

The p68 RNA Helicase

**A thesis submitted by Richard Derek Iggo
for the degree of Doctor of Philosophy
at the University of London**

**Imperial Cancer Research Fund
and
University College London**

April 1991

ProQuest Number: 10610961

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10610961

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

The p68 protein is immunologically related to large T antigen, the dominant transforming oncogene of simian virus 40. The deduced amino acid sequence of human p68 contains amino acid sequence motifs characteristic of a large family of putative RNA helicases called "DEAD box" proteins. The family derives its name from the sequence of the most conserved motif: aspartate-glutamate-alanine-aspartate.

Human p68 protein was immunoprecipitated from HeLa cells and shown to possess RNA-dependent ATPase activity. As expected for an RNA helicase, the ATPase activity is stimulated preferentially by RNA lacking secondary structure.

Cell staining shows that the sub-nuclear location of mammalian p68 changes dramatically during the cell cycle. p68 is present in the nucleoplasm and excluded from the nucleoli in interphase. This pattern is reversed in telophase, when p68 appears to be concentrated in prenucleolar bodies.

The human p68 cDNA was completed by anchor PCR and the human p68 gene was mapped to chromosome 17q23. Putative p68 homologues were cloned from yeast and named *DBP2* in *S.cerevisiae* and *dbp2* in *S.pombe*. The yeast genes were shown by gene disruption and tetrad analysis to be essential for viability.

The p68, *DBP2* and *dbp2* genes contain an intron at the same site in helicase motif V. The intron is 1.2 kb long in p68, 1 kb long in *DBP2*, and 700 bp long in *dbp2*. The unusual length, position and conservation of the intron in *DBP2* suggest that it may have a function. The possibility that the intron autoregulates *DBP2* expression was examined by Northern and Western blotting in *S.cerevisiae*.

Acknowledgements

I wish to thank my external supervisor, David Lane, and the other members of the Molecular Immunochemistry Laboratory, notably Julian Gannon and Jenny Southgate, for their help, advice and encouragement. Among the numerous other individuals at Clare Hall who provided invaluable assistance Michelle Ginsburg and John Nicholson deserve particular mention for help with computing and photography respectively. The yeast work presented in chapter seven would not have been possible without the help of Derek Jamieson, Stuart MacNeill, Jean Beggs and Paul Nurse. Finally, I would like to thank the Imperial Cancer Research Fund for financial support.

Contents

Abstract	2
Acknowledgements	3
Contents	4
Tables and Figures	7
Chapter 1 Introduction	9
The identification and cloning of p68	9
SV40 T antigen is an RNA helicase	11
The Hodgman/Gorbalenya helicase motifs	14
DEAD and DEAH proteins	15
The DEAD family	17
eIF4A	17
<i>S.cerevisiae</i> DEAD genes involved in RNA splicing	20
<i>MSS116</i>	20
<i>PRP5</i>	21
<i>PRP28</i>	21
The <i>DED1</i> , <i>DBP1</i> , PL10 and An3 family	21
PL10	22
An3	22
<i>DED1</i>	22
<i>DBP1</i>	23
The p68, RM62 and <i>DBP2</i> family	23
RM62	23
Other DEAD genes	24
<i>SPB4</i>	24
<i>srmB</i>	24
<i>vasa</i>	25
CA3, CA4 and CA5/6	26
The DEAH family	26
<i>PRP16</i>	27
<i>PRP22</i>	27
<i>PRP2</i> , JA1 and JA2	27
Functional implications	28
Scope of this study	29
Chapter 2 Materials and Methods	30
Introduction	30
Molecular Biology (Chapter 3)	30
Plasmids	30
DNA sequencing	30
Primer extension	34
Polymerase chain reaction techniques	34
PCR mutagenesis	35
RT-PCR	35
anchor PCR	35

conventional PCR sequencing	37
asymmetric PCR sequencing	37
inverse PCR	37
PCR mapping	38
Sequence analysis	38
Antibodies and Cell Biology (Chapters 4 and 5)	39
Cell lines and Antibodies	39
Hybridoma fusion	40
Cell staining	41
Microinjection	41
Transfection	41
Electrophoresis and Immunoblotting	42
Biochemistry (Chapter 6)	42
Preparation of extracts	42
ATPase Assays	43
pCp labelling	43
Yeast Genetics and RNA analysis (Chapter 7)	44
<i>S.cerevisiae</i> genotypes	44
<i>S.cerevisiae</i> techniques	44
<i>S.pombe</i> techniques	46
Northern blotting	46
Chapter 3 DNA Cloning	48
Introduction	48
Human p68	48
Completion of the human p68 cDNA sequence	48
Chromosome mapping by PCR	53
Discussion	57
Identification of yeast genes related to p68	60
<i>S.pombe dbp2</i>	60
<i>S.cerevisiae DBP2</i>	60
Analysis of the p68, <i>dbp2</i> and <i>DBP2</i> sequences	66
p68, <i>DBP2</i> and <i>dbp2</i> contain a conserved intron	71
Discussion	73
<i>E.coli dbpA</i>	75
Discussion	80
General discussion	83
Chapter 4 Antibodies	85
Introduction	85
The p68 cDNA clones encode authentic p68	85
Preparation of new anti-p68 antibodies	85
MaD1 binds to a cryptic epitope	90
MaD1 can detect p68 in whole cell extract	92
<i>dbp2</i> lacks the PAb204 epitope	92
Discussion	92
Chapter 5 Cell Biology	95
Introduction	95

Cell cycle variation in p68 distribution	95
Microinjection of PAb204	98
Discussion	102
Chapter 6 Biochemistry	105
Introduction	105
p68 is an RNA-dependent ATPase	105
Small nuclear RNA does not coprecipitate with p68	111
Discussion	113
Chapter 7 Functional analysis of DBP2 and dbp2	115
Introduction	115
Disruption of the <i>DBP2</i> and <i>dbp2</i> genes is lethal	115
Viability is restored by expression of a <i>DBP2</i> cDNA	117
The conserved intron may regulate <i>DBP2</i> expression	120
Discussion	124
Chapter 8 Discussion	127
References	137

Tables

1	The DEAD and DEAH families	16
2	Derivation of plasmids	31

Figures

1.1	PAb204 peptides	10
1.2	Putative helicase motifs	13
1.3	Alignment of DEAD and DEAH proteins	18
1.4	p68 amino acid sequence motifs	19
2.1	Asymmetric PCR sequencing strategy	36
3.1	Human p68 sequence	49
3.2	Anchor PCR cloning strategy	51
3.3	Human p68 anchor PCR products	52
3.4	Southern blot of anchor PCR products	52
3.5	PCR amplification of human p68 cDNA	52
3.6	Primer extension using the RT primer	52
3.7	Zooblot of genomic DNA using a human p68 probe	54
3.8	Human p68 PCR products using primers 6 and 7	54
3.9	Human p68 PCR products using primers 8 and 9	54
3.10	Introns in eIF4AII, <i>vasa</i> and human p68	55
3.11	Human p68 intron sequence between primers 6 and 7	56
3.12	Assignment of human p68 to chromosome 17	58
3.13	Localization of human p68 to 17q23	59
3.14	<i>S.pombe dbp2</i> sequence	61
3.15	<i>dbp2</i> cDNA sequence around the first intron	63
3.16	Inverse PCR cloning strategy	64
3.17	A, <i>S.cerevisiae</i> Southern blot; B, inverse PCR products	65
3.18	<i>S.cerevisiae DBP2</i> sequence	67
3.19	Alignment of p68, <i>DBP2</i> and <i>dbp2</i>	69
3.20	Dotplot comparison of p68, <i>DBP2</i> and <i>dbp2</i>	70
3.21	Conserved splice sites in p68, <i>DBP2</i> and <i>dbp2</i>	72
3.22	Third codon preference plots of <i>DBP2</i> and <i>dbp2</i>	74
3.23	W3110 chromosome map showing <i>dbpA</i>	76
3.24	<i>E.coli dbpA</i> sequence	77
3.25	Alignment of <i>dbpA</i> , p68 and <i>srmB</i>	78
3.26	Dotplot comparison of <i>dbpA</i> , p68 and <i>srmB</i>	79
3.27	Dotplot analysis of <i>dbpA</i> cloning	81
3.28	Dotplot analysis of <i>dbp2</i> and <i>DBP2</i> cloning	82
3.29	Immunoblot of <i>E.coli</i> W3110 probed with MaD1	84
4.1	Immunoblots probed with MF7	86
4.2	Zooblots probed with MaD1	87
4.3	<i>argU</i> increases p68 expression in <i>E.coli</i>	89
4.4	Immunoblot of bacteria overexpressing <i>DbpA</i>	89
4.5	MaD1 immunoprecipitation	91

4.6	Immunoblot of Hela extract probed with MaD1	91
4.7	Immunoblot of <i>Xenopus</i> extract probed with MaD1	94
5.1	Immunoperoxidase staining of p68 in CV1 cells	96
5.2	Anti-p68 and anti-nucleolar staining of C6 cells	97
5.3	C6 cells treated with actinomycin D	99
5.4	Antibodies used for the microinjection experiments	99
5.5	Microinjection of PAb204 and BG2 into 3T3 and C6 cells	101
6.1	p68 immunoprecipitated from Hela cells (150 mM NaCl)	106
6.2	p68 immunoprecipitated from Hela cells (500 mM NaCl)	106
6.3	p68 is an RNA-dependent ATPase	108
6.4	Kinetic analysis of p68's ATPase activity	109
6.5	p68 is not an RNA-dependent GTPase	110
6.6	Secondary structure inhibits p68's ATPase activity	110
6.7	Immunoblot of Hela immunoprecipitates	112
6.8	pCp labelling of RNA	112
7.1	<i>S.cerevisiae</i> tetrad dissection	116
7.2	PCR identification of <i>DBP2::URA3</i> yeast transformants	116
7.3	Northern blot of <i>DBP2::URA3</i> yeast transformants	116
7.4	Western blot of PY26 cells probed with MaD1	118
7.5	Northern blot of <i>DBP2::URA3</i> cells rescued with pIG70	119
7.6	Western blot of <i>DBP2::URA3</i> cells rescued with pIG70	119
7.7	Northern blot using an intron probe	121
7.8	Northern blot comparing pIG70 with pIG78	123

Chapter 1 Introduction

Manipulation of RNA secondary structure is essential for the proper execution of a large number of processes in the cell. The recent discovery of a large family of putative RNA helicases underlines the importance which the cell attaches to this problem. These proteins share extensive regions of amino acid sequence homology, are ubiquitous in living organisms and are involved in many aspects of RNA metabolism including splicing, translation and ribosome assembly. One of the most conserved motifs in the primary sequence of the proteins, Asp-Glu-Ala-Asp, has given the family its distinctive name: "DEAD box" proteins. The human p68 nuclear antigen, the subject of this thesis, was the first protein in which the DEAD motifs were identified. It was originally discovered by David Lane in the course of his immunological analysis of SV40 large T antigen. Subsequent immunological and biochemical research has revealed both similarities and differences between p68 and T antigen which leave open the possibility that p68 is a cellular "T-equivalent protein".

This introduction will analyse these developments from a historical perspective, starting with the discovery of p68. After assessing the link between p68 and T antigen it will go on to describe the identification of the Hodgman/Gorbalenya helicase motifs and finish with a survey of the present state of the DEAD box protein field.

The identification and cloning of p68

p68 was first identified by Lane and Hoeffler (1980) using a monoclonal antibody against SV40 large T antigen. The antibody was originally called DL3C4 but is now called PAb204. It crossreacts with a nuclear antigen of 68 kD which is present in a wide range of mammalian cell lines. The antigen is barely detectable by immunofluorescence in quiescent cells, although the same cells stain brightly during exponential growth. On the basis of the crossreaction, the growth induction and some preliminary evidence that the antibody inhibits the ATPase activity of T antigen, Lane and Hoeffler suggested that p68 and T

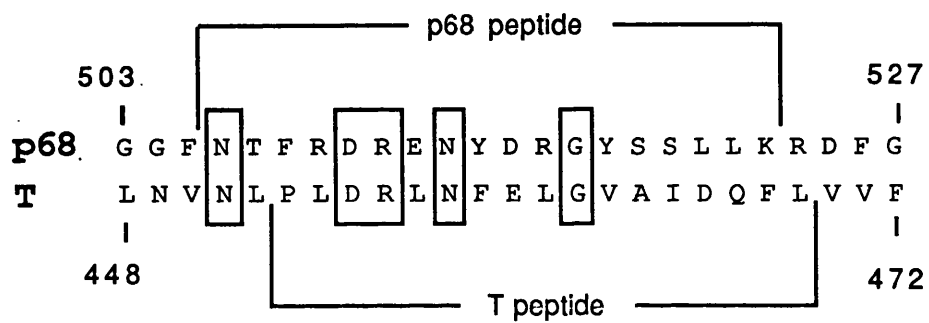


Figure 1.1 Alignment of p68 and T antigen peptides to which PAb204 can bind (Ford et al., 1988).

antigen might share a common function, perhaps involving the initiation of DNA replication.

The biological significance of the crossreaction is disputed. Since a monoclonal crossreaction involves a single B cell clone one can not assume that it is biologically meaningful just because it is highly specific. Another argument against the PAb204 crossreaction being intrinsically significant is a statistical one: if one makes a conservative estimate of the number of proteins in the cell and assumes that the average epitope is made up of four contiguous amino acids then there is a good chance that every anti-T antigen monoclonal antibody will crossreact with a cellular protein (Crawford, et al., 1982).

Ford, et al. (1988) cloned p68 and mapped the PAb204 epitope on p68 and T antigen using short synthetic peptides. The minimal regions of T antigen and p68 to which PAb204 can bind have four amino acids in common (figure 1.1). Apart from the PAb204 epitope the homology between the p68 and T antigen sequences is limited to the presence of consensus nucleotide binding motifs in both proteins (Walker, et al., 1982). The location of the PAb204 epitope only a few residues away from the type A nucleotide binding motif in T antigen and the ability of PAb204 to inhibit the T antigen helicase assay (Stahl, et al., 1986) are consistent with the view that PAb204 is directed against a catalytic helicase motif in T antigen. However, the epitope is located hundreds of amino acids away from the type A motif in p68.

p68 and SV40 T antigen are RNA helicases

The p68 amino acid sequence is 32% identical to that of eIF4A, a well characterised RNA helicase (Ford, et al., 1988). The sequence motifs shared by p68 and eIF4A define a family of putative RNA helicases which is described in detail below. T antigen is an ATP-dependent DNA helicase (Stahl et al., 1986). The indirect link between T antigen and eIF4A led Sheffner et al. (1989) to test T antigen for RNA helicase activity. They found that T antigen is able to unwind RNA in the presence of UTP, GTP or CTP. That the RNA unwinding is inhibited by ATP and 100 mM KCl may indicate that a cofactor is missing. T antigen is

known to interact with many different cellular proteins, including DNA polymerase α (Smale and Tjian, 1986; Gannon and Lane, 1987), p105Rb (DeCaprio, et al., 1988), p107 (Ewen, et al., 1989; Dyson, et al., 1989), p53 (Lane and Crawford, 1979; Linzer and Levine, 1979) and AP2 (Mitchell, et al., 1987). Other processes which might affect T antigen's RNA helicase activity include phosphorylation (Prives, 1990), covalent coupling to RNA (Carroll, et al., 1988), oligomerisation and DNA binding (Mastrangelo, et al., 1989). At present the interaction which appears most likely to control the nucleic acid preference of T antigen's helicase activity is p53 binding. p53 increases T antigen's GTPase activity (Tack, et al., 1989) and inhibits its DNA helicase activity (Wang, et al., 1989; Sturzbecher, et al., 1988). The most convincing way to demonstrate a biological function for T antigen's RNA helicase activity would be to isolate mutants which had selectively lost the ability to unwind RNA, show that they gave rise to defective virus and identify a relevant defect in RNA processing. In practice this could probably be achieved by mutagenising the putative guanine binding motif in T antigen ("NKRT", Scheffner, et al., 1989).

Various groups have looked for RNA targets which T antigen might unwind. Michel and Schwyzer (1982) found that up to 10% of T antigen was present in cytoplasmic messenger ribonucleoprotein particles. Weil's group found that the bulk of T antigen-bound RNA is derived from non-viral, non-coding, double-stranded heterogeneous-nuclear RNA (Khandjian, et al., 1982; Darlix, et al., 1984). Carroll, et al. (1988) identified RNA covalently linked to a carboxy-terminal serine residue in T antigen. Westphal (personal communication) has cloned several RNAs which bind to T antigen *in vitro*. One clone appears to transform primary cells when expressed in one orientation and suppresses T antigen-induced transformation when expressed in the opposite orientation. However, the clone contains no significant open reading frames and doubts persist about the specificity of the interaction between T antigen and the RNA.

The RNA helicase activity of T antigen could be used to regulate alternate splicing. In the context of SV40 infection this might influence the ratio of large T to small t antigen produced

Motif I	Motif Ia	Motif II	Motif III
+.....G.GKT	++..o..+..o+	++++DE.H	+.TAT...o
S		D	SGS
 Motif IV	 Motif V		 Motif VI
F..S.o.o	o..o..+++..To+...G.o+		o..o..Q..GR..R
Y T	S S		H

Figure 1.2 Putative helicase motifs identified by Gorbalenya in superfamily 2. +: V,L,I,M,W,Y,F. o: S,T,D,E,R,K,N,Q.

or direct the synthesis of host proteins beneficial to the virus. Adenovirus capsid protein IIIa messenger RNA is produced from an alternate splice which is used late in infection. Failure to use this splice in monkey cells is overcome by transfection of T antigen (Delsert, et al., 1989).

Hay, et al. (1982) proposed that a sequence near the start of transcription of the SV40 agnoprotein may be able to adopt two conformations and called the sequence the SV40 attenuator by analogy with the *trp* operon attenuator (Yanofsky, 1981). Hay, et al. postulate that when the sequence adopts one conformation it leads to translation of the agnoprotein and VP1, and when it adopts the alternative conformation it leads to premature transcription termination. It is conceivable that T antigen could regulate capsid protein production by using its RNA helicase activity to alter the conformation of the switch sequence in the attenuator.

The case for linking T antigen with p68 now rests as much on functional as on immunological evidence. The discovery that p68 belongs to a large family of putative RNA helicases suggests that, irrespective of any link with T antigen, p68 is of fundamental importance in its own right. It is therefore appropriate to describe in detail the family of proteins to which p68 belongs.

The Hodgman/Gorbalenya helicase motifs

Hodgman (1988) and Gorbalenya, et al. (1988) identified putative helicase motifs in proteins from 21 organisms including RNA viruses, DNA viruses, bacteria and yeast ("superfamily 1"). Gorbalenya, et al. (1989) extended the analysis to define a second, large putative helicase family ("superfamily 2", figure 1.2) which includes two subfamilies: "DEAD box proteins" (Linder, et al., 1989) and "DEAH box proteins" (Gorbalenya, et al., 1989; Company, et al., 1991). In superfamily 1 DNA helicase activity has been demonstrated for helicase IV (Wood and Matson, 1989), *uvrD* (Kumura and Sekiguchi, 1984) and *rep* (Kornberg, et al., 1978). In superfamily 2 DNA helicase activity has been demonstrated for RAD3 (Sung, et al., 1987) and RNA helicase activity has been demonstrated for eIF4A (Rozen, et al.,

1990) and p68 (Hirling, et al., 1989). In superfamily 1 DNA-dependent ATPase activity has been demonstrated for recB (Hickson, et al., 1985). In superfamily 2 RNA-dependent ATPase activity has been demonstrated for plum pox potyvirus CI protein (Lain, et al., 1991), SrmB (Nishi, et al., 1988) and PRP16 (Schwer and Guthrie, 1991). For some of the viruses there is genetic evidence that the putative helicase genes are required for viral replication [reviewed by Gorbalenya, et al. (1989) and Gorbalenya, et al. (1990)]. SV40 T antigen possesses a different set of putative helicase motifs ("superfamily 3", Gorbalenya and Koonin, 1989; Gorbalenya, et al., 1990).

DEAD and DEAH box proteins

All members of the DEAD and DEAH families have the same general structure. The conserved motifs lie in a 300 amino acid domain in the centre of the molecule. The degree of conservation varies for different proteins, but "unrelated" members of the family have around 30% identical amino acids in the central domain. The terminal domains, which are highly divergent, probably provide specific RNA and protein binding sites. Two conserved motifs have a defined function: GKT and DEAD are specialised forms of the type A and B nucleotide binding motifs (Walker, et al., 1982). Comparison with known structures suggests that elements of these motifs bind the phosphates in ATP. Comparison with adenylate kinase (Fry, et al., 1986) and ras (Schlichting, et al., 1990) suggests that the first aspartate in the DEAD motif probably coordinates a magnesium ion either directly or through a water molecule. Sequence analysis suggests that the region around motif V forms an RNA binding version of a helix-turn-helix motif (Leroy, et al., 1989).

It is possible to identify subfamilies of DEAD proteins in different organisms on the basis of sequence similarity. In subfamilies defined in this way the similarity in the central domain increases to give long runs of absolute identity, and in some cases the similarity extends to the terminal domains (eg, An3 and PL10). In only one case has a biochemically equivalent protein been examined in highly divergent species (mouse eIF4A and yeast TIF1), and in that case 65% of the deduced amino acid

Protein	Organism	Function	Biochemistry	Reference
TIF1&2	<i>S.cerevisiae</i>	Translation initiation	RNA helicase	Linder and Slonimski, 1988
SPB4	<i>S.cerevisiae</i>	25S rRNA processing		Sachs and Davis, 1990
MSS116	<i>S.cerevisiae</i>	Mitochondrial splicing		Seraphin, et al., 1989
DED1	<i>S.cerevisiae</i>	Suppresses splicing mutant		Jamieson, et al., 1991
DBP1	<i>S.cerevisiae</i>	Suppresses DED1 mutant		Jamieson and Beggs, 1991
DBP2	<i>S.cerevisiae</i>	This work		Iggo, et al., 1991
PRP5	<i>S.cerevisiae</i>	Nuclear splicing		Dalbadie-McFarland and Abelson, 1990
PRP28	<i>S.cerevisiae</i>	Nuclear splicing		Strauss and Guthrie, 1991
CA3,4,5/6	<i>S.cerevisiae</i>	? (degenerate PCR clones)		Chang, et al., 1990
CA9,10,11	<i>S.cerevisiae</i>	? (degenerate PCR clones)		J.Abelson, personal communication
PRP2	<i>S.cerevisiae</i>	Nuclear splicing*		Chen, et al., 1990
PRP16	<i>S.cerevisiae</i>	Nuclear splicing*	RNA-dep. ATPase	Schwer and Guthrie, 1991
PRP22	<i>S.cerevisiae</i>	Nuclear splicing*		Company, et al., 1991
JAI,2	<i>S.cerevisiae</i>	? (degenerate PCR clones)*		Company, et al., 1991
dbp2	<i>S.pombe</i>	This work		Iggo, et al., 1991
srmB	<i>E.coli</i>	L24 suppressor		Nishi, et al., 1988
dbpA	<i>E.coli</i>	This work	RNA-dep. ATPase	Iggo, et al., 1990
3 rhl genes	<i>E.coli</i>	?		M.Cashel, personal communication
vasa	<i>Drosophila</i>	<i>Grandchildless-knirps</i> gene		Lasko and Ashburner, 1988
RM62	<i>Drosophila</i>	?		Dorer, et al., 1990
ME31B	<i>Drosophila</i>	?		de Valoir, et al., 1991†
An3	<i>Xenopus</i>	?		Gururajan, et al., 1991
eIF4A	Mouse	Translation initiation		Nielsen, et al., 1985
PL10	Mouse	Spermatogenesis	RNA helicase	Leroy, et al., 1989
p68	Man	This work	RNA helicase	Ford, et al., 1988
Table 1	The DEAD and DEAH families.		* DEAH proteins	† cited by Wassarman and Steitz, (1991)

sequences are identical (Linder and Slonimski, 1989). eIF4A/TIF has almost the shortest terminal domains of any DEAD box protein, which has the effect of decreasing the relative contribution of the least conserved parts of the molecule to the overall percentage identity. Defining homology on the basis of sequence similarity runs the risk that two proteins with different functions may be considered homologous. Defining homology in genetic terms can also give misleading results. For example, eIF4A will not complement *TIF* mutants (Prat, et al., 1990). Defining homology in biochemical terms is also confusing. For example, TIF will substitute for eIF4A in unwinding assays (Jaramillo, et al., 1990) but eIF4A will not substitute for TIF in translation assays (Prat, et al., 1990). Ultimately, two proteins should be considered homologous if their normal cellular function is the same, but for the majority of DEAD proteins the normal cellular function is unknown. Table 1 lists known DEAD/DEAH genes. Figure 1.3 shows an alignment of deduced amino acid sequences of selected proteins. Only one member of a subfamily (defined by sequence) has been shown, and where possible an *S.cerevisiae* gene is given. Figure 1.4 shows the conserved motifs using the sequence of p68 as an example.

The DEAD family

eIF4A

eIF4A is the best characterised member of the DEAD box family. It is the only member of the family whose helicase activity was inferred from biochemical rather than DNA sequence information. It forms part of the messenger RNA cap binding complex of proteins which loads messenger RNA onto ribosomes [reviewed by Rhoads (1988) and Sonenberg (1988)]. This complex has been extensively studied using in vitro translation systems. It consists of eIF4A (Nielsen, et al., 1985), eIF4B (Milburn, et al., 1990), eIF4E (Altmann, et al., 1987; Rychlik, et al., 1987) and eIF4F (a complex of eIF4A, eIF4E and p220). eIF4B, which contains the RNP1 and RNP2 consensus RNA binding motifs, is essential for RNA unwinding catalysed by

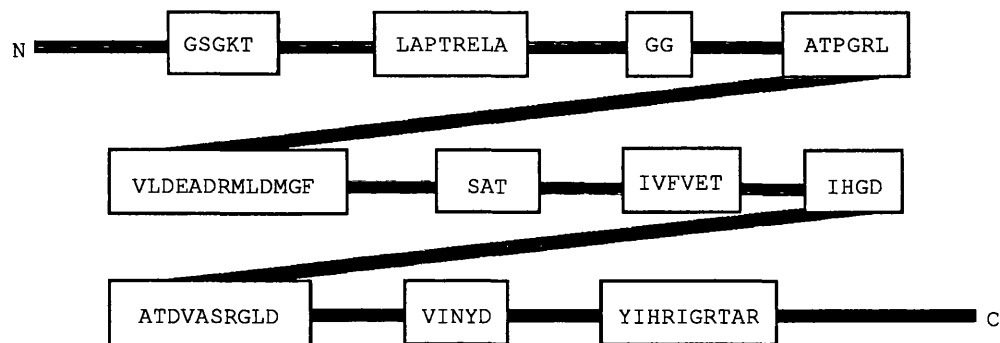


Figure 1.4 Diagram showing the most conserved regions of human p68.

eIF4A and eIF4F (Rozen, et al., 1990). eIF4E is the m⁷GpppX cap binding protein (Altmann, et al., 1987; Rychlik, et al., 1987).

eIF4A has been shown to induce ATP-dependent increases in RNase sensitivity (Ray, et al., 1985) and to possess ATP-dependent RNA binding (Abramson, et al., 1987), RNA-dependent ATPase (Grifo, et al., 1984) and RNA-dependent helicase activity (Rozen, et al., 1990). There are two closely related eIF4A genes in mouse (Nielsen and Trachsel, 1988), yeast (Linder and Slonimski, 1988) and rabbits (Merrick, et al., 1990). In the mouse the two genes show tissue specific variation in levels of expression (Nielsen and Trachsel, 1988). Compared with other DEAD box proteins, eIF4A has a very small amino-terminal domain. One interpretation is that the missing region is an RNA binding domain which is provided in *trans* to eIF4F by eIF4E (for cap binding) and to free eIF4A by eIF4B (through its RNP domains). eIF4F's helicase activity is cap-dependent but eIF4A's helicase activity is cap-independent (Rozen, et al., 1990). The cap-independent unwinding is bidirectional. Bidirectional unwinding would not be surprising if the helicase complex bound initially to double-stranded regions in the template, but this is apparently not the case because completely double-stranded templates are not unwound. It is possible that catalytic activity is strictly unidirectional but that the helicase complex can only initiate unwinding from a single-stranded region. The complex could bind to single-stranded RNA, slide freely until it reached a double-stranded region and then, if necessary, switch strand before invading the duplex.

***S.cerevisiae* DEAD box genes involved in RNA splicing**

MSS116

MSS116 is a yeast nuclear gene which is required for splicing of several mitochondrial introns (Seraphin, et al., 1989). By testing respiratory competence, Seraphin et al identified introns which could not be spliced out of cytochrome b (*cob*) and cytochrome c oxidase subunit 1 (*cox1*) in *mss116-1* or *mss116-2* containing strains. Specifically, *MSS116* is required for excision of *cox1* a1; *cox1a5α* and/or *cox1a5β*; *cobb1*, *cobb2* and/or *cobb3*.

cox1a1 and *cobb1* are group II introns; the rest are group I introns. Deletion of all of the mitochondrial introns does not confer respiratory competence on yeast containing another *MSS116* allele, *mss116-3*, which suggests that *MSS116* has additional functions. The authors suggest that some self-splicing introns adopt inactive conformations which require an RNA helicase to convert them to an active form.

PRP5

PRP5 is required for spliceosome assembly (Lin, et al., 1987). It was originally identified as a temperature sensitive mutant defective in macromolecular synthesis (Hartwell, 1967). Dalbadie-McFarland and Abelson (1990) cloned it by complementation of a temperature sensitive *prp5* mutant. Since U2 small nuclear RNA binding requires functional PRP5, they suggest that PRP5's helicase activity may be used to modify U1, U2 or pre-messenger RNA secondary structure to allow U2 binding.

PRP28

PRP28 was identified by screening cold-sensitive yeast for splicing defects (Strauss and Guthrie, 1991). Double mutants containing conditional alleles of *PRP28* and *PRP24* are inviable. Strauss and Guthrie also identified a temperature sensitive suppressor of the cold sensitive phenotype of the *prp28-1* allele as a new allele of *PRP8*. An attractive model which explains these findings is that the helicase activity of *PRP28* is used in the splicing reaction to separate the U4 from the U6 small nuclear RNA (Strauss and Guthrie, 1991).

The *DED1*, *DBP1*, PL10 and An3 family

PL10, An3, *DED1* and *DBP1* form a subfamily defined by sequence homology. *DED1* and *DBP1* are yeast genes which have been linked to pre-messenger RNA splicing by suppressor analysis. PL10 and An3 are developmentally regulated mouse and *Xenopus* genes respectively, although conclusions about their function rest on little more than circumstantial evidence about the timing and location of their expression.

PL10

Leroy, et al. (1989) identified the PL10 gene by screening a testis cDNA library with a probe derived from the mouse Y chromosome. In addition to clones (called PL5) corresponding to the probe they obtained clones (called PL10) which map to chromosome 1. *In situ* hybridisation and northern blotting show that PL10 expression is restricted to the testis. Expression peaks in spermatocyte cells in prophase of the first meiotic division, although high levels of transcript are still detectable in subsequent early haploid stages of spermatogenesis. The authors speculate that PL10 may play a role in translational control of stored messages.

An3

Rebagliati, et al. (1985) identified a maternal RNA which is concentrated in the animal hemisphere of unfertilised *Xenopus* eggs and remains there through the early blastula stage. Gururajan, et al. (1991) sequenced the cDNA and found that it was 74% identical to PL10. Unlike PL10, which is expressed only in the male germ line, the An3 message is broadly distributed in adult tissues.

DED1

PRP8 is one of the RNA processing genes originally identified by Hartwell (1967). It is a 280 kD component of the U5 small nuclear ribonucleoprotein particle (snRNP). It associates with the U4/U6 small nuclear RNA/snRNP to form a multi-snRNP complex which is thought to be important for spliceosome assembly (Lossky, et al., 1987). Jamieson, et al. (1991) isolated cold-sensitive suppressors of a temperature sensitive *PRP8* mutant, *prp8-1*, which map to a previously identified DEAD box gene of unknown function, *DED1* (Struhl, 1985). Struhl called the gene *DED1* before the DEAD motifs were identified. It stands for *Defines Essential Domain*, and refers to the observation that deletion of the gene is lethal (Struhl, 1985). Jamieson, et al. (1991) proposed that *DED1* may be involved in the separation of U4 from U6, but the ability of *PRP8* mutants to suppress *PRP28*

mutants calls into question the role of *DED1* in splicing (Derek Jamieson, personal communication). *DED1* is 62% identical to PL10. Taken at face value the conclusion that An3 and PL10 control translation of stored messages conflicts with the yeast data implicating the *DED1*, *DBP1*, PL10 and An3 subfamily in splicing. The involvement of An3 and PL10 in oogenesis and spermatogenesis might suggest a role for *DED1* and *DBP1* in meiosis but there is no evidence for this (Derek Jamieson, personal communication).

DBP1

This non-essential gene was cloned by Jamieson and Beggs (1991) when they attempted to clone *DED1* by complementation of the cold sensitive defect conferred by the suppressors of *prp8-1* described above (ie, *DBP1* suppresses a suppressor). The deduced amino acid sequences of *DBP1* and *DED1* are 72% identical. Although *DBP1* can suppress cold sensitive *DED1* mutants if present on low copy number plasmids, it can only suppress null *DED1* mutants at high copy number. This is probably because *DBP1* is normally expressed at a much lower level than *DED1* (Jamieson and Beggs, 1991).

The p68, RM62 and *DBP2* family

The prototypic member of this family is p68. The *S.cerevisiae* *DBP2* and *S.pombe* *dbp2* genes are the subject of this thesis. The family is defined by sequence homology; functional complementation has not been attempted.

RM62

Dorer, et al. (1990) cloned RM62 while studying the *Tripliolethal* gene. *Tripliolethal* is the only locus in the *Drosophila* genome which is lethal when present in three copies. Dorer et al used P-M hybrid dysgenesis to generate new alleles of *Tripliolethal*. They identified a P element integration close to the *Tripliolethal* locus, and cloned the flanking DNA. The P element integration site lies within an intron which splits the initiator ATG of a previously unknown gene, RM62. When they sequenced

the gene they found that it encoded a DEAD box protein 66% identical to p68 in the central part of the molecule (between residues 95 and 520). As well as the intron in the initiator ATG, they identified a 107 bp intron at amino acid 371 (KEEK/LKTL), and a 1.7 kb intron in conserved motif V (at amino acid 451, RGLD/VDGI) [Alan Christensen, personal communication]. p68 contains introns at the latter two sites, and the yeast genes *DBP2* and *dbp2* contain an intron in motif V (see chapter three).

Subsequent matings revealed that the P element was not responsible for the "Triplolethality" and further work has ruled out any link between RM62 and *Triplolethal* (Alan Christensen, personal communication).

Other DEAD genes

SPB4

Deletion of the poly(A) binding protein gene (*PAB1*) in yeast leads to inhibition of translation initiation, resulting in a decrease in polyribosome content and accumulation of free ribosomal subunits (Sachs and Davis, 1989). *spb4-1* is a cold sensitive suppressor of a temperature sensitive poly(A) binding protein mutant, *pab1-F364L* (Sachs, et al., 1987). *spb* mutations reduce the amount of the 60S ribosomal subunit (Sachs and Davis, 1989). Pulse chase analysis shows that maturation of 25S ribosomal RNA is delayed and the amount of mature 25S ribosomal RNA is reduced 2.5 fold in the *spb4-1* mutant (Sachs and Davis, 1990). The authors argue that *SPB4* is unlikely to be part of mature ribosomes but may be part of preribosomal particles.

srmB

srmB is a high copy number suppressor of a defect in *E.coli* ribosome assembly caused by a temperature sensitive mutation in the ribosomal protein L24 (Nishi, et al., 1988). The wild type function of *srmB* is unknown. Nishi et al speculate that L24 normally destabilises an RNA structure which impedes ribosome assembly, and that high levels of SrmB can cooperate with mutant L24 to achieve the same effect.

srmB is one of the few DEAD box proteins for which any biochemical data is available. Nishi et al overexpressed and purified the protein and tested it for RNA binding and ATP hydrolysis. It has no ATPase activity in the absence of nucleic acid. ATP hydrolysis is stimulated by double-stranded DNA (29 fmol/ μ g/s), single-stranded DNA (67 fmol/ μ g/s) and a variety of RNAs (101-447 fmol/ μ g/s). RNA binding is ATP sensitive: addition of ATP reduces poly(A) binding by 25-50%.

vasa

Anti-horseradish peroxidase sera stain the nervous systems of *Drosophila* and grasshoppers (Jan and Jan, 1982). To generate monoclonal antibodies against neural antigens Hay, et al. (1988a) made hybridomas from mice immunised with horseradish peroxidase and screened them on tissue sections. They obtained an antibody (Mab46F11) which recognises germ line cells but not horseradish peroxidase. The localisation of the Mab46F11 antigen in polar granules suggested that the gene for it might belong to the *grandchildless-knirps* class of maternal effect gap genes. Analysis of *grandchildless-knirps* mutants revealed that the Mab46F11 antigen is absent in *vasa* mutants. A cDNA encoding the Mab46F11 antigen was cloned from a λ gt11 library and found to contain the DEAD motifs (Hay, et al., 1988b). Simultaneously, Lasko and Ashburner (1988) cloned the same gene (*vasa*) by chromosome walking. Cell staining and in situ hybridisation show that *vasa* protein is synthesised in nurse cells and transferred to the oocyte, where it becomes localised in polar granules and then pole cells. Thereafter, it is confined exclusively to germ line structures (Hay, et al., 1988a; Lasko and Ashburner, 1990; Hay, et al., 1990). *Vasa* is required for oocyte maturation, pole cell formation, and abdominal segmentation [reviewed by Lasko and Ashburner (1990)]. Although a speculative hierarchy of genes involved in these processes has been established it is far from clear what part *vasa*'s putative RNA helicase activity may play. As an RNA binding protein it might have a role in localising the transcripts of other posterior group genes. It could then use its catalytic activity to alter their expression. Two pole cell specific transcripts are known: cyclin B

and *nanos*. At least one *oscar* mutant localises cyclin B but not *vasa*, suggesting that *vasa* does not anchor cyclin B (J. Raff, personal communication). However, *vasa* is required for *nanos* localisation. Failure to localise *nanos* would result in a failure to inactivate the maternal *hunchback* product which would in turn lead to a failure to activate the zygotic gap gene *knirps*, thus giving rise to *vasa*'s abdominal segmentation phenotype (Lasko and Ashburner, 1990).

