The EMBO Journal vol.4 no.5 pp.1251-1255, 1985

Human p53 cellular tumor antigen: cDNA sequence and expression in COS cells

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

Department of Chemical Immunology, The Weismann Institute of Science, Rehovot 76100, Israel

Communicated by D.Givol

A 2.5-kb cDNA clone for human p53 tumor antigen has been isolated. This clone contains the entire coding region including 135 bp upstream of the first ATG. Comparison of the nucleotide sequence of human p53 and mouse p53 demonstrates that the first ATG in human p53 corresponds to the second ATG (codon No. 4) in mouse p53. The human p53 comprises 393 residues and is longer than the mouse p53 due to six additional codons present at the region corresponding to exon 4 of the mouse p53 gene. The DNA sequence homology between the coding regions of mouse and human p53 is 81% and the conservation of homology is not equally distributed along the molecule. When inserted into SV40-based expression vectors the human p53 cDNA successfully directs the production of a polypeptide with an apparent mol. wt. of 55 kd which can be precipitated by monoclonal antibodies to p53.

Key words: p53 tumor antigen/cDNA/expression/COS cells

Introduction

The p53 cellular tumor antigen was found in elevated levels in a variety of mammalian transformed cell lines and tumors (Crawford, 1983). Various experiments suggested that p53, a nuclear protein, may play a role in regulating the cell cycle, probably in the transition from G_0 to G_1 (Milner and Milner, 1981; Mercer *et al.*, 1982, 1984; Reich and Levine, 1984). Mouse p53-specific cDNA and the corresponding genes were characterized (Oren *et al.*, 1983; Zakut-Houri *et al.*, 1983; Bienz *et al.*, 1984; Pennica *et al.*, 1984; Jenkins *et al.*, 1984b). The data indicate that p53 is composed of 390 amino acids and possesses an acidic N-terminal domain and a relatively basic C-terminal one. When introduced into appropriate expression vectors both p53 cDNA and genomic DNA express authentic murine p53 upon transfection into COS cells (Pinhasi and Oren, 1984). Recently, it was shown that a p53 encoding gene can cooperate with activated Ha-*ras* in the transformation of rat embryonal cells (Eliyahu *et al.*, 1984; Parada *et al.*, 1984; Jenkins *et al.*, 1984a). To understand the structure-function relationship in p53 it is desirable to compare the mouse p53 sequence with that of other species. Furthermore, to study the relevance of p53 to human tumors it is highly advantageous to obtain human p53 clones. Matlashewski *et al.* (1984) reported the isolation and sequence of a human p53 cDNA clone and showed that the cDNA in this clone (p102) starts at a position homologous to codon 101 of mouse p53. We report here the isolation of a 2.5-kb long human cDNA clone which starts 135 bp upstream to the initiator ATG codon and contains the entire coding region as well as the 3'-untranslated region. This p53-specific DNA, when carried on SV40-based expression vectors successfully directs in COS cells the production of authentic human p53 with an apparent mobility of 55 kd. The sequence of the entire coding region when compared with that of mouse p53 demonstrates the existence of unequally conserved regions within this protein.

Results

Sequence divergence between human and mouse p53

The isolation of a human p53 cDNA clone (php53c1) from a cDNA library of SV40 transformed fibroblasts (Okayama and Berg, 1983) was described elsewhere (McBride *et al.*, in preparation). The plasmid vector used in the construction of this library contains a *Bam*HI site at each end of the cloning site. Digestion of php53c1 with *Bam*HI yielded a 0.55-kb fragment and a 2.1-kb fragment, only the latter hybridizing to the mouse p53 cDNA probe (McBride *et al.*, in preparation). Restriction enzyme analysis and partial DNA sequencing demonstrated that php53c1 is identical to plasmid p102 recently described by Matlashewski *et al.* (1984) except that it extends ~400 bp upstream to the 5' end of p102 and should therefore contain the entire coding region. To verify this we sequenced the 5' segment of php53c1 employing the strategy shown in Figure 1. Our sequence data overlap the first 17 bp of p102 (Matlashewski *et al.*, 1984) and extend 435 bp upstream to it.

