Zeev,A. (1983) *EMBO J.,* **2**, 1041-1047.

Rotter,V., Wolf,D., Pravtcheva,D.D. and Ruddle,F.H. (1984) *Mol. Cell. Biol.,* **4**, 383-385.

Ruley,H.E. (1983) *Nature,* **304**, 602-606.

Saito,H., Hayday,A.C., Wiman,K., Hayward,W.S. and Tonegawa,S. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 7476-7480.

Southern,E.M. (1975) *J. Mol. Biol.,* **98**, 503-517.

Tinoco,J., Borer,P.N., Dengler,B., Levine,M.D, Uhlenbeck,O.C., Crothers, D.M. and Gralla,J. (1973) *Nature,* **246**, 40-41.

Weiss,R. (1982) *Nature,* **299**, 9-10.

Weissman,S.M. (1980) *Mol. Cell. Biochem.,* **35**, 29-38.

Zakut-Houri,R., Oren,M., Bienz,B, Lavie,V., Hazum,S. and Givol,D. (1983) *Nature,* **306**, 594-597.

Zakut-Houri,R., Givol,D., Bienz,B., Rogel,A. and Oren,M. (1984) *CSH Conferences on Cell Proliferation and Cancer,* Vol. **11**, Cold Spring Harbor Laboratory Press, NY (in press).

Received on 15 March 1984; revised on 6 June 1984