

p72: a human nuclear DEAD box protein highly related to p68

Gábor M. Lamm, Samantha M. Nicol¹, Frances V. Fuller-Pace¹ and Angus I. Lamond^{2,*}

Research Institute of Molecular Pathology, Dr. Bohr-Gasse 7, A-1030 Vienna, Austria, ¹Department of Molecular and Cellular Pathology, University of Dundee Medical School, Ninewells Hospital, Dundee DD1 9SY, Scotland and ²University of Dundee, Department of Biochemistry, Dundee DD1 4HN, Scotland

Received June 20, 1996; Revised and Accepted August 12, 1996

DDBJ/EMBL/GenBank accession no. U59321

ABSTRACT

P72, a novel human member of the DEAD box family of putative RNA-dependent ATPases and ATP-dependent RNA helicases was isolated from a HeLa cDNA library. The predicted amino acid sequence of p72 is highly homologous to that of the prototypic DEAD box protein p68. In addition to the conserved core domains characteristic of DEAD box proteins, p72 contains several N-terminal RGG RNA-binding domains and a serine/glycine rich C-terminus likely involved in mediating protein-protein interactions. A p72-specific probe detects two mRNAs of approximately 5300 and 9300 bases which, although ubiquitously expressed, show variability in their expression levels in different tissues. Purified recombinant p72 exhibits ATPase activity in the presence of a range of RNA moieties. Immunocytochemical studies of p68 and p72 show that these proteins localise to similar locations in the nucleus of HeLa cells, suggesting their involvement in a nuclear process.

INTRODUCTION

The D-E-A-D box protein family (1) of putative RNA helicases includes over 40 proteins from a wide range of organisms, spanning bacteria to humans, that share a group of conserved motifs including the sequence Asp-Glu-Ala-Asp (D-E-A-D) which provides their name [for review see refs (2–5)]. These proteins are implicated in diverse cellular functions including splicing, ribosome assembly, translation initiation, spermatogenesis, mRNA stability, embryogenesis and cell growth and division.

The DEAD box family is characterised by a core region represented by eIF-4A [eukaryotic (translation) initiation factor 4A] and contains eight conserved amino acid regions, one of which is the D-E-A-D motif [also called DEAD box (1,4)]. The conserved core region is flanked by N- and C-terminal extensions which share little sequence homology and are probably involved in mediating specialised functions of the individual proteins. The DEAH (Asp-Glu-Ala-His) sub-family includes the *Saccharomyces cerevisiae* gene products PRP2, PRP16 and PRP22 involved in pre-mRNA splicing [reviewed in refs (6,7)], the *Escherichia coli*

hrpA gene product of hitherto unknown function (8) and HRH1, the putative human homologue of PRP22 (9). Interestingly, all of these proteins are exceptionally large. The *Drosophila* Maleless (Mle) protein involved in X chromosome dosage compensation and its human orthologue, RNA helicase A, are most similar to the DEAH box proteins although they have the D-E-I-H motif (10,11). Another sub-family, containing a D-E-x-H motif, includes RNA helicases of positive strand RNA viruses (12).

To date, four human members of the DEAD box family have been reported. Apart from the prototypic member, p68 (13), there are currently three other human DEAD box proteins: p54, which was cloned from a human lymphoid cell line (14); NP52, isolated from a HeLa expression library due to a cross reaction with a monoclonal antibody raised against human aldolase A (15) and DDX1, which was found amplified in two retinoblastoma cell lines (16). P54, NP52 and DDX1 have not been further characterised biochemically and their function(s) remains unknown, although DDX1 has been found amplified in some primary neuroblastomas (17). In the case of p68, the purified protein has been shown to exhibit RNA-dependent ATPase activity and functions as an RNA helicase *in vitro* (18,19).

In this paper we describe the identification and characterisation of p72, a novel human nuclear DEAD box protein, which shows a striking homology to p68. We demonstrate that p72 is an ATPase activated by a variety of RNA species but not by dsDNA. The localisation and possible functional roles of p72 are discussed and compared with other DEAD box proteins.

MATERIALS AND METHODS

cDNA cloning and sequencing

The cDNA clone #461 coding for the N-terminal part of p72 was isolated from a random primed expression library of HeLa poly(A)⁺ RNA prepared in pUEX (20) (gift of Dr T. Kreis) due to a cross reaction with an unrelated monoclonal antibody. Clone #461 was subcloned into the *KpnI* site of Bluescript KS (Stratagene) and sequenced on both strands using oligonucleotide primers either with the dideoxy chain termination method (21) using [α -³⁵S]dATP or with fluorescent primers by the EMBL sequencing service. Radiolabelled clone #461 was then used as a probe to screen a λ Zap HeLa cDNA library (Stratagene) to

* To whom correspondence should be addressed

isolate additional clones spanning the missing 3' terminus of the p72 cDNA. A cDNA encoding full-length p72 was assembled from the resulting clones and subcloned into the *Sma*I site of pBluescript SK(-). This construct is henceforth referred to as p72-pBS SK.

Construction of expression vectors

To express the recombinant p72 in the *E. coli* strain BL21(DE3) a *Bam*HI-*Bam*HI fragment of p72 (purified from p72-pBS SK) which encoded for the full-length cDNA was cloned in-frame with the poly His-tag into the T7 driven pRSET expression vector (Invitrogen). For antibody production a fragment of p72 containing amino acids 1-343, was subcloned into pRSET in-frame with the poly His-tag.

For expression in HeLa cells the p72 cDNA was subcloned in frame into the *Bam*HI site of the eukaryotic expression vector pSG5 (22) containing a myc-tag (MEQKLISEEDL) (23). In all cases, correct orientation of the constructs was confirmed by restriction digestion analysis and DNA sequencing.

Growth and induction of bacteria expressing p72

Fresh overnight cultures of BL21(DE3) containing p72 cDNAs in the pRSET plasmid under the IPTG-inducible T7 promoter (24) were diluted 30-fold, grown to an OD (650 nm) of 0.3-0.4 at 37°C and induced by the addition of 0.75 mM IPTG. The cultures were transferred to 26°C and grown for a further 4 h before being harvested by centrifugation. The cell pellets were washed in 50 mM Tris-HCl (pH 7.4), harvested by centrifugation and stored at -70°C.