CA3, CA4 and CA5/6

These *S.cerevisiae* genes were identified by polymerase chain reaction (PCR) amplification of genomic DNA using degenerate primers corresponding to the DEAD and HRIGR motifs (Chang, et al., 1990). Their functions are unknown.

The DEAH family

Eukaryotic DEAH proteins have only recently been identified. They contain the motifs originally identified by Gorbalenya et al. (1989) in helicase superfamily 2. Apart from DEAH, the proteins contain recognisable GKT and SAT motifs. The final helicase motif (HRIGRTGR in DEAD proteins) is QR-GRAGR in the DEAH family. Given the large divergence between the DEAD and DEAH motifs and the fact that no DEAH protein has been shown to unwind RNA it would be premature to ascribe helicase activity to this family. The three DEAH genes identified as splicing mutants (*PRP2*, *PRP16*, *PRP22*) all function *in vitro* in a way compatible with them having a kinetic proofreading role (Wassarman and Steitz, 1991). It is envisaged that they act at different steps in splicing as switches which must be thrown to allow the spliceosome to progress to the next step of the splicing reaction. How the spliceosome distinguishes *bona fide* splice sites from other almost identical sequences is an unsolved problem. Kinetic proofreading lends itself to detecting small differences in affinity. It works by inserting a short delay before an irreversible catalytic step. In that time incorrect substrate molecules are given the opportunity to leave the active site. The inbuilt delay is set by the intrinsic rate of hydrolysis of ATP by the kinetic proofreading protein.

PRP16

PRP16 was first identified in a screen for suppression of an A to C mutation in the lariat branch point sequence, TACTAAC (Burgess, et al., 1990). Genomic DNA libraries from a *prp16* mutant were screened for clones with suppressor activity. The wild type *PRP16* allele was then cloned by gap repair and shown to be essential for viability. Burgess, et al. (1990) argued that PRP16 might have a kinetic proofreading function. Schwer and Guthrie (1991) have subsequently shown that PRP16 is required for the second step of splicing. Consistent with the kinetic proofreading model, they found that the protein interacts transiently with the spliceosome in an ATP-dependent fashion. In addition they showed that the purified protein has RNA-dependent ATPase activity.

PRP22

Mutants in this gene accumulate pre-messenger RNA and introns in lariat form (Vijayraghavan, et al., 1989). Company, et al. (1991) showed that spliced RNA and lariat intron accumulate in large splicing complexes at the non-permissive temperature in extracts from a *prp22* strain and suggest that the wild type gene is required for release of spliced product from the spliceosome. They cloned *PRP22* by complementation of the temperature sensitive defect. As in other members of the DEAD/DEAH family the amino and carboxy termini are highly divergent. The argument that these domains provide specific RNA and protein binding sites is supported by the observation that in *PRP22* the unique amino terminal domain contains putative RNA binding motifs of the type first identified in *E.coli* ribosomal protein S1 and polynucleotide phosphorylase (Company, et al., 1991).

PRP2, JA1 and JA2

These genes all belong to the DEAH family. *PRP2* is required to convert pre-messenger RNA to linear exon 1 and lariat-exon 2 in the spliceosome (Lin, et al., 1987; Chen and Lin, 1990). JA1 and JA2 were identified by degenerate PCR (Company, et al., 1991).

Functional implications

There are at least five DEAD genes in *E.coli* (Michael Cashel, personal communication) and twenty DEAD/DEAH genes in *S.cerevisiae* (John Abelson, personal communication). In seeking to explain the existence of so many putative RNA helicases one can invoke two distinct arguments. The "housekeeping" argument dictates that wherever one sees labile basepairing, for example small nuclear RNA binding to pre-messenger RNA, one should expect to find a DEAD or DEAH protein. Since RNA tends to fold itself into complex secondary structures, this argument can be extended to include any situation where RNA sequence needs to be read directly. Cech's kinetic analysis of the ribozyme illustrates what can happen in the absence of an RNA helicase (Herschlag and Cech, 1990). The ribozyme is a perfect enzyme: its rate of catalysis is limited only by the speed at which new substrate can diffuse to its active site. Although DNA never acts as a physiological substrate, and kinetic analysis shows that the k_c for DNA is one thousandth that for RNA, the overall rate of cleavage of DNA is only a tenth that of RNA. The reason for the discrepancy is that the ribozyme lacks a helicase. The price DNA pays for not fitting properly into the active site is a low k_c , but the benefit is that cleaved product is quickly released. In nature the ribozyme is a suicide enzyme but when used artificially as a biochemical tool it must cut many times and in practice it is release of product that limits turnover. There is thus considerable scope to improve ribozyme technology by using a DEAD protein to remove cleaved product. If ribozymes really were the primordial nucleic acid polymerases (Doudna and Szostak, 1989) one can reasonably ask how they overcame the problem of turnover. At the very least it is not surprising that helicases should be as ancient as polymerases.

The alternative to the "housekeeping" argument is the "gene control" argument. The only DEAD gene which is known to be involved in gene control is the *Drosophila* gene *vasa*, but the mechanism of its action is obscure. The level of activity of the *trp* operon (Yanofsky, 1981) and the ColE1 replicon (Polisky, 1988) depends on the ratio of competing RNA secondary

structures in the *trp* attenuator and replication primer respectively. It is possible that regulation of alternate splicing in eukaryotes also involves the formation of competing secondary structures. The existence of many different RNA secondary structures within a few percent of the lowest ΔG (Zuker, 1989) suggests that RNA may be particularly amenable to allosteric modification. The question posed by the discovery of a large family of RNA helicases is the extent to which this plasticity is deliberately manipulated as a mechanism for gene control.

Scope of this study

When the work described in this thesis began, the only DEAD box proteins which had been identified were p68 and eIF4A. It was therefore important to examine p68 for biochemical evidence of RNA helicase activity. Data is presented in chapter six which shows that immunoaffinity-purified p68 possesses RNA-dependent ATPase activity. p68 is now one of the best understood members of the DEAD box family in a biochemical sense but its biological function is unknown. A broad range of techniques has been used in an attempt to identify p68's biological function. Chapter five presents cell staining, transfection and antibody microinjection data which links p68 with the nucleolus. Chapter three describes the identification and cloning of putative yeast p68 homologues and a new *E.coli* gene unrelated to p68. The yeast genes are called *DBP2* in *S.cerevisiae* and *dbp2* in *S.pombe*. Chapter seven includes data showing that the *DBP2* and *dbp2* genes are essential for viability. The p68, *DBP2* and *dbp2* genes all contain an intron at the same site in conserved motif V. The unusual features of the intron suggest that it may have a function in *S.cerevisiae*. Data is presented in chapter seven describing the effect of the *DBP2* intron on *DBP2* expression. This was examined by northern blotting and by immunoblotting using a monoclonal antibody against the sequence LVLDEADRMLDMGFEP. The production of the antibody is described in chapter four.

Supplementary Information

Suppliers

Where no source is given materials were purchased from Sigma. Bacillus Calmette Guerin was purchased from Evans Medical. Tuberculin purified protein derivative was purchased from Statens Seruminstitut, Copenhagen. Radiochemicals were purchased from Amersham, except for 5'-³²P cytidine 3', 5' diphosphate (pCp) which was purchased from ICN.

References

More detailed descriptions of general immunological and polymerase chain reaction techniques can be found in "Antibodies: a laboratory manual" by Harlow and Lane (1988) and "PCR protocols" by Innis et al. (1990).

Radiochemicals

Radiochemicals were used within one half-life of their activity date. The specific activity of ³²PαdATP and ³²PαdCTP was >3000 Ci/mmol. The specific activity of ³²PγATP was >5000 Ci/mmol. The specific activity of ³⁵SαdATP was >600 Ci/mmol. The specific activity of 5'-³²P pCp was 2500 Ci/mmol. For the ATPase assays described in chapter six trace amounts of radiolabelled nucleotide were used. The specific activity was different in each set of reactions because the ATP concentration varied. In the experiment shown in figure 6.4, for example, 62.5 nCi was included in each 25 μl reaction at each ATP concentration (33 μM, 100 μM, 300 μM, 900 μM and 2.7 mM).

Chapter 2 Materials and Methods

Introduction

To avoid repetition all information about materials and methods is presented in this chapter. Since the overall structure of this thesis lends itself to subdivision on technical grounds it has been possible to group technical information according to chapter. Chapter three deals with DNA cloning, chapter four with antibody production, chapter five with cell biology, chapter six with biochemistry and chapter seven with yeast genetics and RNA analysis. These subdivisions are clearly marked below to facilitate cross-referencing.

Molecular Biology (Chapter 3)

Plasmids

Routine cloning was done using standard techniques, as described by Sambrook, et al. (1989). Plasmids created in the course of this work are listed in table 2. All sequencing of prototypic clones was done after subcloning the relevant DNA in Bluescript (Stratagene). Important features of specialised plasmids employed for expression of cloned genes are described in the text. In general, the authenticity of constructs was verified by sequencing the cloning sites.

DNA sequencing

DNA sequencing was done by the dideoxy technique with $^{35}\text{S}\alpha\text{dATP}$ using a Sequenase kit (United States Biochemical) from either PCR templates (see below) or double-stranded plasmid templates. Plasmids were denatured for 5 minutes at 37° in 0.2 M NaOH and then precipitated with ethanol. Primer was annealed to template for 15 minutes at 37° . Sequencing reaction products were electrophoresed through denaturing 5% polyacrylamide field gradient and 6% polyacrylamide salt gradient gels and analysed by autoradiography (Sambrook, et al., 1989).

E.coli dbpA fragments cloned in Bluescript

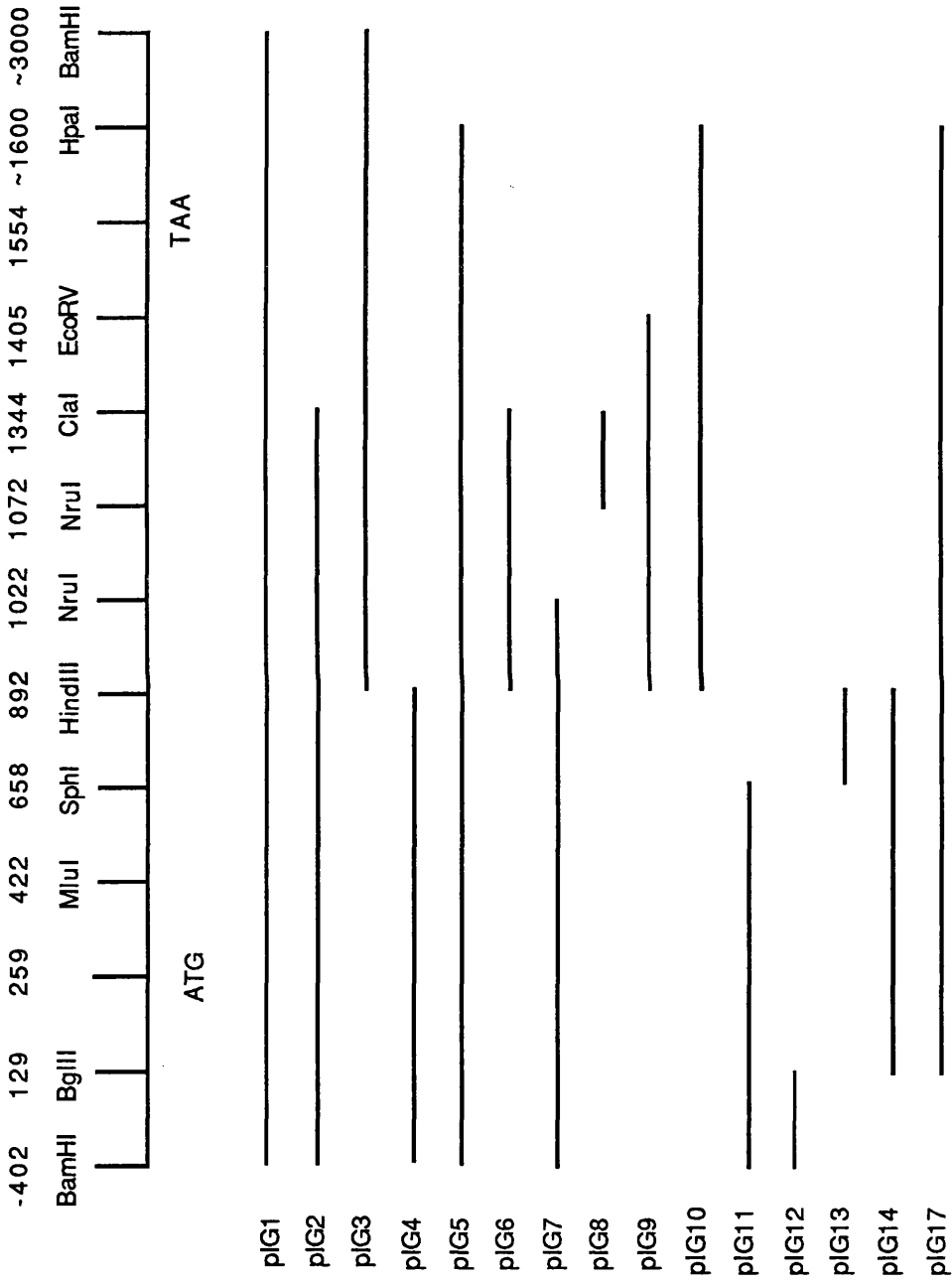


Table 2 Derivation of plasmids

B means fragment cloned into Bluescript
donor and target sites are identical unless specified
Δ means deletion between named sites
base numbers refer to sequences submitted to embl

E.coli

λ2B5	λ clone from which dbpA 3.4 kb BamHI fragment cut
pIG1B	dbpA 3.4 kb BamHI fragment
pIG2B	dbpA 1.7 kb BamHI/ClaI fragment
pIG3B	dbpA 2.1 kb BamHI/HindIII fragment
pIG4B	dbpA 1.3 kb BamHI/HindIII fragment
pIG5B	pIG1ΔHpaI/EcoRI (polylinker EcoRI site)
pIG6B	pIG2ΔBamHI/HindIII
pIG7B	pIG2ΔClaI/NruI
pIG8B	pIG2ΔBamHI/NruI
pIG9B	pIG3ΔBamHI/EcoRV
pIG10B	pIG3ΔBamHI/HpaI
pIG11B	pIG4ΔHindIII/SphI
pIG12B	pIG4ΔHindIII/BglII
pIG13B	pIG4ΔBamHI/SphI
pIG14B	pIG4ΔBamHI/BglII
pIG17B	pIG5ΔBamHI/BglII

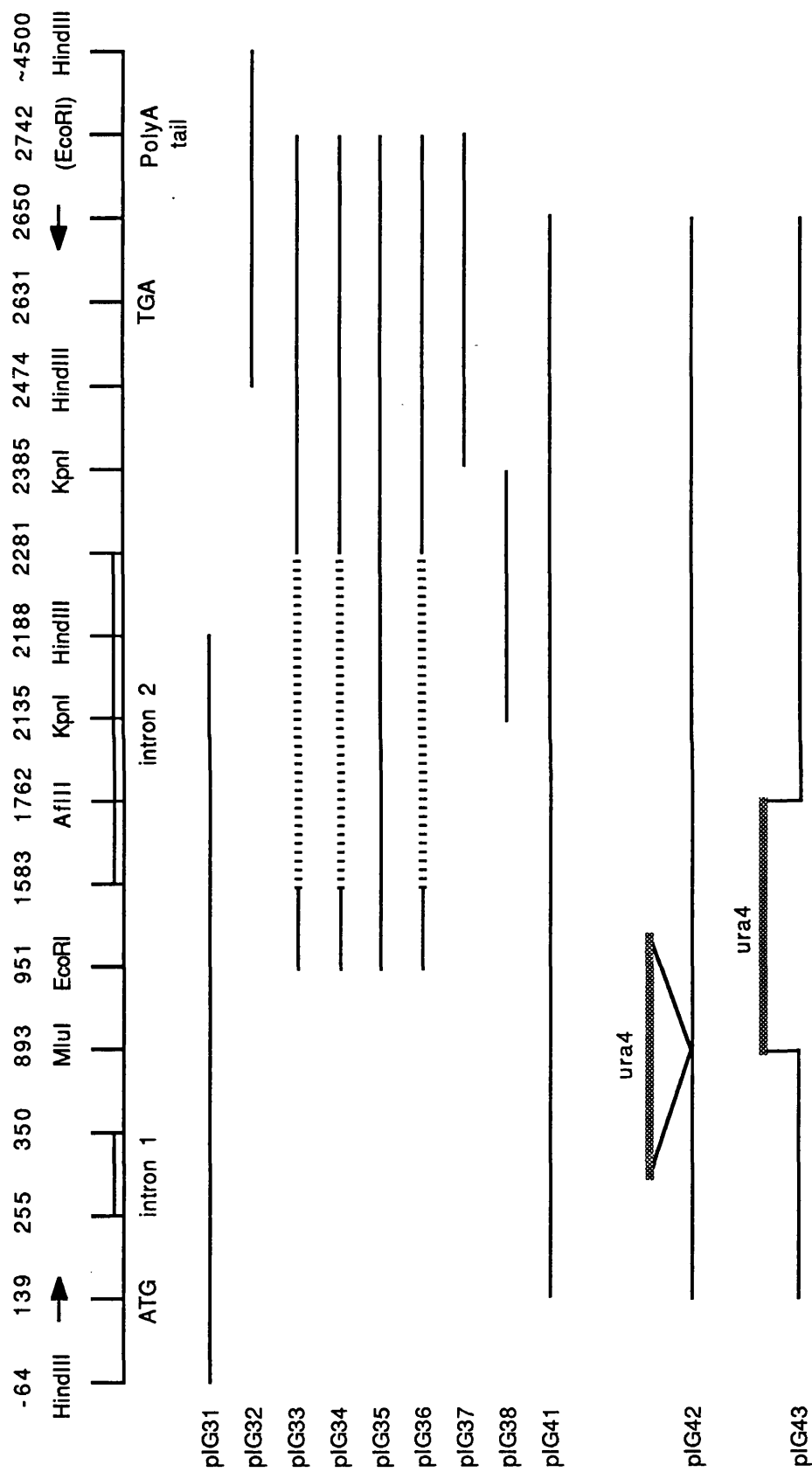
S.pombe

pIG30	dbp2 cosmid
pIG31B	pIG30 HindIII fragment -64 to +2188
pIG32B	pIG30 2 kb HindIII fragment +2474 to 3' site
pIG33B	dbp2 λgt11 cDNA EcoRI fragment +951 to tail (wt1)
pIG34B	dbp2 λgt11 cDNA EcoRI fragment +951 to tail (wt7)
pIG35B	dbp2 λgt11 cDNA plus intron EcoRI fragment +951 to tail (op6)
pIG36B	dbp2 λgt11 cDNA EcoRI fragment +951 to tail (op8)
pIG37B	pIG35 KpnI/EcoRI fragment +2385 to tail
pIG38B	pIG35 KpnI fragment +2135 to +2385
pIG39	pIG33 EcoRI insert in pURB
pIG40	pIG35 EcoRI insert in pURB
pIG41B	dbp2 PCR clone in Bluescript EcoRV (note 1)
pIG42B	pIRT2U ura4 HindIII into pIG41 MluI
pIG43B	pIRT2U ura4 HindIII into pIG41 MluI/AflII

S.cerevisiae

λg3	DBP2 λgt11 genomic DNA fragment (Clontech YL1001)
λg11	DBP2 λgt11 genomic DNA fragment (Clontech YL1001)
pIG60B	λg3 2 kb EcoRI fragment +2157 to 3' site
pIG61B	λg3 0.8 kb EcoRI fragment 3' of pIG60
pIG62B	λg3 0.5 kb EcoRI fragment 3' of pIG60
pIG63B	BclI/BglII inverse PCR clone (note 2)

S.pombe dbp2 fragments cloned in Bluescript



pIG33, 34 & 36 are derived from normally spliced mRNA. They do not contain the intron (.....).
 Arrows represent PCR primers (see Table 2, note 1).
 The EcoRI site at 2742 is derived from the cloning linker.

pIG64B BclI/BglII inverse PCR clone (note 3)
 pIG65B pFL38 URA3 BglII into pIG60 BamHI/BglII
 pIG66B pIG63 SacII/EcoRV into pIG65 SacII/NotI
 pIG67B DBP2 SacII/XbaI PCR cloned into pIG60
 pIG68B DBP2 XbaI/BglII RT-PCR cloned into pIG67
 pIG69B DBP2 SacI/SacII 5'-adaptor into pIG68 (note 4)
 pIG70 pIG69 XhoI/NsiI into P2 XhoI/PstI
 pIG71B BamHI/BglII inverse PCR clone (note 5)
 pIG72B BamHI/BglII inverse PCR clone (note 5)
 pIG73B pIG71 0.7 kb SacII/XbaI into pIG60 (note 6)
 pIG74 pIG73 SacI/SalI into pFL39
 pIG75 pIG74 after BamHI/NsiI gap repair
 pIG76 pIG75 EcoRV/BglII into pRS313 EcoRV/BamHI
 pIG77B DBP2 PCR clones (note 7)
 pIG78 pIG75 SacII/ClaI into pIG70 SacII/SmaI

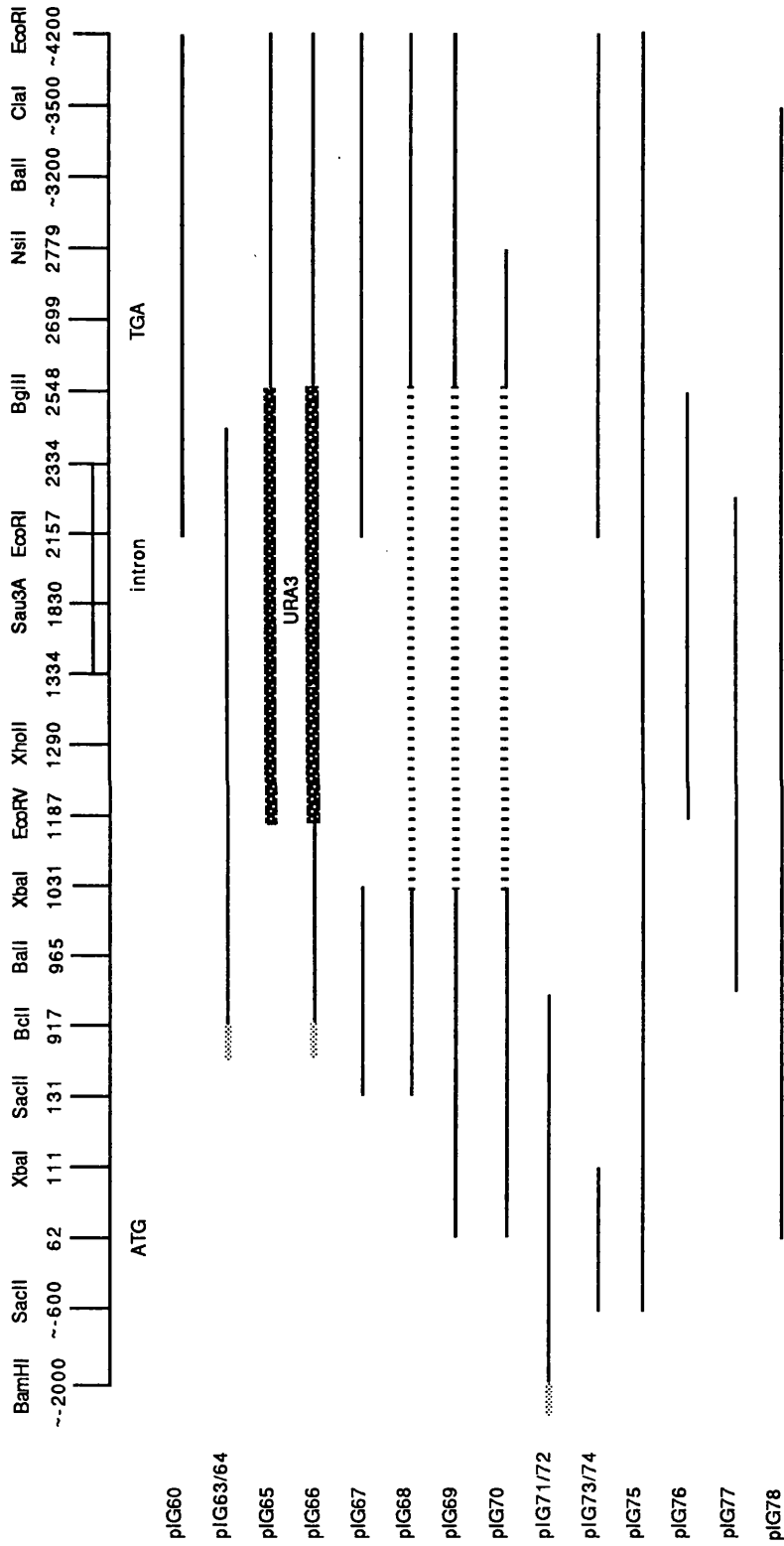
Human plasmids

pIG100 B2000 EcoRI insert in pNGS8+
 pIG101 Hela NdeI/ApaI plus pIG100 ApaI/HindIII into pT7.7 (note 8)
 pIG102 pIG101 XbaI/HindIII into pCIS5
 pIG103 pIG101 XbaI/HindIII into pCIS7 (note 9)
 pIG104B Hela RT-PCR SalI/ApaI (note 10)
 pIG105 pIG104 XbaI/ApaI into pIG102
 pIG106B OkBerg orthodox 1.1 kb BamHI/EcoRI fragment (note 11)
 pIG107B OkBerg alternate 1.98 kb BamHI/EcoRI fragment (note 11)
 pIG108B OkBerg alternate 0.56 kb BamHI/EcoRI fragment (note 11)
 pIG109B A800 0.7 kb EcoRI fragment
 pIG110B Conserved (motif V) intron PCR clone (note 12)

Other plasmids

pIG150 pZA1 PstI/HindIII into pUR291
 pZA1 PJ Nielsen (mouse eIF4AI 1.4 kb EcoRI fragment in pTZ19)
 A800 M Ford (λ gt11 human p68 0.7 kb cDNA including poly(A) tail)
 B2000 M Ford (λ gt11 human p68 2 kb cDNA in pUR290)
 pUBS520 U Brinkmann (*argU* 540 bp DpnI fragment in pFDX500)
 Bluescript Stratagene
 pT7.7 derived from pT7.1 (Tabor and Richardson, 1985)
 pIRT2U Carr, et al., 1989
 pNGS8+ Stoker, et al., 1989
 pFL38 F Lacroute
 pFL39 F Lacroute
 pRS313 Sikorski and Hieter, 1989
 P2 G Micklem (derived from pRS314, Sikorski and Hieter, 1989)
 pCIS5 Gorman, et al., 1990
 pCIS7 Gorman, et al., 1990
 pUR290 Ruther and Muller, 1983
 pUR291 Ruther and Muller, 1983
 pURB D Harrison (derived from pUR292, Ruther and Muller, 1983)

S.cerevisiae DBP2 constructs



..... The 5'-end of inverse PCR clones begins with DNA from the 3'-end of the gene. This is the sequence from the iPCR1 primer to the BglII site (see fig 3.16). The inverse PCR products were subcloned into the EcoRV site of Bluescript.

All sites shown above are genomic. The following sites were derived from polylinker DNA: BamHI at the 5'-end of URA3; SacII and NotI in pIG66; SacI and SalI in pIG73 and pIG74; SmaI in pIG78.

The Ball fragment from pIG66 was used for the DBP2 gene disruption. pIG66 contains inverse PCR sequences and reduplicated polylinker 5' of the Ball fragment.

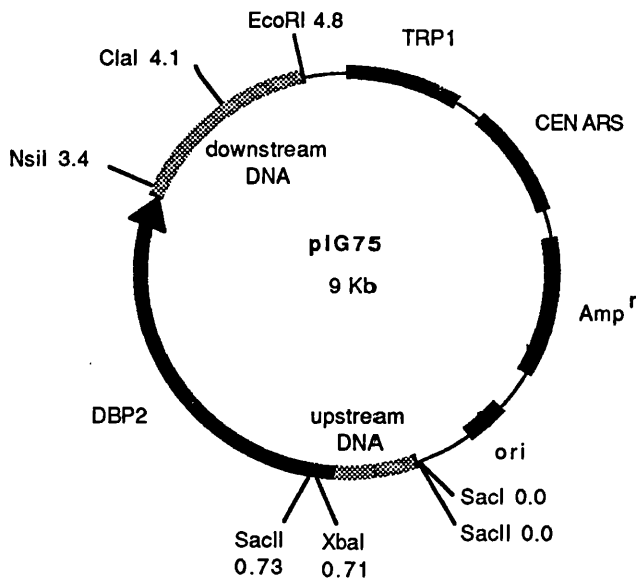
Apparently discontinuous sequences in pIG67, 73 & 74 are joined by Bluescript polylinker DNA between the relevant identical restriction sites. The BamHI site used for gap repair lies in this intervening polylinker sequence in pIG74.

pIG68, 69 & 70 do not contain the intron. The region marked (.....) is derived from normally spliced mRNA by RT-PCR.

The XhoI site used to transfer the DBP2 insert from pIG69 to pIG70 is adjacent to the initiator methionine in the adaptor (CTCGAGcaATG; see Table 2, note 4).

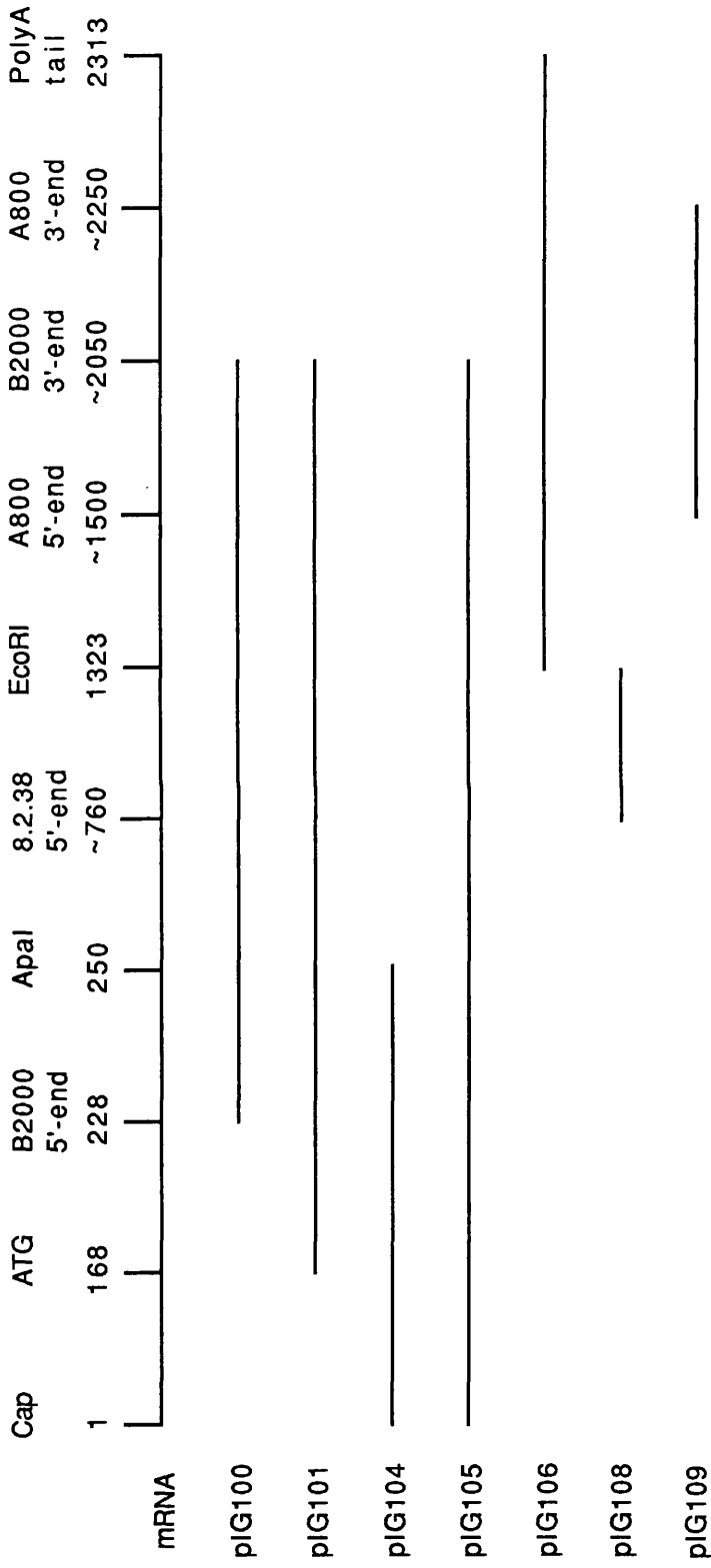
Notes

1. PCR from genomic DNA using ATGTCTTACAGAGATAACGAA and TACGCGGATTGTTGATCACCA.
2. iPCR from genomic DNA using CCACAAATTCATGTTTGGACT and ATTATGTTTACAGAATCGGTA.
3. iPCR from genomic DNA using CAAATTTTCCCTGGCTTGAAG and CTGCTATATCTTTCTTACCG.
4. Double-stranded SacI/SacII adaptor giving the sequence GAGCTCGAGCATATGACTTACGGTGGTAGAGATCAGCAATATACAAGACTAACTACAAGTCTAGAGGTGGCGACTTCCGCGG.
5. iPCR from genomic DNA using GAGTAATATTATGGGAGGCAG and CTGCTATATCTTTCTTACCG.
6. The XbaI site is at +111. The SacII/XbaI fragment overlaps the initiator methionine.
7. PCR from genomic DNA using CCTGATAGACAAACCTTGATG and CCACAAATTCATGTTTGGACT.
8. Mutagenic RT-PCR from Hela RNA to create an NdeI methionine using CATTGACCATATGTCGGGTTATTTCGAGTGA.
9. The pCIS5 polylinker is inverted in pCIS7.
10. anchor PCR from Hela RNA. SalI is in the RACE adaptor: GACTCGAGTCGACATCGATTTTTTTTTTTTTTTTTT.
11. Cloned from Okayama Berg pCD library by M Ford. "Alternate" refers to an alternate 5'-splice site in the conserved intron.
12. PCR from genomic DNA using CAACAAGAGCGTGACTGGGTT and ATCCTCTGAGGAGTTAGGGTA.



pIG75 is the prototypic genomic DBP2 clone. It was obtained by gap repair of pIG74 linearised with Nsil and BamHI (=XbaI above). SacI is plasmid unique. XbaI, SacI, Nsil, Clal and EcoRI are not unique. DBP2 can be excised on a 4.1 kb SacI/Clal fragment.

Human p68 constructs



The HindIII site in pIG100 was derived from the pNGS8+ polylinker. An NdeI site was introduced at position 168 in pIG101 by PCR mutagenesis (see table 2, note 8). The XbaI site in pIG101 was derived from the pT7.7 polylinker. pIG102 and pIG103 contain the same insert as pIG101. The XbaI site in pIG104 was derived from the Bluescript polylinker. The BamHI sites in pIG106, 107 and 108 were derived from the cloning linker. pIG106 was derived from an Okayama Berg clone called 14.3. pIG107 and 108 were derived from an Okayama Berg clone called 8.2.38. pIG107 is identical with pIG106 except that it contains the alternately spliced form of the conserved intron in helicase motif V (see p 71).

Primer extension

100 ng of primer was end-labelled with T4 polynucleotide kinase to a specific activity of ~500 cpm/pg and purified on a NENSORB column (Dupont). RNA from cells lysed in guanidinium isothiocyanate was purified by centrifugation through a caesium chloride step gradient (Ausubel, et al., 1987). *S.pombe* RNA was prepared from guanidinium lysed spheroplasts. Polyadenylated Hela cell RNA was affinity purified on oligothymidylic acid-cellulose resin ("oligo(dT)-selected") as described by Ausubel, et al. (1987). ~1µg of oligo(dT)-selected Hela cell RNA or ~10 µg of total *S.pombe* RNA was mixed with $\sim 1.25 \times 10^4$ cpm primer in a total volume of 10 µl of annealing buffer (400 mM NaCl, 10 mM piperazine bisethanesulfonic acid pH 6.4, 0.25 units/µl human placental ribonuclease inhibitor), heated to 85° for 15 minutes then incubated overnight at 55° in a capillary tube.

