Volume 12 Number 14 1984 Nucleic Acids Research

Cloning and expression analysis of full length mouse cDNA sequences encoding the transformation associated protein p53

John R.Jenkins, Karran Rudge, Shelagh Redmond* and Alison Wade-Evans

Marie Curie Memorial Foundation Research Institute, The Chart, Oxted, Surrey RH8 OTL, UK

Received 17 May 1984; Revised and Accepted 3 July 1984

ABSTRACT

We have cloned and sequenced overlapping cDNA fragments which together encode the entire mouse protein p53. Using these cDNA's we have reconstructed the full length coding region for the protein, and have analysed its coding potential by expression in vitro, both as a full length sequence and as a subfragment contained in a fusion protein. The predicted amino acid sequence contains no obvious homologies to any known oncogenes but includes a possible tyrosine kinase acceptor site.

INTRODUCTION.

p53 is a cellular, transformation related phosphoprotein that is frequently present at higher levels in transformed cells than in the corresponding untransformed parental lines (1-4). p53 has been detected at elevated levels in cells transformed by viruses (5-6) chemicals and x-rays (2). In cells transformed by Simian Virus 40 (SV40) or adenovirus, p53 forms a specific, stable complex with a viral oncogene product (5-8). In mouse 3T3 cells p53 has a half life of around 30 minutes, while on transformation by SV40 this increases to more than 22 hours (9). Since the level of p53 specific mRNA in both cases is similar (9), the high level of this protein in the SV40 transformed cells is presumably due to its extended half-life. p53 synthesis is also significantly elevated in concanavilin A stimulated lymphocytes (10,11), and microinjection of p53 specific monoclonal antibodies into the nuclei of quiescent 3T3 cells inhibits the entry of these cells into S-phase following serum stimulation (12,13).

The association of p53 with cell proliferation has led to the suggestion that it resembles the hypothetical R-protein (14) regulating the commitment of cells to S-phase and division. This model proposes that a short half-life protein (R-protein) accumulates during G_1 and that cells become committed to S-phase when a critical level is reached. Stabilisation of such a protein could cause uncontrolled proliferation. Since p53 is stabilised in SV40

and adenovirus transformed cells and its expression at elevated levels is associated with cell proliferation, it may well play a role in the regulation of normal cell growth and in maintenance of the transformed phenotype.

In order to further study the control of p53 levels in the cell and its possible function(s) we have cloned cDNA sequences coding for the complete mouse p53 protein. We have expressed these sequences in vitro using an appropriate vector and show that they encode a polypeptide which migrates on SDS polyacrylamide gels with the appropriate (apparent) molecular weight of 53 K daltons, and generate epitope(s) recognised by several p53 specific monoclonal antibodies. Study of the nucleotide and predicted amino acid sequence reveals no obvious general homology with any known oncogene. We show that mouse p53 protein contains three clear domains and possesses a putative tyrosine kinase acceptor site.

MATERIALS AND METHODS

Enzymes

Restriction enzymes, T4 DNA ligase, E.coli RNA Polymerase Holoenzyme and synthetic linkers were obtained from New England Biolabs and used according to the manufacturer's instructions. DNA polymerase 1 (Klenow fragment) was obtained from P-L laboratories. RNasin was obtained from Biotec and Reverse transcriptase was a gift from J.W. Beard. S1 nuclease was obtained from Boehringer.

Plasmid Preparations

Supercoiled plasmid DNA was isolated using the alkali method of Birnboim and Doly (15) followed by equilibrium banding in caesium chloride.

cDNA Cloning and Plasmid Constructions.

c terminal clone p27.1a. Immunoselected Poly A^+ p53 mRNA prepared as previously described (16) was annealed to a single stranded 103 nucleotide DNA primer derived from clone 9, a short cDNA clone containing p53 specific sequences (16) in 50 mM Hepes pH 6.5 150 mM NaCl, 1 mM EDTA 0.1% SDS at 65° for 2 hours. Following the addition of Dextran (20 μg) the nucleic acids were precipitated by the addition of sodium acetate (NaOAc) to 0.3 M followed by 3 volumes of ethanol. After chilling to -70° for 5 minutes the reactions were spun for 5 minutes in a MSE microfuge. Following a reprecipitation from 0.3 M NaOAc/ethanol, the pellet was washed with absolute ethanol, air dryed and then taken up in 10 mM Tris HC1 pH 8.1, 1 mM EDTA (TE).

Reverse transcription of the hybrids was carried out using the

conditions described by Retzel et al (17) with the addition of RNasin (4U/λ) to the reaction. Reactions were terminated by the addition of Stopmix (5 M ammonium acetate 20 mM EDTA) to give a final ammonium acetate concentration of 2 M. Following repeated extractions with phenol chloroform (1:1) the nucleic acids were precipitated twice with ethanol as before.

The pellet was taken up in 50 λTE and digested for 60 minutes at 37° in the presence of 10 μg DNase free RNase (Boehringer). The reaction was terminated by the addition of Stopmix followed by two ethanol precipitations.

Second strand synthesis was performed using the conditions described by Wickens et al (18) followed by two ethanol precipitations as before.

S1 digestion of the double stranded cDNA (dscDNA) was performed as described (19).

End repair of the dscDNA was carried out under second strand synthesis conditions except that the Klenow fragment concentration was 40U/ml. Reactions were terminated by addition of Stopmix followed by two phenol chloroform extractions and ethanol precipitations.