Purification of p72

The bacterial pellet was resuspended in ice cold Buffer A (1 ml per 100 mg pellet) containing 6 M guanidine-HCl, 0.1 M NaPi (pH 8.0), 10 mM Tris-HCl (pH 8.0), 5 mM imidazole, 1 mM phenylmethylsulphonylfluoride (PMSF), 1 mM benzamidine, 2 µg/ml leupeptin, 2 µg/ml aprotinin and placed in an ice/salt water bath for 30 min with intermittent vortexing. The resuspended bacterial pellet was then sonicated twice for 30 s to shear DNA and the insoluble material was pelleted by centrifugation. A denaturing protocol was necessary for the purification of p72 as the protein was found in bacterial inclusion bodies. The supernate was then incubated on a rotating wheel at 4°C with Ni²⁺-NTA-Agarose (3 ml packed volume for every 10 ml of supernate) for 3 h, was washed once with Buffer A and then resuspended again in Buffer A and poured into a disposable BioRad column. The resin was washed with 10 column volumes of Buffer A followed by 2 column volumes of Buffer B (identical to Buffer A but containing 10 mM imidazole). The recombinant protein was eluted with 2.5 column volumes Buffer C (as Buffer A but containing 200 mM imidazole) and eluates were collected as 1 ml fractions. Fractions containing recombinant p72 (determined by running an aliquot on SDS-PAGE and Coomassie staining) were pooled and dialysed at 4°C overnight into Buffer D [20% glycerol, 500 mM KCl, 50 mM Tris-HCl (pH 8.0), 0.5 mM EDTA, 1M guanidine-HCl, 1 mM DTT, 1 mM benzamidine, 1 mM PMSF] (using 0.25 litres/1 ml fraction) and then 3 h into Buffer E (same as Buffer D but containing 250 mM KCl). The protein was then concentrated to 1/10 of its original volume over an Amicon filter column, diluted 1:20 into Buffer F [15% glycerol, 50 mM KCl,

50 mM Tris-HCl (pH 8.0), 1 mM DTT, 1 mM benzamidine], loaded onto poly(U)-Sephacrose swollen in Buffer F and eluted using 100 mM KCl steps from 0.5 to 1 M KCl. The eluate was again concentrated over an Amicon filter column to 1/5 of its original volume and stored in aliquots in liquid nitrogen. Immediately prior to use in functional assays the purified protein was diluted in Buffer F to 50 ng/µl.

Antibody production

BL21(DE3) cells expressing a fragment of p72 corresponding to amino acids 1-343 were treated as described above for the purification of full-length recombinant p72. After Ni²⁺-NTA-Agarose chromatography the pooled fractions were electrophoresed through an SDS-PAGE gel, Coomassie stained and the p72 fragment excised from the gel. The acrylamide slice was macerated and 300 µg of recombinant p72 was mixed with 2 vol Freund's complete adjuvant (Sigma) and injected into rabbits. Further injections were carried out at three week intervals using 300 µg protein and Freund's incomplete adjuvant.

ATP hydrolysis assays

ATP hydrolysis assays were carried out as described in (25) containing RNA or DNA species as described in the appropriate figure legends. The amount of phosphate hydrolysed from [γ -³²P]ATP was determined by counting the relevant areas of the TLC plate (as Cerenkov counts) in a liquid scintillation counter. *E. coli* 16S and 23S rRNA was purchased from Boehringer.

In vitro transcription

Uniformly labelled, capped rabbit β -globin pre-mRNA and wild-type adenovirus pre-mRNA were transcribed as described in (26).

Northern blotting

The *Bam*HI-*Ssp*I 5' fragment of p72 was radiolabelled by random priming (27) and used to probe a commercial multiple tissue Northern blot of human poly(A)⁺ RNA (Clontech) as according to the manufacturer's recommendations.

SDS-PAGE and Western blotting

SDS-PAGE gel analysis was performed according to (28) and transferred onto nitrocellulose membrane (Schleicher and Schuell). Membranes were blocked in 2% non-fat milk powder in phosphate buffered saline (PBS), incubated with the primary antibody for 2 h at room temperature, washed and incubated with the appropriate secondary antibody (Amersham) coupled to horseradish peroxidase. Immunoblots were developed with the ECL detection kit (Amersham) as according to the manufacturer's recommendations.

Cell culture, transfection and immunofluorescent microscopy

HeLa cells were grown on coverslips at 37°C with 5% CO₂ in Dulbecco's modified Eagle's medium (Gibco BRL) supplemented with 10% foetal calf serum, 100 U/ml penicillin and streptomycin (Gibco BRL) and 1% glutamine. The myc-tagged p72 construct in pSG5 was transfected with LipofectAMINE transfection reagent (Gibco BRL) according to the manufacturer's protocol and the cells were fixed with 3.7% paraformaldehyde in CSK buffer [100 mM

NaCl, 300 mM sucrose, 10 mM PIPES (pH 6.8), 3 mM MgCl₂, 1 mM EGTA] for 10 min at room temperature. The cells were permeabilised with 0.5% Triton X100 in CSK buffer for 15 min at room temperature. Using immunofluorescence analysis we observed that routinely 30–40% of cells were transfected.

Immunofluorescent labelling was carried out as described (29) and analysed on a Zeiss Axiophot Epifluorescence microscope. Excitation wavelengths of 476 nm (FITC) and 529 nm (TexasRed) were used. The two channels were recorded independently and pseudo-coloured images were generated and superimposed. The pictures were printed on a Canon 700 Colour Laser Copier.

The following antibodies were used: rabbit anti-p80 coilin polyclonal serum 204/5 (dilution 1:350) (30), rabbit anti-p68 peptide antibody 2907 (dilution 1:300), mAb 9E10 (dilution 1:500) (23). TexasRed and fluorescein (FITC) conjugated anti-rabbit or anti-mouse secondary antibodies were purchased from Dianova and diluted 1:500.

Sequence analysis

The compilation and analysis of DNA sequences was done using the University of Wisconsin Genetics Computer Group (UWGCG) programmes (31) on a Vax computer cluster at EMBL, Heidelberg. The molecular weight and amino acid composition of p72 was determined using the Peptidesort programme (31). The TFasta or BLAST (32) programmes were used to search for homologies between p72 and the GenEMBL data banks. The CLUSTAL V programme (33) was used to search for amino acid homologies in the Swissprot database. The Motifs programme (31) was used to search for p72 protein motifs in the ProSite data bank.

RESULTS

Cloning and structural organisation of p72

A 1.3 kb cDNA fragment encoding the N-terminal portion of p72 was isolated from a HeLa expression library during expression screening with an unrelated antibody. The cDNA fragment was used to further screen HeLa cDNA libraries and a 1.1 kb fragment encoding the C-terminal portion of p72 was isolated. The overlapping cDNAs contain an open reading frame (ORF) of 1950 bp capable of encoding a protein with a predicted molecular mass of 71.9 kDa and an isoelectric point of 8.73 (Fig. 1). The presumed ATG initiation codon is 259 bp from the 5'-end of the isolated cDNA and the upstream sequence contains stop codons in all three reading frames (data not shown). The 3' untranslated sequence is at least 59 bp in length and contains neither a poly(A) tail nor a consensus polyadenylation signal (34), suggesting that p72 mRNA contains additional 3' untranslated sequence. *In vitro* translation of the assembled p72 cDNA in a reticulocyte lysate system yields a labelled translation product that migrates with an apparent molecular weight of 79 kDa on SDS-polyacrylamide gels (data not shown) and an antibody raised against recombinant p72 specifically detects a protein migrating at 79 kDa on Western blots of HeLa nuclear or cytoplasmic extracts (Fig. 4B, lane 7). This indicates that both recombinant and endogenous p72 migrates aberrantly at 79 kDa on SDS-PAGE.