Complementary DNA (cDNA) synthesis was performed by expelling the contents of the capillary tube into 90 µl of extension buffer (final concentration 50 mM tris (hydroxy methyl) aminomethane (Tris) pH 8.2, 10 mM dithiothreitol, 6 mM MgCl₂, 0.5 mM deoxynucleotides, 0.25 units/µl human placental ribonuclease inhibitor), adding 10 units of avian myeloblastosis virus reverse transcriptase (Super RT, Anglian Biotec) and incubating at 42° for 1 hour (Williams and Mason, 1985). The product was ethanol precipitated and analysed by denaturing polyacrylamide gel electrophoresis and autoradiography.

Polymerase chain reaction techniques

DNA was amplified by the polymerase chain reaction [PCR, Saiki, et al. (1988)] using Amplitaq DNA polymerase (Perkin Elmer Cetus). The best results were obtained using buffer containing 10 mM Tris pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 0.1% Triton-X100, 200 µM deoxynucleotides, 10% dimethyl sulfoxide, 100 µM tetramethylammonium chloride and 5 µg/ml primers. A typical reaction with 21-mer primers used the following temperature cycles: 95° for 5 minutes to denature the template, lyse cells and inactivate proteases (where relevant) before addition of 1 unit of Taq polymerase; then 20-30 cycles of 92°

for 30 seconds, 55° for 30 seconds and 72° for 2 minutes. Despite the manufacturers claims it was rarely necessary to titrate the Mg²⁺, nucleotide or enzyme concentrations or vary the temperatures unless primers less than 21 nucleotides in length were used.

PCR mutagenesis

Mutagenesis was done by priming PCR with non-homologous primers. A typical primer is shown in table 2, note 8. It is a 30-mer with three point substitutions which create a new restriction site (NdeI) in the middle of the primer. The smallest number of PCR cycles (usually 20) was used to introduce the mutations and amplify the DNA, and the smallest possible region was cut from the PCR product. The mutagenised DNA was subcloned into plasmids using the nearest available restriction sites and then checked by DNA sequencing.

RT-PCR

cDNA was synthesised as above, except that 500 ng of primer was used and the annealing step was omitted. 10-20% of the product was then added to a normal PCR reaction for DNA amplification. To control for contamination by genomic DNA, plasmid or extraneous PCR products reverse transcriptase was omitted from control reactions. Endogenous reverse transcriptase activity did not give rise to specific DNA amplification using the conditions described above.

anchor PCR

See figure 3.1 for the position of the primers. 25 pg of RT primer was annealed to 10 µg of total cellular RNA and cDNA was synthesised as above. The cDNA was ethanol precipitated, resuspended in 20 µl of tailing buffer (100 mM K cacodylate pH 7.2, 2 mM CoCl₂, 0.2 mM dithiothreitol, 0.2 mM dATP), incubated with 12.5 units terminal transferase at 37° for 10 minutes, heated to 65° for 15 minutes and then diluted in 300 µl water. 10 µl of tailed cDNA was amplified for 40 cycles using primer 1 and RACE primer (GACTCGAGTCGACATCGATTTTTTTTTTTTTTTTTT, Frohman, et al., 1988). The anchor PCR product was gel purified,

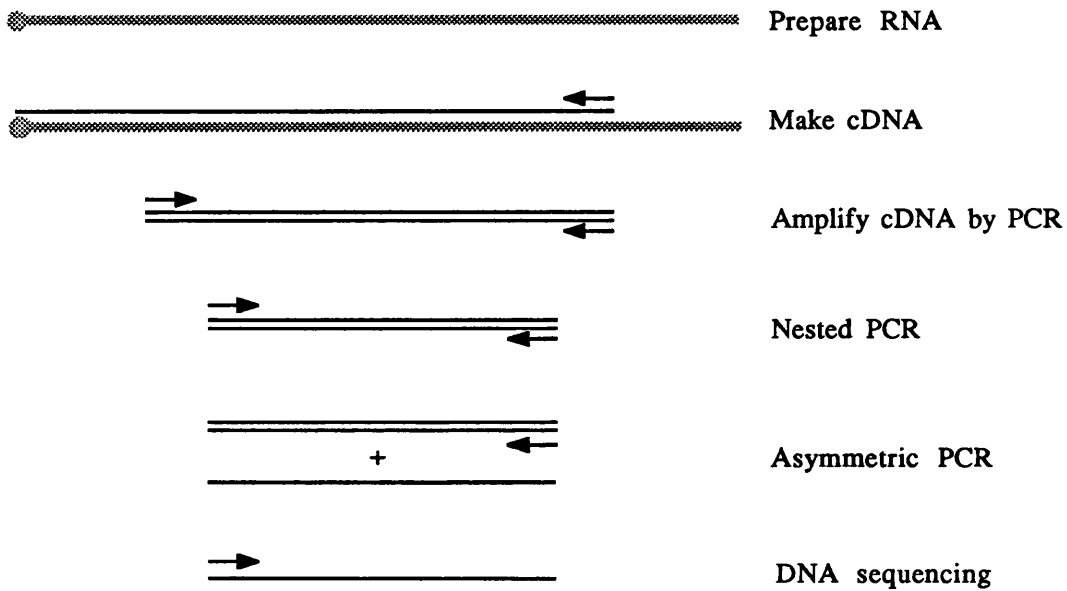


Figure 2.1 Preparation of single-stranded sequencing templates by PCR. The first two steps can be omitted when starting from genomic DNA.

digested with *Apal* and *Sall* and subcloned into Bluescript (Stratagene). 100 transformants were gridded out and colony blotted with primer 2 (Buluwela, et al., 1989). The 33 positive colonies were picked directly into PCR tubes and inserts were amplified by PCR with Bluescript forward and reverse sequencing primers (Stratagene) to identify plasmids with the longest inserts. Plasmid minipreps of the two longest clones were sequenced.

conventional PCR sequencing

DNA was amplified by PCR for 30 cycles. The product was extracted with phenol/chloroform, spun through a Sephadex G50 column, ethanol precipitated and suspended in 20 μ l of 10 mM Tris, 1 mM ethylenediamine tetra-acetic acid (EDTA) pH 8.0. 100 ng of sequencing primer was end-labelled with 32 P using T4 polynucleotide kinase, spun through a Sephadex G50 column, dried down and resuspended in 20 μ l water. 2 μ l of PCR product was mixed with 3 μ l of labelled primer in 10 μ l of Sequenase buffer, boiled for 5 minutes and placed on ice. Thereafter sequencing was carried out as for plasmid templates (omitting labelled dATP).

asymmetric PCR sequencing

See figure 2.1. DNA was initially amplified by PCR for 15 cycles. 1/20th of the product was then taken and reamplified with nested primers for 20 cycles. The product was extracted with phenol/chloroform and precipitated with 2M ammonium acetate/50% isopropanol to eliminate the primers. Half the product was reamplified with a single nested primer [asymmetric PCR, Gyllensten and Erlich (1988)] for 20 cycles to generate single-stranded DNA sequencing templates. The asymmetric PCR product was extracted with phenol/chloroform and ethanol precipitated. Thereafter sequencing was done with a Sequenase kit using 35 S α dATP.

inverse PCR

See figure 3.16. For inverse PCR (Triglia, et al., 1988) genomic DNA restriction digests were diluted to 3 μ g/ml and

self-ligated. No special steps are required for DNA amplification by inverse PCR. Although linearisation of the self-ligated minicircles with a restriction enzyme cutting between the 5'-ends of the primers increases the efficiency of the initial reactions it limits the choice of sequences which can be amplified. As pointed out by Triglia, et al. (1988) linearisation is unnecessary, and it was not done for amplification of *DBP2*.

PCR mapping

The derivation of the hybrid cell lines listed in figure 3.12 is: 1, Jones, et al., 1976; 2, Solomon, et al., 1976; 3, Swallow, et al., 1977; 4, Martin, et al., 1987; 5, Bai, et al., 1982; 6, Spurr, et al., 1986; 7, Solomon, et al., 1979; 8, Povey, et al., 1980; 9, Van Heyningen, et al., 1975. TRID62, PJT2/A1, KLT-8, PLT3, PLT-6B, PLT7 and PLT-8 are described by Xu, et al. (1988). P7A2.7 is described by Voss, et al. (1980). 1 ng of primer 7 (figure 3.1) was end-labelled with ³²P by T4 polynucleotide kinase and added to the PCR reaction mix before amplification of the intron between primers 6 and 7 (figure 3.8).

Sequence analysis

Intelligenetics programs were used for sequence data entry and manipulation. EMBL and Genbank DNA searches were performed using Intelligenetics FASTDB. University of Wisconsin Genetics Computer Group programs were used for sequence analysis. *S.cerevisiae* introns were identified using the programs "Find" and "Strings" to identify "TACTAAC" and "IVS" respectively. Fickett analysis was done using the program "Testcode" with a window of 300bp (Devereux, et al., 1984; Fickett, 1982). "Compare" and "Dotplot" were used for dotplots. Sequence alignment was done with "Lineup". PIR database searches were done on an Active Memory Technology Distributed Array Processor using the program PROSRCH (JF Collins and AFW Coulson, University of Edinburgh).

Antibodies and Cell Biology (Chapters 4 and 5)

Cell lines and Antibodies

C6 cells (rat embryo fibroblasts transformed by activated ras and a temperature sensitive p53 mutant [Michalovitz, et al., 1990]) were provided by Dr M Oren. 293 cells (adenovirus DNA transformed human embryonic kidney cells [Graham, et al., 1977]) were provided by Dr P Gallimore. HFF (human foreskin fibroblast), 3T3 (mouse fibroblast) and CV1 (african green monkey kidney epithelial [Jensen, et al., 1964]) cells were provided by ICRF cell production. Anti-nucleophosmin monoclonal antibody hybridoma supernatant (Ochs, et al., 1983) was provided by Dr PK Chan. Human anti-nuclear antibody (nucleolar pattern: ANA-N) was purchased from Sigma. Mouse monoclonal hybridomas PAb204 (Lane and Hoeffler, 1980), PAb240 (Gannon, et al., 1990), BG2 (anti- β -galactosidase) and CAT2 (anti-chloramphenicol acetyl-transferase) were produced in this laboratory. Serum free hybridoma culture medium (HB101) was purchased from Dupont.

MF7 is pooled serum from mice immunised with a gel-purified p68/ β -galactosidase fusion protein made in *E.coli* by expressing a human p68 cDNA cloned into the EcoRI site of pUR290 (B2000, table 2). Affinity-purified rabbit anti-p68 immunoglobulin was prepared by acid elution (0.1 M glycine pH 3) from a p68 column of serum from a rabbit (MF6) immunised with gel-purified p68/ β -galactosidase fusion protein.* The p68 column was made by incubating Hela cell extract with PAb204 and immobilising the immune complexes on a rabbit anti-mouse immunoglobulin column by coupling with dimethyl pimelidate*. The p68 immunoprecipitation protocol is given below (see chapter 6). The rabbit anti-mouse immunoglobulin column (coupling ratio: 8 mg rabbit anti-mouse immunoglobulin per ml of column) was prepared by reacting Dako rabbit anti-mouse immunoglobulins with Pharmacia cyanogen bromide activated sepharose 4B. To reduce contamination with antibodies against β -galactosidase and other bacterial proteins, the MF6 serum was preincubated overnight with a crude extract of *E.coli* containing pUR291 induced with isopropyl- β -D-thiogalactopyranoside*

(IPTG). PAb204 leakage from the column can be ruled out because a peptide corresponding to the PAb204/p68 epitope (figure 1.1) abolished immunofluorescent cell staining with PAb204 but had no effect on affinity purified MF6 staining. The β -galactosidase fusion proteins injected into the MF7 mice and the MF6 rabbit were prepared by Martin Ford and Gerald Gough.

Hybridoma fusion

A peptide with the sequence LVLDEADRMLDMGFEP was coupled to ovalbumin, albumin and tuberculin purified protein derivative (PPD) using glutaraldehyde (Harlow and Lane, 1988). The peptide was made by the ICRF Peptide Synthesis Laboratory. Balb/c mice were immunised four times with 250 μ g ovalbumin-peptide intraperitoneally. The first dose of ovalbumin-peptide was mixed with Freund's complete adjuvant; subsequent doses were given in Freund's incomplete adjuvant. Mice immune to the ovalbumin-peptide conjugate were vaccinated with Bacillus Calmette-Guerin (BCG) subcutaneously and given a final boost of 250 μ g PPD-peptide intravenously three days before spleens were harvested (Lachmann, et al., 1986). Spleen cells from an immune animal were fused with SP2 cells using polyethylene glycol and plated over 1000 wells. The cell fusion was done by Jenny Southgate. Hybridomas were selected by incubation in azaserine/hypoxanthine medium (Harlow and Lane, 1988). Approximately 5000 drug resistant colonies developed and were screened on albumin-peptide coated microtitre plates. Plates were coated with antigen by incubation overnight with 50 μ l per well of 100 μ g/ml peptide conjugate in 100 mM bicarbonate buffer pH 9.8. Plates were blocked by incubation for a few minutes in 20% dried milk in phosphate buffered saline. Primary antibody was detected with 1/1000 horseradish peroxidase-conjugated rabbit anti-mouse immunoglobulins (Dako) in 2% dried milk in phosphate buffered saline. Horseradish peroxidase was detected with tetramethylbenzidine (Harlow and Lane, 1988). Positives which blocked with peptide were tested on immunoblots of a p68/ β -galactosidase fusion protein and on immunoblots of human p68 immunoprecipitated from Hela cells by PAb204. The MaD1 antibody was

subtyped with Amersham isotyping strips and with a BioRad kit on albumin-peptide plates.

Cell staining

Cells were grown on cover slips or plastic dishes in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. All figures show cells fixed in 50% acetone in methanol for 5 minutes.* Fixed cells were air-dried, incubated in primary antibody, rinsed in phosphate buffered saline, incubated in conjugated secondary antibody diluted in 10% fetal bovine serum in phosphate buffered saline and rinsed in phosphate buffered saline. Horseradish peroxidase conjugates were developed with diaminobenzidine. For immunofluorescence, primary antibodies were visualised with commercial fluorescein and rhodamine conjugated anti-immunoglobulin sera, with the exception of figure 5.2c (biotinylated PAb204 plus streptavidin-fluorescein), figure 5.5f (directly conjugated CAT2-fluorescein) and figure 5.5e (biotinylated PAb204 plus streptavidin-texas red).

Microinjection

PAb204 was purified from serum-free hybridoma supernatant by chromatography on protein A sepharose in the presence of 3.3 M NaCl (Harlow and Lane, 1988). BG2 was purified from monoclonal ascites by chromatography on a rabbit anti-mouse immunoglobulin column (Harlow and Lane, 1988). Antibody at 10 mg/ml was injected in buffer containing 57 mM KCl, 7.5 mM NaCl, 2.75 mM MgCl₂, 5 mM Tris pH 7.0. Cells in hydroxyethylpiperazine ethanesulfonic acid buffered Dulbecco's Modified Eagle's Medium (ie, bicarbonate free medium) were maintained at 35° C on a heated microscope stage and injected under phase contrast using a Zeiss micromanipulator MR and an Eppendorf microinjector 5242.

Transfection

293 cells were transfected with pIG105 (table 2), which contains a full length human p68 cDNA under the control of the cytomegalovirus major immediate early promoter, using the

Nitrocellulose sheets (BA85, Schleicher and Schuell) were used for immunoblotting.

calcium phosphate method as described by Gorman, et al. (1990). PAb204 can not distinguish endogenous from transfected p68. pIG105 was therefore cotransfected with pCIS-CAT (which contains the chloramphenicol acetyl-transferase gene under the control of the cytomegalovirus major immediate early promoter) so that transfected cells could be identified by double labelling with CAT2 and PAb204. Virtually all cells staining with CAT2 gave intense nuclear staining with PAb204. Control transfections with pCIS-CAT alone showed that expression of chloramphenicol acetyl-transferase does not alter the pattern of endogenous p68 staining. pCIS5 and pCIS-CAT (Gorman, et al., 1990) were supplied by Dr C. Gorman.

Electrophoresis and Immunoblotting

Denaturing polyacrylamide gel electrophoresis and immunoblotting were performed as described by Harlow and Lane (1988) using BioRad minigel apparatus. Blots were air-dried after transfer and then soaked in 3% trichloroacetic acid/0.02% Ponceau S and rinsed in water to identify molecular weight markers. After staining, blots were blocked for a few minutes in 2% dried milk in phosphate buffered saline. Blots were washed in 1% nonidet P40 in phosphate buffered saline for 30 minutes at appropriate steps to remove free antibody. Antibodies were either used neat or diluted in blocking buffer plus 0.1% nonidet P40, as required. Primary antibodies were detected with Dako alkaline phosphatase conjugated anti-immunoglobulin sera and bromochloroindolyl phosphate nitroblue tetrazolium (Harlow and Lane, 1988).

Biochemistry (Chapter 6)

Preparation of extracts

Protein G sepharose was purchased from Pharmacia. Ribonuclease (RNase) free deoxyribonuclease I (DNase I) was purchased from Stratagene. Rotavirus genomic RNA was provided by Dr Malcolm McCrae. SV40 large T antigen was prepared by the method of Simanis and Lane (1985).

Tissue culture cells were sonicated in lysis buffer (500 mM NaCl, 50 mM Tris pH 8.0, 5 mM EDTA, 1% nonidet P40, 1 mM dithiothreitol, 1mM phenylmethylsulfonyl fluoride) and incubated for 30 minutes on ice. The cell extract was centrifuged at 100,000 x g for 30 minutes and the pellet was discarded. For figure 6.1 the cells were lysed in 150 mM NaCl. To eliminate endogenous RNA, 0.1 mg/ml ribonuclease A was added to the lysis buffer. Monoclonal antibody and protein G were added, either sequentially or as preformed complexes. After incubation of the mixture overnight at 4° the beads were washed in lysis buffer and ATPase buffer (150 mM NaCl, 5 mM KCl, 1.5 mM MgCl₂, 20 mM hydroxyethylpiperazine ethanesulfonic acid pH 7.0, 1 mM dithiothreitol, 1mM phenylmethylsulfonyl fluoride).

ATPase Assays

ATPase and GTPase activity is expressed as P_i released from $\gamma^{32}\text{P}$ labelled ATP or GTP. Assays were performed in 30 μl ATPase buffer containing approximately 10 μl of protein G sepharose beads bound to PAb204 or PAb240. Each assay contained approximately 50 ng p68 (judged by densitometry of Coomassie stained polyacrylamide gels). Unless otherwise stated the assays were incubated for 30 minutes at 37°. 300 μl of 7% activated charcoal in 50 mM HCl, 5 mM H₃PO₄ was then added and the mixture centrifuged at 10,000 x g for 10 minutes to remove unreacted ATP. The supernatant (containing free $^{32}\text{P}_i$) was mixed with Aquasol liquid scintillation cocktail and counted in a Packard scintillation counter.

pCp labelling

The anti-trimethylguanosine and Kung anti-Sm antibodies were supplied by Dr J Beggs. Immunoprecipitations were performed as described above, except that antibodies were prebound to protein G beads which were then washed extensively to remove contaminating ribonucleases. Extracts were made from Hela cells lysed in 150 mM NaCl buffer. After incubation of extract plus antibody for 150 minutes at 4° the beads were washed briefly in lysis buffer. The immuno-

precipitated material was then digested with proteinase K for 30 minutes at 37°, extracted with phenol/chloroform and ethanol precipitated. RNA was labelled with ³²P by incubation overnight at 4° with 10 µCi 5'-³²P cytidine 3',5'-diphosphate (pCp), 2 U T₄ RNA ligase and 0.5 U human placental ribonuclease inhibitor in 5 µl of 50 mM hydroxyethylpiperazine ethanesulfonic acid pH 7.5, 10 mM MgCl₂, 3 mM dithiothreitol, 10% dimethyl sulfoxide, 25 µM ATP, 0.1 mg/ml bovine serum albumin (England, et al., 1980). The product was digested with proteinase K for 30 minutes at 65°, extracted with phenol/chloroform and ethanol precipitated. The pellet was taken up in 4 µl of 5 M urea, 0.5 x Tris/borate/EDTA loading buffer and run on a denaturing 6% polyacrylamide gel (Sambrook, et al., 1989). The unfixed gel was fluorographed at -70°.

Yeast Genetics and RNA analysis (Chapter 7)

S.cerevisiae genotypes:

DJY84: *MATa/α ura3-52/ura3-52 his3-Δ1/his3-Δ200*

ADE2/ade2-101 LYS2/lys2-801 trp1-289/trp1-Δ1

leu2-3,112/LEU2

DJY98: *MATa/α DBP2/DBP2::URA3 ura3-52/ura3-52*

his3-Δ1/his3-Δ200 ADE2/ade2-101 LYS2/lys2-801

trp1-289/trp1-Δ1 leu2-3,112/LEU2

PY26: *Mata ura3-52 trp1Δ prb1-1122 prc1-407 pep4-3*

leu2-3,112 nucl::LEU2

yIG4: *DBP2::URA3 ura3-52 his3 ade2-101 LYS2 trp1 LEU2*

[TRP1 GAL1::dbp2]

yIG5: *DBP2::URA3 ura3-52 his3 ADE2 lys2-801 trp1 leu2-3,112*

[TRP1 GAL1::dbp2]

yIG6: *DBP2 ura3-52 his3 ADE2 lys2-801 trp1 leu2-3,112*

S.cerevisiae techniques

Deletion of the *S.pombe* and *S.cerevisiae* p68-like genes was accomplished by one step gene replacement (Rothstein, 1983). The *S.cerevisiae* *URA3* gene was cloned into the EcoRV

and BglII sites of *S.cerevisiae* *DBP2* (between amino acids 375 and 497). A BglI fragment (from nucleotide 963 to a BglI site 0.7 kb downstream of the BglII site) was gel purified and transformed into DJY84 cells by the lithium acetate technique (Ito, et al., 1983). Stable diploid *ura*⁺ colonies were picked and analysed by PCR (from primer 11, figure 3.18, to a *URA3* primer, GGTAGAGGGTGAACGTTACAG) and Southern blotting to identify transformants in which one copy of *DBP2* had been disrupted. Cells from the strain (DJY98) heterozygous for the disruption were sporulated and tetrads were dissected and allowed to germinate on YPD plates (Sherman, et al., 1982). To demonstrate rescue of the inviable spores a full length *DBP2* cDNA (from nucleotides 61 to 2776) under the control of the *GAL1* promoter on a *CEN6/ARSH4/TRP1* plasmid was transformed into diploid DJY98 cells. The vector used was pRS314 with the *GAL1/10* promoter cloned into the ApaI site and the *CYC1* terminator cloned into the SacII site (Sikorski and Hieter, 1989). The resulting plasmid is called pIG70 (table 2). The strain transformed with pIG70 was sporulated and tetrads were dissected as above. Spores giving rise to viable strains were analysed for uracil prototrophy (indicating the presence of the disrupted chromosomal copy of the *DBP2* gene) and tryptophan prototrophy (indicating the presence of the plasmid copy of the *DBP2* cDNA) as well as PCR to confirm the integrity of the recombinant DNA molecules in the recovered strains. PCR reactions were performed with four different pairs of primers: PCR from a *GAL1* promoter primer (ACCTCTATACTTTAACGTCAA GGAG) to a *DBP2* primer near the 5'-end of the coding sequence; PCR from a *URA3* primer to a *DBP2* primer external to the linear fragment used for the disruption; PCR between two primers within the *DBP2* intron; and PCR between two primers on either side of the *DBP2* intron. These reactions check for the presence of the expression plasmid (pIG70), the disruption, the genomic copy of *DBP2* and the cDNA copy of *DBP2* respectively.

For gap repair 0.7 kb of genomic DNA upstream and 1.4 kb downstream of the *DBP2* coding sequence was cloned into pFL39 (a centromeric plasmid carrying the *trp1* gene). Specifically, the gap repair plasmid (pIG74) contains the pIG71 SacII/XbaI

fragment which overlaps the initiator methionine and the pIG60 λ gt11 EcoRI fragment. Only 70 bp of the *DBP2* coding sequence remained after the plasmid was cut with BamHI and NsiI. Cut vector was gel purified away from the *DBP2* insert and transformed into DJY84 cells by the lithium acetate technique. Transformants were plated on trp⁻ plates. Whole yeast from trp⁺ colonies were checked by PCR from vector primers to gene specific primers to confirm that the plasmid had been repaired correctly. The repaired plasmid (pIG75) was then rescued from spheroplasts using a standard bacterial plasmid preparation protocol (alkaline lysis followed by binding to a Quiagen column, [Diagen Inc]), transformed into bacteria and checked by restriction analysis and partial DNA sequencing.

S.pombe techniques

The *S.pombe dbp2* gene (nucleotides 187-2648) with the *S.pombe ura4* gene inserted between the MluI and AflII sites (from amino acid 205 to a position within the intron) was cloned into the EcoRV site of Bluescript (Stratagene). ~10 μ g of insert DNA (from the BamHI to the Sall sites in Bluescript) was gel purified and transformed by the method of (Beach and Nurse, 1981) into a *leu1-32/leu1-32 ura4-D18/ura4-D18 ade6-M210/ade6-M216 h⁺/h⁻* strain (Moreno, et al., 1990). Following seven days of growth on minimal medium plus leucine, stable diploid ura⁺ transformant colonies were picked and analysed by PCR (from primer 10, figure 3.14, to a *ura4* primer, CCACAAATT CATGTTTGGACT) to identify transformants in which the *dbp2* gene had been partially deleted as intended. Diploids containing the deletion were sporulated on malt extract medium and tetrads were dissected using a micromanipulator (Moreno, et al., 1990).

Northern Blotting

RNA was prepared from *S.cerevisiae* by hot phenol extraction (Schmitt, et al., 1990), electrophoresed through morpholinopropanesulfonic acid buffered, 6% formaldehyde, 1.25% agarose gels (Ausubel, et al., 1987), transferred to Hybond N (Amersham) by capillary transfer and cross-linked by

illumination with ultraviolet light. Lanes containing RNA size markers (GIBCO BRL) were cut from the blot and visualised by methylene blue staining (Sambrook, et al., 1989). Prehybridisation and hybridisation were done at 37° in 50% formamide, 10 x Denhardt's solution, 1% sodium dodecylsulfate, 5 x SSC and 10 mg/ml sonicated salmon sperm DNA and blots were washed in 1% sodium dodecylsulfate and 0.1 x SSC at 65° (Sambrook, et al., 1989). 10 x Denhardt's solution is 0.2 % bovine serum albumin, 0.2% Ficoll, 0.2% polyvinylpyrrolidone. 20 x SSC is 3 M NaCl, 0.3 M trisodium citrate. Strand specific probes were made by asymmetric PCR (Gyllensten and Erlich, 1988). A typical labelling reaction used a single primer for 20 cycles in standard conditions (see above, chapter 3) except that the normal nucleotide mix was replaced with 50 µCi ³²P αdCTP, 50 µM dGTP, 50 µM dATP, 50 µM TTP and the total volume was limited to 20 µl. The input DNA was a double-stranded PCR product which had been precipitated with 2M ammonium acetate/50% isopropanol to eliminate the primers. The specific activity of asymmetric single-stranded probes is difficult to calculate because a large amount of double-stranded DNA is present at the end of the reaction. It is unlikely to be less than the specific activity of double-stranded PCR probes, which is at least 10⁹ cpm/µg using the conditions described above. A satisfactory means to strip asymmetric probes was not found: incubation for several hours in 96% formamide at 65° or in boiling water had little effect. Where actin loading controls are shown the blots were probed sequentially without stripping. Typical exposure times for blots of oligo(dT)-selected RNA were 1-2 days for *DBP2* and 1-2 hours for actin.

Introduction

Four cloning projects are described in this chapter. The first is the human p68 project. This represents an extension of work started by Martin Ford. The prototypic p68 cDNA clone he isolated was incomplete. DNA encoding twenty extra amino acids at the 5'-end of the human p68 sequence was cloned by anchor PCR. A new approach to human chromosome mapping is also presented. Using this technique it was possible to map p68 to chromosome 17q23. The next two cloning projects described were undertaken in an attempt to identify p68-like genes in genetically manipulable organisms. Both were successful, and the sequences of probable p68 homologues identified in *S.pombe* and *S.cerevisiae* are presented. Finally, the search for p68-like genes was extended to *E.coli*. The result was the identification of a new DEAD box gene in *E.coli*.

The library clones for the yeast and bacterial genes were obtained by low stringency screening of phage libraries. Since the library screening was done by Jane McPheat that part of the work is not described in detail.

Human p68

Completion of the human p68 cDNA sequence

The p68 sequence published by Ford, et al. (1988) was completed (figure 3.1) by anchor PCR (aPCR, figure 3.2 [Frohman, et al., 1988]). The anchor PCR products (figure 3.3) were analysed by Southern blotting (figure 3.4). Anchor PCR using the RT primer gives a smear of DNA with and without reverse transcriptase because residual RT primer which is not incorporated into cDNA is tailed and forms a primer dimer with the RACE primer (figure 3.3, lanes 4 and 8). The semi-degenerate nature of the reaction permits amplification of non-specific DNA (figure 3.3, lanes 1-4). Primers based on the anchor PCR plasmid sequence (primers 4 and 5 in figure 3.2) were used first to check that the expected pattern of bands was amplified by PCR (figure 3.5) and then for direct RT-PCR sequencing to rule out cloning

```

          primer5>                                primer4>
TCA TTC ATT TCT ACC GGT CTC TAG TAG TGC AGC TTC GGC TGG TGT CAT CGG TGT 54
CCT TCC TCC GCT GCC GCC CCC GCA AGG CTT CGC CGT CAT CGA GGC CAT TTC CAG 108
CGA CTT GTC GCA CGC TTT TCT ATA TAC TTC GTT CCC CGC CAA CCG CAA CCA TTG 162

ACG CCA TGT CGG GTT ATT CGA GTG ACC GAG ACC GCG GCC GGG ACC GAG GGT TTG 216
  M  S  G  Y  S  S  D  R  D  R  G  R  D  R  G  F  G  17

          primer2>
GTG CAC CTC GAT TTG GAG GAA GTA GGG CAG GGC CCT TAT CTG GAA AGA AGT TTG 270
  A  P  R  F  G  G  S  R  A  G  P  L  S  G  K  K  F  G  35

<primer3>                                <primer1>
GAA ACC CTG GGG AGA AAT TAG TTA AAA AGA AGT GGA ATC TTG ATG AGC TGC CTA 324
  N  P  G  E  K  L  V  K  K  K  W  N  L  D  E  L  P  K  53

          <RTprimer>
AAT TTG AGA AGA ATT TTT ATC AAG AGC ACC CTG ATT TGG CTA GGC GCA CAG CAC 378
  F  E  K  N  F  Y  Q  E  H  P  D  L  A  R  R  T  A  Q  71

AAG AGG TGG AAA CAT ACA GAA GAA GCA AGG AAA TTA CAG TTA GAG GTC ACA ACT 432
  E  V  E  T  Y  R  R  S  K  E  I  T  V  R  G  H  N  C  89

GCC CGA AGC CAG TTC TAA ATT TTT ATG AAG CCA ATT TCC CTG CAA ATG TCA TGG 486
  P  K  P  V  L  N  F  Y  E  A  N  F  P  A  N  V  M  D  107

ATG TTA TTG CAA GAC AGA ATT TCA CTG AAC CCA CTG CTA TTC AAG CTC AGG GAT 540
  V  I  A  R  Q  N  F  T  E  P  T  A  I  Q  A  Q  G  W  125

GGC CAG TTG CTC TAA GTG GAT TGG ATA TGG TTG GAG TGG CAC AGA CTG GAT CTG 594
  P  V  A  L  S  G  L  D  M  V  G  V  A  Q  T  G  S  G  143

GGA AAA CAT TGT CTT ATT TGC TTC CTG CCA TTG TCC ACA TCA ATC ATC AGC CAT 648
  K  T  L  S  Y  L  L  P  A  I  V  H  I  N  H  Q  P  F  161

TCC TAG AGA GAG GCG ATG GGC CTA TTT GTT TGG TGC TGG CAC CAA CTC GGG AAC 702
  L  E  R  G  D  G  P  I  C  L  V  L  A  P  T  R  E  L  179

TGG CCC AAC AGG TGC AGC AAG TAG CTG CTG AAT ATT GTA GAG CAT GTC GCT TGA 756
  A  Q  Q  V  Q  Q  V  A  A  E  Y  C  R  A  C  R  L  K  197

AccI
AGT CTA CTT GTA TCT ACG GTG GTG CTC CTA AGG GAC CAC AAA TAC GTG ATT TGG 810
  S  T  C  I  Y  G  G  A  P  K  G  P  Q  I  R  D  L  E  215

AGA GAG GTG TGG AAA TCT GTA TTG CAA CAC CTG GAA GAC TGA TTG ACT TTT TAG 864
  R  G  V  E  I  C  I  A  T  P  G  R  L  I  D  F  L  E  233

AGT GTG GAA AAA CCA ATC TGA GAA GAA CAA CCT ACC TTG TCC TTG ATG AAG CAG 918
  C  G  K  T  N  L  R  R  T  T  Y  L  V  L  D  E  A  D  251

ATA GAA TGC TTG ATA TGG GCT TTG AAC CCC AAA TAA GGA AGA TTG TGG ATC AAA 972
  R  M  L  D  M  G  F  E  P  Q  I  R  K  I  V  D  Q  I  269

          primer6>
TAA GAC CTG ATA GGC AAA CTC TAA TGT GGA GTG CGA CTT GGC CAA AAG AAG TAA 1026
  R  P  D  R  Q  T  L  M  W  S  A  T  W  P  K  E  V  R  287

GAC AGC TTG CTG AAG ATT TCC TGA AAG ACT ATA TTC ATA TAA ACA TTG GTG CAC 1080
  Q  L  A  E  D  F  L  K  D  Y  I  H  I  N  I  G  A  L  305

TTG AAC TGA GTG CAA ACC ACA ACA TTC TTC AGA TTG TGG ATG TGT GTC ATG ACG 1134
  E  L  S  A  N  H  N  I  L  Q  I  V  D  V  C  H  D  V  323

          <>intron
TAG AAA AGG ATG AAA AAC TTA TTC GTC TAA TGG AAG AGA TCA TGA GTG AGA AGG 1188
  E  K  D  E  K  L  I  R  L  M  E  E  I  M  S  E  K  E  341

          <primer7>
AGA ATA AAA CCA TTG TTT TTG TGG AAA CCA AAA GAA GAT GTG ATG AGC TTA CCA 1242
  N  K  T  I  V  F  V  E  T  K  R  R  C  D  E  L  T  R  359

GAA AAA TGA GGA GAG ATG GGT GGC CTG CCA TGG GTA TCC ATG GTG ACA AGA GTC 1296
  K  M  R  R  D  G  W  P  A  M  G  I  H  G  D  K  S  Q  377

          primer8>
AAC AAG AGC GTG ACT GGG TTC TAA ATG GTA AAC ATT TCA AAT GAA GTA TTT TTC 1350
  Q  E  R  D  W  V  L  N  E  < intron 385

```

Figure 3.1 Human p68 sequence. Amino acids are placed under the first nucleotides of their cognate codons. Accession number X15729