Synthetic Bgl II and Pst I linkers were ligated to the dscDNA by the method of Kurtz and Nicodemus (20). Following extensive digestion with Bgl II and Pst I enzymes, the reaction was made up to 20% Formamide 0.01% Bromophenol blue 10 mM EDTA and loaded directly onto a 1% low gelling temperature agarose gel (LGT gel) Gel buffer was 40 mM Tris-Acetate, 2mM EDTA, 0.3 μg/ml ethidium bromide. After electrophoresis in the presence of λHind III DNA markers, gel fragments corresponding to the size ranges 2.5 kb - 1.5 kb, 1.5 kb - 0.5 kb, 0.5 kb - 0.2 kb were excised, Dextran (20 μg) was added and the dscDNA recovered by two Phenol extractions of the molten gel in 0.5 M ammonium acetate followed by two ethanol precipitations as before. The pAT 153 derived vector plasmid protem Bgl, (J.R. Jenkins and E. Harlow unpublished) which contains a Bgl II and a Pst I site in the region originally occupied by the RTEM-1 β-lactamase gene, was cut with these enzymes and the vector fragment isolated on an LGT gel as before. Linkered dscDNA was ligated to this vector fragment using conditions recommended by the T4DNA Ligase enzyme suppliers (Biolabs). Optimal DNA ratios were determined empirically. Ligated DNA was transfected into the host strain LE 392 (19) using the protocol of Hanahan and Meselson (21) and plated out on LB agar containing 10 μg/ml Tetracycline. Recombinant colonies were spotted on to nitrocellulose filters and screened (21) using as probe p^{32} labelled insert from clone 9 (16). Positive clones were grown up as 5 ml cultures in LB medium + 10 μg/ml Tetracycline, plasmid DNA isolated (15) and analysed by

restriction digest and agarose gel electrophoresis. A recombinant clone, p27.1a, was identified which contained a cDNA insert of approximately 900 bp. Subsequent analysis demonstrated that clone p27.1a contained sequences encoding the C terminal region of p53.

Protem Hind. This plasmid (Fig. 1) contains the (RTEM1) β-lactamase promoter and has a unique Hind III insertion site 2 bp downstream from the β-lactamase AUG initiation codon. It is a derivative of pAT 153 in which the Hind III site at position 29 has been eliminated by Hind III digestion, T4 DNA polymerase repair and religation. One of the Mbo II sites (position 3452) has also been eliminated (shown bracketed) by partial Mbo II digestion, T4 DNA polymerase repair and religation with a Hind III linker, thus generating a new, unique Hind III site at position 3452. This intermediate construct (not shown) was linearised by Pst I digestion (position 2903) and treated with SI nuclease to remove the carboxy-terminal region of the β-lactamase gene. A Pst I linker was ligated to the truncated product, thus generating a Pst I site at position 2710 (leaving 118 bp of the β-lactamase coding region). The β-lactamase coding region (all except terminal 118 bp) was removed by Hind III-Pst I digestion and replaced by a Hind III-Pst I fragment of polyoma virus acting as a stuffer fragment. This construct allowed us to insert DNA fragments between the Hind III-Pst I sites, whose expression would be under the control of the β-lactamase promoter, provided that the inserted DNA occupied the same reading frame as the β-lactamase AUG.

pSV10. This plasmid (Fig. 1) was generated by introducing the Hind III-Taq I fragment of SV40 DNA (nucleotides 5154-4722), containing the majority of the coding region for SV40t, into the Hind III-Cla I sites of pAT153.

pAli 4. pSV10 was cut by Eco RI digestion, repaired with T4 DNA polymerase, then ligated together with a Pst I linker (Fig. 1). The SV40t coding sequences were excised from SV10 by Hind III-Pst I digestion and transferred into the corresponding sites of Protem Hind (removing the polyoma stuffer fragment). In order to insert the SV40t coding sequences in the same reading frame as the β-lactamase AUG, it was necessary to modify the Hind III site in the construct. To do this the small fragment (203 bp) Alu I (isoschizomer of Hind III)-Eco RI from Protem Hind was ligated to a repaired Hind III site (using T4 DNA polymerase)-Eco RI vector fragment from pAli 4. This manipulation eliminated the Hind III site in pAli 4 by the addition of 2 bp (Fig. 1).

pAli 4,7 & pAli 4,3. These plasmid constructs (Fig. 1) were obtained from pAli 4 by Pst I digestion, T4 DNA polymerase repair and religation with a

pair of linkers, Hind III and Bgl II. This intermediate construct was then digested with Hind III and Bgl II, and a Hind III-Bgl II fragment of p53 cDNA from p27.1a (920 bp) was inserted. The resulting constructs were selected for insertion of the p53 gene in both orientations by restriction enzyme analysis, using Hind III-Eco RI double digests. pAli 4,7 has both SV40t and p53 genes in the same orientation, whereas pAli 4,3 has the p53 gene inserted in the opposing orientation.

pAli 4,73. This construct (Fig. 1) was obtained from pAli 4,7 by Hind III digestion, T4 DNA polymerase repair and religation, thus eliminating the Hind III site and placing the p53 gene in the same reading frame as the SV40t gene.

N terminal clone plb. A cDNA library for isolating p53 amino terminal sequences was constructed essentially as above, with the following differences.

(a) Total Poly A + mRNA was annealed with a primer prepared as follows: the cDNA insert from p27.1a was excised with Bgl II and Pst I enzyme, and isolated on a LGT agarose gel. Insert DNA was digested first with Pvu II, Acc I and Hae III, and then with exonuclease III using the conditions described by Smith (22) such that remaining DNA was single stranded. This single stranded DNA was then used to prime cDNA synthesis as before.

(b) Blunt ended dscDNA was treated with Eco RI methylase under conditions recommended by the suppliers (Biolabs) followed by two phenol chloroform extractions and two ethanol precipitations prior to linker addition.

(c) Eco RI and Hind III linkers were added instead of Bgl II and Pst I.

(d) The pAT 153 derived vector plasmid Protem Hind was used (Fig. 1) which contains an Eco RI and a Hind III site in the region originally occupied by the β-lactamase gene.

(e) Transfections were plated out at high density on PALL Biodyne filters and screened as before using p^{32} labelled p27.1a insert as a probe. Positive clones were picked and rescreened at low density, grown up as minipreps and tested for the presence of a Pst I site in the cDNA insert. Of a number of clones tested, one such clone designed plb was found on subsequent analysis to contain the region encoding the remaining amino terminal sequence of p53 (Fig. 7).

pP53-5. As shown in Fig. 3, clone plb DNA was partially digested with Pst I followed by complete digestion with Hind III. The appropriate DNA fragment corresponding to the 450 bp Hind II - Pst I fragment was isolated on a 1% LGT gel and ligated in approximately equimolar amounts to Pst I/ Bgl II ended p.27.1a insert and Hind III/Bgl II ended protem Hind vector

fragment. The ligation was transfected into Host Strain HB101 (19) and plated on LB agar containing 10 μg/ml tetracycline. Colonies were grown up as 5 ml minipreps in LB medium plus 10 μg/ml Tetracycline and plasmid DNA isolated for restriction analysis. A clone, pP53-5 was isolated which contained the appropriate construct.