The deduced amino acid sequence of p72 demonstrates that it is a new member of the DEAD box family of proteins containing all the conserved domains which are hallmarks of this family (Fig. 1). In addition to the conserved core domain, p72 contains

```

1  MRQQGFQDRD RDRDRGGGFGA RGGQGLPPPK FQMPQERLRK KKMDSLSELPK
51  FERNFYVERP EVARLTPEYV DELARRKKEIT VRQDDVCPKP VFAPFHAMFP
101  QYVMQVLMQD SFTEPTFIQC QGFPLALSGR DMVGRVQTGS GKTLAYLLFA
151  IVEIMHQPYL ERGDGPICLV LAPTRRLAQD VQQAADDYGR CSRLESTCIY
201  GGAPKQPQIR DLERGVEICI AEPGNLIDFL ISGKTHLRRC TYLVIEADRA
251  MLDMGFSPQI RKIVDQIRPD RGTLMNBAVM PKEVRQLAED FLRDYTGQNV
301  GWLELSANKN ILQIVDVCME SERDKRLIQL MEEIMARKEN KTIIVVEIKR
351  RCDLTRRMR RDGWFAMCIE GDRSQPERDM VLNEFRSOKA PILIATDVAS
401  RGLLVVDVKF VINYDYPNS EHYVHRIGRF ARSTNKGTAY TFPPTGNLQK
451  ARELIKVLEE ANQAINEPKLM QLVDRHQGGG QQQGRSRYRT TSSANMPLM
501  YQDECDRLM GVKDGGRRDS ASYDRSESTD RAGYANGSGY GSPNSAFGAQ
551  AGQYTYGQQT YGAAAYGTSS YTAQEQYGAQT YGASSTSTG RSSQSSSQGF
601  SGIGRSQQQP QPLMSQQFAQ PFGATNMIGY MGCATYQYEF RRRRRRPRK

```

Figure 1. Amino acid sequence of p72. The motifs typical of the DEAD box family of helicases are boxed. The N-terminal RGG boxes and the C-terminal glycine hinge region and proline tract are underlined. The p72 DNA sequence has been deposited in GenBank—accession number U59321.

N- and C-terminal extensions. The N-terminus contains four repeats of the RGG box originally identified as an RNA binding motif in the hnRNP U protein (35). A run of seven consecutive glycines separate the last conserved DEAD box family domain (HRIGR) and the serine/glycine rich (13.2 and 17.8%, respectively) C-terminus of p72. The extreme C-terminus of p72 additionally contains nine consecutive prolines. Serine/glycine rich regions have been shown to mediate protein–protein interactions in cytokeratins (36) and proline rich motifs appear to fulfil a similar function in hnRNP, snRNP and poly(A)-binding proteins (37–39) as well as in several transcription factors (40,41). We conclude that p72 encodes a novel human DEAD box protein which, in addition to the conserved core motifs, contains domains that may modulate p72–RNA and p72–protein interactions.

p72 shows striking homology to p68

The deduced amino acid sequence of p72 was used to carry out a BLAST (basic local alignment search tool) search of the Swiss-Prot database. This search revealed a striking homology between p72 and p68, a prototypic member of the DEAD box family (13). A multiple alignment of the first 481 amino acids of p72, encompassing the conserved DEAD box motifs, with the translated, most closely related DEAD box protein entries in the DDBJ/EMBL/GenBank database is presented in Figure 2. Out of 650 residues in p72, 453 residues (69.7%) are identical in human p68 and 457 residues (70.3%) are identical in mouse p68. An additional 53 residues in human and 52 residues in mouse p68 are similar amino acid substitutions (77.6 and 78.4% similarity, respectively). Within the region spanning the conserved motifs characteristic of this family (2) the homology between p72 and p68 is ~90%, which is considerably higher than that seen between other members of the family. However, C-terminal to the last conserved DEAD box domain (HRIGR) (Fig. 1) the identity between human p68 and p72 drops to 27.5%, suggesting that these proteins have different functions in the cell. This also supports the established view that DEAD box proteins have a similar core region encompassing the conserved domains but have N- and C-terminal extensions which endow the proteins with specialised functions [for review see refs (3,4)].

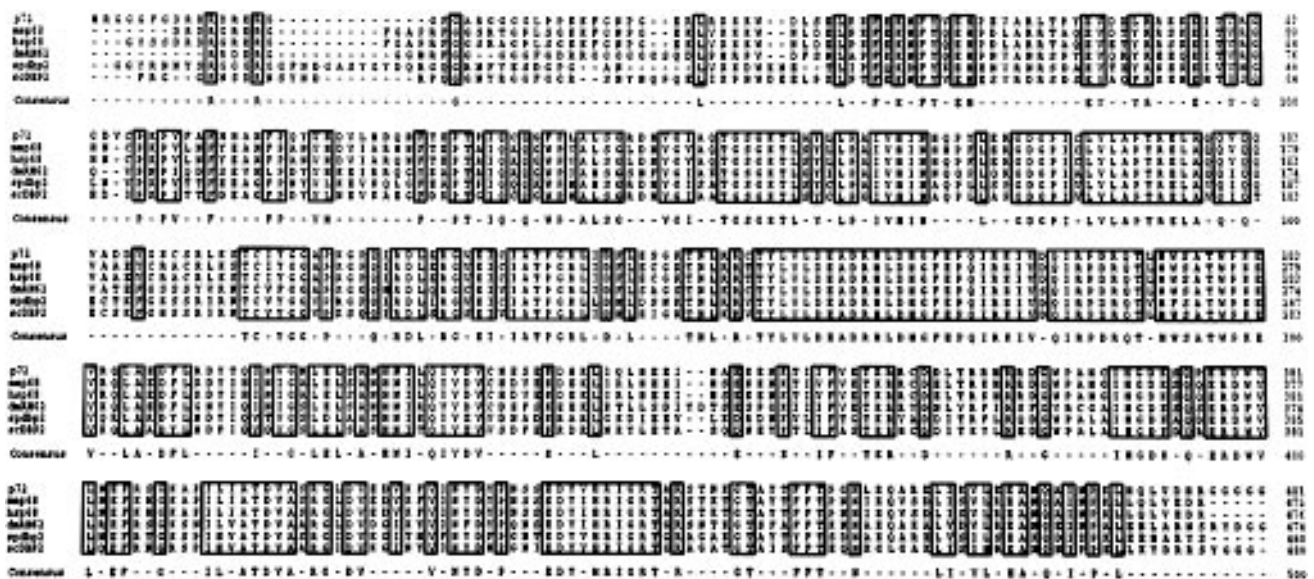


Figure 2. Amino acid sequence alignment of the first 481 amino acids of p72 with other DEAD box proteins. Multiple alignment was done using the CLUSTAL V programme (33) and generated by the PRETTYPLOT programme. Consensus positions are calculated from at least two most often occurring residues at a particular position taking into account amino acid similarity values. Boxed residues represent plurality values above 4.0. The accession numbers of the aligned sequences are: mmp68 (X65627), hsp68 (X15729), dmRM62 (X52846), spdbp2 (L11574) and scDBP2 (X55993).

p72 is also very similar to the *Drosophila* RM62 protein (60% identity, 73% similarity) (42). The DEAD box proteins DBP2 and dbp2, which are the putative *S.cerevisiae* and *S.pombe* homologues of p68 (43), also show strong similarity with p72 (Fig. 2). Interestingly, full-length p72, in comparison to p68, appears to be slightly more similar to both dbp2 and DBP2 (p72/dbp2 = 58.2%, p68/dbp2 = 54.4%, p72/DBP2 = 55.3%, p68/DBP2 = 53.2%). A search of the complete *S.cerevisiae* genome sequence (Martinsried Protein Sequence database) for p72-like sequences found only DBP2, suggesting that either (a) there is some redundancy in the function of these two proteins, or (b) multicellular organisms require both proteins.