CCC CTT ACT TAA CCT AGC TAG AAT TCT GCT CAG ATA ATT GAT CAT GTA TAT GCC 1404
 TTC CTT TGT AGA ATT CAA ACA TGG AAA AGC TCC TAT TCT GAT TGC TAC AGA TGT 1458
 intron > F K H G K A P I L I A T D V 400
 GGC CTC CAG AGG GCT AGG TTA GTA CAA ACT CGC ATT CAT GGC TTG GTT TCC CAG 1512
 A S R G L D< intron 406
 AAG ATC TCC ATT TAA CTT TTT TAA AGA AAG TTT ATT GCT TTC TTT AAC CTG CAT 1566
 TTT TTC TAA GTT TTT TTT CAC ATA AAG GTG CTG TCT TTG TGG CAA GGC TAG GCA 1620
 TGA CAA TCG GAG GAC TCG AGG GGG ATG GAG GAC TAG TGA TCG GCT GGC TGC TTC 1674
 CAG TCG ATT AGA GAG GTG AAA AAG CTG AAC GTG TGC CAG TAA TCT TCA AAA GGC 1728
 AGA ACA TAT CAC CTC TGC CCC GTA AAC TGT TCT CTC CGA GGG AAA AAA TGG AAG 1782
 TTA TCC TCA CAG TTC ACT GCC GTG GTA TTT CTT CTG TCC CAT GCT TTC GAT GAC 1836
 TGC CAT GGT ACA GCC TTG TTT CAA ACT GTT CAC TGT GAT CTG TGG GTC TTT GAG 1890
 TTT CAG TGA GTT TGC TGA AAT GTC GAA GAA GTA GTT CCA AAC TTC AAT GTT CAA 1944
 TGA AAT TTT TGT TCA AGT TTG AAA TGG AGA GAG CAG CTT TAA AAG GTA CAA AGC 1998
 CTT TTA CAA ATT GGT GAG TAC TGG CAC ATG AGA TCT AGA GCA GGA GCA ACT TCT 2052
 CAC ACA TAG TAA GTG GGA AAA GAA AGT GCT TTG AAA GTT CCT CCC TCA CCT ACA 2106
 CAG TAG TCG TCA TGT CGA GAC CTG CCA GAG AGA GGC ACA TTC TCA AGT GAA TCC 2160
 TGG CTT CTT GGA AGC GCT TGC CTA GAC GAG ACA CAG TGC ATA AAA ACA ACT TTT 2214
 GGG GGA CAG GTA TGT TTT CTT GCA GCT GCG GTT GTA AGG TCT TGG CAA GGC AAG 2268
 CAG TGT GGC CAG AAT TTT GAA CTT CTG ATG AAT GTG TAA TGC AAA GGA CCT TGT 2322
 ACA TTT TTT TGT TTC AAG GTC CTC AAA ATG AGC ACA TGA AGA GGT GGC TGT GAA 2376
 ACT TTA AGT GGC CCT ACT CCG CAG AAG CAT TCA GAT GTC ACT TGA TGA TCT GTA 2430
 AGG GAA CTT GCT GAT TTG GGA ATG TGC TTA GGG AAC ACA CAT TCC TTT TGA CAG 2484
 GGT CTG TCA CTG GGT GGG TGA TGG ATT ATA CAG ATG ACA TGT GGA CTT TTT TTT 2538
 CTT TTT TCA ACC TCA ATG GTA TTC CTA CAG GAA ATG GAT AAC CAT TTT AAC TGT 2592
 ATT TTT TGC AGC CCG TAC CTT CTT GGG AAT ACA ATT GTC TAA CTT TTT ATT TTT 2646
 GGT CTG GCT GTT GTG GTG TGC AAA ACT CCG TAC ATT GCT ATT TTG CCA CAC TGC 2700
 AAC ACC TTA CAG ATG TGG AAG ATG TGA AAT TTG TCA TCA ATT ATG ACT ACC CTA 2754
 intron > V E D V K F V I N Y D Y P N 420
 <primer9
 ACT CCT CAG AGG ATT ATA TTC ATC GAA TTG GAA GAA CTG CTC GCA GTA CCA AAA 2808
 S S E D Y I H R I G R T A R S T K T 438
 CAG GCA CAG CAT ACA CTT TCT TTA CAC CTA ATA ACA TAA AGC AAG TGA GCG ACC 2862
 G T A Y T F F T P N N I K Q V S D L 456
 TTA TCT CTG TGC TTC GTG AAG CTA ATC AAG CAA TTA ATC CCA AGT TGC TTC AGT 2916
 I S V L R E A N Q A I N P K L L Q L 474
 TGG TCG AAG ACA GAG GTT CAG GTC GTT CCA GGG GTA GAG GAG GCA TGA AGG ATG 2970
 V E D R G S G R S R G R G G M K D D 492
 ACC GTC GGG ACA GAT ACT CTG CGG GCA AAA GGG GTG GAT TTA ATA CCT TTA GAG 3024
 R R D R Y S A G K R G G F N T F R D 510
 ACA GGG AAA ATT ATG ACA GAG GTT ACT CTA GCC TGC TTA AAA GAG ATT TTG GGG 3078
 R E N Y D R G Y S S L L K R D F G A 528
 CAA AAA CTC AGA ATG GTG TTT ACA GTG CTG CAA ATT ACA CCA ATG GGA GCT TTG 3132
 K T Q N G V Y S A A N Y T N G S F G 546
 AccI
 GAA GTA ATT TTG TGT CTG CTG GTA TAC AGA CCA GTT TTA GGA CTG GTA ATC CAA 3186
 S N F V S A G I Q T S F R T G N P T 564
 CAG GGA CTT ACC AGA ATG GTT ATG ATA GCA CTC AGC AAT ACG GAA GTA ATG TTC 3240
 G T Y Q N G Y D S T Q Q Y G S N V P 582
 CAA ATA TGC ACA ATG GTA TGA ACC AAC AGG CAT ATG CAT ATC CTG CTA CTG CAG 3294
 N M H N G M N Q Q A Y A Y P A T A A 600
 CTG CAC CTA TGA TTG GTT ATC CAA TGC CAA CAG GAT ATT CCC AAT AAG ACT TTA 3348
 A P M I G Y P M P T G Y S Q . 614
 GAA GTA TAT GTA AAT GTC TGT TTT TCA TAA TTG CTC TTT ATA TTG TGT GTT ATC 3402
 TGA CAA GAT AGT TAT TTA AGA AAC ATG GGA ATT GCA GAA ATG ACT GCA GTG CAG 3456
 CAG TAA TTA TGG TGC ACT TTT TCG CTA TTT AAG TTG GAT ATT TCT CTA CAT TCC 3510
 TGA AAC AAT TTT TAG GTT TTT TTT GTA CTA GAA AAT GCA GGC AGT GTT TTC ACA 3564
 AAA GTA AAT GTA CAG TGA TTT GAA ATA CAA TAA ATG AAG GCA ATG CAT GGC CTT 3618
 CCA ATA AAA AAT ATT TGA AGA CTG AAA AAA AAA 3650

Figure 3.1 Human p68 sequence (continued)

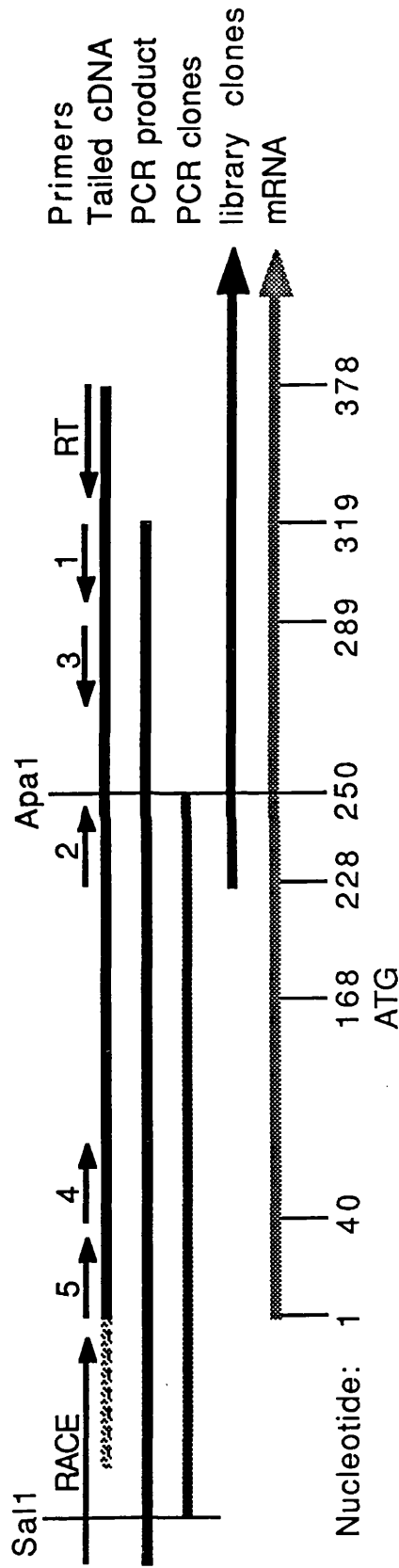


Figure 3.2 aPCR cloning strategy. cDNA synthesised from the RT primer was tailed with dATP by terminal transferase (), amplified by PCR between primer 1 and the RACE primer and subcloned using Sal1 and Apa1. Several clones were sequenced from plasmid to establish a consensus 5' end. The sequence was confirmed by direct sequencing with primers 1 and 4 of cDNA amplified with primers RT and 5. Nucleotides are numbered from the presumptive mRNA cap defined by primer extension (figure 3.6). Library clones obtained by M.Ford started at nucleotide 228.

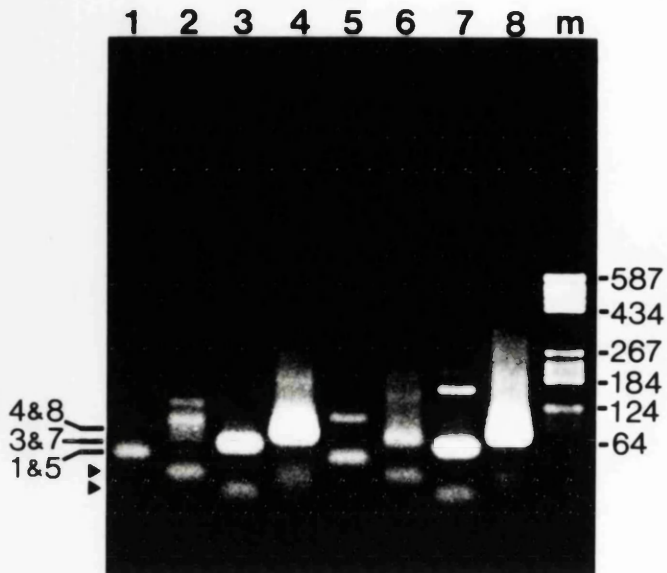


Figure 3.3 aPCR products. Lanes 5-8, tailed cDNA template; lanes 1-4, negative controls (no reverse transcriptase). Primers 1 & 2 (lanes 1 & 5); primers 1 & RACE (lanes 2 & 6), primers RT & 2 (lanes 3 & 7); primers RT & RACE (lanes 4 & 8); m, markers in bp. Solid triangles show the position of the primers. Bars at the left show the position of primer dimers in the lanes specified.



Figure 3.4 Southern blot of aPCR products probed with primer 3. Lanes and markers as in figure 3.3.

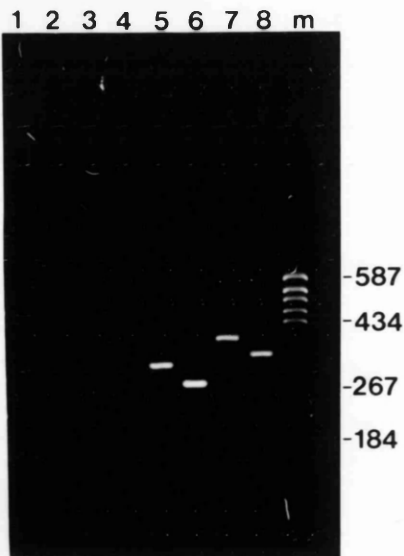


Figure 3.5 PCR amplification of cDNA using primers based on the aPCR sequence. Lanes 5-8, cDNA template; lanes 1-4, negative controls (no reverse transcriptase). Primers 1 & 5 (lanes 1 & 5); primers 1 & 4 (lanes 2 & 6); primers RT & 5 (lanes 3 & 7); primers RT & 4 (lanes 4 & 8); m, markers in bp.



Figure 3.6 Primer extension using the RT primer (A). M13 sequencing ladder (m). Unincorporated primer is seen at 30 nt.

artefacts. The start site identified agrees with that expected from primer extension analysis (figure 3.6) and adds another twenty amino acids to the deduced sequence.

Chromosome mapping by PCR

The ICRF human genetic resources laboratory was unable to map the human p68 gene by conventional Southern blotting of somatic cell hybrids (figure 3.7). However, it proved easy to map the gene by PCR. Primers around putative splice sites (figure 3.10) amplified DNA fragments which showed species specific length differences (figures 3.8 and 3.9), presumably due to intron length variation, which allowed unequivocal identification of the human gene in mouse-human hybrids. The specificity of the PCR reactions was confirmed by DNA sequencing (the intron sequences are shown in figures 3.1 and 3.11). In the initial hybrid screen (primers 6 and 7, figure 3.8) the PCR product was end-labelled with ^{32}P , run on a denaturing polyacrylamide gel and visualised by autoradiography. Lane a shows the PCR product with mouse DNA alone, lane b with human DNA alone. There is a band in the mouse track at 234 nucleotides, which is the size expected for an intronless pseudogene. In both mouse and human DNA there is a band at around 600 nucleotides, but the human p68 PCR product is ~15 base pairs longer. The mouse PCR products act as positive controls. Lanes c-o are from mouse-human somatic cell hybrids. Even on this overexposed film the human (upper) band is clearly absent in lanes c, i and m. Analysis of the results from lanes c-n (figure 3.12) suggests that p68 is on chromosome 17. Lane o is from hybrid PCTBA1.8, whose human DNA is solely derived from chromosome 17, and confirms that p68 is on chromosome 17. Primers 8 and 9 (figure 3.9) gave identical mapping results, including the presence of a pseudogene in the mouse DNA (which gave a 145 bp band). All of the hybrids were tested with both pairs of primers. Lanes e-j are from cell lines containing only fragments of human chromosome 17: lanes e-g are from somatic cell hybrids with translocations of chromosome 17, lanes h-j are from cell lines developed using chromosome mediated gene transfer. The presence of the human (lower) band in lanes e, f, g

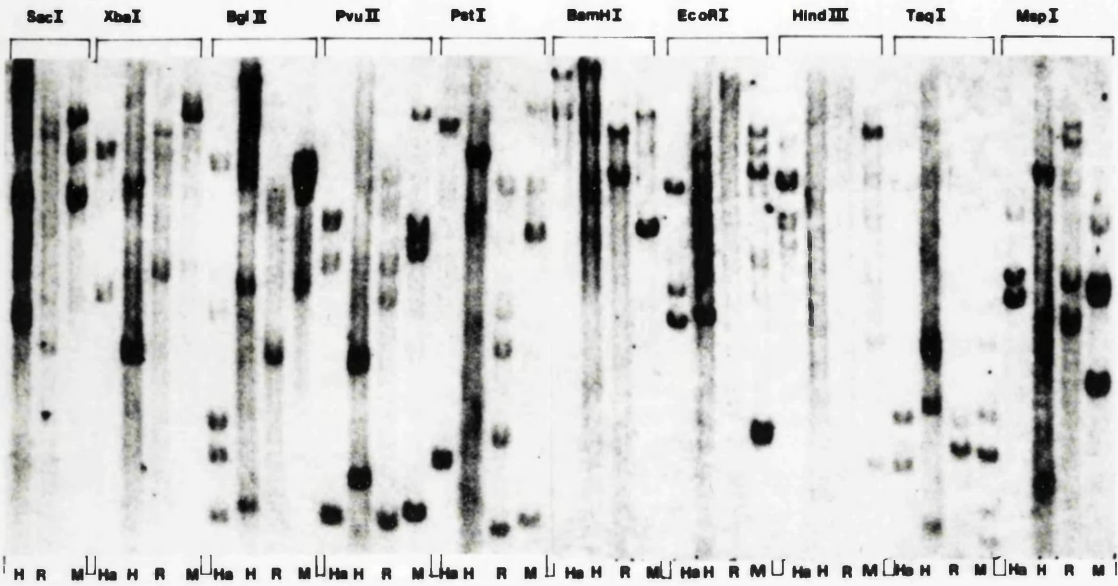


Figure 3.7 Southern blot of genomic DNA probed with an *AccI* fragment of the human p68 cDNA (see figure 3.1). Human (H), mouse (M), rat (R) and hamster (Ha) DNAs were digested with the restriction enzymes indicated. Photograph supplied by Alan Gough.

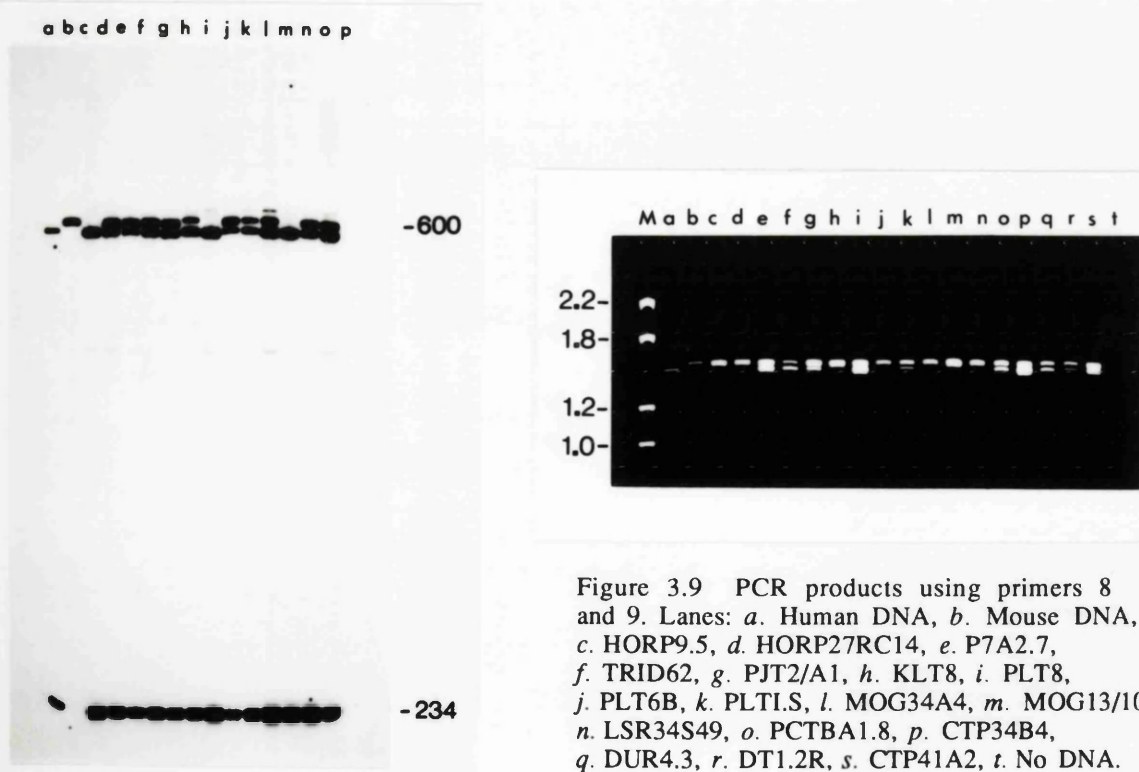


Figure 3.8 PCR products using primers 6 and 7. Lanes: *a*. Mouse DNA, *b*. Human DNA, *c*. MOG13/10, *d*. P7A2.7, *e*. TRID62, *f*. PJT2/A1, *g*. CTP41A2, *h*. PLT1.S, *i*. MOG34A4, *j*. CTP34B4, *k*. DUR4.3, *l*. DT1.2R, *m*. LSR34S49, *n*. DT1.2.4, *o*. PCTBA1.8, *p*. No DNA. Sizes of the DNA fragments (in nucleotides) were measured on an M13 sequencing ladder.

Figure 3.9 PCR products using primers 8 and 9. Lanes: *a*. Human DNA, *b*. Mouse DNA, *c*. HORP9.5, *d*. HORP27RC14, *e*. P7A2.7, *f*. TRID62, *g*. PJT2/A1, *h*. KLT8, *i*. PLT8, *j*. PLT6B, *k*. PLT1.S, *l*. MOG34A4, *m*. MOG13/10, *n*. LSR34S49, *o*. PCTBA1.8, *p*. CTP34B4, *q*. DUR4.3, *r*. DT1.2R, *s*. CTP41A2, *t*. No DNA. Markers are in kb.

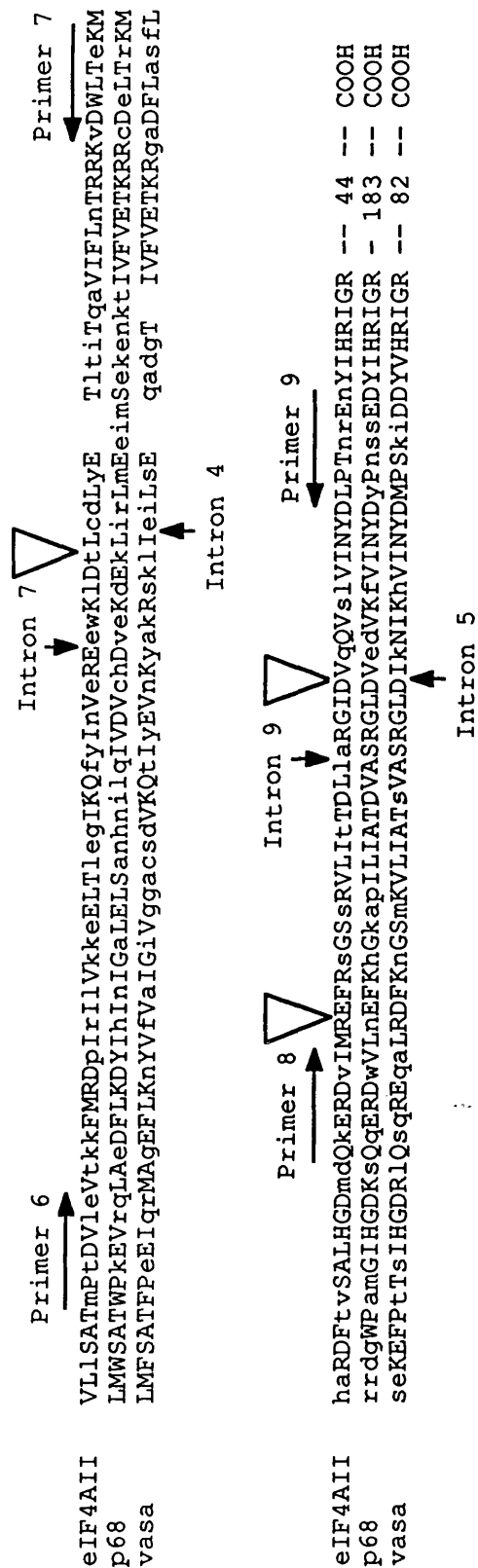


Figure 3.10 Vertical arrows indicate the position of introns in eIF4AII and vasa. Open triangles show the position of introns identified in p68 by PCR. Closely related amino-acids (A&G; S&T; Q&N; D&E; R&K; W,Y&F; V,L,I&M) shared by two sequences are shown in capitals.

5' splice
 site 10 20 30 40 50 60 70
 AAA^GTAAGT TTTATTAAGT CTGTTATATT TGCTTCCTAA CAACTTTGCT GTAAAATTGA GGGATCATTTG
 80 90 100 110 120 130 140
 TTTGGTGAGT TGTTTTAGGT TATTTTCAGTT GGTGTGATTT CATTAGTTA GCCTACTAAT CCGAAAAATT
 150 160 170 180 190 200 210
 TCTTGAATCC TTCAAATAAT GGCTGTCACC ATTTATAGCT TTCCTATAGA AGGAATTTCAT GTGTCCCCCTG
 220 230 240 250 260 270 280
 GTTGACTTAA GGACCAAGGG TCGAACTGCT CGATAAGTGG ATTAGCAGGC GTTTCCTCTT TGACTTCCAG
 290 300 310 320 330 340
 CCATGTAAAT TGAACCTAAT GTTTTGCTGA CCATAAATGT GTGGCCCTAG CAATGGTCTT TTAAA ~35 bp to
 3' splice site

Figure 3.11 Incomplete sequence for the intron in p68 between primers 6 and 7.

and i localises p68 to the distal part of the long arm of chromosome 17 in the region 17q23-q25. Negative results with PLT3 suggest that p68 lies at 17q23 (figure 3.13).

Discussion

Anchor PCR is now established as the method of choice for cloning the ends of cDNA molecules. The technique was relatively new when it was used to complete the p68 sequence. It was therefore reassuring to find that the sequence obtained is in complete agreement with that of a conventional cDNA subsequently cloned by Hloch, et al. (1990).

Chromosome mapping is normally done by Southern analysis of DNA from hybrid cell lines of known cytogenetic composition. This technique depends on the ready identification of species specific restriction fragments of the gene of interest. This is frequently not possible if cDNA probes are used because these probes detect pseudogenes and may fail to distinguish between conserved genes in different species. Intron probes overcome all of these problems but classically they can only be obtained by screening genomic libraries. The PCR mapping technique described here not only avoids recourse to libraries but also involves a great deal less work than Southern blotting. Although eIF4AII and vasa provided clues to the location of introns in p68, this information is probably unnecessary. The modal length of internal introns is 80-99 bp and the mean length of internal exons is 137 bp (Hawkins, 1988). Therefore, any pair of PCR primers separated by several hundred base pairs should amplify an intron and still remain within the size range which can easily be amplified by PCR, provided the primers lie within the coding sequence. Although 19% of internal introns exceed 1600 bp these introns are unlikely to be a problem if several pairs of primers are used.

Loss of heterozygosity has been found with a 17q23-25.3 probe in ovarian cancer (Eccles, et al., 1990; Russell, et al., 1990) and breast cancer (Coles, et al., 1990). However, more detailed mapping suggests that, at least for early-onset familial breast cancer, the relevant gene maps to 17q21 (Hall, et al., 1990). At

Hybrid	Human Chromosome																		p68	Reference					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18			19	20	21	22	X
CIP34B4	+	+	+	-	+	+	+	-	-	-	+	-	+	+	-	+	+	+	-	-	-	-	+	+	1
CIP41A2	-	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	+	+	1
DUR4.3	-	-	+	-	+	-	-	-	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	2
DT1.2R	-	-	+	-	-	-	-	-	+	+	-	-	-	-	-	-	+	+	+	+	+	+	+	+	3
DT1.2.4	-	-	+	-	-	-	-	-	+	+	-	-	-	-	-	-	+	+	+	+	+	+	+	+	3
PLTI.S	+	-	+	-	-	+	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	4
PCTBAL.8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	5
LSR34S49	+	+	+	-	+	+	-	-	+	-	-	+	+	+	+	+	-	-	+	+	+	+	+	+	6
MOG34A4	+	-	+	+	+	+	+	-	+	+	+	+	+	+	+	+	-	-	+	+	+	+	+	+	7
MOG13/10	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8
HORP9.5	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	9
HORP27RC14	-	-	-	+	-	-	+	-	-	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	9

Figure 3.12 Assignment of p68 to chromosome 17. +/- indicates whether a hybrid contains the listed human chromosome (blank = unassigned). Under p68, +/- indicates whether a hybrid yielded the human p68 PCR product. The final column lists references describing the hybrids (see Methods section for details).

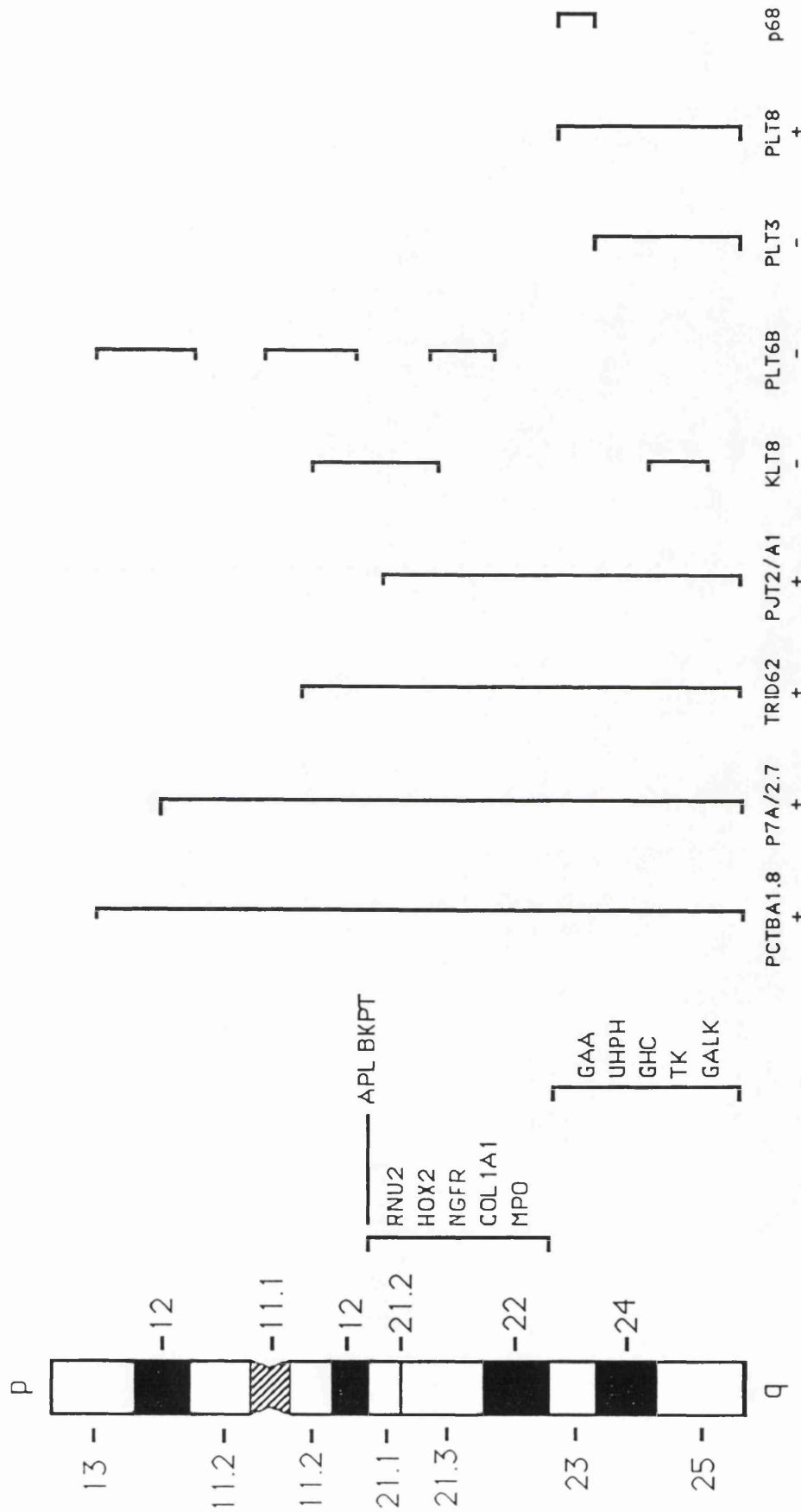


Figure 3.13 This shows the localization of p68 to 17q23. +/- indicates whether a cell line yielded the human p68 PCR product. The gene order is taken from Xu et al (1988): RNU2 (U2 snRNA), HOX 2 (Homeobox region 2), NGFR (nerve growth factor receptor), COL1A1 (collagen a1[1]), MPO (myeloperoxidase), GAA (acid α glucosidase), UMPH (uridine 5' mono-phosphate phosphohydrolase), GHC (growth hormone complex), TK (thymidine kinase) and GALK (galactokinase) and the APL (acute promyelocytic leukaemia) breakpoint.

present no diseases are known to be linked to the p68 locus on 17q23.

Identification of yeast genes related to p68

S.pombe dbp2

A λ gt11 library containing wild type *S.pombe* cDNA (from Dr V.Simanis) was screened with a human p68 cDNA probe (an AccI fragment, fig 3.1) using the low stringency hybridisation conditions of Miller, et al. (1988) according to the method of Huynh, et al. (1985). All four positive clones stop at an internal EcoRI site (nucleotide 952): three are identical and the fourth encodes the same open reading frame interrupted by a 700 bp intron. The library was constructed with EcoRI linkers; internal EcoRI sites should have been protected with EcoRI methylase, but the fact that all four clones stop at an internal EcoRI indicates that methylation of internal sites was ineffective. Cosmids containing the complete *dbp2* gene (figure 3.14) were obtained by screening a pcos8 library (from Dr A.Craig) with the 1.2 kb EcoRI insert from one of the normally spliced λ gt11 clones. Intron 1 was identified by direct asymmetric nested RT-PCR sequencing (figures 2.1 and 3.15). Nucleotide 136 was identified as the predicted mRNA start site by primer extension analysis.

S.cerevisiae DBP2

A λ gt11 library (Clontech, YL1001) containing *S.cerevisiae* X2180 genomic DNA was screened using the low stringency conditions described above. The two positive clones obtained both stop at an EcoRI site within a 1 kb intron (nucleotide 2157). Full length genomic clones were obtained by inverse PCR (iPCR [Triglia, et al., 1988]). Restriction enzymes leaving GATC sticky ends were used to create an inverse PCR deletion series (figure 3.16). S2886C genomic DNA was digested with BglII (nucleotide 2548) and compatible enzymes (Sau3AI, XhoII, BclI and BamHI), analysed by Southern blotting (figure 3.17A) with an EcoRI-BglII probe (nucleotides 2157-2548) and self ligated. Amplification with iPCR-1 and iPCR-2 primers (fig 3.17B) gave products of the

```

AAT AAT ATT TTA CAT TTT TTT AAA TTA TTC GGA GTA TAT TTT CTT GAA GTT AAT 54
TGC TAC TAT TAT CTG ACT GAA AAT TTT TAT TTT GCA ATC TAC TAT ATA AGG CGC 108
CTT ACC GCC TGT TAA TTA AGC AAA CCG TTT ACC ATA AAC CAC CAC ATC TCT TCA 162
>>pIG41>>
GTT TTA TAG TAA AAA GAA ATC ATA ATG TCT TAC AGA GAT AAC GAA TAT AGT GGA 216
M S Y R D N E Y S G 10
AAT TAC AAT GGA AAA GAG GAT GGC TAT AAC AGC AGA GGG TAA GAA TTC ACA ATT 270
N Y N G K E D G Y N S R G - intron 1 - 23
GGG CGT GCC CCT GTT CAG TAC AAA ATT ATT GAT ACT GAT TTA CCT TTT TTT TTA 324
AAA TAA TTA ATA TTA ACA TTT TAC AGT CGT TAT GGT GGC GGA TAC CGC AAT AAC 378
R Y G G G Y R N N 32
TAT AGT CGT GGC GGC GGT CGT GGA GGT TTC AAT GAT GGC GCT TCT TAT GGC TAC 432
Y S R G G G R G G F N D G A S Y G Y 50
GAT CAA CGA GGA CAA GGA AGA AAT TTT TAC GAG TCA GAT GGT CCC GGT GCT AAT 486
D Q R G Q G R N F Y E S D G P G A N 68
TTG GTA AAG AAG GAT TGG AAA AAT GAA ACT TTG ATT CCG TTT CAA AAA GAT TTT 540
L V K K D W K N E T L I P F Q K D F 86
TAT AAA GAA CAT GAA AAT GTT CGC AAT CGT TCA GAC GCT GAA GTC ACC GAA TAC 594
Y K E H E N V R N R S D A E V T E Y 104
CGT AAA GAA AAA GAA AAT GTC GTG CAT GGA TTA AAT GTT CCT AAA CCT GTT ACG 648
R K E K E I V V H G L N V P K P V T 122
ACA TTT GAA GAG GCT GGT TTT CCT AAC TAT GTA CTA AAG GAA GTC AAA CAA TTG 702
T F E E A G F P N Y V L K E V K Q L 140
GGG TTC GAA GCT CCT ACT CCT ATC CAA CAA CAA GCA TGG CCG ATG GCT ATG TCT 756
G F E A P T P I Q Q Q A W P M A M S 158
GGT AGA GAT ATG GTA GGC ATT TCA GCT ACG GGT TCT GGT AAG ACA CTT TCT TAT 810
G R D M V G I S A T G S G K T L S Y 176
TGT CTT CCT GCT ATC GTA CAT ATT AAT GCC CAA CCA CTA CTT TCT CCA GGC GAT 864
C L P A I V H I N A Q P L L S P G D 194
GGT CCA ATT GTA CTC GTT CTT GCT CCA ACG CGT GAA TTA GCT GTC CAA ATT CAG 918
G P I V L V L A P T R E L A V Q I Q 212
EcoRI
CAA GAA TGT ACT AAA TTT GGC AAA TCA TCC CGA ATT CGC AAC ACA TGT GTG TAT 972
Q E C T K F G K S S R I R N T C V Y 230
GGT GGT GTC CCT CTG GGT CCA CAA ATT CTG GAC CTA ATT CGC GGT GTC GAA ATT 1026
G G V P L G P Q I L D L I R G V E I 248
TGC ATT GCA ACA CCT GGT CGT TTA TTG GAT ATG TTA GAC TCT AAC AAA ACC AAC 1080
C I A T P G R L L D M L D S N K T N 266
TTA CGT CGT GTG ACC TAT TTA GTG CTG GAT GAA GCC GAC CGA ATG CTT GAT ATG 1134
L R R V T Y L V L D E A D R M L D M 284
GGT TTT GAA CCT CAA ATT AGA AAG ATT GTT GAT CAA ATT CGA CCA GAC CGT CAA 1188
G F E P Q I R K I V D Q I R P D R Q 302
ACT GTC ATG TTT TCT GCA ACA TGG CCA AAA GAA GTT CAA CGT CTT GCT CGT GAC 1242
T V M F S A T W P K E V Q R L A R D 320

```

Figure 3.14 *S.pombe dbp2* sequence Accession number X52648

TAT TTG AAT GAC TAC ATT CAA GTA ACT GTC GGC TCT TTG GAT TTG GCA GCT TCT 1296
Y L N D Y I Q V T V G S L D L A A S 338
CAT AAC ATT AAG CAA ATC GTA GAA GTT GTT GAC AAT GCT GAT AAA AGA GCT CGT 1350
H N I K Q I V E V V D N A D K R A R 356
CTT GGA AAG GAT ATC GAA GAG GTT TTG AAG GAT AGG GAT AAC AAA GTT CTT ATT 1404
L G K D I E E V L K D R D N K V L I 374
TTT ACA GGC ACA AAA CGT GTT GCT GAT GAC ATT ACT CGC TTC CTT AGG CAA GAT 1458
F T G T K R V A D D I T R F L R Q D 392
GGT TGG CCT GCC CTT GCC ATT CAT GGT GAT AAA GCT CAA GAC GAG CGT GAC TGG 1512
G W P A L A I H G D K A Q D E R D W 410
GTA CTC AAT GAG TTC CGT ACT GGC AAG TCT CCG ATT ATG GTT GCT ACT GAT GTT 1566
V L N E F R T G K S P I M V A T D V 428
GCT TCT AGA GGA ATA GGT AAG TTT TTA TTT AAC AGT TTT TTC CTT TTC GCA GGC 1620
A S R G I D - intron 2 - 434
TTT TTT ATT TGA CGT TAT ACT TTC ACC GGC GTT CTC AAT GTC TCT TTT GAT AAT 1674
AAA AGA TGG ACT GGC GAT CAA CAC CAT TTT AGC TTG AAT TTT GCA AGT TTG TGT 1728
AflII
TGC TTT GAT GAA TAT TTT TGT TTA CGC CCA GTC TTA AGG AAA GTA GAA AAT TCG 1782
GAT GAT TGT TTA GCC TGT TTC TTA CAG GCA CGA ATG GTG TTA CAA AGA GAA ACT 1836
GAT GAT TTC TAT TTC ATG ATC TTG GCC TTT TCA ATA GTA GTC GAG ATA CGT GTG 1890
GTC TGC TAC CTG TTA AAT TTG TTG TCA ACG TTT CCG CGA TCT ATG TAA CCA TTT 1944
GAT TTG CGG TCA ATT TAA AAG GTT TTT AAA CGC TTA TTC TAT TCA GCC TTC AAA 1998
CCA ACT TGA CTC ATT GTT TAC TTT TAG ACC ACA TTA AAT GAT TTA AAA CAT AGT 2052
TTT TCT GTA CTG TTT CCT CCT TTT CTT TAC TTT CTT TTG TCT TTT GTC TTT CTT 2106
TCA TTG TGA ACT TAT TGC TGT TCG GTA CCA TTG CTA AAA TGT TTA TGC AGT GTT 2160
GTT GAC CTA ACT TGT CAA ATT TAA AGA AGC TTA AAG TTG TTG ATA TAT TGG ATA 2214
AAT CCA TGA AAC ATA TAT CTC GAA GCT TAA AGT CAA AAC CCT CTT GGG CAT TAC 2268
TAA CTT TTT ATA GAT GTT AAG GGT ATT ACA CAT GTC TTT AAC TAT GAT TTC CCT 2322
V K G I T H V F N Y D F P 447
GGA AAT ACG GAA GAC TAC GTT CAT AGA ATT GGA CGT ACT GGT CGT GCT GGA GCT 2376
G N T E D Y V H R I G R T G R A G A 465
AAA GGT ACC GCA TAT ACC TAT TTC ACT TCG GAT AAT GCT AAG CAA GCT CGT GAA 2430
K G T A Y T Y F T S D N A K Q A R E 483
CTT GTT TCC ATT CTC TCT GAA GCC AAG CAG GAC ATT GAT CCT AAG CTT GAG GAA 2484
L V S I L S E A K Q D I D P K L E E 501
ATG GCC CGC TAT TCA AGC GGT GGT CGT GGT GGT AAC TAC CGC CGT GGC GGC TAT 2538
M A R Y S S G G R G G N Y R R G G Y 519
GGT CGT GGT GGT TTC CGT CGT GGA GGA GGT TAT GGA AAC CGT AAT AGG GGT TTT 2592
G R G G F R R G G G Y G N R N R G F 537
ACC GGA AGT AAT TCT GCC CCC CTT GCG CGT TCA CGT TGG TGA TCA ACA ATC CGC 2646
T G S N S A P L A R S R W . 550
GTA TTT TGA AAT ATA AGT AAA ATT AGA TTA ATC GGT TGG ATA TAT AAG GTA ATT 2700
<primer10
TTG AGG TTT ACA AGG GTT AAT AAA AAT ATG ATT TTA TTT AAA 2742

Figure 3.14 *S.pombe dbp2* sequence (continued)

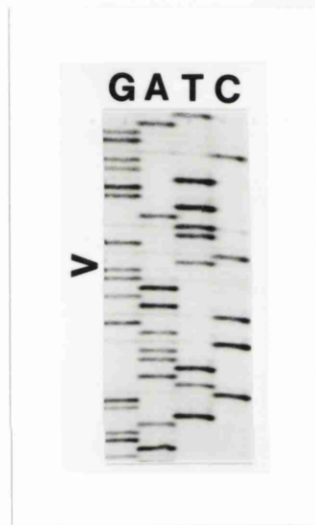


Figure 3.15 Autoradiograph of sequencing gel showing the cDNA sequence of *dbp2* around the site of the first intron. The splice site is marked.