In Vitro Transcription/Translation.

Caesium Chloride gradient purified supercoiled plasmid DNA's were transcribed with E.coli RNA polymerase holoenzyme using the conditions of Dandolo and Blangy (23) with the following modifications: reactions contained 150 mM potassium acetate instead of potassium chloride. Each reaction (100 λ) contained 2 μCi 5-H^3 CTP so that cRNA yields could be assayed by Trichloracetic acid precipitation. Transcription reactions were incubated for 2 hours at 37°C, terminated by the addition of 1/10 volume of Stopmix and precipitated by the addition of 3 volumes of absolute ethanol, followed by 5 minutes at -70°. Samples were thawed to room temperature before centrifugation for 5 minutes in an MSE microcentaur microfuge. After two further precipitations from 0.3 M Potassium acetate, 75% ethanol, samples were taken up in double distilled water and stored as aliquots under liquid nitrogen. Transcription products were heated to 100°C for 60 seconds before addition to translation reactions.

Translation of cRNA products was carried out in the mRNA dependent rabbit reticulocyte lysate system (24) using S^{35}-methionine as label. Reactions were incubated for 60 minutes at 32°C and included 5 mM 2-aminopurine.

Immunoprecipitation of in vitro translation products was carried out as follows: translation reactions were diluted with NET/BSA + 0.5 M NaCl, precleared with washed Staph A (BRL Immunoprecipitin) and incubated with the appropriate monoclonal antibody (in the form of supernatant) at 4°C overnight. The monoclonals were precipitated with 10% Staph A for 15 minutes at room temperature, then washed as follows:

1st wash - NET/GEL + 0.5 M NACL
2nd wash - NET/GEL + 0.1% SDS
3rd wash - 0.1% NP40, 10 mM Tris HC1 pH7.4
(NET/GEL: 0.15 M NACL, 0.05 M TRIS PH7.4, 0.005 M EDTA, 0.02% NaN_3, 0.05% NP 40, 0.25% gelatin
NET/BSA = NET/GEL + 0.2% BSA)

The resulting pellet was taken up in loading buffer containing β-mercaptoethanol and boiled for 5 minutes, then the supernatant was loaded

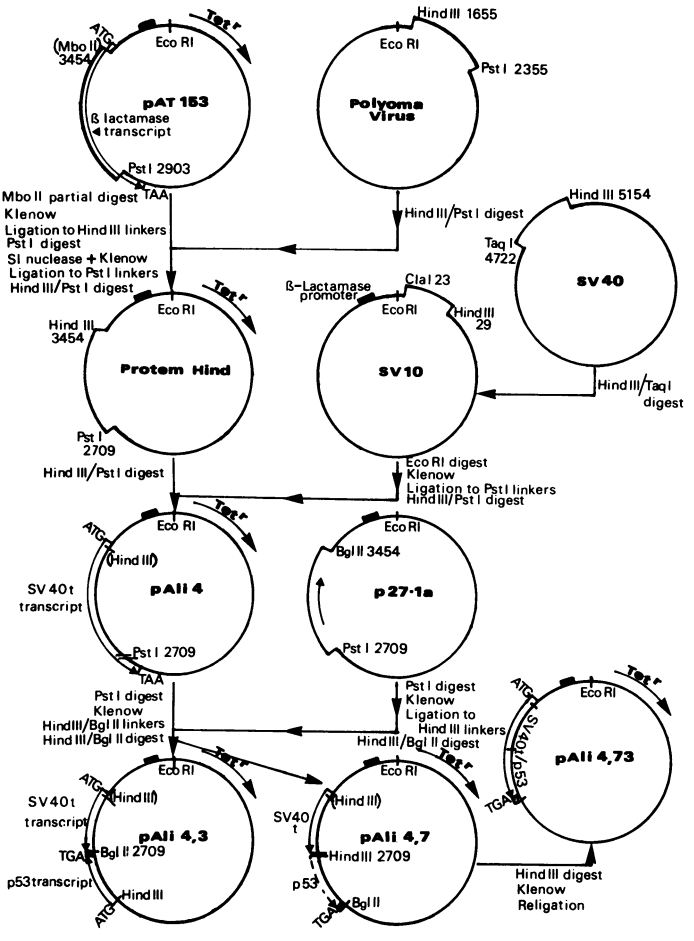

Figure 1. Construction of Plasmid Coding for fusion protein.
Protem Hind - a 700 bp fragment of polyoma virus (Hind III - Pst I) - as stuffer fragment inserted into Mbo II - Pst I site of pAt 153.
pAli - 4 - a 463 bp fragment of SV40 (Hind III - Taq I) cloned into Hind III - Pst I sites of Protem Hind. The insert contains all except 98 bp of the 3' coding region of small t.
pAli - 4,73 - a 920 bp fragment of mouse cDNA cloned into Pst I sites of pAli - 4. The insert contains two-thirds of p53 gene.
pAli - 4,3 - as pAli 4,73 except p53 gene is in wrong orientation.
pAli - 4,7 - as pAli 4,73 except p53 gene is out of sequence with ß-lactamase promoter of Protem Hind.

onto a 15% SDS polyacrylamide gel according to the method described by Laemmli (25) with the modification of being run at a constant voltage of 150V for a period of 5 hours. The gel was fixed for 1 hour in 45% methanol, 7% acetic acid, 30 minutes in ENLIGHTENING (NEN), dried at 80°C under

vacuum for 1 hour and the proteins visualised by autoradiography, using Kodak AR-5 X-Omat film exposed for 16 hours-5 days.

In Vivo Labelling.

Balb-C SV3T3 cells (Flow Laboratories) were labelled overnight with ^{35}S methionine (Amersham) and processed as described by Van Roy et al (26).

DNA Sequencing.

DNA sequences were determined using a combination of Maxam and Gilbert (27) and Sanger dideoxy chain termination (28, 29) techniques. The sequencing strategy is shown in Fig. 6.