Figure 2 provides the sequence of the entire coding region of human p53, including 135 bp upstream to the first ATG, and compares it with the mouse p53 cDNA and protein sequence (Zakut-Houri *et al.*, 1983; Bienz *et al.*, 1984).

Several aspects concerning the divergence and conservation of various regions in human and mouse p53 emerge from this comparison. The coding region of mouse p53 starts with Met-Thr-Ala-Met (nucleotides −9 to +3 in Figure 2), whereas in human p53 this first ATG is replaced by GTC (Val) at position −9 to −7 (Figure 2) and the human p53 sequence should yield Val-Thr-Ala-Met. Since there is no other ATG upstream to this position and a termination codon TGA is present at −49 to −47 (Figure 2), we conclude that the ATG coding for the N-terminal methionine of human p53 corresponds to the ATG coding for the second methionine of mouse p53 (nucleotides 1−3, Figure 2). We cannot exclude the possibility that the mouse p53 also begins at the second methionine (codon No. 4 in mouse p53; Zakut-Houri *et al.*, 1983). In fact, the recently proposed consensus sequence for eukaryotic initiation sites (CCA/GCCAUGG, Kozak, 1984) is compatible with the sequence around the second ATG codon (CTGCCATGG, nucleotides −5 to 4, Figure 2) in mouse p53, rather than with the sequences around the first ATG (TCCGGATGA, nucleotides −14 to −6, Figure 2). Hence, the mouse p53 may be three residues shorter (387 amino acids instead of 390) than described before (Zaku-Houri *et al.*, 1983). In Figure 2 and the following discussion we compare the human and mouse p53 assuming that both start with the same ATG (Figure 2), which would make human p53 (393 amino acids) six residues longer than mouse p53. It appears that all the six extra residues are localized in the region corresponding to exon 4 in the mouse gene (nucleotides 97−375, Bienz *et al.*, 1984) which

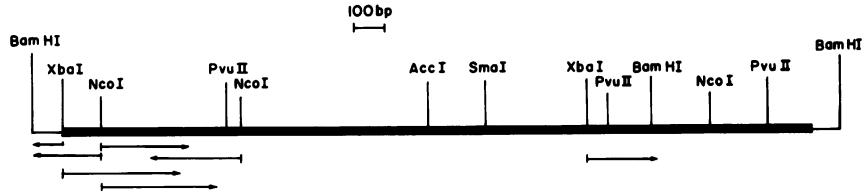

Fig. 1. Restriction enzyme map and sequencing strategy for human p53 cDNA clone php53C1. The *Bam*HI sites and the segments represented by thin lines at both ends of the drawing are from the vector (Okayama and Berg, 1983).