P72 is encoded by two transcripts

A human multiple tissue northern blot of poly(A)⁺ RNA was probed with two non-overlapping cDNA fragments encompassing the 5' half (Fig. 3) and 3' portion of p72 (data not shown). Both cDNA fragments gave identical patterns of mRNA distribution in the different tissues and both recognise mRNA transcripts of approximately 5300 and 9300 bases (Fig. 3). The 5300 transcript appears to be ubiquitously expressed in all tissues tested with similar levels of expression in heart, brain, placenta, lung and liver and apparently higher levels of expression in skeletal muscle, kidney and pancreas. The 9300 transcript is also ubiquitously expressed, although extremely low levels are detected in heart and placenta and the transcript is most abundantly expressed in kidney and pancreas. The ratio between the two transcripts is also highly variable in the different tissues. While in brain, liver, kidney and pancreas the two transcripts are expressed at similar levels, in heart, placenta, lung and skeletal muscle predominantly the 5300 base transcript is present. A cDNA probe encompassing the p68 coding region verified that neither transcript represents a cross-reaction with p68 mRNA (data not shown). Although the smallest transcript detected is

5300 bases, the isolated p72 cDNA sequence only spans 2268 contiguous base pairs which lacks a poly(A) signal and poly(A) tail. It is, therefore, likely that p72 mRNA contains additional downstream and perhaps also upstream untranslated regions. The two p72 transcripts may arise by transcription of independent genes, differential transcription of a common gene or by alternative splicing of a common pre-mRNA. These results suggest that the expression of separate p72 transcripts is regulated in a tissue specific manner. Interestingly, when the same blot was probed for p68 mRNA two transcripts were also observed. However, these p68 transcripts differed from p72 in both their size and tissue distribution (data not shown) and no cross hybridisation was observed between the p72 and p68 probes. This indicates that the expression of p68 mRNA may also be subject to tissue specific regulation.

Purification of p72

Histidine-tagged p72 was expressed in *E.coli* and purified to homogeneity as described in Materials and Methods (Fig. 4A and B). Bacteria transformed with the p72 plasmid and induced with IPTG abundantly express the histidine-tagged protein, as is apparent by the appearance of an extra protein band migrating at 79 kDa on Coomassie stained gels (compare Fig. 4, lanes 1 and 2). Ni²⁺-NTA-Agarose chromatography of the bacterial lysate harbouring recombinant p72 yielded a substantial purification of the protein (Fig. 4, lane 3). A final poly(U)-Sepharose chromatography step yielded recombinant p72 purified to homogeneity (Fig. 4A, lane 4). The additional bands detected after the poly(U)-Sepharose purification step are degradation products of p72 (see below).

In order to verify the purification protocol of recombinant p72, Western blot analysis of the various purification steps was carried out using an anti-p72 antibody and the MAD1 monoclonal antibody (Fig. 4B). The MAD1 monoclonal antibody was raised

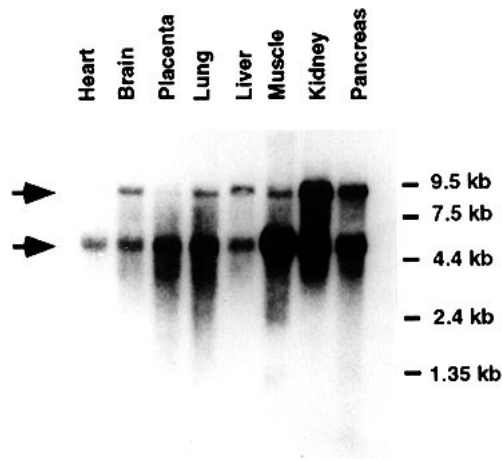


Figure 3. Multiple tissue Northern analysis. Poly(A)⁺ RNA (2 µg/ lane) from human tissues was probed with an *Av*I–*S*spI fragment of p72 cDNA encompassing the 5' portion of p72. Tissues are marked above the lanes, position of migration of RNA size markers are shown on the right and the arrows indicate the p72 transcripts.

against a peptide encompassing the DEAD motif of p68 (44). This region is 100% conserved in p72 and MAD1 should, therefore, also recognise the p72 protein. In HeLa nuclear extracts MAD1 predominantly recognises a 68 and a 79 kDa protein band (Fig. 4B, lane 5). The former band corresponds to p68 as identified by staining with p68-specific antibodies (data not shown). The latter band corresponds to p72 since anti-p72 antibodies detect a protein of similar size in HeLa nuclear extracts (Fig. 4B, lane 7). The MAD1 monoclonal antibody detects recombinant p72 in the lysate of *E. coli* carrying the p72 plasmid (Fig. 4B, lane 2) but not in bacteria transformed with the vector alone (Fig. 4B, lane 1). Although MAD1 detects recombinant p72 as a single protein band of 79 kDa after Ni²⁺-NTA-Agarose chromatography (Fig. 4B, lane 3) it detects several bands after poly(U)-Sephadex chromatography (Fig. 4B, lane 4). These additional bands correspond to degradation products of p72 as after poly(U)-Sephadex purification an identical pattern is observed with an anti-p72 antibody (Fig. 4B, lane 6). This purification protocol routinely yielded 1–1.5 mg of homogenous p72 from a one litre bacterial culture.