RESULTS

Construction and analysis of plasmids encoding an SV40T/p53 fusion protein.

Clone p27.1a (Methods) was isolated using clone 9 (16) a mouse p53 specific clone, as probe, and preliminary sequence analysis confirmed that clone 9 and p27.1a shared common sequence. To investigate in more detail the coding potential of the p27.1a insert, this fragment was excised with Bgl II and Pst I enzymes and inserted downstream of sequences encoding an amino terminal fragment of SV40T/t (Fig. 1). These constructs contain a transcriptional unit in which the RTEM-1 β-lactamase promoter can be used with E.coli RNA polymerase holoenzyme to generate cRNA transcripts consisting of the β-lactamase 5' leader sequence and initiating AUG, SV40 T/t coding sequences in frame with the β-lactamase AUG, and p27.1a derived sequences immediately downstream of the SV40T/t sequences. Any open reading frame in the p27.1a insert, arranged in frame with the SV40 sequences, can be expressed as a fusion protein containing SV40T/t epitopes when this cRNA is translated in the reticulocyte lysate (Methods). p53 coding sequence inserted in the correct reading frame should give rise to a fusion protein containing SV40T epitopes and bona fide p53 protein sequence, with an apparent molecular weight greater than that of the SV40T/t product alone.

It was of interest to see whether such a fusion protein would display any of the p53 epitopes recognised by the available p53 specific monoclonal antibodies. Fig. 2 shows such a result. Tracks A5 and A10 contain S^{35}-methionine labelled products whose molecular weight (49.5 K daltons) is some 32.5 K daltons greater than the T/t peptide by itself (17 K daltons) and is recognised by both α T/t (pAB 419) (30) and αp53 (pAB 421) (30) monoclonal antibodies. The multiplicity of bands in the p53 immunoprecipitations (track A10) may be due to the fact that the monoclonal pAB 421 (30) recognises an epitope immediately adjacent to the carboxy terminus of the p53 derived

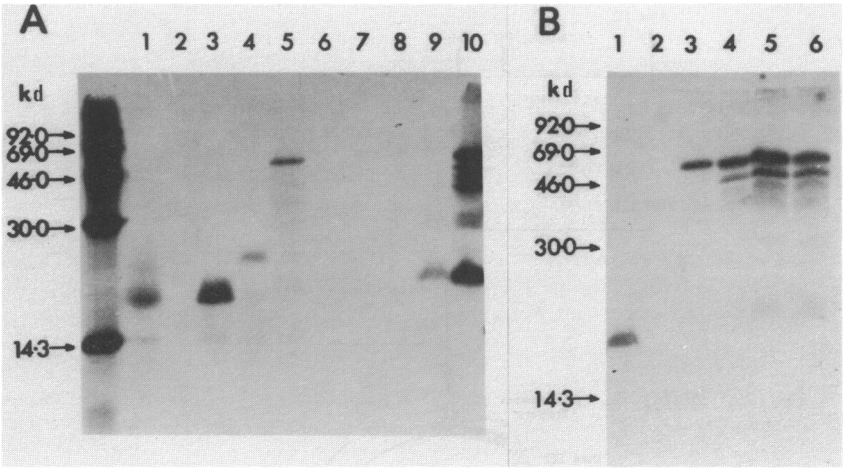

FIGURE 2. Analysis of p53 cDNA coding potential.

A. Immunoprecipitation of SV40 T/t-p27.1a fusion proteins, transcribed and translated in vitro from plasmid DNA, by monoclonal antibody pAb 419 (anti SV40 T/t). 1)pAli 4; 3)pAli 4,3; 4)pAli 4,7; 5)pAli 4,73; by pAb 421 (anti p53) 6)pAli 4; 8)pAli 4,3; 9)pAli 4,7; 10)pAli 4,73; and by pAb 416 (anti SV40 T only). 2) & 7)pAli 4.

B. Immunoprecipitation of in-vitro transcription-translation products of pAli 4 by 1)pAb 419; 2)pAb 416; and of pAli 4,73 by 3)pAb 419; 4)pAb 421; 5)pAb 410; 6)pAb 122.

peptide sequences (Wade-Evans and J.R. Jenkins manuscript in preparation), thus recognising what may be the translation products of internal initiation events. While the αT/t monoclonal pAB 419 (30) recognises an amino terminal epitope and thus only the full length fusion protein is precipitated. The band running at approximately 19 K daltons in tracks A9 and A10 is probably a product of translational initiation within the p53 coding region because it contains a p53 epitope(s) but no T/t epitopes. The fusion protein was also recognised by the monoclonals pAB 410 (30) and pAB 122 (31) (Fig. 2).

These results confirm the correct reading frame of p27.1a insert, and a subsequent comparison of this with the complete sequence (Fig. 7) showed that this reading frame was open from the 5' end (relative to the mRNA sequence) to a position 88 bases from the 3' end.

Therefore the coding sequences in p27.1a include the carboxy terminus of p53 and encode some 280 amino acids.

Preliminary DNA sequencing at the 5' end of the p27.1a insert (Fig. 7)

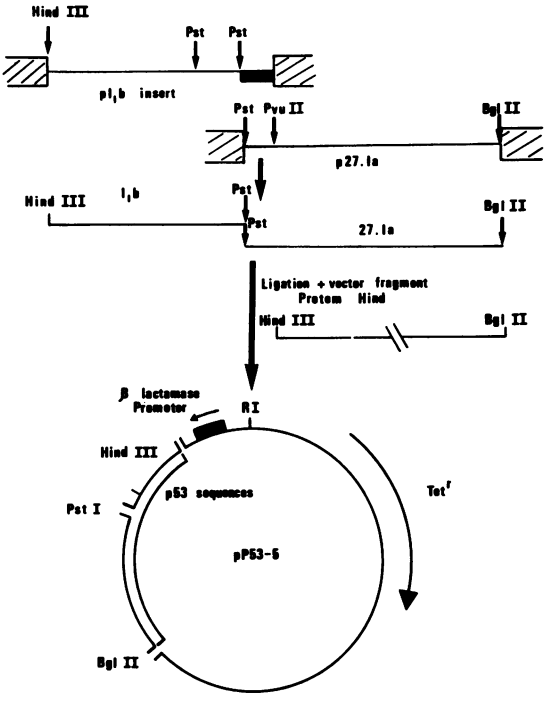

Figure 3. Construction of full length p53 coding sequence. Relevant restriction sites are indicated. Hatched boxes represent flanking vector DNA. Bold line indicates common sequence between pl1b and p27.1a inserts.

indicated that the Pst I site at the vector/insert junction was not derived from a synthetic linker, (linker sequence: $^{5'}$GCTGCAGC$^{3'}$, junction sequence $^{5'}$CTGCAGT$^{3'}$) and was almost certainly inherited from the p53 cDNA sequence. This observation prompted the screening procedure used to identify clone pl1b.