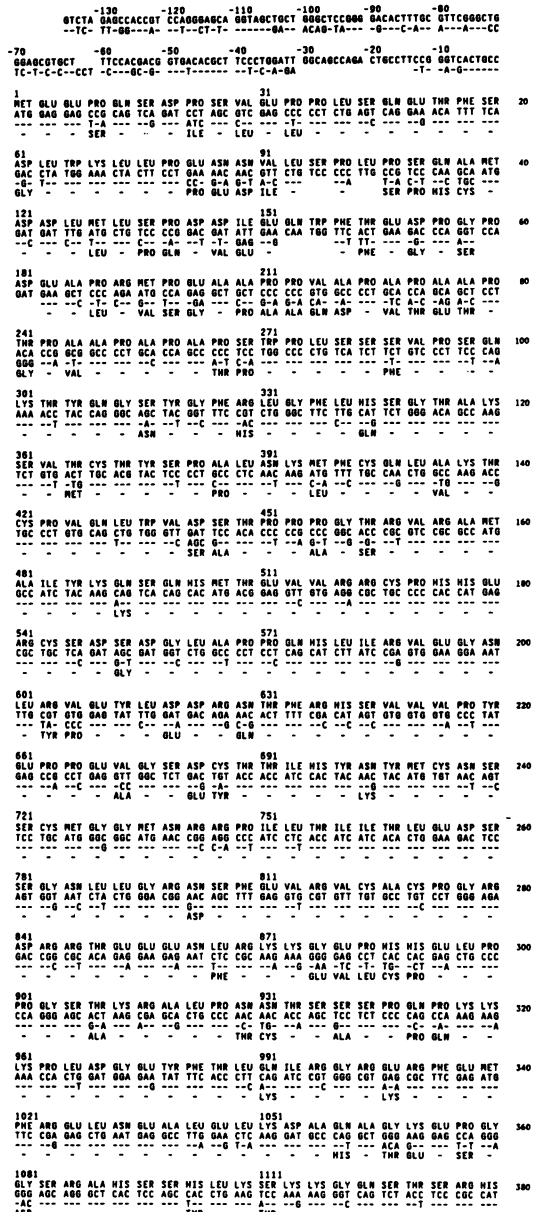

Fig. 2. Nucleotide sequence and deduced amino acid sequence of the entire coding region of human p53 cDNA in comparison with mouse p53 cDNA. Numbers on the right indicate amino acids in human p53 and numbers above the lines refer to nucleotides, starting at the A of the first ATG. Hyphens represent identical nucleotides or amino acids. Deletions were introduced into the sequence in order to obtain maximum homology between human and mouse p53. The sequence downstream to nucleotide 301 is taken from Matlashewski *et al.* (1984); in several sections this sequence was also confirmed by us (see Figure 1).

Table I. Amino acid composition of mouse and human p53

Amino acid	Human	Mouse
Ala	24	22
Arg	26	24
Asn	14	8
Asp	20	17
Cys	10	12
Gln	15	14
Glu	30	32
Gly	23	24
His	12	11
Ile	8	9
Leu	32	35
Lys	20	24
Met	12	10
Phe	11	13
Pro	45	38
Ser	38	35
Thr	22	23
Trp	4	3
Tyr	9	12
Val	18	21
Total	393	387

also appears to be the most divergent between the two species.

Table I compares the predicted amino acid composition of the two proteins. Human p53 contain seven more prolines, and this perhaps may be related to its slightly slower mobility on SDS-polyacrylamide gels (see later). The mouse p53 contains two additional cysteines (12 instead of 10 in human p53) and the sequence indicates that only nine of the mouse p53 cysteines are conserved in human p53; Cys 39, Cys 296 and Cys 311 in mouse are replaced by Ala, His and Asn, respectively, in man, whereas Cys 229 in human p53 is replaced by Tyr in mouse p53 (numbering as in Figure 2). In the coding sequence the homology between mouse and human p53 is 78% at the amino acid level and 81% at the DNA level. The substitutions, however, are not equally distributed along the molecule. The pattern of amino acid substitutions along the coding sequence was determined by calculating the percent amino acid replacements for contiguous blocks of 20 codons as shown in Figure 3. The divergence between human and mouse sequences is highest between codons 28 and 79 (Figures 2 and 3). In this region homology is only 59% and 37% at the DNA and protein level, respectively. Conservation of the precise sequence of this region, which corresponds to mouse exon 3 (a small exon of 7 codons) and a part of exon 4 is therefore likely to be less important for the biological function of p53. The region between codons 156 and 288 shows 87% and 92% homology at the DNA and protein level, respectively. It thus appears likely that the sequence conservation in this region (approximately one third of p53) reflects its importance for the function of this protein.

Expression of human p53 in COS cells

The mol. wt. of human p53, as predicted from the sequence shown in Figure 2, is 44 kd. This is substantially lower than the value