ATPase activity of p72

A common feature of DEAD box protein family members is their ability to hydrolyse ATP in the presence of RNA (2). The ATPase activity of purified p72 was tested by its ability to release radioactive phosphate from [γ -³²P]ATP. P72 hydrolysed ATP in the presence of total HeLa RNA and exhibited no ATPase activity in the absence of RNA (Fig. 5). Moreover, the ATPase activity of p72 was abolished when total HeLa RNA was pre-treated with RNase A indicating that the ATPase activity of the protein was dependent on the added RNA. The *E. coli* DEAD box protein DbpA, which is specifically activated by *E. coli* ribosomal RNA (25), was used as a positive control in these assays. ATPase reactions containing [γ -³²P]ATP and a 10-fold excess of cold nucleoside triphosphates showed competition only by unlabelled ATP, suggesting that only ATP is a substrate for p72 (data not shown). The K_m of p72 for ATP was found to be 170 µM (data not shown). This value is within the range reported for p68

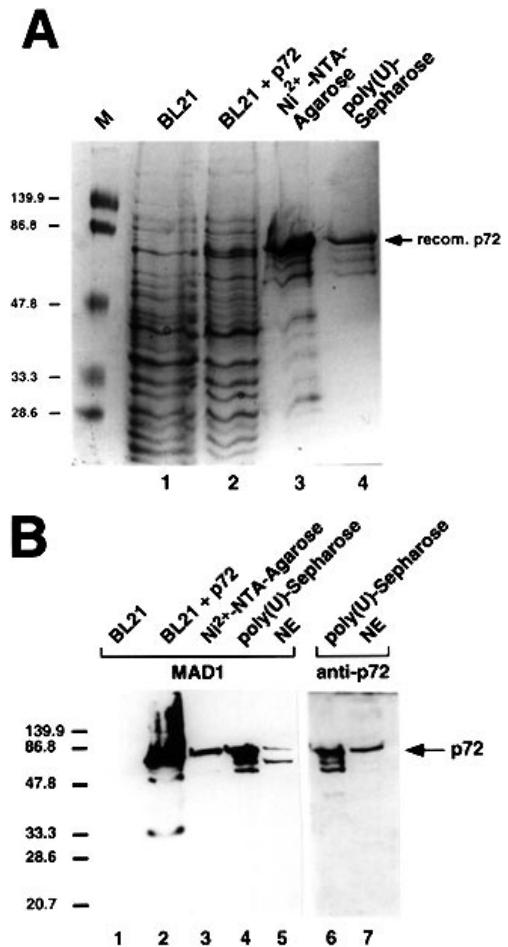


Figure 4. Purification of recombinant p72. Coomassie stain (A) or Western blot analysis (B) of proteins recovered during the various purification steps, separated on 10% SDS-PAGE. (A) BL21, total protein lysate of *E. coli* BL21 transformed with expression plasmid alone (lane 1). BL21 + p72, as above but with expression plasmid containing p72 (lane 2). Proteins present after Ni²⁺-NTA-Agarose chromatography and after poly(U)-Sephadex chromatography are separated on lanes 3 and 4, respectively. (B) Western blot analysis of purification steps probed with both MAD1 monoclonal and anti-p72 antibodies. Abbreviations are as in (A). The antibody used to probe the different lanes is indicated under the brackets. NE, HeLa nuclear extract (lanes 5 and 7). Protein size markers are indicated in kDa on the left (A + B) and recombinant p72 is shown by the arrowhead on the right.

(100–1000 µM) (18), DbpA (150 µM) (25) and eIF-4A (50 µM) (45). Taken together, the results above clearly demonstrate that p72 is an RNA-dependent ATPase.

Further ATPase assays were carried out to determine whether the ATPase activity of p72 could be preferentially stimulated by a specific RNA moiety or by DNA. As shown in Figure 6A, the ATPase activity of p72 was stimulated by a variety of RNAs. These include total RNA and tRNA from HeLa cells, *E. coli* and yeast; rabbit and *E. coli* rRNA; purified *E. coli* 16S and 23S rRNA; and both adenovirus and β -globin pre-mRNA. The amount of released phosphate in each reaction was measured as described in Materials and Methods and is depicted in graphic form in Figure 6B. (Interestingly, ssDNA from phage M13 also stimulated a low level of ATP hydrolysis. The ssDNA preparation was treated with RNase A prior to use to preclude an RNA contamination.) No

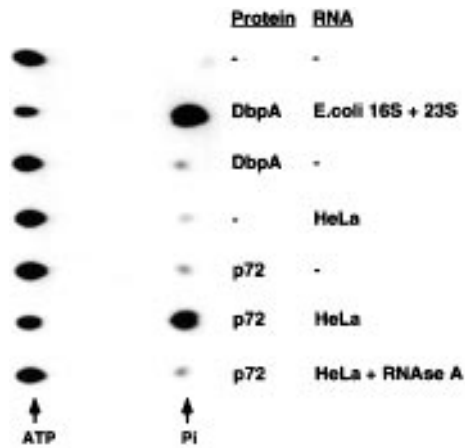


Figure 5. Assay of ATP hydrolysis by p72. Radioactive phosphate released from [γ - 32 P]ATP was separated (migration from left to right) on a polyethyleneimine thin layer chromatography (PEI TLC) plate. The presence or absence of protein and RNA in the reaction is indicated. HeLa, total HeLa RNA. HeLa + RNase A, total HeLa RNA pre-incubated with RNase A. 200 ng of purified p72 or DbpA and 800 ng of either total HeLa RNA or *E.coli* 16S and 23S rRNA were used per reaction.

activity was observed in the presence of total HeLa DNA or poly(U) RNA. The latter observation is particularly relevant since p72 can obviously bind poly(U) RNA as is shown by its purification over a poly(U)-Sepharose column. The ATPase activity of p72 is, therefore, likely to be dependent on RNA secondary structure. We therefore conclude that the ATPase activity of p72 can be stimulated by a variety of RNAs from various species and that this activity appears to require RNA secondary structure.

Sub-cellular localisation of p72

We were interested in determining the sub-cellular localisation of p72. Since the anti-p72 antibody did not give a specific signal in immunolocalisation experiments, we constructed a plasmid in which p72 was fused to a myc-epitope tag at its N-terminus and expressed under the control of the SV40 early promoter. This tagged construct was used in transient expression studies using HeLa cells and detected using a monoclonal anti-myc antibody. Myc-tagged p72 localises to the nucleus of HeLa cells (Fig. 7A, E and I) as determined by co-staining with DAPI (Fig. 7C). Tagged p72 shows a predominantly granular nuclear staining pattern (Fig. 7A) with occasional elevated levels of peri-nucleolar staining (Fig. 7E and I, indicated by arrowheads). Consistent with previous studies untransfected cells are not labelled by anti-myc antibody (data not shown). The high homology between p72 and p68 prompted us to compare the sub-cellular localisation of these two proteins. P68 as previously reported (43) shows a diffuse granular nuclear distribution in interphase cells (Fig. 7F) and colocalises with tagged-p72 (Fig. 7G). Since several DEAH-box proteins from *Saccharomyces cerevisiae* such as PRP2, PRP16 and PRP22 have been shown to be involved in pre-mRNA splicing [for review see ref. (6)] we were interested in whether p72 localises to splicing snRNP-enriched nuclear organelles called 'coiled bodies' [for review see ref. (46)]. HeLa cells transiently expressing tagged p72 show an average of 2–5 coiled bodies (Fig. 7J) and their staining pattern does not overlap with