Construction and coding potential analysis of the full length sequence.

Clone pl1b (Methods) was derived from a p27.1a primed cDNA library using p27.1a as probe. Restriction analysis showed that the insert contained a predicted Pst I site (Methods and above) and preliminary sequence analysis showed that one end contained sequences common with p27.1a and in agreement with those expected from a p27.1a fragment mediated priming event.

To investigate in detail the combined coding potential of pl1b and p27.1a the two inserts were fused together at the appropriate Pst I site via a partial digestion (Fig. 3 and Methods) to give an uninterrupted sequence of 1377 bp. To check that this 1377 bp fusion construct was indeed colinear

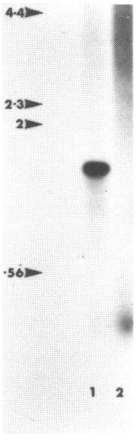

Figure 4. S1 analysis of the reconstructed p53 cDNA.
Tracks (i) insert in mp8 + PA^+ mRNA
(ii) insert in mp9 + PA^+ mRNA. S1 digests contained 1000 u/ml S1 nuclease, and were at 37° x 30 minutes.

with p53 mRNA, and contained no artifactual sequence rearrangements, the fragment with a Hind III site at its 5' end (relative to the coding strand) and a Bgl II site at its 3' end was cloned in M13 mp8 and mp9 using the Hind III and Bam HI sites of this vector (32).

Single stranded phage DNA was isolated, hybridised to mouse poly A + RNA, and digested with S1 nuclease (33). Hybrids were run out on an alkaline 1% agarose gel, transferred to nitrocellulose and probed with p^{32} labelled p27.1a insert. The result shows that a DNA fragment of approximately 1350 nucleotides is protected from S1 digestion (Fig. 4). Since the 5' end of plib contains a synthetic RI linker concatemer and a synthetic Hind III site, (Fig. 7) this corresponds to protection of the entire cDNA derived sequence. By this criterion the cDNA derived sequence in the fusion construct is colinear with p53 mRNA.

This fragment was also cloned into the in vitro expression vector protem Hind (Fig. 3). In this construction (pP53-5) the β-lactamase initiating AUG is out of frame relative to the p53 coding sequence and so any protein products immunoprecipitable with p53 specific monoclonal antibodies will be derived from translation initiations within the p53 specific sequences of the insert. The apparent molecular weight of such immunoprecipitable products can be used as an estimate of cDNA insert coding capacity relative to the bona fide p53 protein. cRNA from plasmid pP53-5 was translated in vitro as described (Methods) and the ^{35}S labelled translation products were immunoprecipitated with monoclonal antibody pAB 421 (30) and run on an SDS PAGE gel. In vivo ^{35}S labelled p53 from SV3T3 cells, immunoprecipitated with pAB 421, was run as a control. Fig. 5 shows the

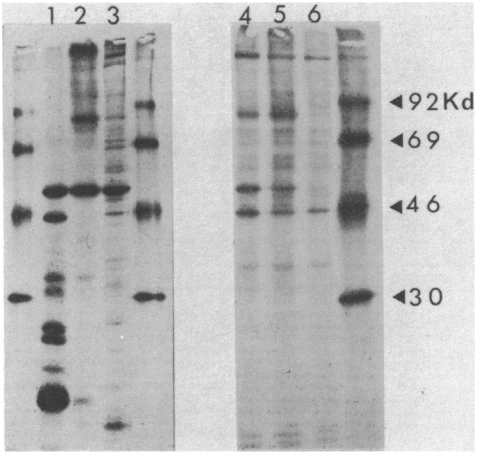

Figure 5. Comparison of pP53-5 *in vitro* translation products with *in vivo* labelled p53.
SV3T3 cells were labelled with ^{35}S-methionine and cell lysates immunoprecipitated (Methods). *In vitro* synthesis and immunoprecipitation of ^{35}S-methionine labelled pP53-5 translation products was as described (Methods).
(1) pP53-5 in vitro translation products immunoprecipitated with pAB 421 (anti-p53).
(3) in vivo labelled cell lysate vs pAB 421 (anti-p53).
(4) in vivo labelled cell lysate vs pAB 421 (anti-p53).
(5) in vivo labelled cell lysate vs pAB 419 (anti SV40 large T/t).
(6) in vivo labelled cell lysate vs normal mouse serum.

[Track 2 is a ^{32}P labelled cell lysate immunoprecipitated with pAB 421 (anti-p53)]

result. pAB 421, which recognises an epitope encoded in the C terminal, p27.1a derived, sequence (see above) brings down four major products, the largest of which migrates with an apparent molecular weight of 53 K daltons, and comigrates with *in vivo* synthesised mouse p53.

This strongly suggests that pP53-5 contains the complete coding sequence for mouse p53.

DNA sequence and coding capacity of the pP53-5 insert.

The DNA sequence of the reconstructed cDNA insert was established using the sequence strategy illustrated in Fig. 6 and as described in Methods. The DNA sequence is shown in Fig. 7. The insert contains 1377 bp and contains an open reading frame running from nucleotide 66 to nucleotide 1289. The first ATG codon in this reading frame starts at nucleotide 123 and the sequence between here and the termination codon at position 1290 encodes a polypeptide of 389 amino acids (Fig. 7) with a predicted molecular weight of

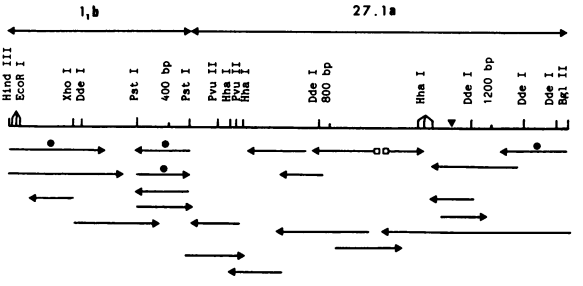

Figure 6. Partial Restriction map and sequencing strategy of p53 DNA. Arrows indicate sequenced DNA fragments (5' end labelled except those denoted by □ - which are 3' end labelled) determined by Maxam and Gilbert (27) apart from those marked by a dot where the method of Sanger (28,29) was employed.