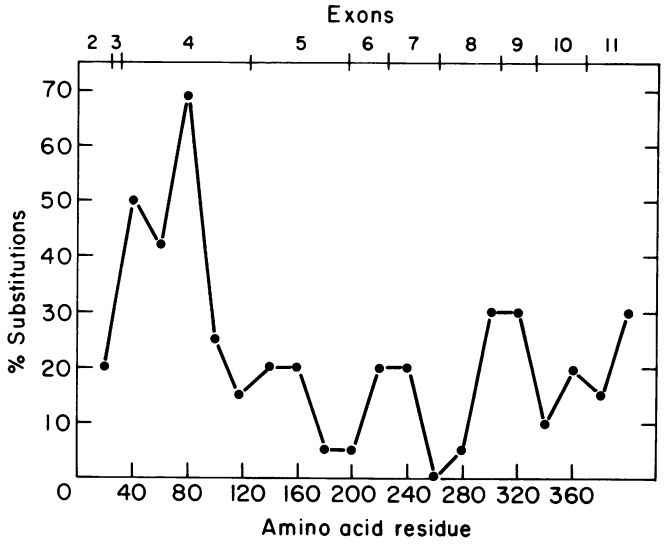

Fig. 3. Divergence plot of mouse and human p53. The percent amino acid substitution at each consecutive block of 20 amino acids was plotted along the p53 polypeptides. Deletions were not taken into account. Bars on top indicate the positions of the corresponding exons in the mouse p53 gene (Bienz *et al.*, 1984).

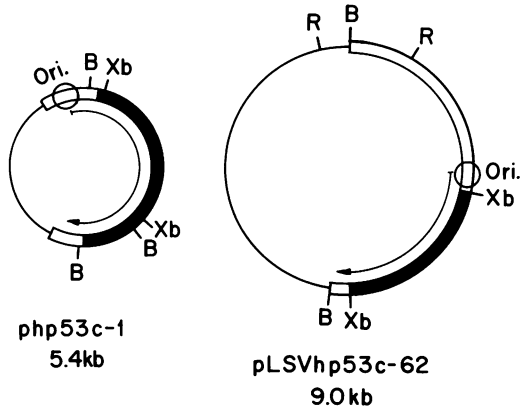

Fig. 4. Expression plasmids employed for the transfection of COS cells. php53c1 is the original cDNA clone, constructed in the Okayama-Berg cloning vehicle (Okayama and Berg, 1983). pLSVhp53c62 contains the internal *Xba*I-*Xba*I fragment of php53c1, inserted into the corresponding site in pLSV (see Materials and methods). Arrows indicate the orientation of transcription from the early SV40 promoter. Thin lines, pBR322 DNA; open bars, SV40 DNA; full bars, human p53 cDNA. Ori denotes the SV40 origin of replication. The following restriction enzyme sites are indicated: B, *Bam*HI; R, *Eco*RI; Xb, *Xba*I.

deduced from SDS-polyacrylamide gels, namely 53–56 kd (Simmons, 1980; Crawford *et al.*, 1981; Thomas *et al.*, 1983). This apparent discrepancy raised the possibility that, due to a cDNA cloning artifact or to mRNA heterogeneity, the protein sequence inferred from our cDNA clone was not fully identical with that of full-size p53. To rule out this possibility, it was necessary to determine whether the DNA stretch present in clone php53c1 was capable of encoding intact human p53. Since the cDNA was cloned in the Okayama-Berg expression vector, which utilizes the SV40 early promoter and origin of replication (Okayama and Berg, 1983), we transfected plasmids php53c1 (Figure 4) directly into COS cells (Gluzman, 1981), thereby allowing the efficient replication and expression of this plasmid. Labelled cell extracts were made from transfected cells, as well as from cells transfected with an SV40-based vector, pLSV (Laub *et al.*, 1983), devoid of any p53-specific sequences. Consequently, the extracts were precipitated with anti-p53 monoclonal antibody to PAb 421, which can recognize human p53 (Harlow *et al.*, 1981). As seen in Figure 5, lane 3b, transfection with php53c1 resulted in the overproduction of conspicuous amounts of a protein co-migrating with monkey p53 (lane 1b), as expected for human p53 (Simmons, 1980; Crawford *et al.*, 1981). The vector used in this experiment can provide an exogenous ATG resulting in a fusion protein (Okayama and Berg, 1983). The 55-kd band observed could in fact represent such a fusion protein rather than a polypeptide initiating at the first internal ATG implicated from the cDNA sequence. To circumvent this objection, we excised the 1.9-kb *Xba*I-*Xba*I fragment (see Figures 1 and 2) containing the entire putative protein coding region and transferred it into pLSV (Laub *et al.*, 1983). The latter plasmid, pLSVhp53c62 (Figure 4) relies exclusively on the use of an internal ATG for synthesis of the cDNA-encoded polypeptide. As seen in Figure 5, lane 4b, this actually resulted in a more efficient overproduction of the 55-kd polypeptide, further establishing the correctness of the deduced protein sequence. Lane 2b displays the results of transfection with a similar plasmid containing mouse p53 DNA, and is in agreement with previous data (Pinhasi and Oren, 1984). The apparently identical size of the monkey and human p53 species presents a problem, since one may argue that the transfection with a replicating SV40 vector could somehow increase the amount of endogenous simian p53. This objection, however, seems irrelevant since a plasmid containing exactly the same sequences except for the p53 cDNA yielded far less 55 kd p53 (lane 1b). Furthermore, analysis of SV40 large T antigen revealed no meaningful differences in the amounts of this protein between cells transfected with either the control pLSV (lane 1c) or with pLSVhp53c62 (lane 4c). This rules out the possibility that the observed results reflect increased accumulation of monkey p53 due to overproduction of large T antigen in pLSVhp53c62-transfected cells.