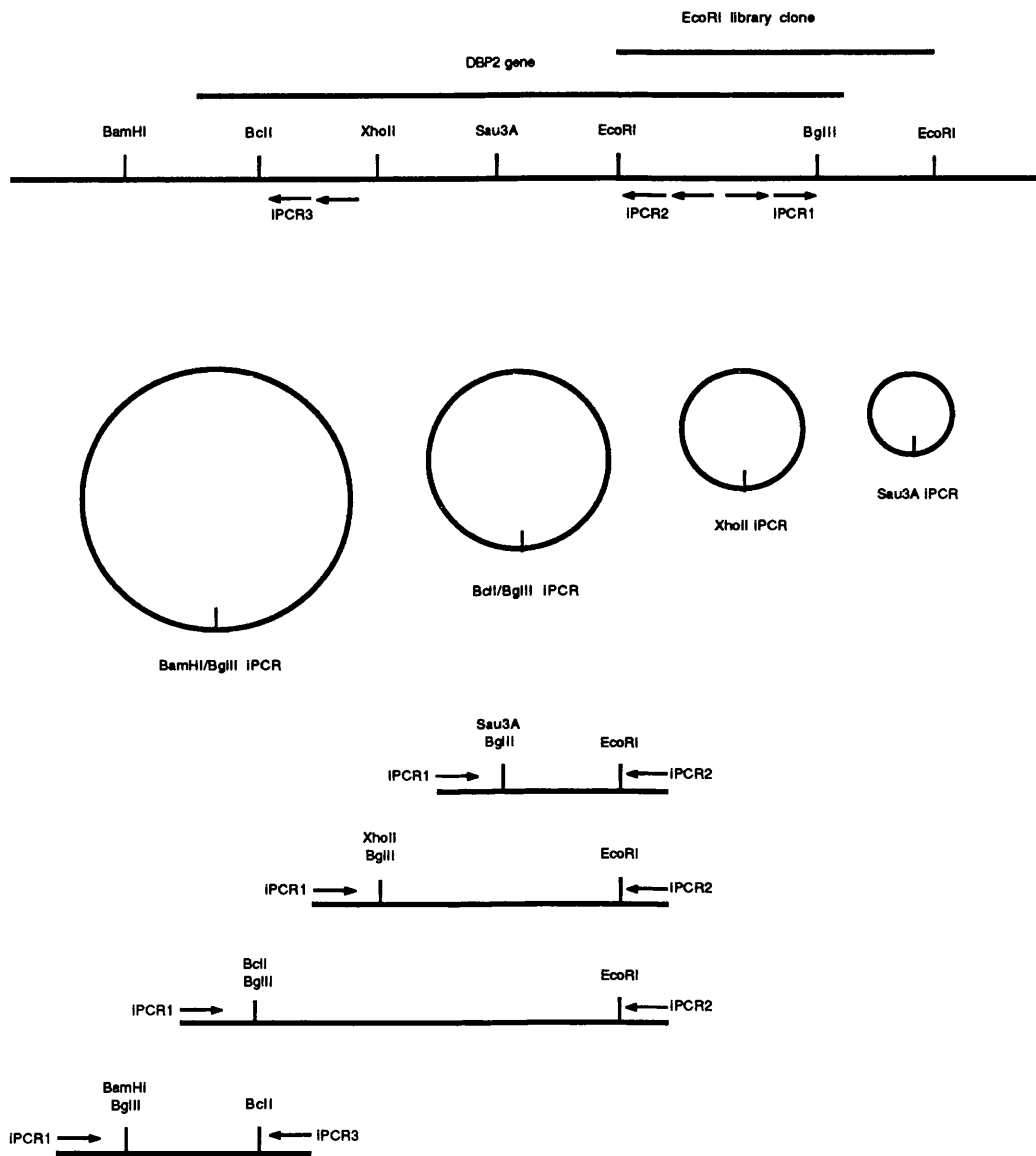


Figure 3.16 iPCR cloning strategy. *S.cerevisiae* genomic DNA was cut with restriction enzymes, self-ligated to create small circular molecules and amplified with the primers shown. Not to scale.

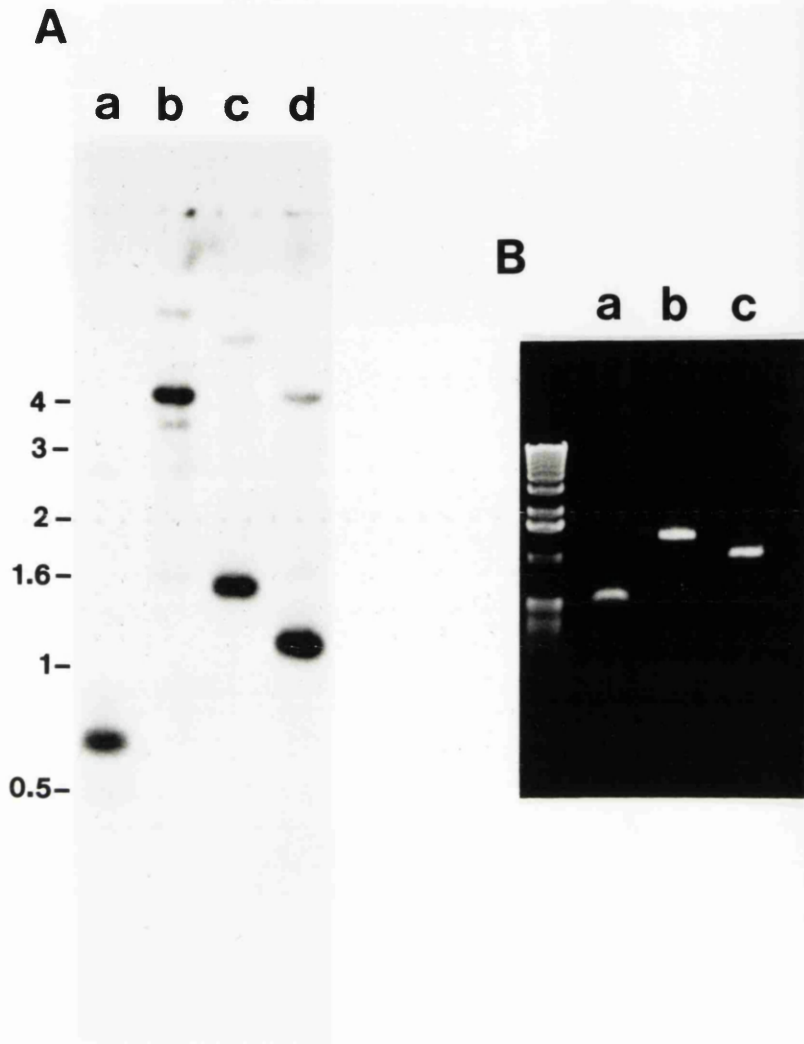


Figure 3.17 A, Southern blot of *S.cerevisiae* genomic DNA hybridised with a *DBP2* probe. a, Sau3A; b, BamHI/BglII; c, BclI/BglII; d, XhoII. Markers in kb.
 B, iPCR products. a, Sau3A; b, BclI/BglII; c, XhoII. Marker is BRL 1 kb ladder.

size expected from the Southern blot: *Sau3AI*, 0.6 kb; *XhoII*, 1.1 kb; *BclI*, 1.5 kb. The *BamHI/BglIII* fragment was too large (~4.5 kb) to amplify with these primers. The *BclI/BglIII* inverse PCR product was subcloned and partially sequenced. An antisense primer near the *BclI* site, iPCR-3, was then used with iPCR-1 to amplify the *BamHI/BglIII* fragment. To obtain clean amplification nested primers were used. Despite this, direct sequencing only worked on the smallest (*Sau3A*) nested inverse asymmetric PCR product. Sequence from subcloned *BamHI/BglIII* inverse PCR product was therefore used to create primers for conventional amplification and sequencing of the entire gene (figure 3.18). Those parts of the gene not sequenced directly were checked in at least three different subclones to rule out PCR artefacts. The intron contains a long run of dT residues (figure 3.18). Since the subclones are all different in this region it was considered prudent to obtain authentic wild type genomic clones by gap repair (see chapter 2 for details).

Derek Jamieson has mapped *DBP2* to chromosome XIII by pulsed field gel analysis (DJJ, personal communication).

Analysis of the p68, *dbp2* and *DBP2* sequences

The yeast genes contain predicted open reading frames encoding 61 kD proteins which are homologous throughout their length, but the human gene differs from the yeast genes at both ends (figure 3.19 and 3.20). 296 amino acids (54%) of the human and *S.pombe* sequences, 301 amino acids (55%) of the human and *S.cerevisiae* sequences and 358 (63%) of the *S.pombe* and *S.cerevisiae* sequences are identical, compared with ~35% for other members of the DEAD box family. The genes were given the names *dbp2* (*S.pombe*) and *DBP2* (*S.cerevisiae*) for DEAD Box Protein 2. The deduced *DBP2* and *dbp2* protein sequences contain all of the motifs characteristic of the DEAD box family.

Their carboxy termini contain the sequences

YGGGHPRYGGGRGGYRRGGYGGGRGGYGGNR (*DBP2*)

and GGRGGNYRRGGYGRGGFRRGGYGNRNRGF (*dbp2*).

PIR database searches identify several RNA binding proteins with related glycine-rich domains, including hnRNPA1 (Cobianchi, et al., 1986) and the nucleolar proteins nucleolin

ATT TTC TTG TCA ACA ACC TGT AAC AGA ATT AAG CAC TAT TAA GGC AAA TTT AGA 54
 GCA AAT ATG ACT TAC GGT GGT AGA GAT CAG CAA TAT AAC AAG ACT AAC TAC AAG 108
 M T Y G G R D Q Q Y N K T N Y K 16
 TCT AGA GGT GGC GAC TTC CGC GGT GGA AGA AAC TCT GAT AGA AAC TCT TAC AAT 162
 S R G G D F R G G R N S D R N S Y N 34
 GAC AGA CCA CAA GGC GGT AAC TAC CGT GGT GGT TTC GGT GGT CGT TCC AAT TAC 216
 D R P Q G G N Y R G G F G G R S N Y 52
 AAC CAA CCC CAG GAA TTG ATC AAA CCA AAC TGG GAT GAA GAA TTA CCC AAA TTG 270
 N Q P Q E L I K P N W D E E L P K L 70
 CCA ACT TTC GAA AAG AAT TTC TAT GTT GAA CAC GAA AGT GTT CGC GAC AGA TCG 324
 P T F E K N F Y V E H E S V R D R S 88
 GAC AGT GAG ATT GCT CAG TTC AGA AAG GAA AAT GAA ATG ACT ATT TCC GGA CAC 378
 D S E I A Q F R K E N E M T I S G H 106
 GAT ATT CCA AAG CCA ATC ACC ACT TTC GAT GAA GCT GGT TTC CCA GAC TAC GTT 432
 D I P K P I T T F D E A G F P D Y V 124
 TTG AAT GAA GTG AAG GCT GAA GGA TTT GAC AAA CCA ACT GGC ATT CAA TGT CAG 486
 L N E V K A E G F D K P T G I Q C Q 142
 GGT TGG CCA ATG GCT TTA TCT GGT AGG GAC ATG GTT GGT ATT GCT GCC ACT GGT 540
 G W P M A L S G R D M V G I A A T G 160
 TCC GGT AAG ACT TTG TCT TAT TGT TTA CCA GGT ATT GTT CAT ATC AAC GCT CAA 594
 S G K T L S Y C L P G I V H I N A Q 178
 primer11>
~~CCA TTA TTG GCT CCA GGC GAT GGA CCA ATT GTT TTG GTT TTG GCT CCA ACT AGA~~ 648
 P L L A P G D G P I V L V L A P T R 196
 GAA TTG GCT GTT CAA ATT CAA ACA GAA TGT TCC AAG TTT GGT CAT AGT TCC AGA 702
 E L A V Q I Q T E C S K F G H S S R 214
 ATC AGA AAT ACC TGT GTC TAC GGT GGT GTT CCA AAA AGT CAA CAA ATC AGA GAT 756
 I R N T C V Y G G V P K S Q Q I R D 232
 TTA TCT CGT GGC TCT GAA ATT GTT ATT GCT ACT CCA GGT CGA CTA ATT GAT ATG 810
 L S R G S E I V I A T P G R L I D M 250
 CTA GAG ATT GGT AAG ACT AAT TTG AAG AGA GTC ACT TAC CTG GTT CTT GAT GAA 864
 L E I G K T N L K R V T Y L V L D E 268
 BclI
 GCT GAT AGA ATG TTA GAT ATG GGT TTT GAA CCT CAA ATC AGA AAG ATT GTT GAT 918
 A D R M L D M G F E P Q I R K I V D 286
 Bali
~~CAA ATC AGA CCT GAT AGA CAA ACC TTG ATG TGG TCT GCC ACT TGG CCA AAG GAG~~ 972
 Q I R P D R Q T L M W S A T W P K E 304
 GTG AAG CAA CTA GCC GCT GAT TAC TTG AAT GAT CCA ATT CAA GTT CAA GTT GGT 1026
 V K Q L A A D Y L N D P I Q V Q V G 322
 <iPCR-3
 TCT CTA GAA CTA TCT GCC TCC CAT AAT ATT ACT CAG ATC GTC GAA GTT GTT TCT 1080
 S L E L S A S H N I T Q I V E V V S 340
 GAT TTC GAA AAG AGA GAT CGT TTG AAC AAG TAC TTA GAA ACA GCC TCT CAA GAC 1134
 D F E K R D R L N K Y L E T A S Q D 358

Figure 3.18 *S.cerevisiae* DBP2 sequence Accession number X52649

EcoRV

AAC GAA TAC AAG ACA TTA ATC TTT GCT TCT ACG AAA AGA ATG TGC GAT GAT ATC 1188
N E Y K T L I F A S T K R M C D D I 376

ACC AAG TAT CTA AGA GAA GAT GGA TGG CCC GCC TTG GCT ATT CAT GGT GAC AAA 1242
T K Y L R E D G W P A L A I H G D K 394

XhoII

GAC CAA AGA GAA CGT GAC TGG GTT CTA CAA GAG TTT AGA AAC GGT AGA TCC CCA 1296
D Q R E R D W V L Q E F R N G R S P 412

ATT ATG GTT GCT ACT GAT GTG GCC GCC AGA GGT ATC GGT ATG TTA AAA TTT TCT 1350
I M V A T D V A A R G I D - intron - 424

CCA TTT TTT TAT TGA TTT TAT TTT TTT TTG TTA CCC CTA ACG ATA TTA CAG CTA 1404
TTT CCT AAT GGC TTT TAA TGA CAT TAA TGA CTT TAT GAC AAC CAT GAT AGT ACA 1458
GAA GAG AGA CCT TTT TTC TTT TTT TTT TTT TCT TTT TTT TTT TTT TTT CTT 1512
TTT TTC TTT TTT TTT TTT TTG AGC CTT TTC GAA TCT AGA CTC TAT GTG AGT CTA 1566
TTC TCG ATG GGG AGT ATC GGA AAT TGA ATT TTA ATT CGA ATG ACT TCT AAT GCA 1620
TCA CTA CAG AAA ACT AAT ATT GGG AGG ATG AGA AAA ATT GAC TTT AAT TAG TCG 1674
TTT TGA GAG ACG GGA AAT TAT AAA CTC GGA GAA GTG TAT TTG TGT TCA TGA TTT 1728
GCA CTT CAT GTC AAA AGA AAT TTC GCT TTT TGA CAT CGG CGC AAA TAA ACA AGG 1782

Sau3AI

AAT TGG CTT TTC AGC TTA TTT CTA GAA CGC ATA CAT ACG CTT CGT TGA TCG TTG 1836
TTT TTT TCA ATG CTT GGC ATT TGT ACA AGG GTA GAT TGT TTT ATT GGA AAA ATA 1890
ATA GTA TAT TCT ACT TTG AAA TGC CGT CAT CCT TCT TGA CTA TTG TTA TTC TCA 1944
TTT TGT GTA GTT TAT GCA TTT TGT AGT TAT ATT GAG ATA CTG TTG CAT CCC AAG 1998
TTC GAA TTA TTA AGA AGT GCT GAT AAA AAT GGA AAA TAA CAA AAT AAA AGG ATT 2052
TCA ACC ATA TTC AGA AAT CAT TTA CTT TGT TTT CCT TTT TTT AAG TGC TAG CTT 2106

EcoRI

TCA TTC AGT TTG AAT AAG GAT TCT GGA GTA TTG ATG ATT AAA TAT TTC GAA TTC 2160
TTT AAT AAA AAT ATA ATT TCT GAT AAT CTT CAA GCC AGG GAA AAT TTG GGA CAT 2214

<iPCR-2

TGT TGG GAG TCC AAA CAT GAA TTT GTG GGG GGC ATG AAA ATA AAG TTC ATG TAC 2268
AAA ATA ACG AAC CAA ATT ACT AAC AGT ATA CTT TGT AAA CGT TTG TTT ACT TCT 2322

TTA TTT TTT CAG ATG TCA AAG GTA TCA ATT ACG TTA TCA ACT ACG ATA TGC CAG 2376
V K G I N Y V I N Y D M P G 439

iPCR-1>

GTA ACA TTG AAG ATT ATG TTC ACA GAA TCG GTA GAA CTG GTA GAG CAG GTG CTA 2430
N I E D Y V H R I G R T G R A G A T 457

CTG GTA CTG CTA TAT CTT TCT TCA CCG AAC AAA ACA AAG GTT TAG GTG CTA AAT 2484
G T A I S F F T E Q N K G L G A K L 475

TAA TCT CTA TCA TGA GAG AAG CTA ATC AAA ATA TTC CTC CCG AAT TAT TGA AAT 2538
I S I M R E A N Q N I P P E L L K Y 493

BglII

ACG ACA GGA GAT CTT ATG GTG GCG GTC ACC CAA GAT ACG GTG GTG GCC GTG GTG 2592
D R R S Y G G G H P R Y G G G R G G 511

GTC GTG GCG GCT ATG GCC GTA GAG GTG GTT ACG GTG GTG GCC GTG GTG GTT ACG 2646
R G G Y G R R G G Y G G G R G G Y G 529

GCG GTA ACA GGC AGA GAG ATG GTG GCT GGG GTA ACA GAG GTC GTT CAA ACT ATT 2700
G N R Q R D G G W G N R G R S N Y . 546

GAA GTC AAA GTG ATT CAA TTC ATT CAT CTC TAT 2733

Figure 3.18 *S.cerevisiae* DBP2 sequence (continued)

```
MTYGGRRQQYNKYKSRGGDFRGRNSDRNSYNDRPPQGNRYRGFGGRSRNYNQFQELIKPNWDEELPKLPFTFEKNFYVEHESVRDRS - S.cerevisiae
MSYRDNEYSGYNGKEDGYNsRGRYGGYRNYSRGGRRGDFNDGASyDQrGGRNFYEsDGPgANLVKKDKWNEILLPFQKDFYKEHENVNRNS - S.pombe
MSGYSSDRDRDRGRDFGAPRFGGSRAGLSGKFGNPGEKLVKKKWNLDELKFEKNFYQEHPLARRT - Human
.....sR.....I.g.ndga..g....I.g.g.....pg..lvkkdw.l..Ip.FeKnFY.EHe.vr.Rs - cons

DSEIAQFRKENEMTISGHDIKPKITTFDEAGFPDYLNEVKAEQKPTGICQCGWPMALSGRDMVGIATSGGKTLsYCLPGIvHINAQPL LAPGD - S.cerevisiae
DAEVEYRKEKEIVVHGLNVPKVITFEAGFNyLVkEVKQLGFEAPPIQQAWPMAMSGRDMVGISATSGKTLsYCLPAIVHINAQPLSPGD - S.pombe
AQEVEYRRSKEITVRGHNCPKVLNFYEA NFANVMVMDVIARQNFTEPTAIQAQWPFVALSGLDMVGVAQTSGKTLsYLLPAlVHINHQPFLERGD - Human

d.Ev..yRkeEItv.Ghn.PKpvtF.EAgFP.yvI.evK..gF..PT.IQ.QgWpMAlSGrDMVGIaATSGGKTLsYcLpAlVHINaQpLL.pGD - cons
motif I

GPIVLVLPATRELAVQIQTECSKFGHSSRIrNtCVYGGVPKsQqIRDLsRGSEiVIAtpGRLIdMLeIGkTNLkRvTYLVlDeADRMlDmgFEpQIR - S.cerevisiae
GPIVLVLPATRELAVQIQECTKFGKSSRIrNtCVYGGVPlGPIldLIrgVEiCiatPGRLIdMlDSnKTNLRRvTYLVlDeADRMlDmgFEpQIR - S.pombe
GPICLVLPATRELAQQVQQAAYECrAcRLKStCIYGGAPkGfQIRDLERgVEiCIATpGRLIdfLEcGkTNLRRtTYLVlDeADRMlDmgFEpQIR - Human

GPIVLVLPATRELAVQIQeC.kfG.ssRiRntCVYGGVPkGpQIRDL.RgVeIciATpGRLIdMLe.gkTNLrRvTYLVlDeADRMlDmgFEpQIR - cons
motif Ia

KIVDQIRPDRQTLmWSATWPEkVQLAADyLNDPIQvQVGSLEsASHNITQIVeVvSDfEKrDRlNKYLEfASQDNEyKTLIFAsTKRMCDdITKY - S.cerevisiae
KIVDQIRPDRQTLmWSATWPEkVQRLARDYLDYIQVTVGSLDLAASHNIkQIVeVVDNADKRRLGkDIEEVLKDRDNKVLIFtGtKRVADdITRF - S.pombe
KIVDQIRPDRQTLmWSATWPEkVRLAEDFLkDYIHINIGALeLSANHIlIQVDVCHdVEkDKLIRLMEEImsEKENKtIVfVETKRRCDeLTrK - Human

KIVDQIRPDRQTLmWSATWPEkV.qLAdYLndYIqv.vGSLeLSASHNI.QIveVv.d.eKr.rL.k..Ee...d.enK.liF..TKR.cdDiTr. - cons
motif III

LREDGWPALAIHGDKQQRERDwVLQEFrNgrSPImvATDVAARgIDVKGiNyVINDMPGNIeDYVHRIGRTGRAGATGTAISFFTEQNKGLGAKLI - S.cerevisiae
LRQDGPALAIHGDKAQDERDwVlNEfRtGKSPImvATDVASRGIDVKgITHVFNDFPGNTeDYVHRIGRTGRAGAKGTAYTfTSDNAKQARELY - S.pombe
MRRDGWPAMGIHGDKSQQRERDwVlNEfKHKAPILlATDVASRGIDVEdvKfVINYDYPNSSeDIHRIGRTARStKtGTAYTfTFPNNIKQVSDLI - Human

IR.DGWPALAIHGDK.Q.ERDwVlNEFr.GksPImvATDVASRGIDVkgI..ViNYD.Pgn.EDYVHRIGRTgRagatGTAYtffT..N.kq...Li - cons
motif V

SIMREANQNIPELLkYDRRsyGGHPryGGRRGgYRRGyGGRRgYGGNRQRDRDGGWGNRGRSNI - S.cerevisiae
SILSEAKQIDPKLEEMARySSGRRGgYRRGgYRRGgYGNRNRGfTGSNSAPLARSW - S.pombe
SVLREANQAInPKLLQlVEDRGSRSRRGGMKDRDRYsAgKRgGfNfRDRENyDRGYSLLKkrDFGAkTQNGVysAAnyTNGSfGSNFVSAgi - Human

SILREANQ.I.PkIIL...r.s.gGr...yggg...gFgg..r.GgyGg.n.g.g.n....g.s.s.. - cons

QTSFRFGNPTGTYQNGYDSTQOYGSNVPMHNGMNQOAYAYPATAAPMIGYPMPGTGYSQ - Human
```

Figure 3.19 Deduced amino acid alignment of p68, DBP2 and *dbp2*. Amino acids identical in all three sequences are shown in capitals in the consensus line, those identical in only two sequences are in lower case. Sequences matching the Hodgman/Gorbalenya helicase motifs are indicated.

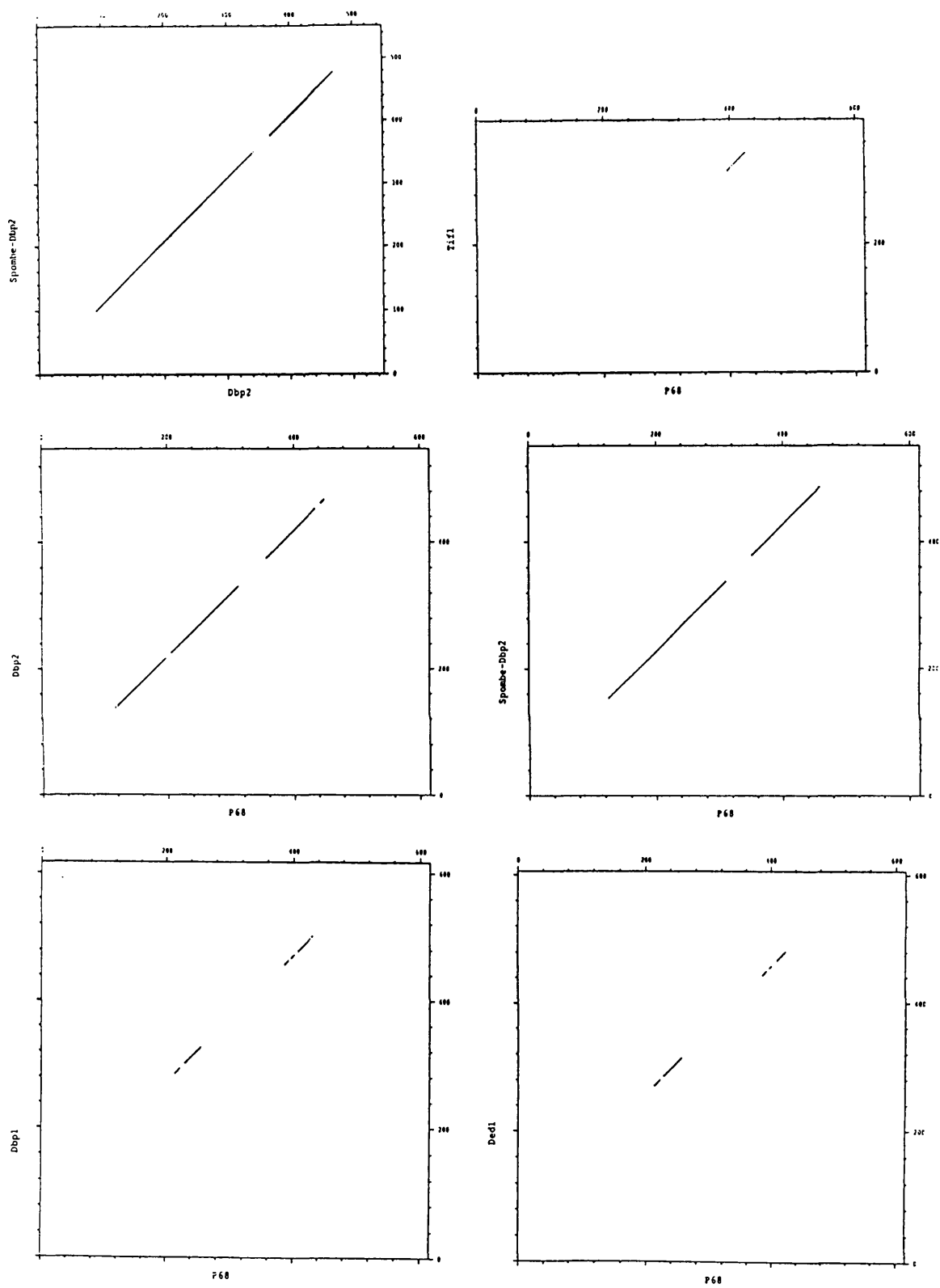


Figure 3.20 Dotplot comparison of the deduced amino acid sequences of *S.cerevisiae* DBP2 ("*Dbp2*"), *S.cerevisiae* DED1, *S.cerevisiae* DBP1, *S.cerevisiae* TIF1, *S.pombe* *dbp2* and human p68 using a window of 40 and a stringency of 39.

(Lapeyre, et al., 1987), fibrillarin (Schimmang, et al., 1989) and yeast SSB1 (Jong, et al., 1987). Among the DEAD box proteins the *Drosophila* gene *vasa* (Lasko and Ashburner, 1988; Hay, et al., 1988b) contains a glycine-rich fivefold heptad repeat (F/SRGGE/QGG) at the amino terminus. Nucleolin, hnRNPA1 and SSB1 contain the RNP1 and RNP2 consensus RNA binding domains (Brandziulis, et al., 1989), which are absent in p68, *DBP2* and *dbp2*. Northwestern blotting of nucleolin fragments suggests that the RNP domains bind RNA but the glycine-rich domain does not (Bugler, et al., 1987). However, the high arginine and aromatic amino acid content of the glycine-rich domains may permit electrostatic and base stacking interactions which modify the specificity of nucleic acid binding (Jong, et al., 1987).

p68, *DBP2* and *dbp2* contain a conserved intron

The p68, *DBP2* and *dbp2* genes all contain an intron at exactly the same position in motif V (the "ARGID" motif, figures 3.19 and 3.21). The intron is 700 bp long in *S.pombe*, 1 kb long in *S.cerevisiae* and 1.2 kb long in man. One clone obtained from an Okayama Berg library apparently underwent alternate splicing: the sequence of this cDNA is identical with that of the genomic DNA up to nucleotide 2428, where an alternate 5' splice site is used (CT/GTAAGG). This introduces a stop codon into the human sequence immediately after the prototypic 5' splice site (nucleotide 1476). Translation of this variant would produce a truncated protein which would almost certainly be non-functional because it lacked part of motif V and all of motif VI. Small introns are not unusual in *S.pombe* genes, but the database appears to contain no *S.pombe* introns larger than that in *dbp2*. No *S.cerevisiae* genes were identified containing introns longer than the 551 bp intron in ribosomal protein S16A (Molenaar, et al., 1984; Hawkins, 1988) and only one intron was identified near the 3'-end of a gene (the 52 bp second intron in *MATa1* [Miller, 1984]). The intron in *DBP2* is thus remarkable both because of its size and because of its location. One reason for conservation of an intron is the existence of an intron-encoded protein. However, the nucleotide sequences of the p68, *DBP2* and

human GCCTCCAGAGGGCTAG GTTAGT---1155bp---CTAAC-----63bp---GCAACACCTTACAG ATGTGGAAGATGTGAAAT
A S R G L D V E D V K F

S.cerevisiae GCCGCCAGAGGTATCG GTATGT---952bp---TACTAAC--28bp---CTTTATTTTTTCAG ATGTCAAAGGTATCAATT
A A R G I D V K G I N Y

S.pombe GCTTCTAGAGGAAATAG GTAAGT---685bp-----CTAACTTTTTATAG ATGTTAAGGGTATTACAC
A S R G I D V K G I T H

Figure 3.21 Human p68, *S.pombe dbp2* and *S.cerevisiae DBP2* contain an intron at the same site in helicase motif V. The predicted human branchpoint sequence was identified by homology to the higher eukaryotic branchpoint consensus (YNRAY).

dbp2 introns are not closely related, their deduced amino acid sequences contain no long open reading frames and database searches identify no closely related sequences. Third codon preference plots (Fickett, 1982) strongly suggest that they are non-coding (figure 3.22).

Discussion

DBP2 is the only member of the DEAD family which contains an intron out of the 15 *S.cerevisiae* DEAD box genes which have been identified so far (John Abelson, personal communication). The intron is present in all of the p68-like genes which have been examined (including the *Drosophila* gene RM62 [Alan Christensen, personal communication]). However, the presence of an intron in motif V can not be considered unequivocal evidence that a DEAD gene is a member of the p68 family because an intron is also present in motif V in the *Drosophila* gene *vasa*.