43.48 K daltons. This reading frame is the same as that predicted for p53 from the SV40T/p53 fusion protein data above and (16). The ATG at position 123 is the first available initiation codon in this reading frame and presumably is the initiation site utilised for the 53 K dalton translation product of pP53-5 (Fig. 5). The discrepancy between the predicted molecular weight of 43.48 K daltons for this polypeptide and the apparent molecular weight of 53 K daltons for p53 protein is discussed below.

DISCUSSION.

The data presented show that we have constructed a functional cDNA gene that appears to contain the complete coding sequence for mouse p53. We show that this coding region generates a protein product which is immunoprecipitable with mouse p53 specific monoclonal antibodies and which comigrates on SDS PAGE gels with bona fide mouse p53.

The predicted amino acid composition of mouse p53 is shown in Table 1. An amino acid composition for the EBNA associated 53K dalton protein have been published elsewhere (34,35). The cDNA sequence derived amino acid composition for mouse p53 presented here is markedly different to this, adding weight to the suggestion (36) that the 53K dalton EBNA associated protein is not related to the SV40 T associated cellular protein p53.

A comparison of the amino acid sequence with that of the known oncogenes reveals no obvious general homologies. A nucleotide sequence for mouse p53 cDNA has been recently published (37). We have compared sequences and find that they differ somewhat. A variation between nucleotides 356-378 results in a predicted amino acid sequence of 390 residues compared to our 389.

```
        10          20          30          40          50          60          70
 GGAATTCCGG AATTCCGGAA TTCCATCCTG GCTGTAGGTA GCGACTACAG TTAGGGGGCA CCTAGCATTC
        80          90         100         110         120
AGGCCCTCAT CCTCCTCCTT CCCAGCAGGG TGTCACGCTT CTCCGAAGCT GG ATG ACT GCC
                                                          MET Thr Ala
    137                 152                 167                 182
ATG GAG GAG TCA CAG TCG GAT ATC AGC CTC GAG CTC CCT CTG AGC CAG GAG ACA
MET Glu Glu Ser Gln Ser Asp Ile Ser Leu Glu Leu Pro Leu Ser Gln Glu Thr

           197                 212                 227
TTT TCA GGC TTA TGG AAA CTA CTT CCT CCA GAA GAT ATC CTG CCA TCA CCT CAC
Phe Ser Gly Leu Trp Lys Leu Leu Pro Pro Glu Asp Ile Leu Pro Ser Pro His

242                 257                 272                 287
TGC ATG GAC GAT CTG TTG CTG CCC CGA GAT GTT GAG GAG TTT TTT GAA GGC CCA
Cys MET Asp Asp Leu Leu Leu Pro Arg Asp Val Glu Glu Phe Phe Glu Gly Pro

       302                 317                 332                 347
AGT GAA GCC CTC CGA GTG TCA GGA GCT CCT GCA GCA CAG GAC CCT GTC ACC GAG
Ser Glu Ala Leu Arg Val Ser Gly Ala Pro Ala Ala Gln Asp Pro Val Thr Glu

               362                 377                 392
ACC CCT GGC CAG TGG CCT GCC CCA GCC ACT CCA TGG CCC CTG TCA TCT TTT GTC
Thr Pro Gly Gln Trp Pro Ala Pro Ala Thr Pro Trp Pro Leu Ser Ser Phe Val

    407                 422                 437                 452
CCT TCT CAA AAA ACT TAC CAG GGC AAC TAT GGC TTC CAC CTG GGC TTC CTG CAG
Pro Ser Gln Lys Thr Tyr Gln Gly Asn Tyr Gly Phe His Leu Gly Phe Leu Gln

           467                 482                 497
ICT GGG ACA GCC AAG TCT GTT ATG TGC ACG TAC TCT CCT CCC CTC AAT AAG CTA
Ser Gly Thr Ala Lys Ser Val MET Cys Thr Tyr Ser Pro Pro Leu Asn Lys Leu

512                 527                 542                 557
TTC TGC CAG CTG GCG AAG ACG TGC CCT GTG CAG TTG TGG GTC AGC GCC ACA CCT
Phe Cys Gln Leu Ala Lys Thr Cys Pro Val Gln Leu Trp Val Ser Ala Thr Pro

       572                 587                 602                 617
CCA GCT GGG AGC CGT GTC CGC GCC ATG GCC ATC TAC AAG AAG TCA CAG CAC ATG
Pro Ala Gly Ser Arg Val Arg Ala MET Ala Ile Tyr Lys Lys Ser Gln His MET

               632                 647                 662
ACG GAG GTC GTG AGA CGC TGC CCC CAC CAT GAG CGC TGC TCC GAT GGT GAT GGC
Thr Glu Val Val Arg Arg Cys Pro His His Glu Arg Cys Ser Asp Gly Asp Gly

    677                 692                 707                 722
CTG GCT CCT CCC CAG CAT CTT ATC CGG GTG GAA GGA AAT TTG TAT CCC GAG TAT
Leu Ala Pro Pro Gln His Leu Ile Arg Val Glu Gly Asn Leu Tyr Pro Glu Tyr

           737                 752                 767
CTG GAA GAC AGG CAG ACT TTT CGC CAC AGC GTG GTG GTA CCT TAT GAG CCA CCC
Leu Glu Asp Arg Gln Thr Phe Arg His Ser Val Val Val Pro Tyr Glu Pro Pro

782                 797                 812                 827
GAG GCC GGC TCT GAG TAT ACC ACC ATC CAC TAC AAG TAC ATG TGT AAT AGC TCC
Glu Ala Gly Ser Glu Tyr Thr Thr Ile His Tyr Lys Tyr MET Cys Asn Ser Ser

       842                 857                 872                 887
TGC ATG GGG GGC ATG AAC CGC CGA CCT ATC CTT ACC ATC ATC ACA CTG GAA GAC
Cys MET Gly Gly MET Asn Arg Arg Pro Ile Leu Thr Ile Ile Thr Leu Glu Asp

               902                 917                 932
TCC AGT GGG AAC CTT CTG GGA CGG GAC AGC TTT GAG GTT CGT GTT TGT GCC TGC
Ser Ser Gly Asn Leu Leu Gly Arg Asp Ser Phe Glu Val Arg Val Cys Ala Cys

    947                 962                 977                 992
CCT GGG AGA GAC CGC CGT ACA GAA GAA GAA AAT TTC CGC AAA AAG GAA GTC CTT
Pro Gly Arg Asp Arg Arg Thr Glu Glu Glu Asn Phe Arg Lys Lys Glu Val Leu

           1007                1022                1037
TGC CCT GAA CTG CCC CCA GGG AGC GCA AAG AGA GCG CTG CCC ACC TGC ACA AGC
Cys Pro Glu Leu Pro Pro Gly Ser Ala Lys Arg Ala Leu Pro Thr Cys Thr Ser

052                 1067                1082                1097
GGC TCT CCC CCG CAA AAG AAA AAA CCA CTT GAT GGA GAG TAT TTC ACC CTC AAG
Ala Ser Pro Pro Gln Lys Lys Lys Pro Leu Asp Gly Glu Tyr Phe Thr Leu Lys

       1112                1127                1142                1157
ATC CGC GGG CGT AAA CGC TTC GAG ATG TTC CGG GAG CTG AAT GAG GCC TTA GAG
Ile Arg Gly Arg Lys Arg Phe Glu MET Phe Arg Glu Leu Asn Glu Ala Leu Glu

               1172                1187                1202
TTA AAG GAT GCC CAT GCT ACA GAG GAG TCT GGA GAC AGC AGG GCT CAC TCC AGC
Leu Lys Asp Ala His Ala Thr Glu Glu Ser Gly Asp Ser Arg Ala His Ser Ser

   1217                1232                1247                1262
TAC CTG AAG ACC AAG AAG GGC CAG TCT ACT TCC CGC CAT AAA AAA ACA ATG GTC
Tyr Leu Lys Thr Lys Lys Gly Gln Ser Thr Ser Arg His Lys Lys Thr MET Val

           1277                1292       1302       1312       1322       1332
AAG AAA GTG GGG CCT GAC TCA GAC TGA CTGCCTCTGC ATCCCGTCCC CATCACCAGC CTCCCCCTCT
Lys Lys Val Gly Pro Asp Ser Asp  .
```