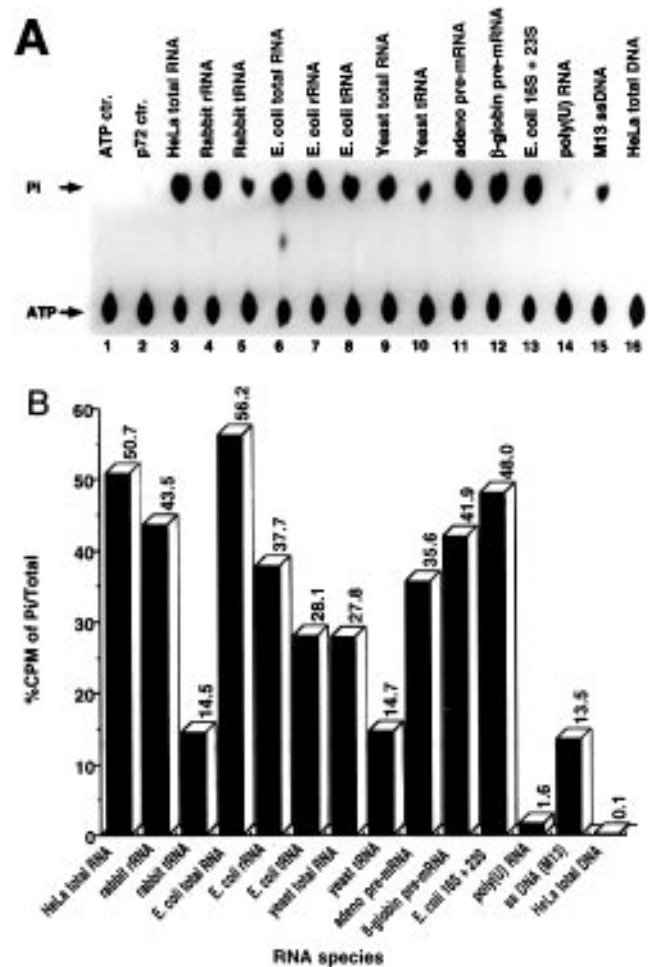


Figure 6. Stimulation of p72 ATPase activity by various RNA substrates. (A) Radioactive phosphate released from [γ - 32 P]ATP was separated (migration from bottom to top) on a PEI TLC plate. The substrates used to stimulate the ATPase activity of p72 are indicated. 'Yeast' indicates *S.cerevisiae*. Controls include incubation of [γ - 32 P]ATP without p72 (lane 1) or with p72 (lane 2) in the absence of nucleic acids. 200 ng of purified p72 and 1500 ng of nucleic acids were used per reaction except for poly(U) RNA (10 000 ng) and β -globin and adeno pre-mRNA (5000 ng). (B) Released phosphate and [γ - 32 P]ATP from each lane above were measured by c.p.m.-Cerenkov and plotted as the percentage of c.p.m.-Cerenkov of Pi/c.p.m.-Cerenkov total. Percentage values are indicated at the top of the columns.

that of tagged p72 (Fig. 7K). In all cases the cells show normal cell morphology as judged by DIC microscopy (Fig. 7D, H and L).

DISCUSSION

In this study we have identified and characterised p72, a novel human member of the DEAD box family of proteins. P72 is a nuclear protein and we have detected p72 mRNA ubiquitously expressed in all human tissues tested. Biochemical studies using recombinant p72 protein expressed in *E.coli* showed that it has RNA-dependent ATPase activity which is stimulated *in vitro* by a range of RNAs, including preparations of tRNA, mRNA and rRNA from bacteria, yeast and mammals.

P72 is strikingly similar to the human p68 protein, which is one of the prototypic DEAD box proteins, originally isolated due to a cross-reaction with a monoclonal antibody (DL3C4/Pab204)