The evolutionary distance between *S.cerevisiae* and *S.pombe* is as great as that between yeast and man (Russell and Nurse, 1986). No other gene is known to contain an intron conserved in *S.cerevisiae*, *S.pombe* and man. It implies that an ancestral intron was present at least 1.2 billion years ago, as has been postulated for the triosephosphate isomerase introns in *Aspergillus*, chickens and man (Gilbert, et al., 1986). Interestingly, translation past codon 189 facilitates splicing of triosephosphate isomerase pre-mRNA, which led Manquat to speculate that the triosephosphate isomerase introns may autoregulate triosephosphate isomerase expression (Cheng, et al., 1990; Daar and Manquat, 1988). Few *S.cerevisiae* genes contain introns, and in several cases those introns have specific functions: mitochondrial introns encode reverse transcriptases (Michel and Lang, 1985), maturases and endonucleases (Delahodde, et al., 1989); L32 prevents splicing of its own message by promoting basepairing between the 5' end of the transcript and the 5' splice site (Warner, 1989); and *MATa1* contains two inefficiently spliced introns which may play a role in the gene conversion events during mating type switching (Fink, 1987). *S.pombe* introns over 180 bp are known to abolish

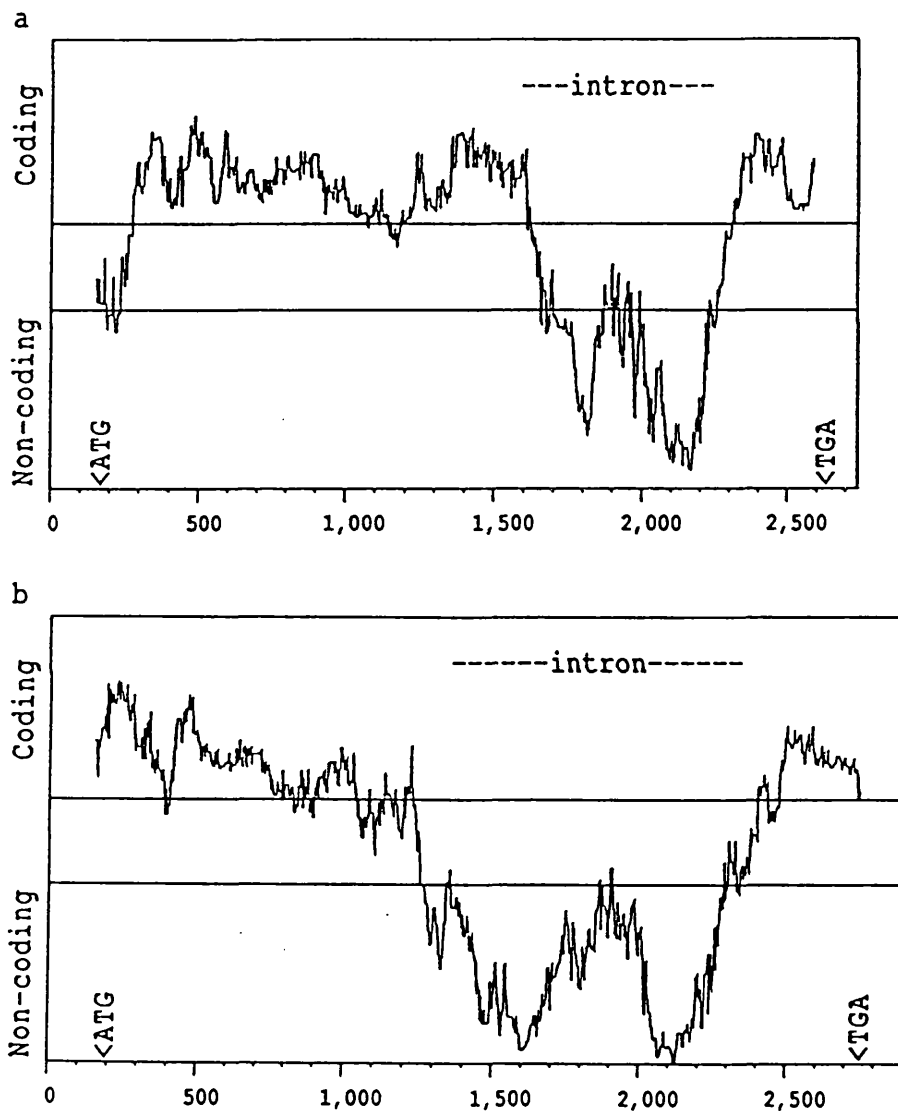


Figure 3.22 Third codon preference plots: a, *S.pombe dbp2*; b, *S.cerevisiae DBP2*. The assignment of coding regions for values of the testcode statistic between the two horizontal lines is uncertain. Units on the x-axis are in base pairs.

splicing in a model system (Gateman, et al., 1989) but the existence of a 700 bp in *dbp2* indicates that this size limit does not operate *in vivo*. Nevertheless, it is perhaps not surprising that one of the *S.pombe* clones was polyadenylated but unspliced. The unusual nature of the introns and the existence of alternately spliced cDNA library clones hint at the possibility of an intriguing form of autoregulation whereby an RNA helicase may control its own expression by modifying the secondary structure of its mRNA. Preliminary data addressing this possibility are presented in chapter 7.

E.coli dbpA

Four λ clones (2A3, 2B2, 2B3 and 2B5) were obtained by low stringency screening of an *E.coli* genomic library with a *dbp2* probe (the 1.2 kb EcoRI insert from one of the normally spliced *dbp2* λ gt11 clones). Hfr crossing of lysogens showed that 2A3, 2B2 and 2B5 map between the origins of Hfr KL99 and KL208, ie between 24 and 31 minutes. The location of 2B3 was not determined (S.Picksley, personal communication).

The *dbp2* probe hybridized with a 3.4 kb BamHI fragment of *E.coli* genomic DNA and an apparently identical fragment of phages 2A3, 2B2 and 2B5. This BamHI fragment was subcloned from 2B5 and located by Southern blotting of genomic DNA at 29.6 minutes on the physical map (Kohara, et al., 1987) of the *E.coli* chromosome (figure 3.23). To confirm the location, DNA from the phage used to create this part of the map were Southern blotted and probed with 2B5 DNA. As expected the only Kohara phage to hybridise with the probe was phage 4A6 (S.Picksley, personal communication).

The 3.4 kb *Bam*H1 fragment contains an open reading frame, which was given the name *dbpA*, for *dead box protein A* (figure 3.24). The deduced amino acid sequence is 432 amino acids long and has a predicted molecular weight of 46.5 kD. Figures 3.25 and 3.26 show a comparison of *dbpA* with human p68 and *srmB*. In the region of overlap 29% of amino acids are identical, a value similar to that for other distinct members of the DEAD box family. *dbpA* contains all the previously identified DEAD box motifs. Despite the fact that it was cloned by

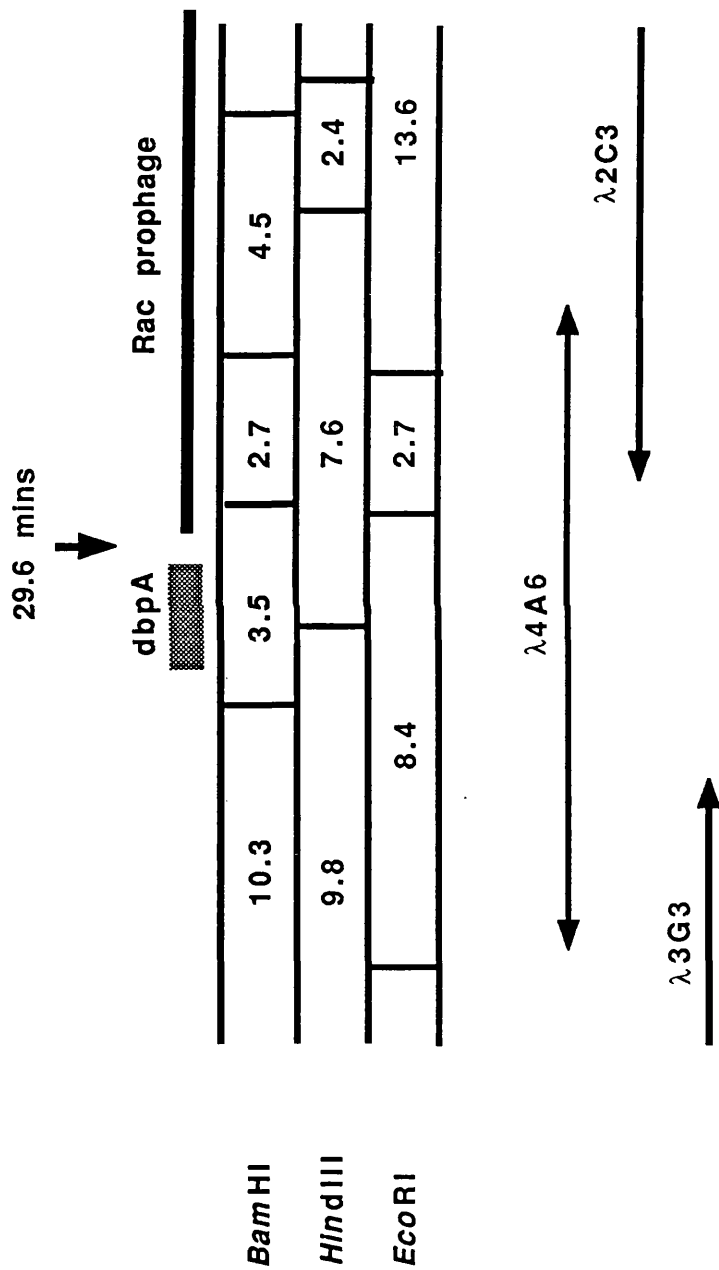


Figure 3.23 Physical map of the W3110 chromosome showing the location of dbpA, the Rac prophage and three phages supplied by Dr Kohara. Restriction fragment sizes are in kb.

GCA AAT TCC TGT CCC TGA ACC GAT TCC TGG CGA TCC TGT TCC AGT GCC CGA TCC
GAT CCC TCG CCC GCA ACC CAT GCC TGA CCC ACC ACC CGA TGA AGA ACC GAT TAA
ATT GTC GCA TCG TGA GCG TAG ATC TGC GAG GAT ACG CGC CTG CTA ACT TTG CGT
CGA TGA CCA CGA GAA TAG ATT GTG ACC GCT TTT TCT ACC CTG AAT GTT TTG CCT
CCC GCC CAA CTC ACG AAC CTT AAT GAG TTG GGT TAT TTA ACC ATG ACG CCG GTG
M T P V
CAG GCC GCC GCG CTT CCG GCG ATC CTT GCC GGA AAA GAT GTT CGC GTG CAG GCG
Q A A A L P A I L A G K D V R V Q A
AAA ACC GGC AGC GGC AAA ACG GCG GCT TTT GGC CTC GGC TTG TTA CAG CAA ATT
K T G S G K T A A F G L G L L Q Q I
GAT GCG TCG CTA TTT CAA ACC CAG GCT TTA GTG CTG TGT CCT ACG CGT GAA CTG
D A S L F Q T Q A L V L C P T R E L
GCG GAT CAG GTG GCA GGT GAA TTG CGT CGG CTG GCG CGT TTT CTG CCA AAT ACC
A D Q V A G E L R R L A R F L P N T
AAA ATT TTG ACG TTG TGC GGT GGT CAA CCG TTC GGT ATG CAG CGT GAT TCG TTG
K I L T L C G G Q P F G M Q R D S L
CAA CAT GCG CCG CAT ATT ATC GTG GCA ACG CCG GGG CGT TTG CTG GAT CAC CTG
Q H A P H I I V A T P G R L L D H L
CAA AAA GGC ACG GTA TCA CTG GAT GCG TTG AAT ACG CTG GTG ATG GAT GAG GCC
Q K G T V S L D A L N T L V M D E A
GAC CGC ATG CTG GAT ATG GGA TTT AGC GAT GCC ATT GAT GAT GTC ATC CGT TTT
D R M L D M G F S D A I D D V I R F
GCG CCT GCA TCT CGA CAG ACG CTT CTG TTT TCG GCA ACC TGG CCG GAA GCC ATC
A P A S R Q T L L F S A T W P E A I
GCT GCA ATC AGC GGA CGA GTG CAA CGC GAT CCT TTG GCG ATT GAA ATT GAC TCA
A A I S G R V Q R D P L A I E I D S
ACA GAT GCT TTG CCA CCC ATT GAA CAA CAA TTT TAT GAG ACA TCC AGC AAA GGC
T D A L P P I E Q Q F Y E T S S K G
AAA ATT CCT CTG TTG CAA CGG TTA TTA AGC TTG CAT CAG CCA TCC TCT TGC GTG
K I P L L Q R L L S L H Q P S S C V
GTG TTT TGC AAT ACC AAA AAA GAT TGC CAG GCT GTC TGC GAC GCG CTG AAT GAA
V F C N T K K D C Q A V C D A L N E
GTA GGG CAA AGT GCA TTG TCA TTA CAC GGC GAT TTG GAG CAA CGC GAT CGC GAT
V G Q S A L S L H G D L E Q R D R D
CAG ACC CTG GTA CGT TTT GCT AAC GGT AGC GCC CGT GTA CTG GTC GCG ACT GAT
Q T L V R F A N G S A R V L V A T D
GTT GCT GCG CGT GGT CTG GAT ATT AAA TCG CTT GAG CTG GTG GTG AAC TTT GAG
V A A R G L D I K S L E L V V N F E
CTG GCG TGG GAC CCT GAA GTT CAT GTA CAT CGC ATC GGT CGT ACA GCT CGT GCA
L A W D P E V H V H R I G R T A R A
GGA AAT AGC GGT CTG GCG ATC AGT TTC TGT GCT CCG GAA GAA GCA CAG CGG GCC
G N S G L A I S F C A P E E A Q R A
AAT ATC ATT TCT GAC ATG TTG CAG ATA AAA CTT AAC TGG CAA ACG CCG CCA GCT
N I I S D M L Q I K L N W Q T P P A
AAT AGT TCC ATT GCG ACG CTG GAA GCA GAA ATG GCA ACG TTG TGT ATC GAT GGC
N S S I A T L E A E M A T L C I D G
GGG AAA AAA GCC AAA ATG CGC CCG GGT GAT GTA TTA GGT GCA CTG ACA GGA GAT
G K K A K M R P G D V L G A L T G D
ATC GGG CTT GAT GGC GCA GAT ATT GGC AAA ATC GCC GTG CAT CCG GCG CAT GTC
I G L D G A D I G K I A V H P A H V
TAT GTC GCG GTC CGT CAG GCT GTT GCT CAT AAA GCA TGG AAA CAG TTA CAG GGC
Y V A V R Q A V A H K A W K Q L Q G
GGG AAG ATT AAA GGA AAA ACG TGC CGG GTG CGG TTA TTA AAA TAA TGA AAT GTT
G K I K G K T C R V R L L K .
GAA TTG CCG GGT GCA AGA GTA AAC ATC TTA TTC GGG ATT GCC GGA TGC GAC GCT
GGC CGC GTC TTA TCC GGC CTC CAT AAG AGT AGC CCG ATA CGC TTG CGC

Figure 3.24 *E.coli dbpA* sequence Accession number X52647

```

MSGYSSDRDRGRDRGFAPRFGGSRAGPLSGKKFGNPGKLVKKKNWLDLDPKFKNFYQEHPLDARRTAQEVETYSRSKE - p68

ITVRGHNCPKVLFNFEANFPANVMDVIARQNFTEPTAIQAQQWPVALSGLDMVGAQTGSGKTLTYLLPAIVHINHQPFLERGDGPI C - p68
MTPVQAAALPAIILAGKDVVQAKTGSKTAFAAGLGLLQQIDASLFTQQA - dbpA
MTVTTTFSELEDESLEALQDKGFTRPTAIQAAAIPALDGRDVLGSAPTGTGKTAAYLLPALQHLDFPRKKSQGPRI - srmb

.v.v.f.e.....ft.pTaiQaaa.P.aL.G.Dv.g.A.TGsgKtaayllpalqhi....pf....g....i. - cons
motif I

LVLPATRELAQQVQAAEYCRACKLKSTCIY GGAPKQPQIRDLERGVETIATPGRLIDFLECGKTLRRTTYLVLDEADRLMDMGF - p68
LVLCPTRELADQVAGELRRLARLPTNKILTCGGQPFQMRDLSQHAPHIIIVATPGRLLDHLQKGTVSLDALNTLVNDEADRLMDMGF - dbpA
LILPTRELAMQVSDHARELAKHTHLDIATIT GGVAYNHAEVFSENQDIIIVATTGRLLQVIKEENFDCRAVETLILDEADRLMDMGF - srmb

LvL.PTRELA.QV...arelar....l....i....GG.p.g....l.....I.vATpGRLLd.l..g....lra..tLvLDEADRLMDMGF - cons
motif Ia
motif II

EPQIRKIVDQIRPDRQTIMWSATWPKVQRQLAEDFLKDYIHINIGALELSANHNILQIVDVCHDVEKDEKLIRLMEEIMSEKENKTIVF - p68
SDAIDDVIRFAPASRQTLIFSATWPEAI AAISGRVQRDPLAIEDSTALPPIEQOQFYETSSKGKIPILQRLLSLHOPSSC VVF - dbpA
AQDIEHIAGETRWRKQTLIFSATLEGDAIQDFAERLLEDPVEVSANPSTRERKKIHQWYIRADDLEHKHTALLVHLLKQPEATRS IVF - srmb

...I...i....r...IQTLlfsatwp....q.a....l.....i....st.a...I.Q.y....d.ek...L.rll.....ivf - cons
motif III
motif IV

VETKRRCDLTRKMRDRGWPAMGIHGDKSQOERDWWLNEFKHGKAPILIAIDVASRGLDVEDVKFVINVDYPNSSSEDIHRIGRTARST - p68
CNTKKCCQAVCDALNEVQSALSLHGDLERDRDQTLVRFANGSARVLVATDVAAARGLDIKSLELVVNFELAWDPEVHVHRIGRTARAG - dbpA
VRKRERVHELANWLREAGINNCYLEGEMVQGRNEAIKRLTEGRVNVLVATDVAARGIDIPDVS HVFNFDMPRSGDTYLHRIGRTARAG - srmb

v.tk.rc.el...lre.G.a..lhGd..Q...Rd..l.rf..G.a.vLvATDVaaRGldi.dv..V.Nfd.p.s.e.y.HRIGRTARag - cons
motif V
motif VI

..G.Aisf..p.....l.e...a.p.l.....se..tg....k...ak.r.....d.k..... - cons

FGAKTONGVYSAANYTNGSFGSNFVSAGIQTSFRGTGNPTGTGYQNGYDSTQQYGSNVNPMHNGMNQQAAYATAAAPMIGYPMPTGYSQ - p68
YVAVRQAVAHKAWKQLQGGKIKGKTCRVRLIK - dbpA
SGTGVPPTTEE - srmb

.ga...q.....a.....g..... - cons

```

Figure 3.25 Deduced amino acid alignment of *dbpA*, *p68* and *srmb*. Amino acids identical in all three sequences are shown in capitals in the consensus line, those identical in only two sequences are in lower case. Sequences matching the Hodgman/Gorbalenya helicase motifs are indicated.

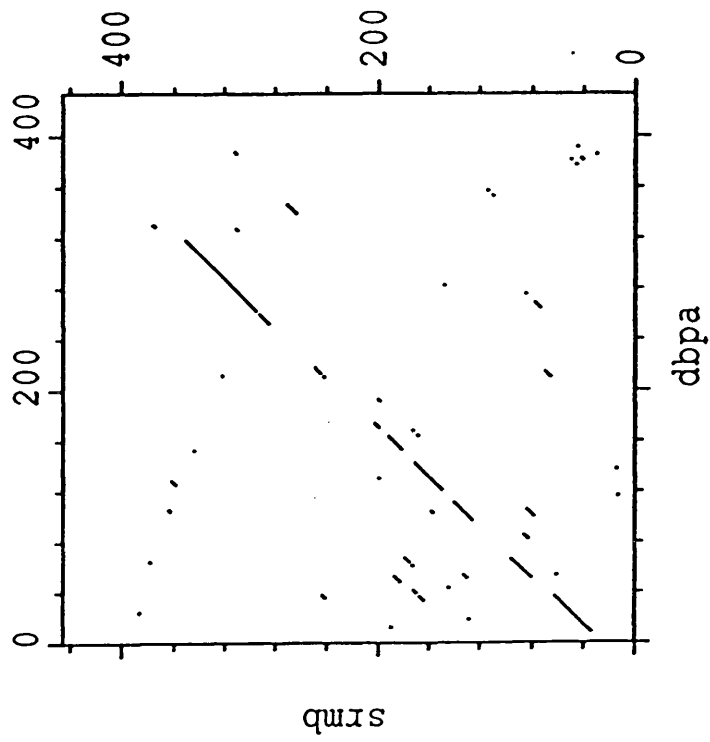
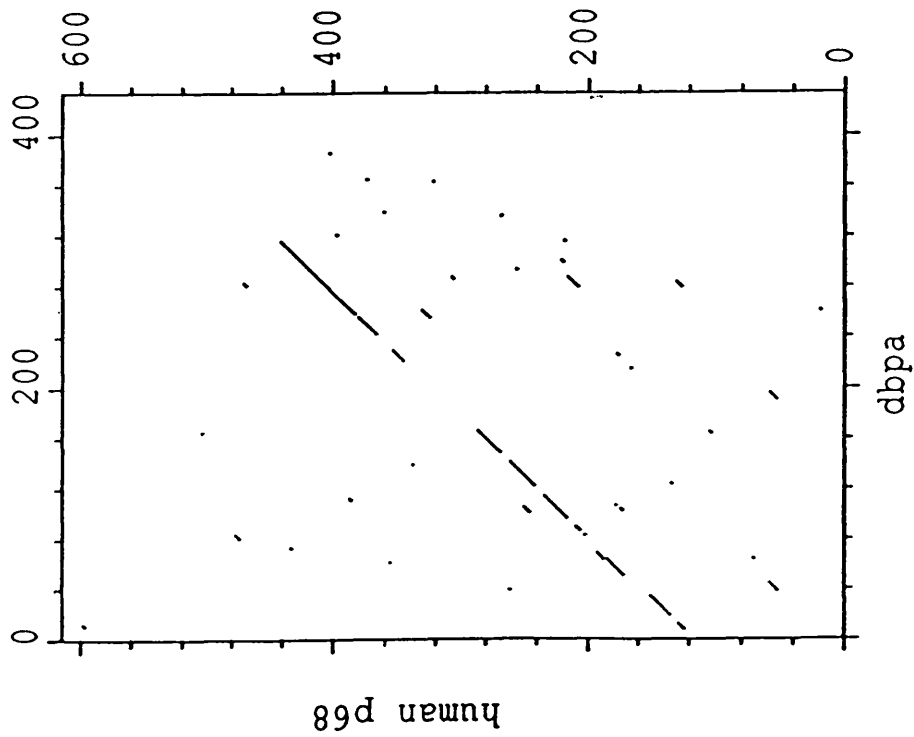


Figure 3.26 Dotplot comparison of the deduced amino acid sequences of *dbpA*, p68 and *srmB* using "Compare" with a window of 15 and a stringency of 10.

hybridisation with *dbp2* DNA, *dbpA* is no more closely related to *dbp2* than it is to any other member of the family. At the DNA level there is scarcely any more homology between *dbpA* and the probe used to screen the library than there is between *srmB* and the *dbp2* probe (figure 3.27). This contrasts with the equivalent analysis of the yeast cloning (figure 3.28).

Overall, the DEAD box proteins have a conserved core region and divergent amino and carboxy termini. *dbpA* has the shortest amino terminal region of any DEAD box protein, with only 25 amino acids before the G-GKT motif. Although it is possible that the amino terminus has been incorrectly identified, this seems unlikely because Fickett (third codon preference) plots strongly predict that the region 5' of the deduced methionine is non-coding; there is an 8 out of 11 match to the consensus Shine Dalgarno sequence ending at position -12 relative to the methionine (underlined in figure 3.24); and finally the sequence was checked by direct asymmetric PCR sequencing from genomic DNA. PIR database searches with the carboxy terminal sequence of *dbpA* using PROSRCH found no significant homology to other proteins. Identification of bacterial promoter elements by sequence analysis is difficult because the motifs are so degenerate. Nevertheless, the promoter probably lies within 130 base pairs of the deduced translation start site because that region is sufficient to cause overexpression of DbpA on copy number alone. The smallest DNA fragment tested which gives DbpA expression comparable to that seen with the 3.4 kb BamHI fragment (figure 4.4) is a 1.5 kb BglIII/HpaI fragment.

Discussion

The function of *dbpA* is unknown, although a role in translation or ribosome assembly would not be surprising. David Shub has postulated the existence of an inducible RNA helicase which controls splicing of the T4 introns (Shub, et al., 1987; Gott, et al., 1988). One of the T4 introns lies in the phage gene encoding the small subunit of ribonucleotide reductase, another lies in the phage thymidylate synthetase gene. Shub's argument is that T4 may try to restrict phage-dependent production of DNA precursors in adverse environmental conditions by not

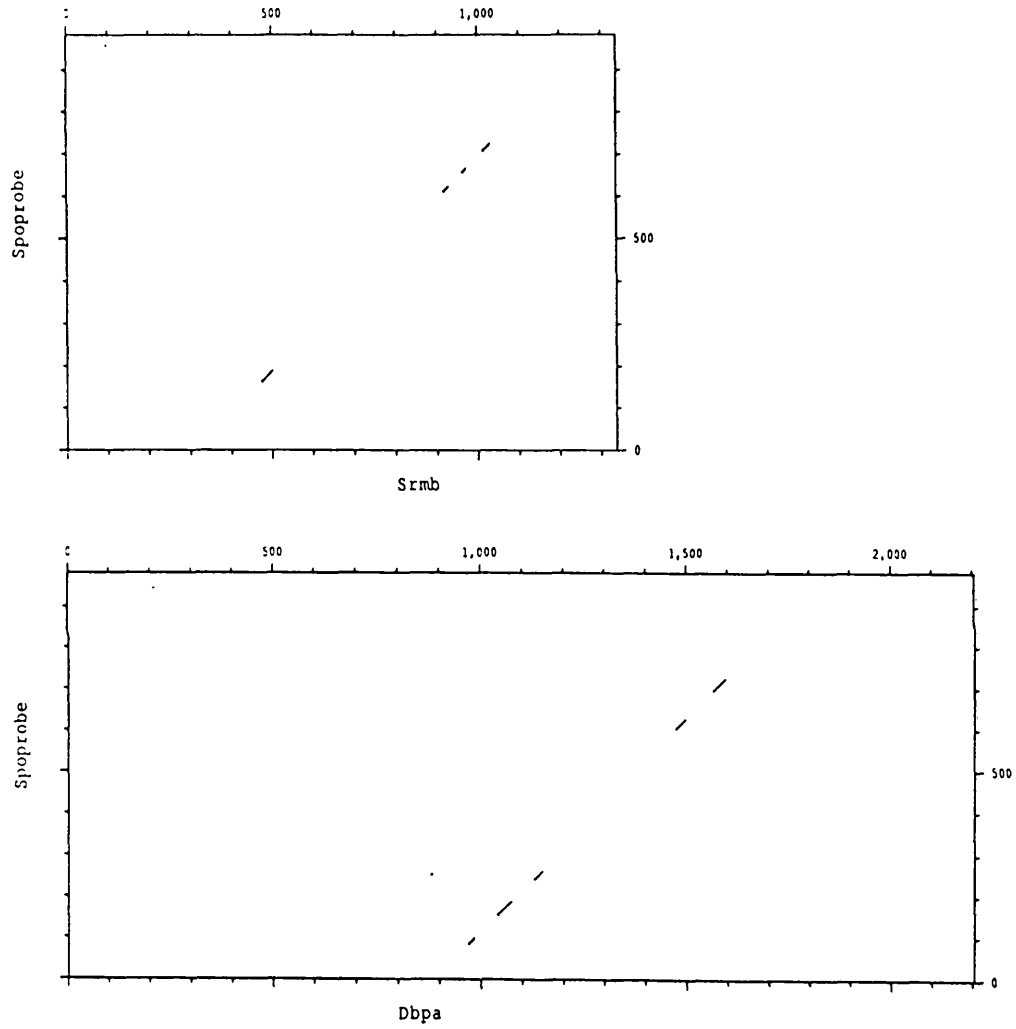


Figure 3.27 Dotplot comparison of the DNA sequence of the *S.pombe* probe and the DNA sequences of *dbpA* and *SrmB* with a window of 30 and a stringency of 20.

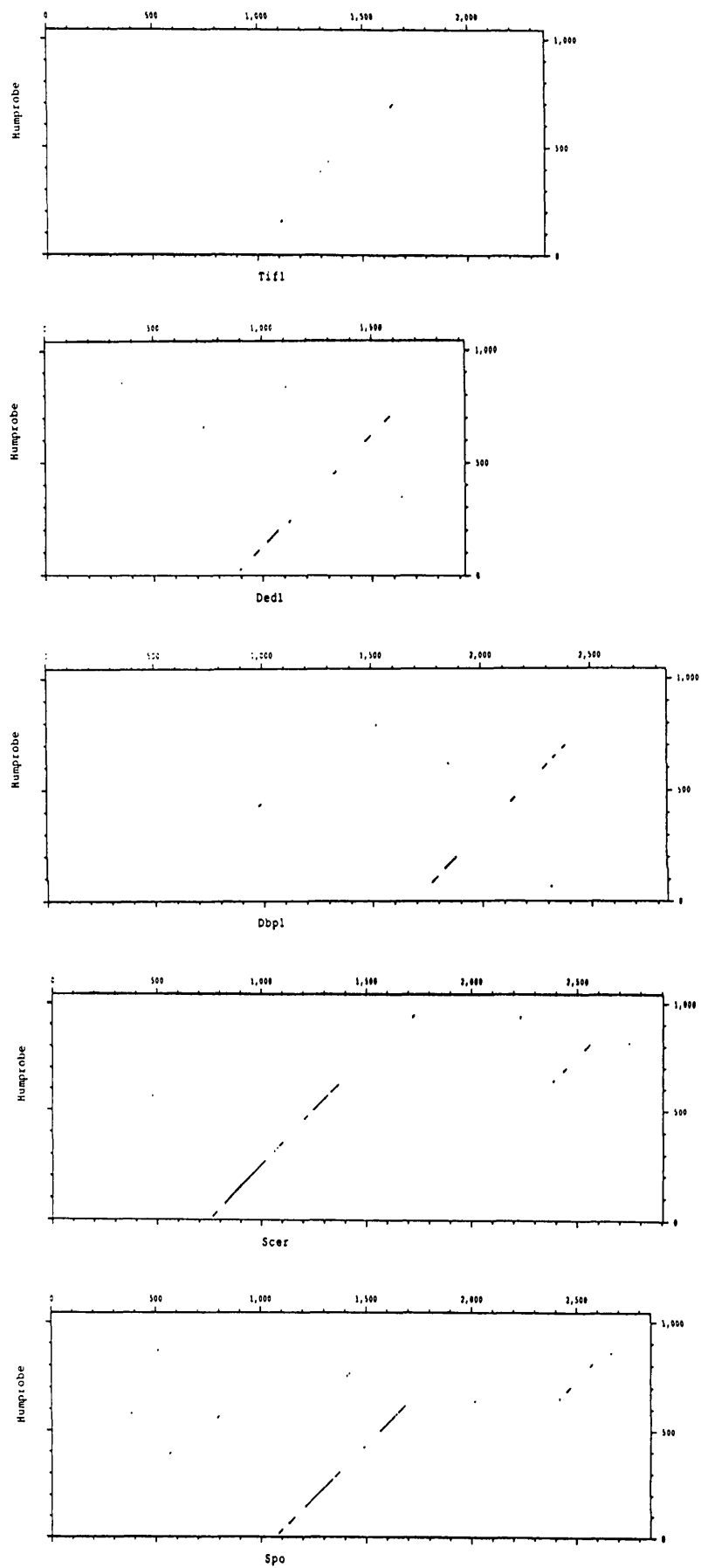


Figure 3.28 Dotplot comparison of the DNA sequence of the human p68 probe and the DNA sequences of *S.pombe dbp2* ("Spo"), *S.cerevisiae DBP2* ("Scer"), *DBP1*, *DED1* and *TIF1* with a window of 30 and a stringency of 20.

splicing out the introns. He postulates that the normal response to environmental stress includes induction of an RNA helicase, and that this helicase abolishes phage intron splicing by modifying the secondary structure of the T4 introns. The anti-DEAD antibody described in chapter 4 provides one means to test this hypothesis (figure 3.29). The most striking change is the appearance of a 40 kD band when bacteria are placed in minimal medium. However, a 40 kD DEAD protein could contain little more than the core helicase motifs. There is a 50 kD band which could be DbpA or SrmB, but it is apparently not induced in any of the situations tested (use of a poor carbon source, induction of the stringent response or oxidant stress). The bands at 89 kD and 116 kD may represent new *E.coli* DEAD box genes.

General Discussion

Cloning the human p68, *S.pombe dbp2*, *S.cerevisiae DBP2* and *E.coli dbpA* genes was essential for much of the remaining work described in this thesis: human p68 was overexpressed in *E.coli* and used to raise polyclonal antibodies (chapters 4 and 5); a peptide sequence conserved in all three p68-like genes was used to make a monoclonal antibody which reacts with a large number of DEAD proteins (chapters 4 and 7); the yeast genes were shown to be essential by gene disruption and tetrad dissection (chapter 7); and finally, the possibility that the conserved intron in motif V autoregulates *S.cerevisiae DBP2* expression was examined (chapter 7).

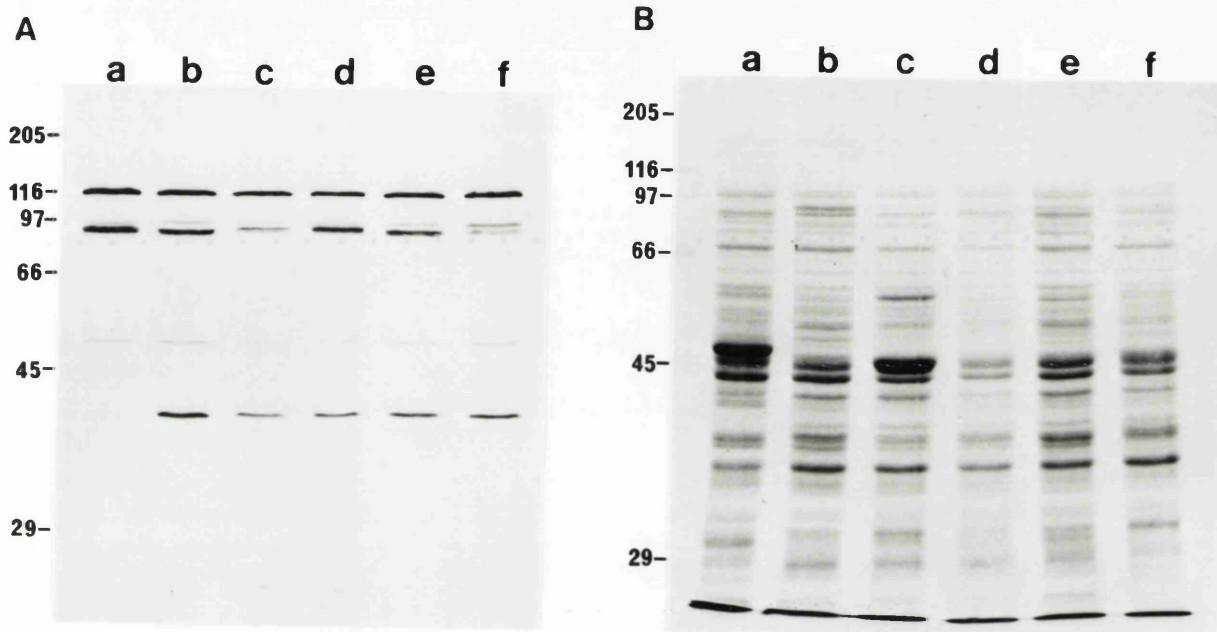


Figure 3.29 Immunoblot with MaD1 (A) and Coomassie stained gel (B) of *E.coli* W3110 grown in different media. a, rich medium (LB); b, minimal medium (M9 salts); c, M9 plus 0.4% acetate; d, M9 plus 500 $\mu\text{g/ml}$ valine; e, M9 plus 6% methanol; f, M9 plus 30 μM then 5 mM H_2O_2 . b, d, e & f contain 0.4% glucose. Acetate is a poor carbon source (Nunn, 1987), valine and methanol induce the stringent response (Cashel and Rudd, 1987) and H_2O_2 produces oxidative damage (Dempfle and Halbrook, 1983).

Introduction

This chapter begins with the demonstration using a polyclonal antibody that the p68 cDNA cloned by Ford, et al. (1988) encodes the p68 protein identified by Lane and Hoeffler (1980). After presenting some interesting data on one of the key factors limiting expression of p68 in *E.coli* the chapter ends with a description of the production of a monoclonal antibody against a short peptide spanning the region around the p68 DEAD motif.

The p68 cDNA clones encode authentic p68

To confirm that the human p68 cDNA clone (Ford, et al., 1988) is authentic p68 rather than an irrelevant cross-reacting protein, antisera from mice immunised with a p68/ β -galactosidase fusion protein were used to probe immunoblots of authentic p68 immunoprecipitated from 293 cells (figure 4.1 a). Because the fusion protein contains the PAb204 epitope, it is formally possible that the serum is directed only against that epitope. This possibility can be ruled out because the serum does not recognise SV40 large T antigen (figure 4.1 b). The PAb204 epitope was mapped by Ford, et al. (1988) to the sequence NTFRDRENYDRGYSSLLK (figure 1.1). Comparison with the epitope in T antigen (PLDRLNFELGVAIDQFL) suggests that the common residues are DRxNxxxG. Figure 4.2 f lane 4 shows that a peptide with the sequence FRDRENYDRG-C is indeed capable of specifically blocking antibody binding.