Figure 7. Nucleotide and predicted amino acid sequence of mouse p53 cDNA. Clones were derived from mouse SV3A1E7 Poly A^+ mRNA.

Table I - Amino acid composition of p53 protein

Name	Number	Percentage
Alanine	23	5.9
Cysteine	12	3.1
Aspartic Acid	17	4.4
Glutamic Acid	32	8.2
Phenylalanine	13	3.3
Glycine	24	6.2
Histidine	11	2.8
Isoleucine	9	2.3
Lysine	24	6.2
Leucine	35	9.0
Methionine	11	2.8
Asparagine	8	2.1
Proline	37	9.5
Glutamine	14	3.6
Arginine	25	6.4
Serine	35	9.0
Threonine	24	6.2
Valine	19	4.9
Tryptophan	4	1.0
Tyrosine	12	3.1
STOP	1	.3
UNKNOWN	0	.0

We have sequenced this region by Maxam and Gilbert (27) and by M13 dideoxy techniques (28), in both orientations, and feel confident of our data.

Similarly we find that residues 265 (G/A), 597 (T/C), 621 (G/A), 706 (A/G) and 821 (G/T) are dissimilar.

The *in vitro* expression data (Fig. 2) show that the p53 specific sequences in p27.1a contribute 32.5K daltons of molecular weight (31.5K daltons predicted) to the T/t p27.1a fusion protein. Since p27.1a encodes some 72% of p53, it appears that the slow migration rate of the full length protein must be largely due to characteristics of the amino terminal 109 amino acids. This region is acidic and is proline rich (13.9%) compared to the complete protein (see above), which may be the reason why a 43.48K dalton polypeptide migrates as a 53K dalton species.

The protein sequence (Fig. 7) contains three clear domains. In addition to the acidic proline rich N terminal region (residues 1-75) there is a hydrophobic region (residues 75-149) and a basic C-terminal region (residues 275-389).

It has been reported that p53 is phosphorylated on serine and threonine residues *in vivo* (38). In the predicted amino acid sequence (Fig. 7) the tyrosine residues at positions 159 and 216 are preceded by arginine at

Table 2. Comparison of p53 amino acid sequences flanking tyrosine residue 159 with the tyrosine phosphorylation sites of tyrosine protein kinase.

	-8	-7	-6	-5	-4	-3	-2	-1		+1	+2	+3	+4	+5	+6
p53	S	R	V	R	A	M	A	I	Y	K	K	S	Q	H	M
p110 Fes	S	R	E	A	A	D	G	I	Y	A	A	S	G	G	L
p85 Fes	S	R	E	E	A	D	G	V	Y	A	A	S	G	G	L
p140 Fps	S	R	Q	E	E	D	G	V	Y	A	S	T	G	G	M
pp60v-Src	A	R	L	I	E	D	N	D	Y	T	A	R	Q	G	A
p90 Yes	A	R	L	I	E	D	N	E	Y	T	A	R	Q	G	A

positions 152 and 209. This conforms with the widely occurring -7 arginine residue at the site of tyrosine phosphorylation in viral transforming proteins (39-42).