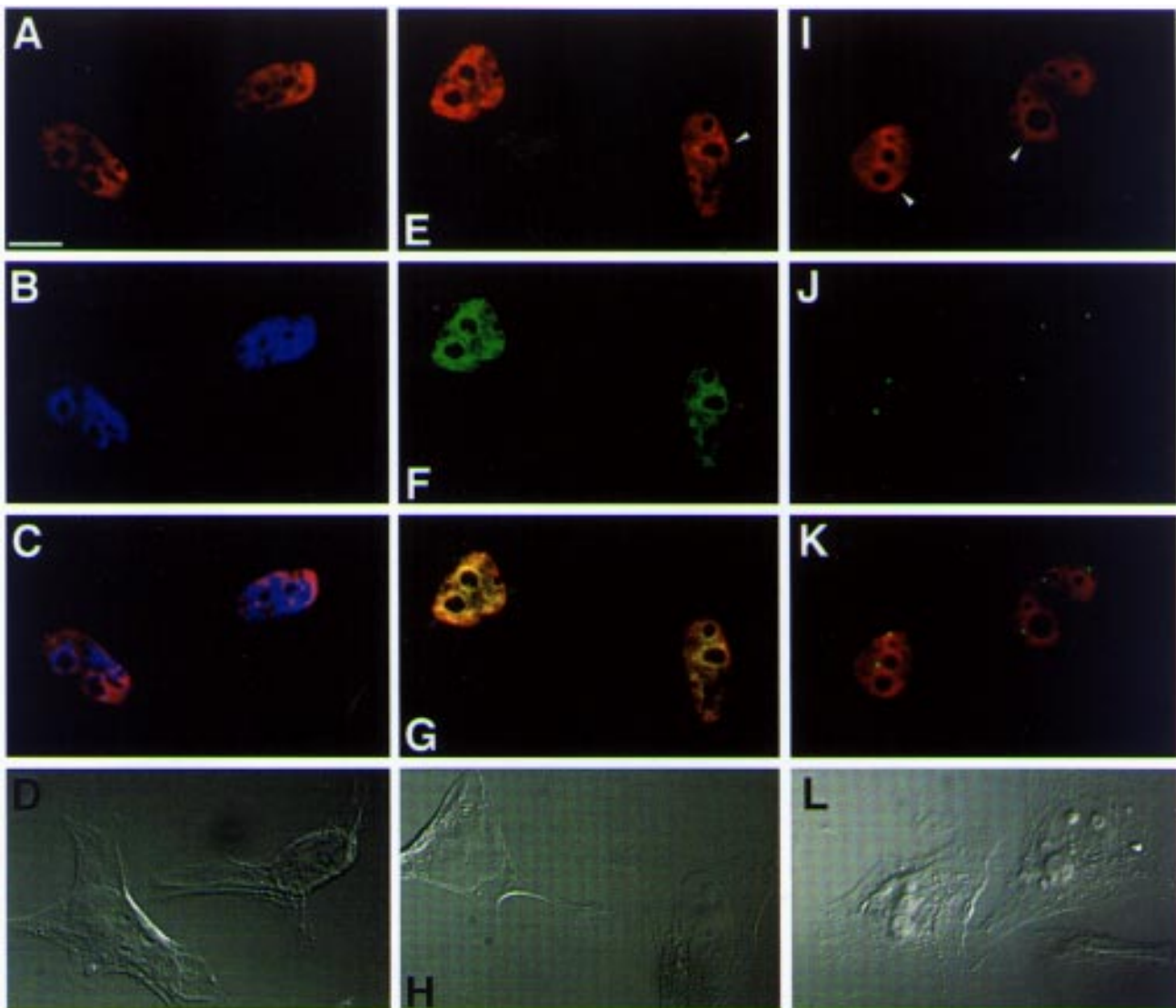


Figure 7. Transient expression of p72 in HeLa cells. Localisation pattern of myc-epitope-tagged p72 was determined by indirect immunofluorescence analysis using anti-myc antibodies on HeLa cells fixed 24 h after transfection. Cells were double labelled with anti-myc antibody (**A**, **E** and **I**) (red) and either DAPI (**B**) (blue), anti-p68 peptide antibody 2907 (**F**) (green) or with anti-coilin antibody (**J**) (green). The respective overlays are shown in (**C**, **G** and **K**). Colocalisation of both antibodies in (**G**) and (**K**) produce a yellow colour. Corresponding DIC images are shown in parallel in (**D**, **H** and **L**). The bar represents 10 μ m, peri-nucleolar localisation of tagged p72 is indicated by arrowheads. Anti-myc antibody was visualised with anti-mouse secondary antibodies coupled to Texas Red and the anti-p68 peptide antibody 2907 was visualised with an anti-rabbit secondary antibody coupled to fluorescein (FITC).

raised against SV40 large T antigen (13). These two proteins are more closely related to each other than to any other members of the DEAD box family analysed to date. Interestingly, *S.cerevisiae* has only one apparent homologue of p68/p72, called DBP2 (43). While DBP2 was previously identified as the yeast homologue of p68, and can be complemented by human p68 (47), it in fact shows a higher degree of sequence homology to p72. We propose that p68 and p72 represent a specific subfamily of the DEAD box protein family. The fact that mammals appear to have at least two members of this subfamily suggests either that there is some functional redundancy or, perhaps more likely, that these proteins exhibit some specialisation in their substrate specificities. This would be consistent with the observation that while they are

~90% identical in the core domains, their N- and C-termini are much less conserved.

The p68 and p72 proteins also show a functional as well as structural relationship. P68 has previously been shown to have RNA-dependent ATPase and RNA helicase activities *in vitro* (18,19). As described above, p72 also exhibits RNA-dependent ATPase activity. However, we have so far been unable to detect RNA helicase activity for p72 (data not shown). There are several possible explanations for this finding: (a) the recombinant p72 does not unwind RNA under the assay conditions used; (b) since the recombinant p72 was purified from bacterial inclusion bodies, it may not have the correct conformation for helicase activity, even though it shows ATPase activity; (c) other factors are

necessary for p72 to unwind RNA [e.g. eIF-4A, which is a much more efficient RNA helicase when part of the eIF-4F complex (48,49), and RhlB which exhibits ATP-dependent RNA helicase activity when part of the 'degradosome' but not as free protein (50)]; or (d) unlike p68, p72 is not actually an RNA helicase *in vivo*. In this regard it is worth noting that relatively few DEAD box proteins have been shown to exhibit helicase activity (2). Although the DEAD box proteins are usually referred to as 'helicases' it may in fact be the case that their common function is actually an ATPase activity, with additional helicase activity being restricted to a subset of the family members.

Both p72 and p68 localise to the nucleus of HeLa cells. Like p68, p72 shows a predominantly granular nucleoplasmic staining pattern, excluding nucleoli. However, it also shows some enhanced peri-nucleolar staining that was not seen with p68. Previous studies have shown that p68 undergoes dramatic changes in nuclear localisation during the cell cycle (18,43). During interphase, p68 is found in the nucleoplasm and is excluded from the nucleoli. However, it transiently enters pre-nucleolar bodies during telophase. We have not observed such a sub-cellular redistribution of p72 during the cell cycle, again suggesting that p68 and p72, although highly homologous, may have different biological functions involving interaction with different RNA targets.

The biological functions of p68/p72 in mammals, and of DBP2 in yeast, remains to be established. The fact that these are nuclear proteins suggests that their *in vivo* substrates are likely to be nuclear RNAs. Analysis of the sequence of p72 shows that in addition to the conserved DEAD box family core domains, it contains N- and C-terminal extensions with additional protein motifs. These include four N-terminal RGG boxes (35), a glycine hinge, a serine/glycine-rich C-terminus and a C-terminal proline tract. The analysis of the nucleolin RGG domain suggests that each RGGF tetrapeptide makes a β -turn and several of these repeats form a β -spiral (51) which, due to the presence of glycines, are extremely flexible and can adopt alternative conformational states (52). P72 is shown here to bind polyU sepharose and to have RNA-dependent ATPase activity. It is possible that the RGG boxes may be involved in the interaction of p72 with RNA. A run of glycines is also seen in the splicing factors U2AF³⁵ (53) and ASF/SF2 (54,55) and may function to flexibly hinge different protein domains. Proline rich motifs have been identified in numerous proteins including hnRNPL (37) the U1 snRNP specific A and C proteins (56,57) and the yeast poly(A)-binding protein (39). The proline/glutamine rich motif in the U1 snRNP-specific C protein has been proposed to stabilise the U1 snRNP-5' splice site interaction (38) and in the transcription factors CTF/NF-1 and Sp-1, this motif has been demonstrated to represent the transcription activation domain (40,41). Thus, proline rich motifs may function as binding sites for additional factors during the assembly of transcription or splicing complexes. The studies above suggest that the C-terminal proline tract of p72 may be involved in mediating protein-protein interactions while the RGG domain may be necessary for the protein to bind to its putative target RNA species.

An important goal for future studies will be to identify the authentic *in vivo* RNA substrates for p68 and p72 and to determine whether they act alone or as part of larger nuclear complexes. It will also be important to characterise in more detail the functional properties of the p68/p72 subfamily of DEAD box proteins and to determine whether there are additional members of this group.

ACKNOWLEDGEMENTS

The authors wish to thank Kerstin Bohmann for the anti-p80 coilin antibody and Dr G. Evan for monoclonal anti-myc antibodies as well as the EMBL sequencing service for technical assistance. We are also especially grateful to Karsten Weis and Joe Lewis for critical reading of the manuscript. Parts of this work were supported by a grant from Boehringer Ingelheim Fonds, an EMBO short-term fellowship to GML and a Medical Research Council Senior Fellowship to FFP.