Preparation of new anti-p68 antibodies

PAb204 does not inhibit p68's ATPase activity; it is also unsuitable for some ELISA assays (eg, to detect p53-p68 complexes [Gannon and Lane, 1987]). It would, therefore, be desirable to have other anti-p68 antibodies. Although sensitive, the mouse and rabbit sera produced by Martin Ford (MF6 and MF7) are not specific. They detect multiple bands on immunoblots of whole cell extract. Production of specific antisera has been hampered by the uniformly low levels of expression of p68 achieved in *E.coli* using a variety of expression systems.

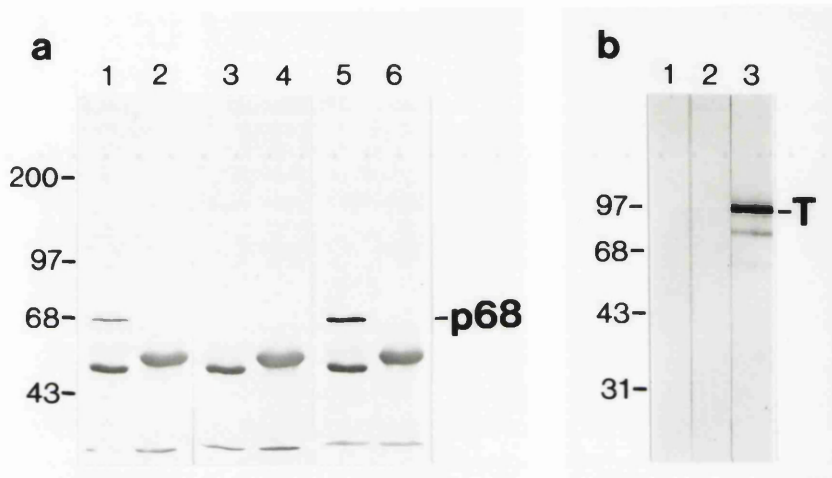


Figure 4.1 a: Immunoblot of immunoprecipitates from 293 cells probed with MF7 anti-p68 serum (lanes 1 and 2), BG2 plus normal mouse serum (lanes 3 and 4) or PAb 204 (lanes 5 and 6). Lanes 1, 3 and 5 are PAb204 immunoprecipitates. Lanes 2, 4 and 6 are BG2 immunoprecipitates. IgG heavy and light chains are visible below the p68 band.
 b: Immunoblot of SV40 large T antigen probed with MF7 (lane 1), BG2 plus normal mouse serum (lane 2) or PAb 204 (lane 3). Markers in kD.

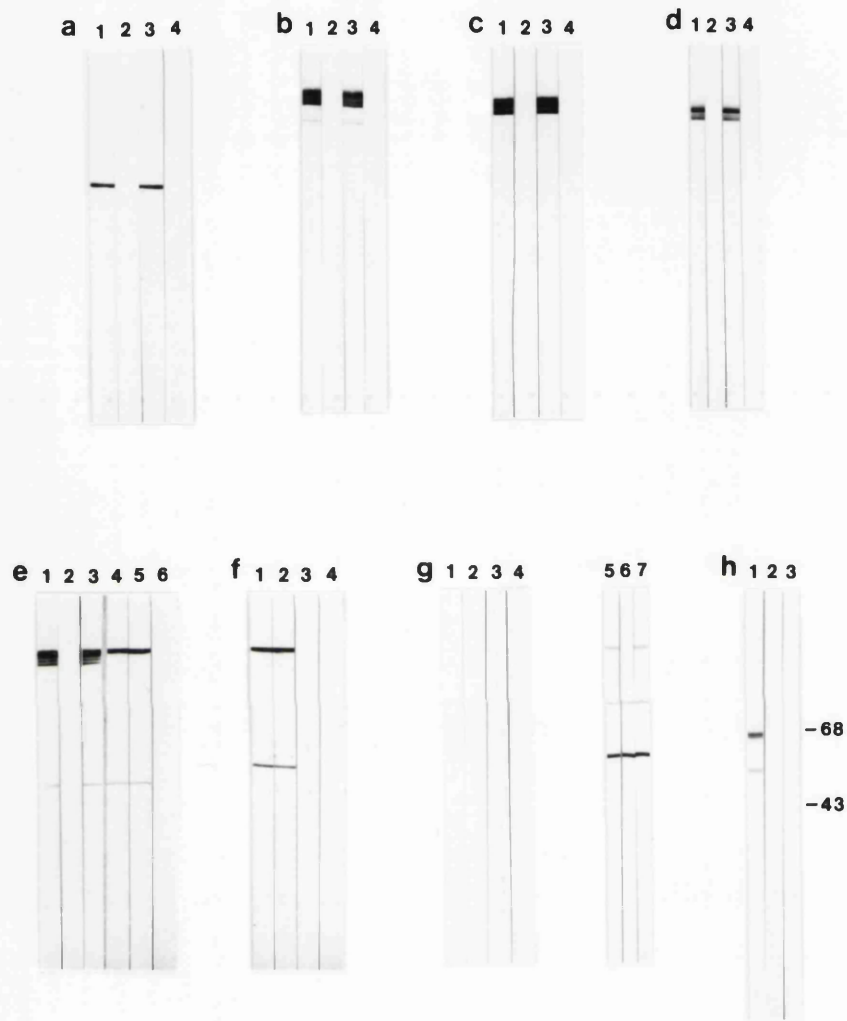


Figure 4.2 Immunoblots: a, *srmB*; b, *DED1*; c, *DBP1*; d, *dbp2*; e & f, p68; g, eIF4A; h, *S. pombe*.
a-e & g: lanes 1, 2 & 3, MaD1; lane 4, PAb204; lane 2, DEAD peptide; lane 3, PAb204/p68 peptide.
e: lanes 4-6, PAb204; lane 5, DEAD peptide; lane 6, PAb204/p68 peptide.
f: lanes 1-4, PAb204; lane 2, DEAD peptide; lane 3, PAb204/p68 peptide; lane 4, truncated p68 peptide.
g: lanes 5-7, R4.4 mouse serum; lane 6, DEAD peptide; lane 7, PAb204/p68 peptide.
h: *S. pombe* whole cell extract; lanes 1 & 2, MaD1; lane 3, PAb204; lane 2, DEAD peptide.
a-g: *E. coli* whole cell extract containing overexpressed DEAD box proteins.
b-g: β -galactosidase fusion proteins. Markers in kD.

Sequence analysis reveals that p68 contains multiple AGG and AGA arginine codons. The *argU* gene encodes the tRNA recognising these codons in *E.coli*. tRNA^{*argU*} is expressed at low levels, but its concentration is not limiting because AGG/AGA codons are not used in abundantly expressed *E.coli* proteins. *argU* was formerly called *dnaY* because a temperature sensitive *argU* mutant produces a slow stop to DNA replication at the non-permissive temperature, probably because the *argU* tRNA is required for translation of DNA replication proteins rich in AGG/AGA codons such as DnaA, DnaG and DnaN (Garcia, et al., 1986). Brinkmann, et al. (1989) have shown that expression of proteins containing more than 3% AGA/AGG codons is greatly increased by concomitant overexpression of *argU*. p68 contains 5.5% AGG/AGA codons. Figure 4.3 shows that p68 expression is dramatically increased by the transformation of an *argU* overexpressing plasmid into cells containing a p68 overexpression plasmid. The upper band in lane b is full length p68, the other bands are degradation products. A trace of p68 is detectable in the cells lacking the *argU* plasmid, but only after prolonged development of the blot. Cell fractionation indicates that p68 is one of the predominant proteins in the bacterial pellet, strongly suggesting that it is sequestered in inclusion bodies. Although a great improvement, this level of expression is not sufficient for trivial purification of wild type p68, and the project was not pursued.

Instead of bacterially expressed p68, a peptide with the sequence LVLDEADRMLDMGFEP was used to immunise mice for the production of monoclonal antibodies. The peptide was initially conjugated to ovalbumin. Although the mice responded well to the peptide in this form they mounted a very strong immune response to the ovalbumin. No anti-DEAD hybridomas were obtained from a fusion of SP2 cells with spleen cells from a mouse which had been immunised solely with the ovalbumin-peptide conjugate. To increase the chance that hybridomas would secrete anti-DEAD rather than anti-carrier antibodies, the Bacillus Calmette-Guerin/tuberculin Purified Protein Derivative (BCG/PPD) system was used (Lachmann, et al., 1986). Although BCG induces a good T cell response, subsequent challenge with

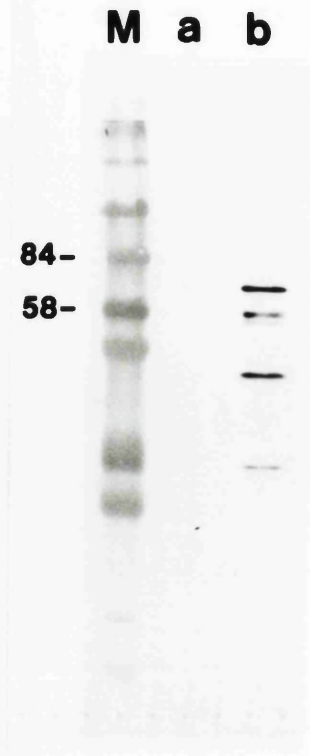


Figure 4.3 Immunoblot of bacterial whole cell extract probed with PAb204. a, *E.coli* BL21/DE3 cells containing two plasmids: a pACYC184 derived vector expressing T7 lysozyme (an inhibitor of T7 RNA polymerase), and a pT7.7 derived vector expressing full length p68 from a T7 promoter. b, as a, but the cells also contain a pACYC177 derived vector expressing *argU*. The cells contain a chromosomal copy of the T7 RNA polymerase under lac repression, and were induced with IPTG for two hours. M, prestained molecular weight markers in kD.

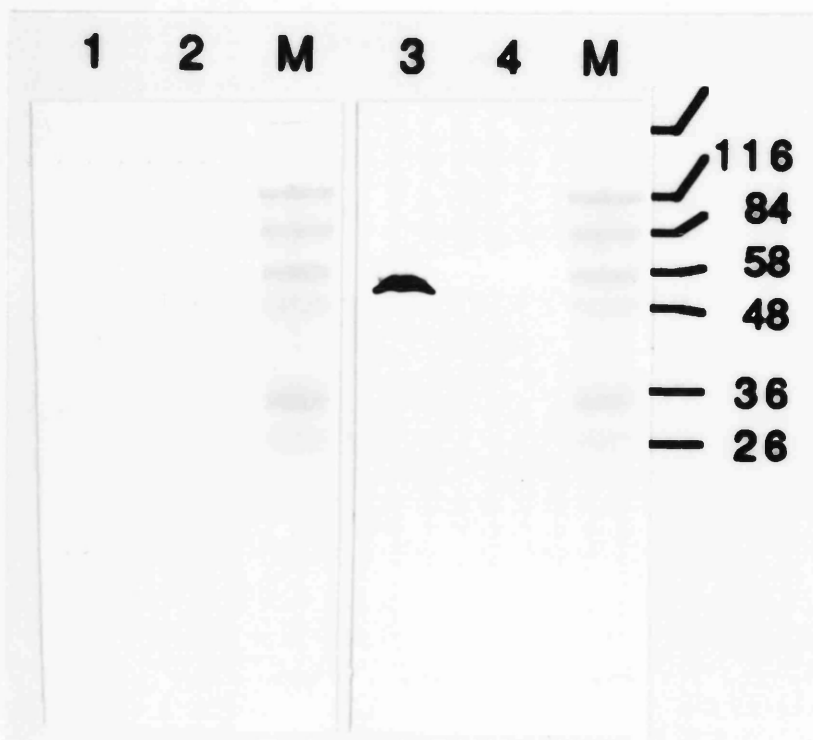


Figure 4.4 Immunoblot of bacteria overexpressing DbpA (lanes 1 and 3) or containing a control plasmid (lanes 2 and 4) probed with MaD1, plus (lanes 1 and 2) or minus (lanes 3 and 4) DEAD peptide. M, prestained molecular weight markers in kD.

PPD fails to elicit an anti-PPD antibody response. Mice which were already immune to the ovalbumin-peptide conjugate were vaccinated with BCG and then given an injection of DEAD peptide conjugated to PPD. Hybridomas from a mouse immunised in this way were screened by ELISA on plates coated with peptide conjugated to albumin. Positives were tested for peptide blocking in the ELISA and then by immunoblotting. Only one well was positive by these criteria. Cells from this well were single cell cloned by limiting dilution until all wells were positive. The cloned hybridoma, which was called MaD1 (Monoclonal anti-DEAD antibody 1), secretes a balb/c SP2 IgG2 κ mouse monoclonal antibody. The antibody is specific for members of the family whose DEAD box resembles that of p68 (figures 4.2 and 4.4). That MaD1 should recognise *DED1*, *DBP1* and *dbp2* is not surprising since they all have the same sequence as the peptide. There was no binding of MaD1 to eIF4A (figure 4.2 g lane 1). This is a strong negative result: there was a visible band of eIF4A by Coomassie and Ponceau S staining. The plasmid expressing the eIF4A DEAD motif was checked as follows: the cloning sites *were* checked by DNA sequencing; the plasmid gave a β -galactosidase fusion protein of the correct size; and anti-DEAD mouse serum R4.4 reacted specifically with the fusion protein (the upper band in figure 4.2 g lane 5, which blocks with DEAD but not irrelevant peptide, lanes 6 & 7).

MaD1 binds to a cryptic epitope

Although MaD1 works well on immunoblots, it does not react with native protein. This was tested in three ways. The antibody is unable to immunoprecipitate p68 (figure 4.5). It fails to inhibit the ATPase activity of immunoprecipitated p68 in the assay described in chapter six, even after preincubation of the p68 immunoprecipitates with a 1000-fold excess of antibody for 30 minutes. Finally, no staining of tissue culture cells is seen unless antibody is used at very high concentrations (~100 μ g/ml) which produce significant background staining. The staining pattern resembles that seen with PAb204 in that prenucleolar bodies are visible (see chapter five). It is possible that harsh fixation conditions would expose the epitope but no conditions

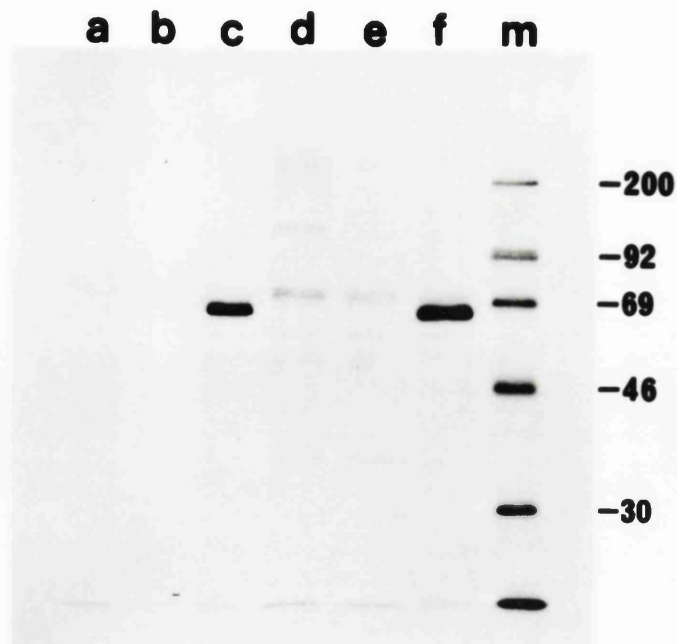


Figure 4.5 Fluorograph of ^{35}S -methionine labelled proteins immunoprecipitated with PAb204 (lanes c and f), MaD1 (lanes b and e) and PAb421 (lanes a and d). HeLa cells were labelled for 17 hours with 50 $\mu\text{Ci/ml}$ ^{35}S -methionine; lysed in 0.5 M NaCl, 50 mM Tris pH 8.0, 5 mM EDTA, 1% NP40, 1 mM PMSF and spun at 100,000 x g. The supernatant was used directly (lanes d-f) or diluted to 100 mM NaCl and preabsorbed with protein G beads (lanes a-c) before addition of antibody. PAb421 (anti-p53) is a negative control. ^{14}C markers in kD.

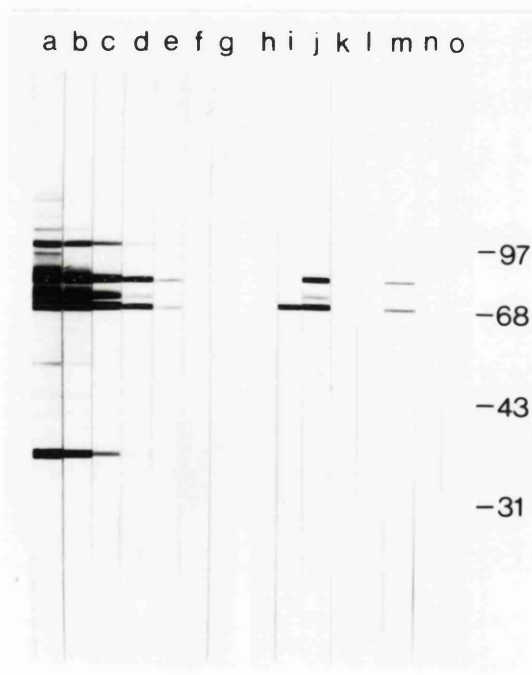


Figure 4.6 Immunoblot of HeLa whole cell extract probed with PAb421 (lane h), PAb204 (lane i) and MaD1 (all other lanes). Lanes a-g show ten-fold dilutions of MaD1 starting at 10 $\mu\text{g/ml}$. Lanes j-l and m-o show peptide blocking: j-l, 100ng/ml MaD1; m-o, 10 ng/ml MaD1; j&m, no peptide; k&n, 20 $\mu\text{g/ml}$ DEAD peptide; l&o, 200 $\mu\text{g/ml}$ DEAD peptide. Markers in kD.

tested (acetone/methanol, methanol, formaldehyde or acetic acid/ethanol) gave good results.

MaD1 can detect p68 in whole cell extract

Figure 4.6 shows that MaD1 recognises a 68 kD protein in whole cell extract. Although this protein has not been formally shown to be p68 in HeLa cells, for example by immunodepletion with PAb204, the corresponding band is absent from immunoblots of *S.cerevisiae* after deletion of the *DBP2* gene (figure 7.6). Antibody concentration has an important effect on the number of bands seen. At 10-100 ng/ml of MaD1 two bands are visible in all species tested (*E.coli* figure 3.29; *S.pombe* figure 4.2 h; *S.cerevisiae* figure 7.4; *Xenopus* figure 4.7; *H.sapiens* figure 4.6 j & m). The most reasonable interpretation is that the upper band is a member of the *DED1*, *DBP1*, An3, PL10 family and the lower band is a member of the *DBP2*, *dbp2*, RM62, p68 family. Other bands appear if higher concentrations of MaD1 are used or after prolonged development of the blot. The significance of these other bands is unknown, although they all block specifically with peptide. In HeLa cells the main extra bands have relative molecular weights of ~100, 75 and 36 kD (figure 4.6).

***dbp2* lacks the PAb204 epitope**

Sequence analysis suggests that the PAb204 epitope is not present in *S.pombe dbp2*. Figure 4.2 d lane 4 shows that PAb204 does not bind to a β -galactosidase/*dbp2* fusion protein. Immunoblots of *S.pombe* whole cell extract (figure 4.2 h lane 3) strongly suggest that PAb204 does not recognise any p68-like molecules in *S.pombe*. By contrast, MaD1 readily detects a ~60 kD band which could be *dbp2* (the lower band in figure 4.2 h lane 1).

Discussion

MaD1 recognises proteins whose DEAD motifs share the following residues: LxxDEADRMLDMGF. It does not bind to eIF4A, which has the sequence FVLDEADEMLSRGFKD. Specificity is apparently conferred by LxxxxxxRxxDM (ie, residues required for binding but different in eIF4A). The recruitment of new bands as antibody concentration rises is provocative: either it

indicates that the cell contains multiple uncharacterised proteins with p68-like DEAD motifs or it means that antibody specificity is lost. Neither result would be surprising. To detect p68 cleanly in whole cell extract the antibody should be used at 10-100 ng/ml. For specific detection of bacterially overexpressed protein when the band is visible by Ponceau S staining 1- 10 ng/ml of antibody is sufficient. Satisfactory peptide blocking of MaD1 used at 10-100 ng/ml on whole cell extract can be achieved with 100 μ g/ml DEAD peptide, which is equivalent to a \sim 10,000 fold molar excess of peptide over antibody.

The DEAD motif is predicted (by analogy with the structures described by Walker et al. [1982]) to lie on the surface of the molecule at a turn between a beta sheet and an alpha helix . The peptide LVLDEADRMLDMGFEP was chosen with this in mind in the hope that antibodies to it would recognise the protein in the native conformation. Furthermore, it was anticipated that these antibodies would inhibit the ATPase, because the DEAD motif is part of the ATP binding site. These hopes were not realised. Only one antibody was obtained from a total of three fusions. That antibody is typical of anti-peptide monoclonals in that it works well on immunoblots but does not recognise protein in the native state: it does not immunoprecipitate p68; it does not cell stain; and it does not inhibit p68's ATPase activity. Nevertheless, the antibody is a convenient tool for checking constructs expressing a wide range of different DEAD proteins and it can be used to ask rather general questions about DEAD protein expression. For example, it was very easy to look for the induction of *E.coli* DEAD proteins predicted by David Shub (see figure 3.29) or the appearance of An3 following fertilisation of *Xenopus* eggs (figure 4.7). The *Xenopus* experiment tests the hypothesis that the burst of helicase activity which follows fertilisation (Fu, et al., 1991) results from translation of stored An3 message (Gururajan, et al., 1991).

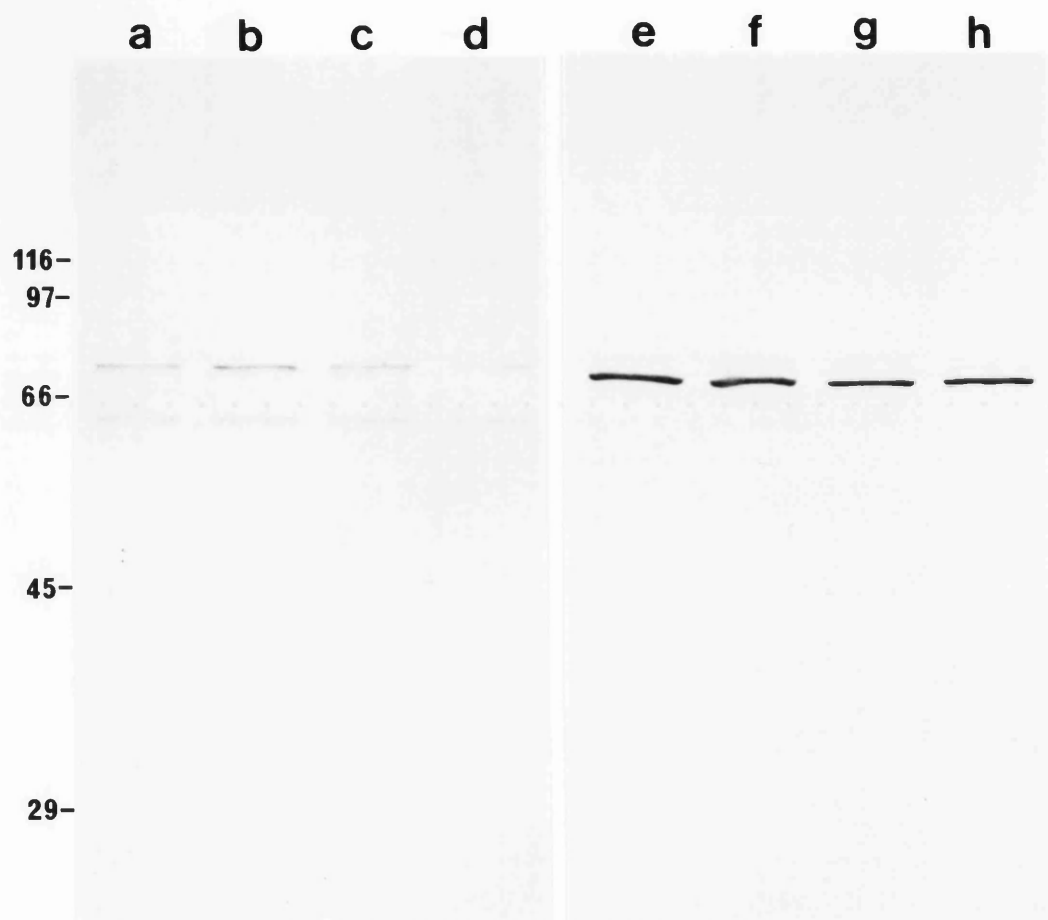


Figure 4.7 Immunoblot of *Xenopus* extract probed with MaD1 (a - d) and the anti-hsp 70 monoclonal antibody 7.10 (e - h). a & e, unfertilised; b - d and f - h, activated extract. b & f, time 0; c & g, 60 minutes; d & h, 100 minutes. The upper band in a - d could be An3. Markers in kD. 7.10 was included as a loading control.

Introduction

Two striking results are presented in this chapter: first, that p68 transiently enters the nucleolus at the end of mitosis; second, that anti-p68 antibody injected into the cytoplasm is specifically transported to the nucleoli after mitosis. The chapter ends with a discussion of these results mainly as problems of cellular localisation, although the possibility that p68 is required for an RNA processing event specific to the nucleolus in telophase is also considered.

Cell cycle variation in p68 distribution.

Immunohistochemical examination of mammalian tissue culture cells with PAb204 reveals a weak granular staining pattern which is localised exclusively in the cell nucleus but spares the nucleoli (figure 5.1). In mitosis PAb204 does not stain chromosomes on the metaphase plate although weak perichromosomal staining is observed. The most striking phenomenon is the appearance of a few intensely staining bodies in the nuclei of cells in telophase. Double label immunofluorescent cell staining with anti-p68 and anti-nucleolar antibodies (figure 5.2) shows that these structures are prenucleolar bodies. The prenucleolar staining is unlikely to be a simple fixation artefact because it is seen whether the cells are fixed in acetone/methanol, methanol alone or buffered formaldehyde. Nucleophosmin is an abundant protein in the granular component of the interphase nucleolus (Spector, et al., 1984). Although detectable in prenucleolar bodies by electron microscopy (Ochs, et al., 1985), it is not seen by light microscopy until early G1 (Ochs, et al., 1983). Double labelling of acetone/methanol fixed cells shows that there is very little overlap between the p68 and nucleophosmin staining patterns: nucleolar PAb204 staining is at its most intense when the majority of nucleophosmin is in the cytoplasm (figure 5.2 c & d).

The proliferating cell nuclear antigen (PCNA) gives transient S-phase nucleolar staining (Bravo, et al., 1981) apparently due to masking of the dominant epitope on PCNA in

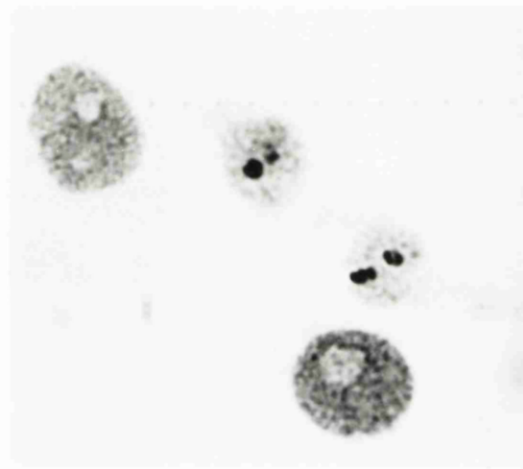
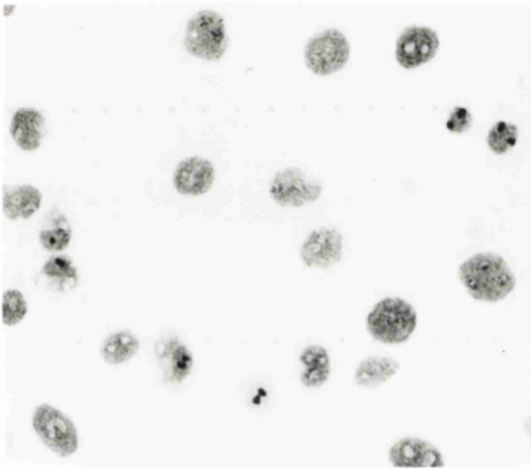
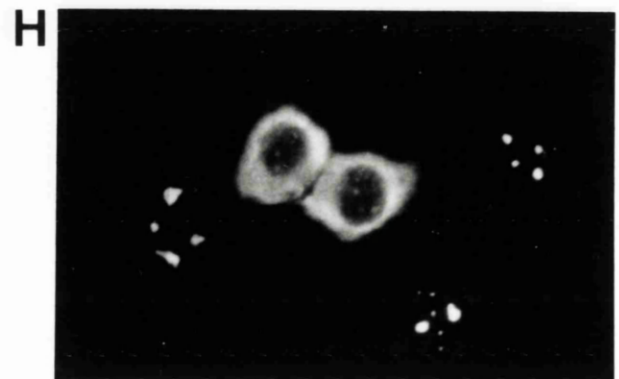
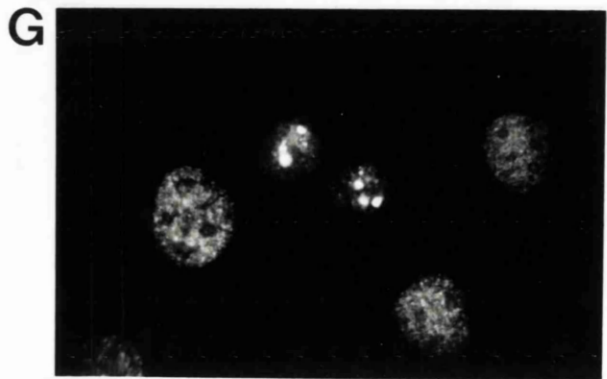
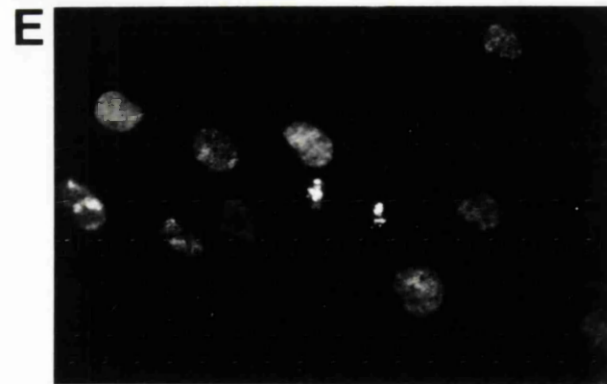
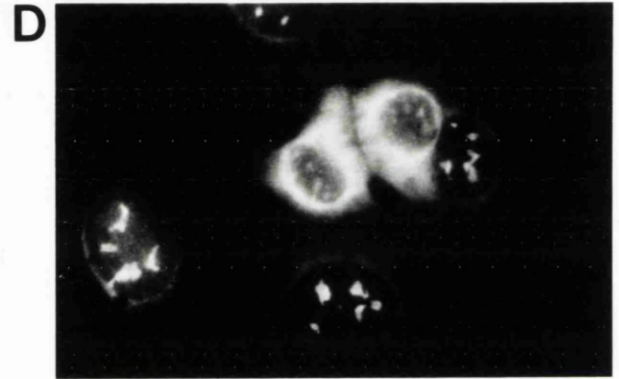
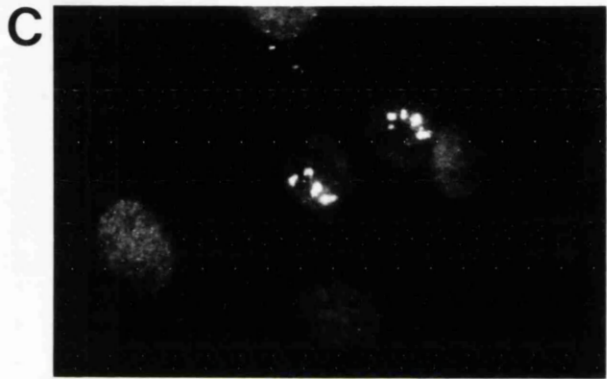
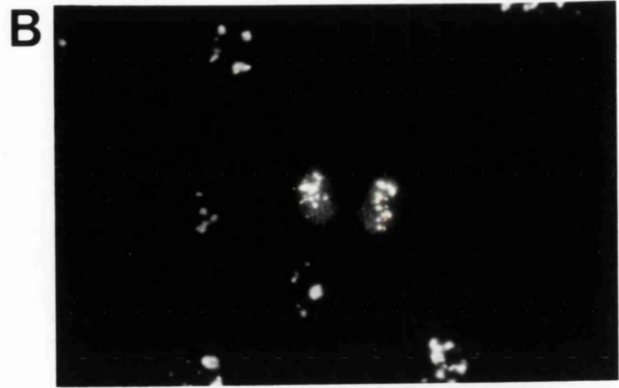
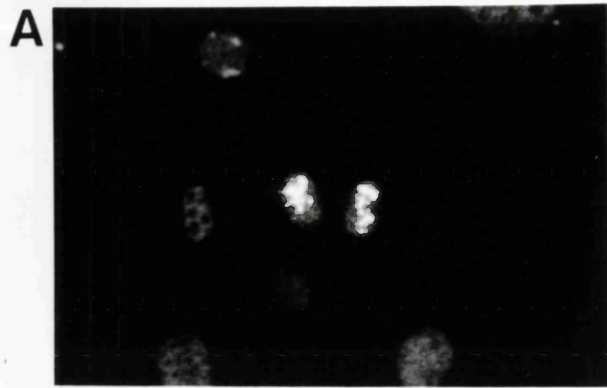


Figure 5.1 Immunoperoxidase staining of p68. CV1 cells fixed in acetone/methanol were incubated sequentially in PAb204, rabbit anti-mouse immunoglobulins conjugated to horseradish peroxidase and diaminobenzidine.

Figure 5.2 Double-label immunofluorescence showing anti-p68 (a,c,e,g) and anti-nucleolar (b,d,f,h) staining of C6 cells. a&c, PAb204; e&g, MF6; b&f, ANA-N; d&h, anti-nucleophosmin. a,b,e&f 375x; c,d,g&h 590x.



the nucleolus during the rest of the cell cycle (Waseem and Lane, 1990). To establish whether the cell cycle variation in p68 staining reflects epitope masking or bulk movement of p68 the staining was repeated with affinity-purified anti-p68 serum (MF6) depleted of antibodies to the PAb204 epitope. Figure 5.2 e-h shows that affinity purified MF6 gives prenucleolar body staining which is identical to that seen with PAb204, suggesting that the cell cycle changes are not simply a consequence of epitope masking.

Treatment of cells with actinomycin D has been reported to prevent fusion of prenucleolar bodies as cells emerge from mitosis (Ochs, et al., 1985). However, it has also been reported to produce "translocation" of nucleophosmin from the nucleoli to the nucleoplasm (Yung, et al., 1985). Incubation of C6 cells with 10 µg/ml of actinomycin D resulted in the progressive disappearance of nucleolar nucleophosmin staining. Initially there was crescentic or cup like staining of the nucleoli as described by Ochs, et al. (1985), but this was subsequently replaced by a diffuse nucleoplasmic stain. At the same time the p68 staining pattern became more coarsely granular until the majority of p68 was clustered in 5 to 10 brightly stained spots (figure 5.3). However, these spots neither colocalised with ANA-N nor with monoclonal anti-Sm antibody (Y12, [Lerner, et al., 1981]).

Microinjection of PAb204

The strong conservation of the PAb204 epitope, both across species and between p68 and SV40 T antigen, suggests that it may be part of a protein or RNA binding site (Iggo and Lane, 1989). According to this model, microinjection of PAb204 should significantly interfere with p68's normal function. In practice 3T3, HFF, C6 and CV1 cells all continue to grow and divide apparently normally following injection (figures 5.4 and 5.5 a & c). The antibody was injected into the cytoplasm but should have gained free access to nuclear p68 molecules following nuclear envelope breakdown at mitosis. Even in post-mitotic cells antibody never accumulated to high levels in the nucleoplasm, although p68 could still be detected there by

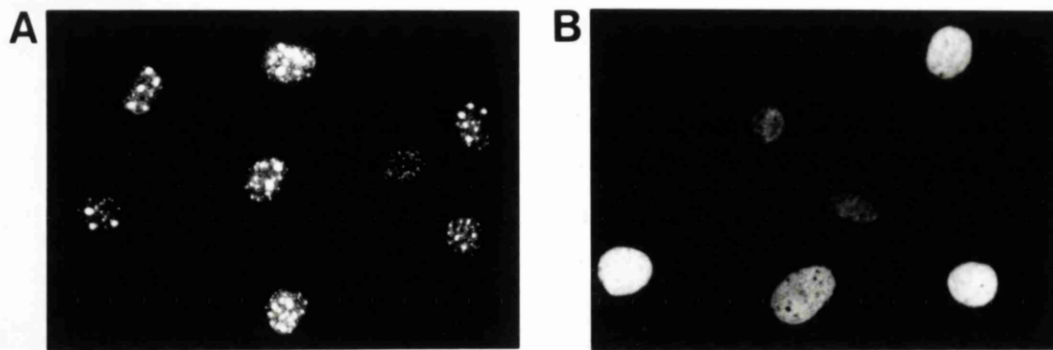


Figure 5.3 C6 cells treated with actinomycin D at 10 $\mu\text{g/ml}$ for 12 hours and stained with PAb204 (A) and anti-nucleophosmin (B) antibodies.