The tyrosine residues at positions 225 and 323 are also preceded by the acidic residues aspartic acid or glutamic acid and Tyr 323 is preceded by lysine at position -7. A more detailed comparison (Table 2) reveals that Tyr 159 is embedded in a 15 amino acid string showing more widespread correspondence. Of the 7 common residues (excluding Tyr) 5 are identical to their equivalents in p85/p110 fes (Table 2). Tyrosine phosphorylation is a rare modification in vivo (42,43) that has been implicated in growth control (44,45) and specifically in the immediate events following stimulation by epidermal growth factor (46). p53 appears to play an obligate role in the early events following serum stimulation of quiescent cells (12,13) and it will be interesting to see if further study reveals tyrosine phosphorylation as a step in the action of p53 in normal or transformed cells.

ACKNOWLEDGEMENTS

We thank Kathy Barrett for skilled technical assistance, Veronique Poirier for providing data in Fig. 5, and we are grateful to the Imperial Cancer Research Fund for allowing access to their computing facilities.

*Visiting worker

REFERENCES

1. Crawford, L.V., Pim, D.C., Gurney, E.G., Goodfellow, P. and Taylor-Papadimitriou, J. (1981) Proc. Natl. Acad. Sci. USA 78, 41-45.

2. DeLeo, A.B., Jay, G., Appella, E., Dubois, G.C., Law, L.W. and Old, L.J. (1979) Proc. Natl. Acad. Sci. USA 76, 2420-2424.
3. Rotter, V. (1983) Proc. Natl. Acad. Sci. USA 80, 2613-2617.
4. Rotter, V., Boss, M.A. and Baltimore, D. (1981) J. Virol. 38, 336-346.
5. Lane, D.P. and Crawford, L.V. (1979) Nature 278, 261-263.
6. Linzer, D.I.H. and Levine, A.J. (1979) Cell 17, 43-52.
7. McCormick, F. and Harlow, E. (1980) J. Virol. 34, 213-224.
8. Sarnow, P., Ho, Y.S., Williams, J. and Levine, A.J. (1982) Cell 28, 387-394.
9. Oren, M., Maltzman, W. and Levine A.J. (1981) Mol. Cell Biol. 1, 101-110.
10. Milner, J. and McCormick, F. (1980) Cell Biol. Int. Rep. 4, 663-667.
11. Milner, J. and Milner S. (1981) Virology 112, 785-788.
12. Mercer, W.E., Nelson, D., DeLeo, A.B., Old, L.J. and Baserga, R. (1982) Proc. Natl. Acad. Sci. USA 79, 6309-6312.
13. Mercer, W.E., Avignolo, C. and Baserga, R. (1984) Mol. Cell Biol. 4, 276-281.
14. Pardee, A.B., Medrano, E.E. and Rossow, P.W. (1981) The Biology of Normal Growth, p. 59-69, Ed. M. Ritzen. Raven Press, New York.
15. Birnboim, H.C. and Doly, J. Nucleic Acids Res. (1979) 7, 1513-1523.
16. Benchimol, S., Harlow, E., Crawford, L.V., Pim, D.C., Williamson, N.M. and Jenkins, J.R. (1983) Mary Lasker Memorial Symposium. The Cancer Cell. Cold Spring Harbor Laboratory (in Press).
17. Retzel, E.F., Collet, M.S. and Faras, A.J. (1980) Biochemistry 19, 513-518.
18. Wickens, M.P., Buell, G.N. and Schimke, R.T. (1978) J. Biol. Chem. 253, 2483-2495.
19. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, a Laboratory manual. Cold Spring Harbor Laboratory.
20. Kurtz, D.T. and Nicodemus, C.F. (1981) Gene 13, 145-152.
21. Hanahan, D. and Meselson, M. (1980) Gene 10, 63-72.
22. Smith, A.J.H. (1979) Nucleic Acids Res. 6, 831-849.
23. Dandolo, L. and Blangy, D. (1980) J. Virol. 33, 927-935.
24. Pelham, R.B. and Jackson, R.J. (1976) Eur. J. Biochem. 67, 247-256.
25. Laemmli, U.K. (1970) Nature 227, 680-685.
26. Van Roy, F., Fransen, L. and Fiers, W. (1981) J. Virol. 40, 28-44.
27. Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
28. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5476.
29. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
30. Harlow, E., Crawford, L.V., Pim, D.C. and Williamson, N.M. (1981) J. Virol. 39, 861-869.
31. Gurney, E.G., Harrison, R.O. and Fenno, J. (1980) J. Virol. 34, 752-763.
32. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
33. Favaloro, J., Treisman, R. and Kamen, R. (1980) Methods in Enzymology 65, 719-749.
34. Luka, J., Jornvall, H. and Klein, G. (1980) J. Virol. 35, 592-602.
35. Jornvall, H., Luka, J., Klein, G. and Appella, E. (1982) Proc. Natl. Acad. Sci. USA, 79, 287-291.
36. Luka, J., Sternas, L., Jornvall, H., Klein, G. and Lerner, R. (1983) Proc. Natl. Acad. Sci. USA, 80, 1199-1203.
37. Zakut-Houri, R., Oren, M., Bienz, B., Lavie, V., Hazum, S. and Girol, D. (1983) Nature 306, 594-597.
38. Van Roy, F., Liebaut, G., Fransen, L. and Fiers, W. (1981) J. Virol. 40, 28-44.
39. Neil, J.C., Ghysdael, J., Vogt., P.K., and Smart, J. (1981) Nature (Lond) 291, 675-677.

40. Smart, J.E., Oppermann, H., Czernilofsky, A.P., Purchio, A.F., Erikson, R.L., and Bishop, J.M. (1981) Proc. Natl. Acad. Sci. USA 78, 6013-6017.
41. Patschinsky, T., and Sefton, B. (1981) J. Virol. 39, 104-114.
42. Patschinsky, T., Hunter, T., Esch, F.S., Cooper, J.A., and Sefton, B.M. (1982) Proc. Natl. Acad. Sci. USA 79, 973-977.
43. Hunter, T., and Sefton, B.M. (1980) Proc. Natl. Acad. Sci. USA 77, 1311-1315.
44. Sefton, B.M., Hunter, T., Beemon, K., and Eckhart, W. (1980) Cell 20, 807-816.
45. Ushiro, H., and Cohen, S. (1980) J. Biol. Chem. 255, 8363-8365.
46. Hunter, T., and Cooper, J.A. (1981) Cell, 24, 741-752.