The marked difference between the apparent and predicted sizes of human p53 is most likely due to the abundance of proline residues within this molecule. This effect, previously observed for mouse p53 (Zakut-Houri *et al.*, 1983), is further accentuated in the human protein, since the latter possesses 45 prolines as compared with 38 in the murine protein.

Discussion

The comparison of the entire coding sequence of mouse p53 (Zakut-Houri *et al.*, 1983; Bienz *et al.*, 1984) and human p53 (Matlashewski *et al.*, 1984, and this paper) indicates that p53 cannot be considered highly conserved. The homology between the DNA in the coding regions is 81% and only 52% of the substitutions are silent, whereas 48% are replacement substitutions. The homology at the 5′-untranslated regions is 63%, but at the 3′-untranslated region the homology is markedly lower (Matlashewski *et al.*, 1984). The deduced protein sequence of human p53 is probably six residues longer than that of mouse p53 and by alignment of the two sequences to obtain maximum homology it appears that these are due to the addition of six codons internally without changing the reading frame. All these additions are in a region coresponding to exon 4 of the mouse gene (Bienz *et al.*, 1984). Another remarkable feature in this region is an exact direct repeat of 15 nucleotides GGCCCCTGCACCAGC which code for Ala-Pro-Ala-Pro-Ala and appear twice at nucleotides 219–233 and 249–263 (Figure 2), whereas in mouse p53 this sequence appears only