REFERENCES

- Linder,P., Lasko,P.F., Ashburner,M., Leroy,P., Nielsen,P. J., Nishi,K., Schnier,J. and Slonimski,P.P. (1989) *Nature*, **337**, 121-122.
- Fuller-Pace,F.V. (1994) *Trends Cell Biol.*, **4**, 271-274.
- Gorbalenya,A.E. and Koonin,E.V. (1993) *Curr. Opin.*, **3**, 419-429.
- Schmid,S.R. and Linder,P. (1992) *Mol. Microbiol.*, **6**, 283-291.
- Wassarman,D.A. and Steitz,J.A. (1991) *Nature*, **349**, 463-464.
- Beggs,J. (1995) In Lamond,A.I. (ed.), *Pre-mRNA Processing*. Springer-Verlag, Heidelberg, pp. 79-95.
- Lamm,G.M. and Lamond,A.I. (1993) *Biochim. Biophys. Acta*, **1173**, 247-265.
- Moriya,H., Kasai,H. and Isono,K. (1995) *Nucleic Acids Res.*, **23**, 595-598.
- Ono,Y., Ohno,M. and Shimura,Y. (1994) *Mol. Cell. Biol.*, **14**, 7611-7620.
- Kuroda,M.I., Kernan,M. J., Kreber,R., Ganetzky,B. and Baker,B.S. (1991) *Cell*, **66**, 935-947.
- Lee,C.G. and Hurwitz,J. (1993) *J. Biol. Chem.*, **268**, 16822-16830.
- Koonin,E.V. (1991) *Nature*, **352**, 290.
- Ford,M.J., Anton,I.A. and Lane,D.P. (1988) *Nature*, **332**, 736-738.
- Lu,D. and Yunis,J.J. (1992) *Nucleic Acids Res.*, **20**, 1967-1972.
- Kitajima,Y., Yatsuki,H., Zhang,R., Matsuhashi,S. and Hori,K. (1994) *Biochem. Biophys. Res. Commun.*, **199**, 748-754.
- Godbout,R., Hale,M. and Bisgrove,D. (1994) *Gene*, **138**, 243-245.
- Squire,J.A., Thorne,P.S., Weitzman,S., Maggi,J.D., Dirks,P., Doyle,J., Hale,M. and Godbout,R. (1995) *Oncogene*, **10**, 1417-1422.
- Iggo,R.D. and Lane,D.P. (1989) *EMBO J.*, **8**, 1827-1831.
- Hirling,H., Scheffner,M., Restle,T. and Stahl,H. (1989) *Nature*, **339**, 562-564.
- Bressan,G.M. and Stanley,K.K. (1987) *Nucleic Acids Res.*, **15**, 10056.
- Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
- Green,S., Issemann,I. and Sheer,E. (1988) *Nucleic Acids Res.*, **16**, 369.
- Evan,G.I., Lewis,G.K., Ramsay,G. and Bishop,J.M. (1985) *Mol. Cell. Biol.*, **5**, 3610-3616.
- Tabor,S. and Richardson,C.C. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4767-4771.
- Fuller-Pace,F.V., Nicol,S.M., Reid,A.D. and Lane,D.P. (1993) *EMBO J.*, **12**, 3619-3626.
- Konarska,M.M. and Sharp,P.A. (1987) *Cell*, **49**, 763-774.
- Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6-13.
- Laemmli,U.K. (1970) *Nature*, **227**, 680-685.
- Carmo-Fonseca,M., Pepperkok,R., Sproat,B.S., Anson,W., Swanson,M.S. and Lamond,A.I. (1991) *EMBO J.*, **10**, 1863-1873.
- Bohmann,K., Ferreira,J.A. and Lamond,A.I. (1995) *J. Cell. Biol.*, **131**, 817-831.
- Devereux,J., Haerberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387-395.
- Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D. J. (1990) *J. Mol. Biol.*, **215**, 403-410.
- Higgins,D., Bleasby,A. and Fuchs,R. (1992) *CABIOS*, **8**, 189-191.
- Wahle,E. and Keller,W. (1992) *Annu. Rev. Biochem.*, **61**, 419-440.
- Kiledjian,M. and Dreyfuss,G. (1992) *EMBO J.*, **11**, 2655-2664.
- Steinert,P.M., Rice,R.H., Roop,D.R., Trus,B.L. and Steven,A.C. (1983) *Nature*, **302**, 794-800.
- Pinol,R.S., Swanson,M.S., Gall,J.G. and Dreyfuss,G. (1989) *J. Cell. Biol.*, **109**, 2575-2587.
- Heinrichs,V., Bach,M., Winkelman,G. and Luhrmann,R. (1990) *Science*, **247**, 69-72.
- Sachs,A.B., Bond,M.W. and Kornberg,R.D. (1986) *Cell*, **45**, 827-835.
- Courey,A.J. and Tjian,R. (1988) *Cell*, **55**, 887-898.
- Tanese,N., Pugh,B.F. and Tjian,R. (1991) *Genes Dev.*, **5**, 2212-2224.

- 42 Dorer,D.R., Christensen,A.C. and Johnson,D.H. (1990) *Nucleic Acids Res.*, **18**, 5489–5494.
- 43 Iggo,R.D., Jamieson,D. J., MacNeill,S.A., Southgate,J., McPheat,J. and Lane,D.P. (1991) *Mol. Cell. Biol.*, **11**, 1326–1333.
- 44 Iggo,R., Picksley,S., Southgate,J., McPheat,J. and Lane,D.P. (1990) *Nucleic Acids Res.*, **18**, 5413–5417.
- 45 Grifo,J.A., Abramson,R.D., Satler,C.A. and Merrick,W.C. (1984) *J. Biol. Chem.*, **259**, 8648–8654.
- 46 Lamond,A.I. and Carmo-Fonseca,M. (1993) *Mol. Biol. Rep.*, **18**, 127–133.
- 47 Barta,I. and Iggo,R. (1995) *EMBO J.*, **14**, 3800–3808.
- 48 Ray,B.K., Lawson,T.G., Kramer,J.C., Cladaras,M.H., Grifo,J.A., Abramson,R.D., Merrick,W.C. and Thach,R.E. (1985) *J. Biol. Chem.*, **260**, 7651–7658.
- 49 Rozen,F., Ederly,I., Meerovitch,K., Dever,T.E., Merrick,W.C. and Sonenberg,N. (1990) *Mol. Cell. Biol.*, **10**, 1134–1144.
- 50 Py,B., Higgins,C.F., Krisch,H.M. and Carpousis,A.J. (1996) *Nature*, **381**, 169–172.
- 51 Ghisolfi,L., Joseph,G., Amalric,F. and Erard,M. (1992) *J. Biol. Chem.*, **267**, 2955–2959.
- 52 Matsushima,N., Creutz,C.E. and Kretsinger,R.H. (1990) *Proteins*, **7**, 125–155.
- 53 Zhang,M., Zamore,P.D., Carmo-Fonseca,M., Lamond,A.I. and Green,M.R. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 8769–8773.
- 54 Krainer,A.R., Mayeda,A., Kozak,D. and Binns,G. (1991) *Cell*, **66**, 383–394.
- 55 Ge,H., Zuo,P. and Manley,J.L. (1991) *Cell*, **66**, 373–382.
- 56 Sillekens,P.T., Habets,W. J., Beijer,R.P. and van Venrooij,W.J. (1987) *EMBO J.*, **6**, 3841–3848.
- 57 Sillekens,P.T., Beijer,R.P., Habets,W. J. and van Venrooij,W.J. (1988) *Nucleic Acids Res.*, **16**, 8307–8321.