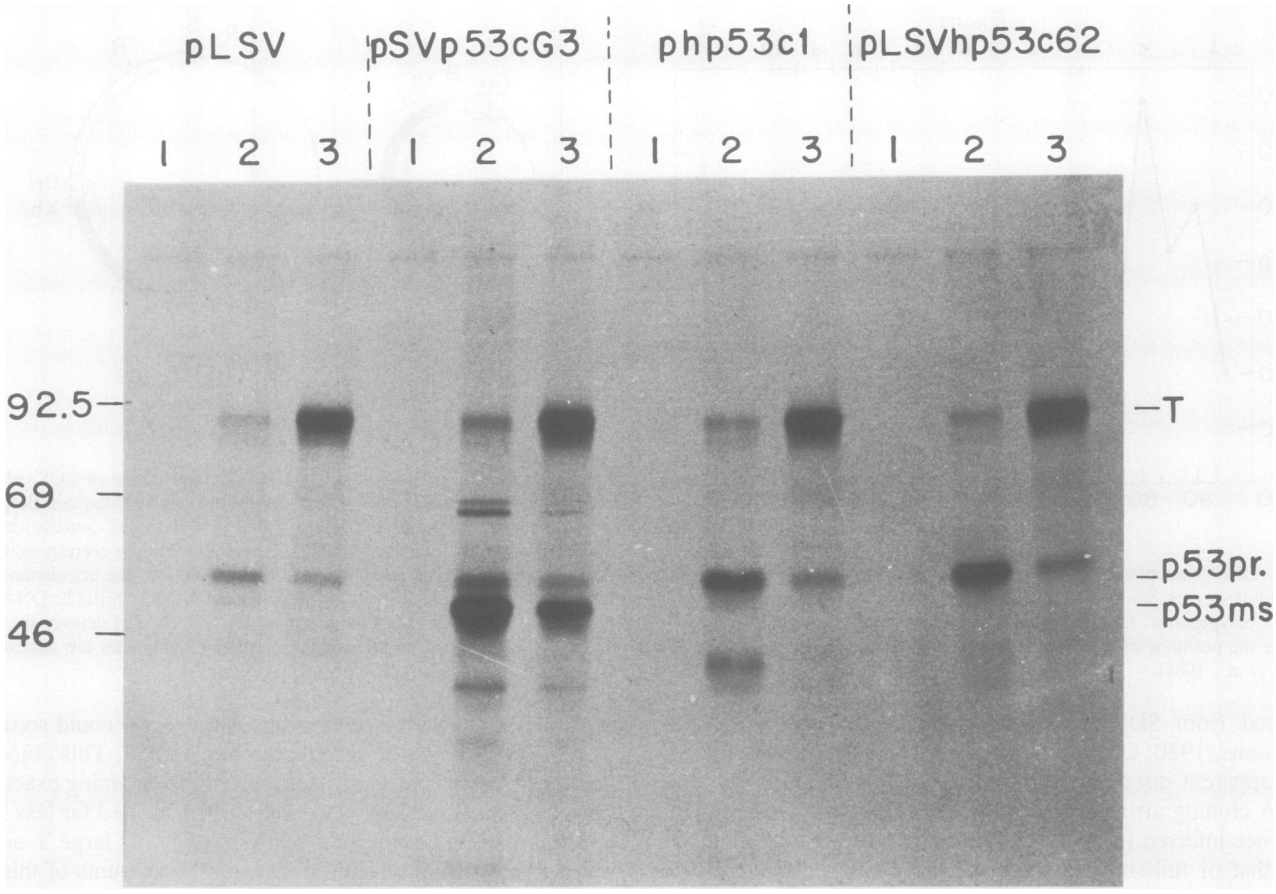

Fig. 5. Expression of p53 in transfected COS cells. Subconfluent COS cells were transfected with the indicated plasmids by the DEAE-Dextran method. 72 h later proteins were labeled for 4 h with [^{35}S]methionine (25 μCi/90 mm dish), extracted and immunoprecipitated with either control antibodies (**1**), anti-p53 monoclonal antibody PAb 421 (**2**) or anti-SV40 large T antigen monoclonal antibody clone 412 (**3**). Each reaction employed 2.1 x 10^6 c.p.m. of TCA-insoluble material. Numbers on the left denote the positions and sizes (in kd) of co-electrophoresed mol. wt. markers. T, SV40 large T antigen; p53 pr, primate p53; p53 ms, mouse p53.

once. The amino acid replacements along the predicted protein sequence are not equally distributed (Figure 3). It appears that two regions (codons 156–201 and 230–288) are highly conserved and within these 105 codons there are only four amino acid replacements (4%). On the other hand, the region containing codons 28–79 is the most divergent and contains 63% replacements of deduced amino acids, not counting the six extra codons of human p53 that are also present in this region. The rest of the coding sequence shows ~20% divergence. The present knowledge does not allow us yet to relate the conservation and divergence of the p53 sequence to functional regions of the molecule. However, as pointed out by Zakut-Houri *et al.* (1983), mouse p53 possesses a very acidic N-terminal domain, while its C-terminal part has a high proportion of basic residues. Both these features are well conserved in the human polypeptide. Thus, in the region extending between residues 319 and 393, human p53 consists of 27% basic residues (Lys, Arg, His) as compared with 29% in the murine analogue. Zakut-Houri *et al.* (1983) have suggested that this part of the molecule may be involved in the DNA binding activity of p53 (Lane and Gannon, 1983). Perhaps even more remarkable is the comparison of the N-terminal segment (residues 1–80). Although numerous amino acid differences exist between human and mouse p53 over this segment (see Figure 3), its acidic nature is strictly maintained. Accordingly, it contains 16 acidic residues and three basic ones in the murine polypeptide, *versus* 17 acidic and two basic residues in the human counterpart. These facts strengthen the case for the negative nature of the N-terminal domain being required for p53 function.

Both human and mouse p53 show a discrepancy between the mol. wt. values calculated from the DNA sequence and those determined by SDS-polyacrylamide gel electrophoresis. Although the significance of this finding is unclear, it should be noted that a similar behavior is exhibited by a number of nuclear transforming proteins, including the adenovirus E1a proteins (Gingeras *et al.*, 1982), the polyoma large T antigen (Soeda *et al.*, 1980), c-*myc* (Persson *et al.*, 1984; Hann and Eisenman, 1984) and *fos* (Verma *et al.*, 1984). It is therefore possible that this peculiar electrophoretic mobility is due to a common structural feature of all these polypeptides, which may reflect functional similarities.

This work also demonstrates the synthesis of full-size human p53 in COS cells transfected by an appropriate expression plasmid. Although it might be expected that the presence of higher levels of p53 in those cells would result in more of this protein complexed to the SV40 large T antigen, this does not appear to be the case (Figure 5, lane 3). Possibly, for some reason, the human p53 may not efficiently form complexes with this viral protein in COS cells. Alternatively, however, the data may imply that only a specific subset of T antigen molecules is capable of forming a relatively tight complex with primate p53, and that this subset is already saturated by monkey p53 in COS.

Finally, the ability to induce the expression of human p53 *via* transfection may prove very useful. Athough vectors directing

the expression of mouse p53 have been successfully used before, the relatively high degree of divergence between human and mouse proteins makes the latter less likely to function efficiently in human systems. Now that it is possible to over-express the human p53, one can start looking for the effects of p53 overproduction on the behavior of various types of human cells.

Materials and methods

Isolation and sequencing of human p53 cDNA clone php53c1

The isolation of the human p53 cDNA clone from the human fibroblast cDNA library (Okayama and Berg, 1983) was described elsewhere (McBride *et al.*, in preparation). Nucleotide sequence determination was carried out by the method of Maxam and Gilbert (1980).

Plasmids used for transfection

Plasmid php53c1 and plasmid pSVp53cG3 have already been described. The former contains human p53 cDNA cloned in the Okayama-Berg vector (McBride *et al.*, in preparation) and the latter is a chimera of mouse p53 cDNA and genomic DNA carried on an SV40 replacement vector (Pinhasi and Oren, 1984). To construct pLSVhp53c62, a 1.85-kb *Xba*I fragment was excised from php53c1. This fragment, including the entire coding region of human p53 (see Figure 2), was inserted into the polylinker of plasmid pLSV (Laub *et al.*, 1983; Pinhasi and Oren, 1984), which positioned it between the early promoter and the early polyadenylation site of SV40 (Figure 4).

Transfection and protein analysis

COS cells were transfected by the DEAE-dextran method as described previously (Pinhasi and Oren, 1984), employing 20 μg of plasmid DNA per 90 mm dish. Incubation with DNA was for 7 h. Transfected cells were labeled, 72 h post-transfection, with 25 μCi per dish of [^{35}S]methionine (Amersham International) in 2.5 ml of methionine-free medium containing 2% dialysed fetal calf serum. Following 4 h labeling, cells were harvested and labeled proteins extracted and analysed as described before (Pinhasi and Oren, 1984). The following monoclonal antibodies were employed: PAb 421 (L21), reactive with both rodent and primate p53 (Harlow *et al.*, 1981) and clone 412, directed against the SV40 large T antigen (Gurney *et al.*, 1980).

Acknowledgements

We are indebted to Dr. H.Okayama for the DNA sample of his library. We thank Shula Hazum and Orit Pinhasi-Kimhi for expert technical assistance. This work was supported in part by grants from the Leukemia Research Foundation of Chicago and the Leo and Julia Forcheimer Center for Molecular Genetics.

References

Bienz,B., Zakut-Houri,R., Givol,D. and Oren,M. (1984) *EMBO J.*, **3**, 2179-2183.
Crawford,L.V. (1983) *Int. Rev. Exp. Pathol.*, **25**, 1-50.
Crawford,L.V., Pim,D.C., Gurney,E.G., Goodfellow,P. and Taylor-Papadimitriou,J. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 41-45.
Eliyahu,D., Raz,A., Gruss,P., Givol,D. and Oren,M. (1984) *Nature*, **312**, 646-649.
Gingeras,T., Sciaky,D., Gelinas,R., Bing-Dong,J., Yen,C., Kelly,M., Bullock,P., Parsons,B., O'Neill,R. and Roberts,R. (1982) *J. Biol. Chem.*, **257**, 13475-13491.
Gluzman,Y. (1981) *Cell*, **23**, 175-182.
Gurney,E.G., Harrison,R.O. and Fenno,J. (1980) *J. Virol.*, **34**, 752-763.
Hann,S.R. and Eisenman,R.N. (1984) *Mol. Cell. Biol.*, **4**, 2486-2497.
Harlow,E., Crawford,L.V., Pim,D.C. and Williamson,N.N. (1981) *J. Virol.*, **39**, 861-869.
Jenkins,J.R., Rudge,K. and Currie,G.A. (1984a) *Nature*, **312**, 651-654.
Jenkins,J.R., Rudge,K., Remond,S. and Wade-Evans,A. (1984b) *Nucleic Acids Res.*, **12**, 5609-5626.
Kozak,M. (1984) *Nature*, **308**, 241-246.
Lane,D.P. and Gannon,J. (1983) *Cell Biol. Int. Rep.*, **7**, 513-514.
Laub,O., Rall,L.B., Truett,M., Shaul,Y., Standring,D.N., Valenzuela,P. and Rutter,W.J. (1983) *J. Virol.*, **48**, 271-280.
Matlashewski,G., Lamb,P., Pim,D. Peacock,J., Crawford,L.V. and Benchimol,S. (1984) *EMBO J.*, **3**, 3257-3262.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 490-560.
Mercer,W.E., Avignolo,C. and Baserga,R. (1984) *Mol. Cell. Biol.*, **4**, 276-281.
Mercer,W.E., Nelson,D., DeLeo,A.B., Old,L.J. and Baserga,R. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6309-6312.
Milner,J. and Milner,S. (1981) *Virology*, **112**, 785-788.
Okayama,H. and Berg,P. (1983) *Mol. Cell. Biol.*, **3**, 280-289.
Oren,M., Bienz,B., Givol,D., Rechavi,G. and Zakut,R. (1983) *EMBO J.*, **2**, 1633-1639.
Parada,L.F., Land,H., Weinberg,R.A., Wolf,D. and Rotter,V. (1984) *Nature*, **312**, 649-651.
Pennica,D., Goeddel,D.V., Hayflick,J.S., Reich,N.C., Anderson,C.W. and Levine,A.J. (1984) *Virology*, **134**, 477-482.
Persson,H., Henninghausen,L., Taub,R., DeGrado,W. and Leder,P. (1984) *Science (Wash.)*, **225**, 687-693.
Pinhasi,O. and Oren,M. (1984) *Mol. Cell. Biol.*, **4**, 2180-2186.
Reich,N.C. and Levine,A.J. (1984) *Nature*, **308**, 199-201.
Simmons,D.T. (1980) *J. Virol.*, **36**, 519-525.
Soeda,E., Arrand,J., Smolar,N., Walsh,J. and Griffin,B. (1980) *Nature*, **283**, 445-453.
Thomas,R., Kaplan,L., Reich,N., Lane,D.P. and Levine,A.J. (1983) *Virology*, **131**, 502-517.
Verma,I.M., Curran,T., Muller,R., van Straaten,F., MacConnell,W.P., Miller,A.D. and Van Beveren,C. (1984) in Vande Woude,G.F., Levine,A.J., Topp,.C. and Watson,J.D. (eds.), *Cancer Cells 2*, Cold Spring Harbor Laboratory, NY, pp. 309-321.
Zakut-Houri,R. Oren,M., Bienz,B., Lavie,V., Hazum,S. and Givol,D. (1983) *Nature*, **306**, 594-597.

Received on 13 March 1985