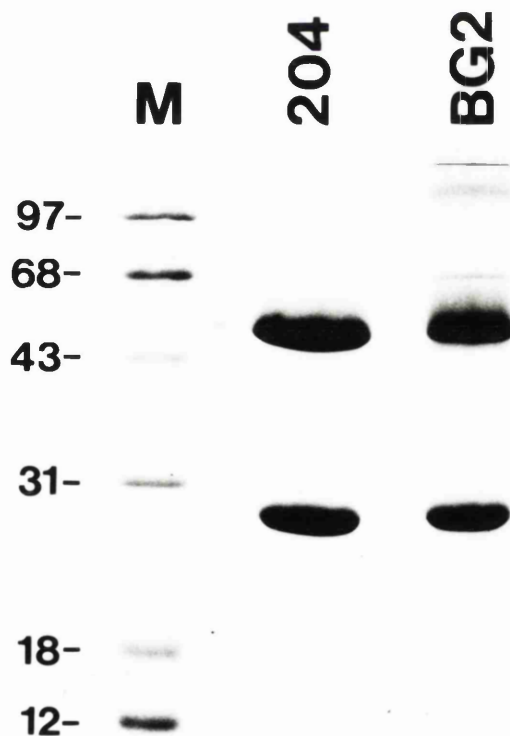
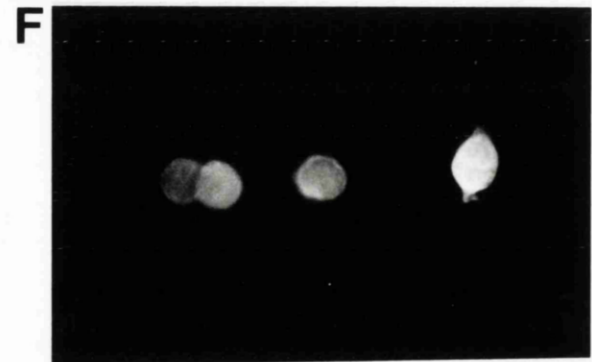
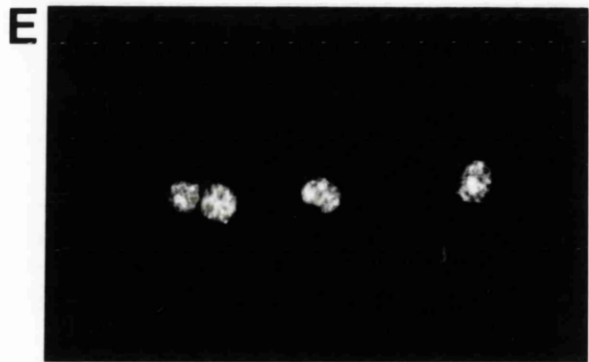
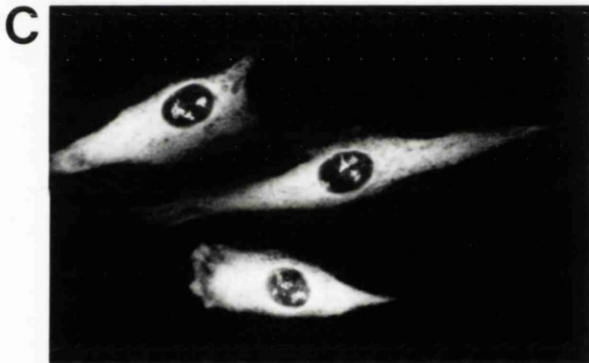
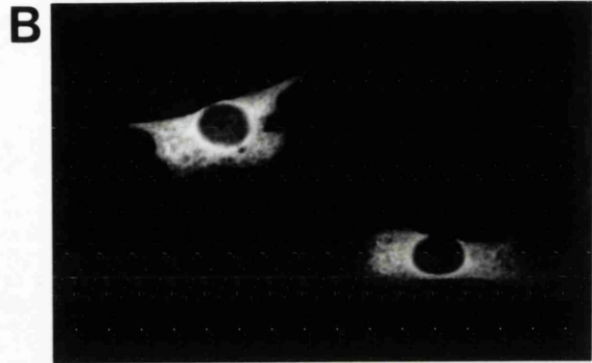
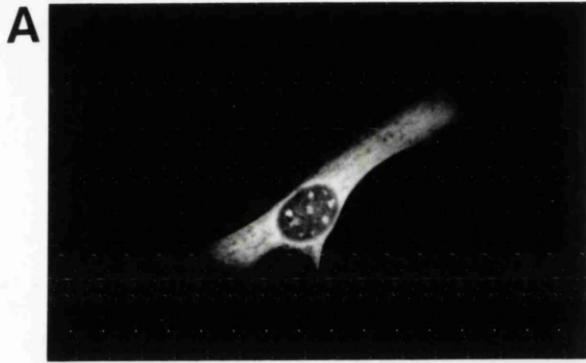


Figure 5.4 Coomassie blue stained SDS polyacrylamide gel showing the antibodies used for the microinjection experiments. PAb204 was purified from tissue culture supernatant, BG2 from monoclonal ascites.

Figure 5.5 Microinjection of PAb204 (a, c, d) and BG2 (b) into 3T3 (a & b) and C6 cells (c & d) fixed after 24 hours and stained with rhodamine-conjugated anti-mouse immunoglobulin (a, b, c) and ANA-N (d, same field as in c). e & f, 293 cells transfected with a p68 overexpression construct: e, PAb204; f, CAT2 (a marker for transfected cells, same field as in e). a, b, e & f 330x; c & d 375x.



double labelling. There are many reasons why microinjected antibody might have no obvious effect: PAb204 may not bind to an essential site (it does not inhibit *in vitro* ATPase [Iggo and Lane, 1989] or helicase reactions [Hirling, et al., 1989]); it may require many rounds of cell division to deplete the cell of p68's target; or the epitope may be masked on the majority of p68 molecules in living cells.

In the three fibroblast cell lines PAb204 was transported to the nucleoli following injection into the cytoplasm (figure 5.5 c & d). This was not an artefact because it was also seen in living cells injected with directly labelled antibody and was not seen with control antibodies (figure 5.5 b). The nucleolar staining was only seen in cells which had divided but, unlike the prenucleolar body staining described above, the PAb204 staining was not restricted to cells in telophase. It was never seen in C6 cells injected at 32°, ie in cells arrested in G1. If the cells were shifted to 37°, PAb204 appeared in the nucleoli immediately after the cells divided. Provided C6 cells are plated at low density at 32° they continue to increase in cell size and contain normal levels of p68 (judged by immunoblotting). Thus the appearance of PAb204 in the nucleoli after temperature shift is unlikely to reflect a sudden burst of p68 synthesis. One possible explanation for the nucleolar transport of PAb204 is that p68 only enters the nucleus at mitosis. This is unlikely because transfected p68 expressed from an exogenous promoter readily enters the nucleus, with no cytoplasmic accumulation in interphase cells (fig 5.5 e & f).

Discussion

When the nucleolus reforms at the end of mitosis the first structures to appear are prenucleolar bodies, which contain mainly dense fibrillar components (Warner, 1990). These components are classically associated with ribosomal RNA processing (Warner, 1990). The delayed maturation of 25S ribosomal RNA in *spb4-1* mutants provides a precedent for the involvement of DEAD proteins in ribosome assembly. *srmB* is another DEAD protein linked with ribosome assembly: when overexpressed it suppresses a temperature sensitive defect in

50S subunit assembly caused by a mutation in ribosomal protein L24. The role of RNA helicases in this situation could be considered analogous to that of chaperonins or foldases in protein folding. The function of small nucleolar RNAs is unknown but by analogy with U1, U2 and U5 it may involve targeting of proteins to specific sites in the ribosome. An obvious function for RNA helicases would then be to remove the small nucleolar RNAs before export of the mature ribosome. Despite the presence of p68 in prenucleolar bodies it is difficult to envisage p68 being directly involved in ribosomal RNA processing because anti-p68 antibodies do not stain the nucleoli of interphase cells. It is conceivable that p68 is present in interphase nucleoli in an immunologically unreactive form, but the failure to detect it with the polyclonal antibody would be surprising. One is left with the challenge of trying to identify an RNA processing event specific to the nucleolus in telophase, a daunting task.

Although p68 could be required for nucleolar assembly, it seems equally plausible that the staining in telophase may reflect a role for the nucleolus in nuclear transport. This possibility is made more attractive by the microinjection data. The migration of antibody from the cytoplasm to the nucleoli is intriguing. Several models could account for it: nucleolar shuttling (Borer, et al., 1989) is unlikely given the requirement for cell division; and mitotic loading (Blow and Laskey, 1988) is unlikely given the failure of p68 to accumulate during interphase in the cytoplasm of transfected cells. Nucleolar transport signals (Siomi, et al., 1988; Dang and Lee, 1989) resemble extended but otherwise unremarkable nuclear transport signals. There seems no *a priori* reason why nucleolar targeting could not be achieved by entirely separate nuclear transport and nucleolar binding signals. Hence, the fact that the two are so tightly coupled suggests that the nucleolus may play a greater role in nuclear transport than has previously been suspected. Neither the signals which direct ribosomes to the cytoplasm, nor the mechanism by which this is achieved are understood. The sheer volume of transport out of the nucleus may have been underestimated. Quite apart from the traffic in messenger RNA, transfer RNA and ribosomes, there must be a

mechanism to rid the nucleus of cytoplasmic components after mitosis as well as a housekeeping function which eliminates old or damaged nuclear components. Maintenance of cellular architecture presumably requires a dynamic interplay between signal sequences, anchors, transporters and barriers. Nuclear transport signals are readily identified by mutagenesis and domain swap experiments. Few anchoring proteins have been identified (HSP90, I- κ B, cactus; reviewed by Hunt [1989]). The proteins which can be crosslinked to nuclear transport peptides (Adam, et al., 1989; Yamasaki, et al., 1989) are candidate transporter molecules. Passage through membrane barriers is controlled by pore proteins, in the case of the nuclear membrane a complex multisubunit assembly (the "double iris" [Dingwall, 1990]). An attractive model to account for the role of the nucleolus in nuclear transport would be that nuclear transporters (molecules which collect proteins in the cytoplasm and ferry them into the nucleus) release their cargo in the nucleolus and are then recycled to the cytoplasm on ribosomes. If release is triggered by competition between the transporter and a "docking site" for the nuclear transport signal it is easy to see how a "strong" nuclear transport signal could result in retention of a protein in the nucleolus. The appearance of p68 in the nucleolus in telophase would then reflect a shift in the equilibrium between binding and release. The persistence of nucleolar staining in interphase in injected cells might indicate that the antibody prevented release of p68 from the nucleolus. That the antibody interferes with the function of the transport signal is suggested by the fact that more antibody does not accumulate in the nucleus of cells which have not divided. The nuclear transport signal in p68 has not been identified experimentally but a potential signal which fulfils Laskey's new bipartite criteria (Robbins, et al., 1991) is present between amino acids 32 and 45 (KK...KKK). This region is, however, remote from the PAb204 epitope. Finally, the possibility that the antibody itself has a specific transport signal should not be overlooked.

Introduction

The assays described in this chapter were carried out against a background of uncertainty about the nature of p68's target. The cross-reaction with T antigen suggested a DNA target, whereas the sequence similarity with eIF4A suggested an RNA target. The presence of the Hodgman/Gorbalenya motifs gave no clue because the motifs are present in both DNA and RNA helicases. The results presented here played an important part in establishing at a biochemical level that the DEAD box family targets RNA. The chapter includes a brief description of an unsuccessful attempt to identify the RNA with which p68 appears to copurify and ends with a discussion of other techniques which could be used to identify p68's RNA target.

p68 is an RNA-dependent ATPase

Hela cell p68 was tested for nucleic acid dependent ATPase activity. Since conventional purification is difficult, biochemical activities were tested in PAb204 immunoprecipitates. The PAb 204 binding site on p68 lies outside the region of homology with eIF4A and is distant from the ATPase and helicase motifs in the primary sequence (Ford, et al., 1988). Coomassie blue staining of Hela material run on a polyacrylamide gel reveals a single band at 68 kD which is absent from control immunoprecipitates (figure 6.1). PAb204, which is an IgG1, does not bind to protein A. For all immunoprecipitations described here PAb204 was captured on protein G beads. This represents a substantial improvement over the use of protein A with a bridging layer of rabbit anti-mouse immunoglobulins, which gives multiple non-specific bands. The tight binding of protein G to PAb204 and of PAb204 to p68 means that cells can conveniently be lysed and immunoprecipitations performed in 0.5 M NaCl. This decreases nonspecific binding to the beads and results in almost complete release of p68 from the cell pellet (as judged by immunoblotting). Figure 6.2 shows a p68 immunoprecipitation performed in high salt (other conditions were as in figure 6.1): there is almost equimolar binding of PAb204 to p68.

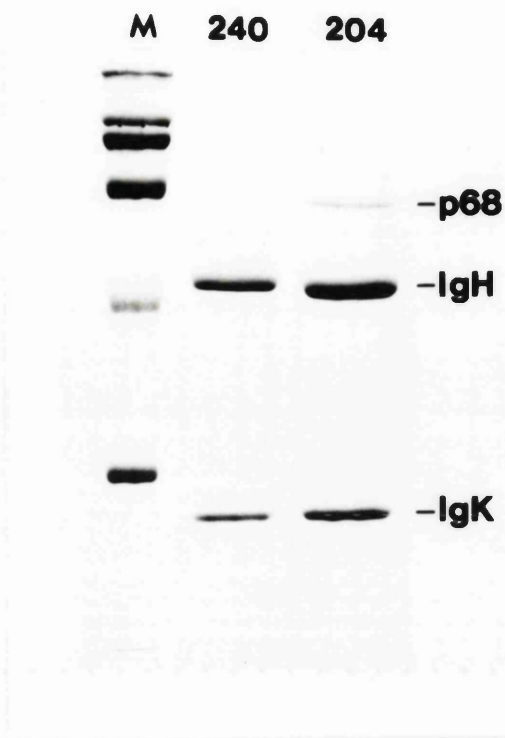


Figure 6.1 SDS polyacrylamide gel stained with Coomassie blue showing p68 immunoprecipitated by monoclonal antibody PAb 204 from HeLa cells lysed at 150 mM NaCl. IgH, immunoglobulin heavy chain; IgK, immunoglobulin kappa light chain; PAb 240, negative control antibody; M, markers in kD: 205, 116, 97, 66, 45 and 29.

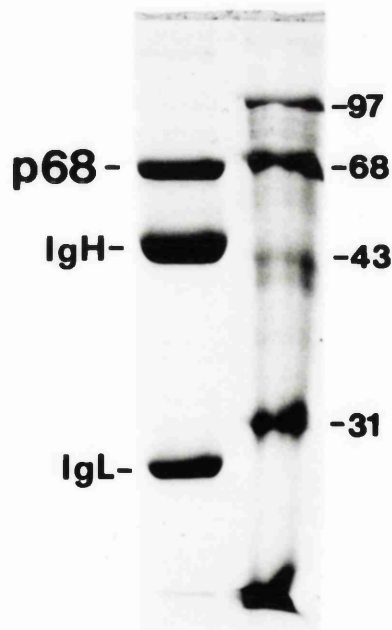


Figure 6.2 p68 immunoprecipitated by PAb204 as in figure 6.1, but the HeLa cells were lysed at 500 mM NaCl.

p68 immunoprecipitated from cells lysed at physiological salt concentration has substantial ATPase activity (measured as release of $^{32}\text{P}_i$ from γ - ^{32}P labelled ATP [Clark, et al., 1981]) which is stimulated by added RNA but not by DNA (figure 6.3). The nucleic acid substrates used for this experiment were Hela cell RNA and boiled sonicated salmon testis DNA (0.2 mg/ml) because these substrates should contain a wide spectrum of secondary structures and sequences. Pretreatment of immunoprecipitates with ribonuclease A (0.2 mg/ml) reduces ATPase activity to near control levels. This suggests that endogenous RNA copurifies with p68 when cells are lysed at 150 mM NaCl. Only RNA can restore activity to ribonuclease treated beads. DNA has no effect, even on deoxyribonuclease I (100 U/ml) treated p68.

Double reciprocal plots (figure 6.4) give a K_m value of 0.6 mM and a V_{\max} of 40 pmol/ $\mu\text{g/s}$ for ATP hydrolysis. This K_m is one order of magnitude higher than that of eIF4A (Grifo, et al., 1984), although still below the cellular ATP level (1-10 mM). The value for V_{\max} was calculated by estimating the amount of p68 by densitometry of a Coomassie stained gel. For technical reasons the first time point (figure 6.4, upper panel) was five minutes. V_0 for the lowest S is clearly an underestimate; this has the effect of increasing $1/V$ and thus K_m . The values for K_m and V_{\max} should be treated as rough estimates. Nevertheless, the K_m is rather high and p68 was therefore tested for GTPase activity (figure 6.5). GTPase activity was not detected with either RNA or DNA, consistent with the absence from p68 of the NKxD/W GTPase motif (Dever, et al., 1987).

Neither different species of cellular RNA nor polyadenylic acid gave significant differences in ATPase activity. The ATPase activity of eIF4A is reduced by secondary structure in the RNA template (Abramson, et al., 1987), which is one of the hallmarks of a helicase. p68 was therefore tested with double stranded ρ otavirus genomic RNA. The viral RNA was divided into equal aliquots, one of which was denatured using the dimethyl sulfoxide technique (which minimises RNA degradation, [McCrae and Joklik, 1978]). In its double stranded form ρ otavirus RNA is

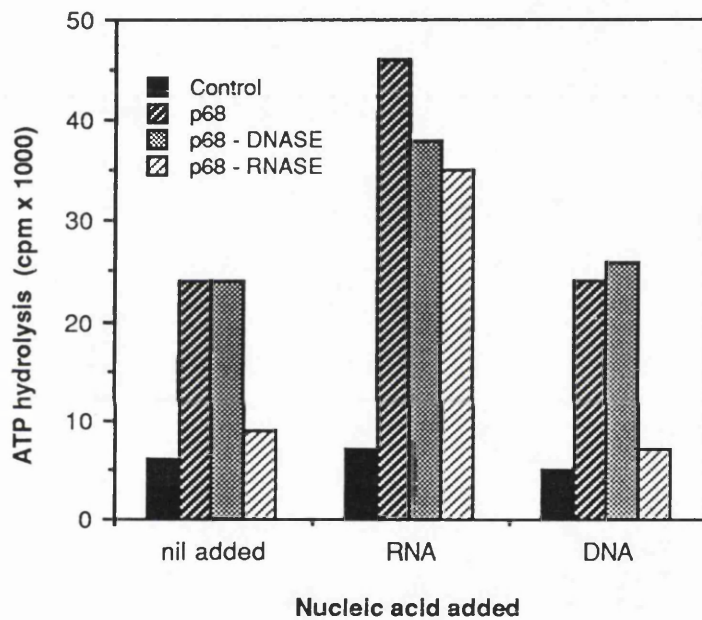


Figure 6.3 ATPase assays showing that p68 is stimulated by RNA but not DNA. Immunoprecipitates with PAb204 (p68) or PAb240 (control) were incubated with 50 μ M γ^{32} P labelled ATP. ATP hydrolysis was measured as release of free 32 P_i (10^3 cpm is approximately 1 fmol/s).

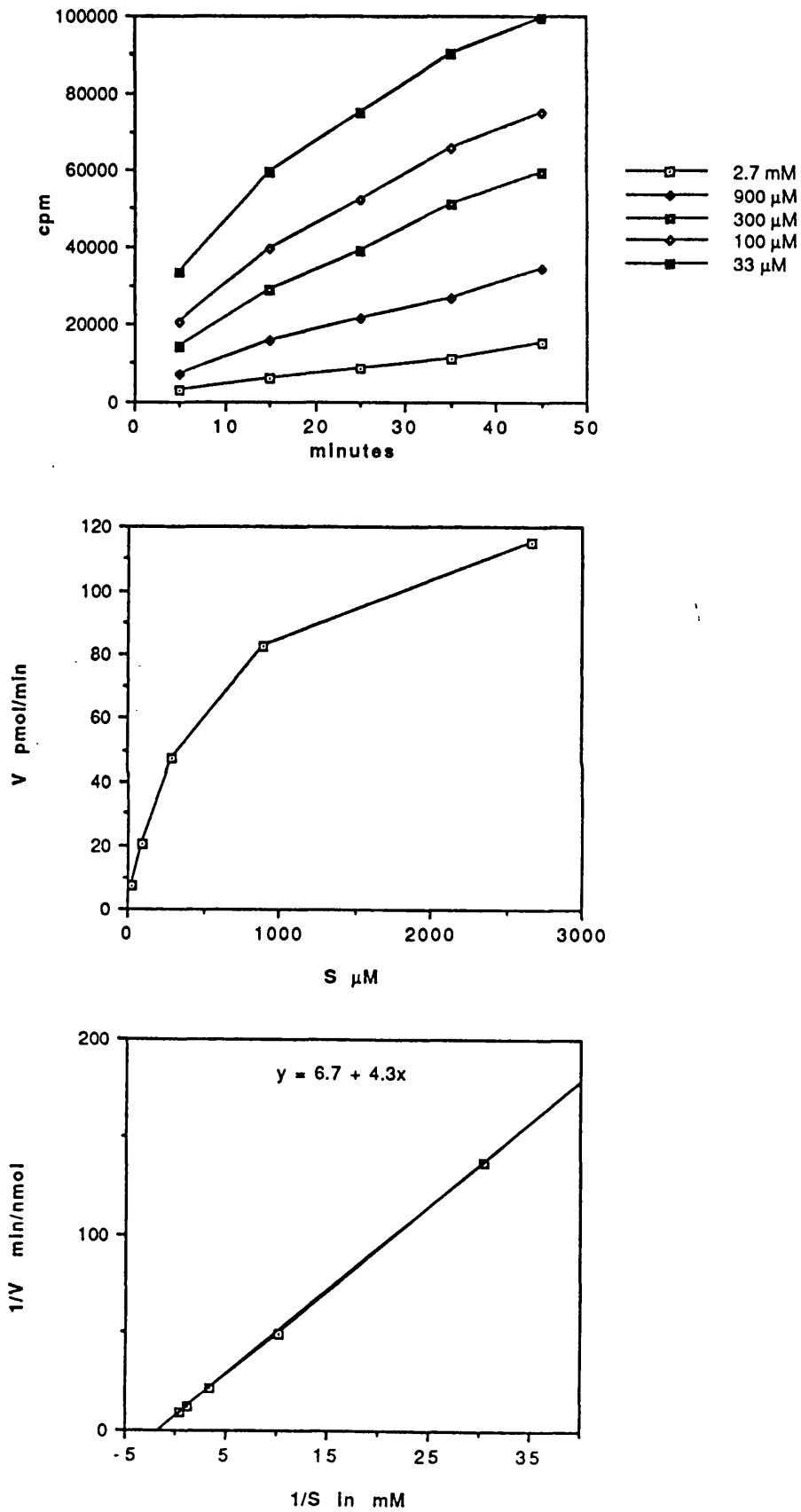


Figure 6.4 Kinetic analysis of p68's ATPase activity. a, crude data. b, plot of rate of ATP hydrolysis (V) against ATP concentration (S). c, double reciprocal plot.

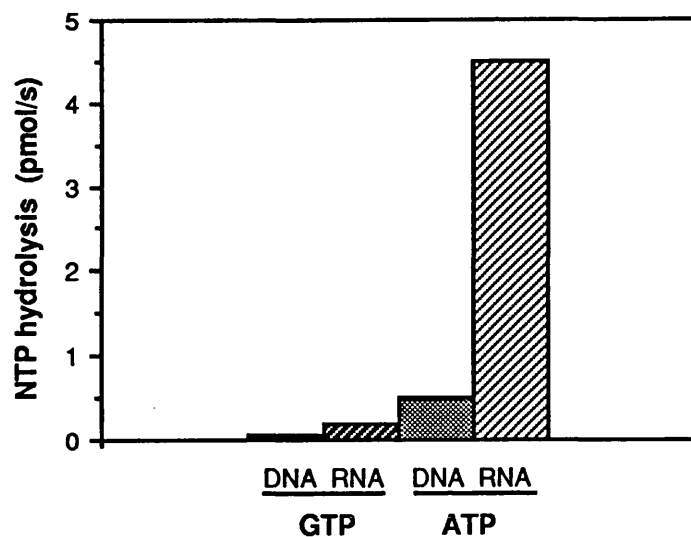


Figure 6.5 ATPase and GTPase assays. Immunoprecipitated p68 was incubated with 1 mM $\gamma^{32}\text{P}$ labelled ATP or GTP and $^{32}\text{P}_i$ release measured in the presence of 0.2 mg/ml Hela cell RNA or salmon testis DNA. The p68 was prepared from cells lysed in high salt and RNase A. Background counts due to spontaneous NTP hydrolysis have been deducted.

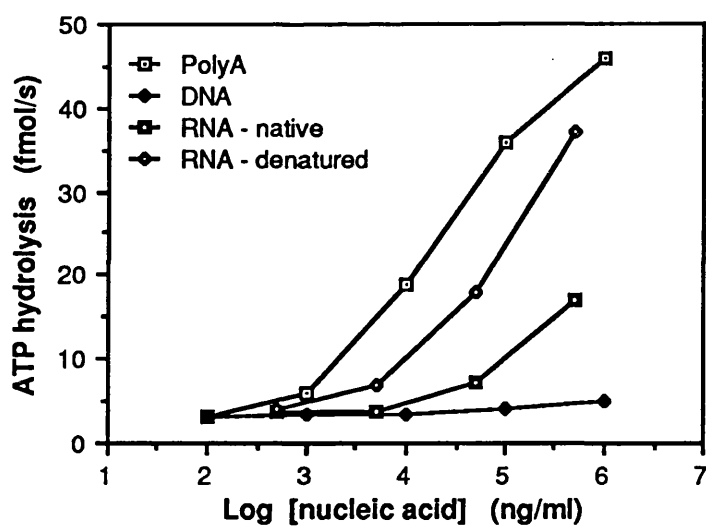


Figure 6.6 ATPase assays showing that p68 is stimulated more effectively by single stranded than double stranded RNA. Immunoprecipitated p68 was incubated with 10 μM ATP and $^{32}\text{P}_i$ release measured in the presence of increasing amounts of polyadenylic acid (PolyA), salmon testis DNA (DNA), native rotavirus genomic RNA (RNA-native) or DMSO denatured rotavirus genomic RNA (RNA-denatured). The immunoprecipitates were RNase treated to reduce endogenous activity.

less effective at stimulating ATP hydrolysis than it is after dimethyl sulfoxide treatment (figure 6.6).

small nuclear RNA does not coprecipitate with p68

One of the most important missing pieces of information about p68 is the nature of its target RNA. The conclusion from the biochemical analysis (figure 6.3) that p68 copurifies with cellular RNA gives some hope that a combination of biochemical and cloning techniques can be used to identify p68's target. Initial attempts to label the 3' end of the RNA with ³²P using T4 RNA ligase (pCp labelling, [England, et al., 1980]) were unsuccessful. p68 was immunoprecipitated from HeLa cells as above, using extracts made from cells lysed at 150 mM NaCl. As a negative control the anti-p53 antibody PAb240 was used. As a positive control affinity purified anti-trimethylguanosine cap immunoglobulin (anti-TMG, prepared by Dr Reinhard Luhrman) was used. To confirm that the immunoprecipitation worked at the protein level, the samples were immunoblotted with PAb204 and an anti-Sm antibody (figure 6.7). Figure 6.8 shows the result of pCp labelling. The positive control (anti-TMG Ig, lane C) shows the position of small nuclear RNA. PAb204 (lane B) did not immunoprecipitate any RNA species capable of being labelled by T4 RNA ligase (the 5S and 5.8S labelling is probably not significant). The labelling in the negative control lane (anti-p53) is interesting. p53 is known to be covalently attached to RNA (Samad and Carroll, 1991). In HeLa cells p53 is rapidly degraded by E6-mediated ubiquitin-dependent proteolysis (Scheffner, et al., 1991). Despite there being so little p53 that it is undetectable unless proteolysis is suppressed it appears that the p53-associated RNA is readily labelled with pCp (the intense smear in figure 6.8 lane A).

Discussion

Soon after the ATPase data above was published (Iggo and Lane, 1989) Stahl's group published a paper on p68's biochemical properties (Hirling, et al., 1989). Stahl initially attempted to purify p68 by PAb204 immunoaffinity chromatography. Elution of p68 from the column required pH13

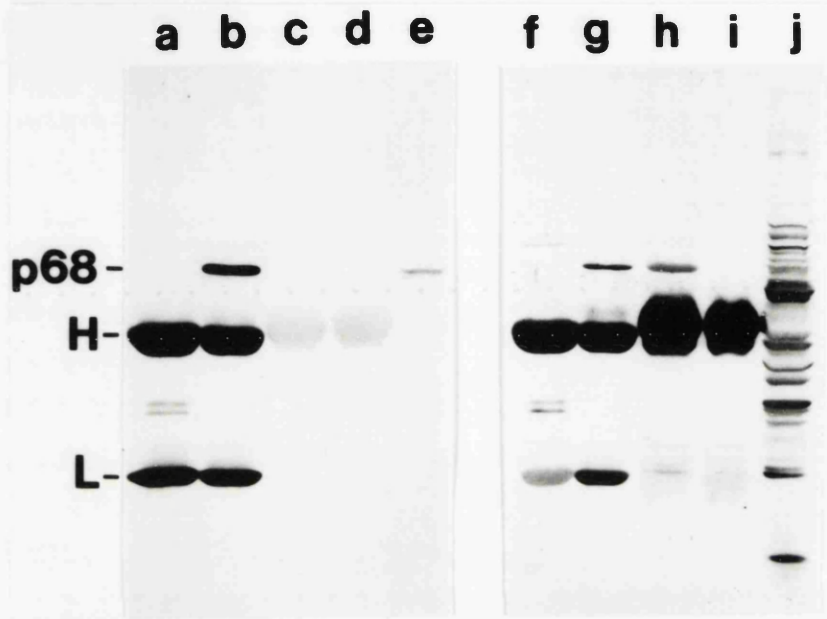


Figure 6.7 Immunoblot of HeLa immunoprecipitates (a&f, PAb240; b&g, PAb204; c&h, anti-TMG), mock immunoprecipitates (d&i, anti-TMG) and HeLa S100 (e&j) probed with PAb204 (a-e) and Kung anti-Sm serum (f-j). H, heavy chain; L, light chain.

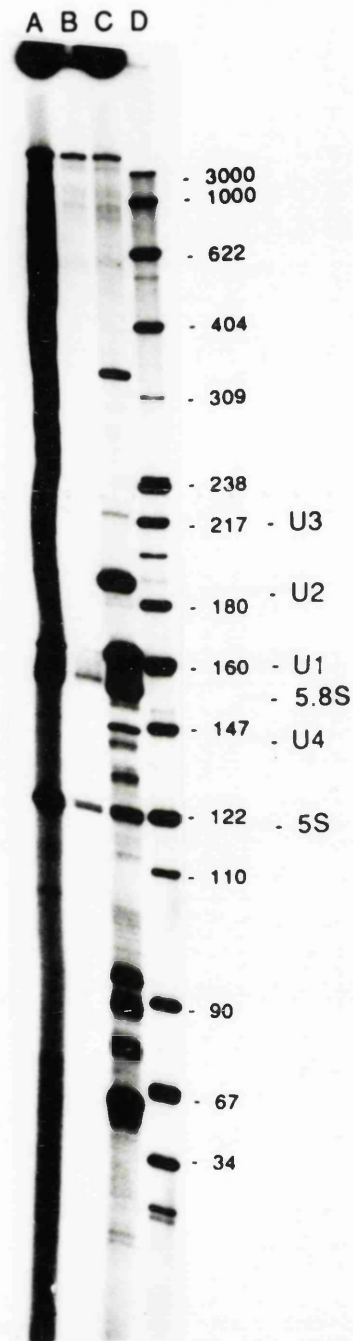


Figure 6.8 pCp labelling of RNA immunoprecipitated with PAb240 (A), PAb204 (B) and anti-TMG (C). Markers are in nucleotides (D).

buffer, which inactivated the ATPase, so Stahl purified p68 conventionally by chromatography of S100 on hydroxyapatite, monoQ and ssDNA cellulose columns followed by centrifugation through a glycerol gradient. The alternative approach, which was used here, was not to attempt elution of the p68 from the antibody. Both approaches have some merit. The problem with using conventional columns is that it is difficult to be absolutely certain that closely related members of the DEAD family have been separated from one another. Although the two studies are complementary, it is a great deal easier to assay pure, soluble protein than it is to work with immunoprecipitates on Sepharose beads and for that reason Stahl's data are intrinsically more reliable. Nevertheless, there is good agreement between the two kinetic analyses: p68 hydrolyses ATP with a V_{max} of 11-40 pmol/ μ g/s. Stahl went on to perform what are now classic RNA-transcript strand-displacement assays. Unfortunately, he did not pursue the work to the point where he could confidently establish the polarity of unwinding. As a result of this work p68 is now one of the best characterised members of the DEAD box family.

There is a striking analogy between RNA helicase biochemistry and kinase biochemistry: it is easy to study catalysis but extremely difficult to identify biologically meaningful ligands. There are two ways to identify RNA targets for p68: biochemistry and genetics. The first steps on the genetic path are described in the next chapter. The observation that p68 copurifies with cellular RNA suggested a direct biochemical way to identify p68's target. The pCp labelling approach used could have yielded two useful results: it could have provided a simple way to make p68 specific probes or it could have identified a specific p68-small nuclear RNA complex. In practice it did neither. It is formally possible that there was no RNA present, because the p68 used for the pCp labelling was not tested for ATPase activity. However, that seems most unlikely. T4 RNA ligase has successfully been used to label a wide range of different substrates (England, et al., 1980). Failure here would be easier to explain if only a single RNA species was present. It would then be possible to imagine that the 3'-end of the RNA was sequestered in a strong secondary structure, or that all of

the RNA had been cleaved by a ribonuclease which left a 3' phosphate group. It is conceivable that p68 itself might prevent labelling if it modified the 3'-end of the RNA. The rather negative outcome of the pCp labelling experiment does not rule out other strategies, such as labelling the RNA *in vivo* by growing cells in $^{32}\text{P}_i$ or cloning the RNA as oligo(dT)-primed or random-primed cDNA. It would also be possible to identify RNA ligands for p68 using affinity-PCR (Ellington and Szostak, 1990; Tuerk and Gold, 1990), but experience with this technique is very limited and it is possible that one might generate an entirely novel ligand which gave no clue to the identity of p68's real cellular target.

Introduction

This chapter begins with some genetic data showing that the yeast p68-like genes described in chapter three are essential for viability. This is the first step on a path that should ultimately lead to the identification of genetic targets for the yeast genes and, by implication, p68. The yeast genes and p68 all have an unusual intron, also described in chapter three, which is likely to have a specific function. The combination of yeast mutants, inducible *DBP2* expression constructs and the anti-DEAD antibody should permit comprehensive analysis of the effect of the intron on *DBP2* expression. Some preliminary data is presented which tests the model that the conserved intron autoregulates *DBP2* expression.

Disruption of the *DBP2* and *dbp2* genes is lethal

The yeast p68-like genes were disrupted by insertion of *ura4* into *dbp2* and *URA3* into *DBP2*, and analysed by tetrad dissection. The disruption and tetrad analysis were carried out by Derek Jamieson in *S.cerevisiae* (figure 7.1) and Stuart MacNeill in *S.pombe*.

Strains were characterised in this laboratory. Integration of the marker into the correct site in the chromosome was confirmed by PCR from a *ura* primer to a gene specific primer external to the linear fragment used for the disruption. Only correctly integrated copies can amplify in this assay (figure 7.2). Northern blotting of the parental *S.cerevisiae* strain (DJY84) and the strain heterozygous for the disruption (DJY98) shows that the disrupted allele produces a truncated message which is processed into mature polyadenylated RNA (figure 7.3). Spore viability in the *S.cerevisiae* tetrad dissection was rather low both for the disrupted strain (figure 7.1 b) and for a control strain (figure 7.1 a). This may be related to the large number of triads seen with these strains (D.Jamieson, personal communication). Nevertheless, it was apparent that for both *S.pombe* and *S.cerevisiae* only two out of four spores were viable in the disrupted strains. Auxotrophic marker analysis showed

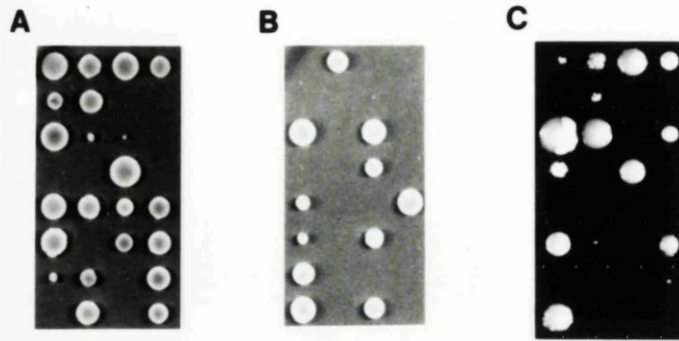


Figure 7.1 *S.cerevisiae* tetrad dissection: a, Random integrant (*DBP2* locus intact); b, *DBP2/DBP2::URA3* cells (DJY98); c, *DBP2/DBP2::URA3* cells (DJY98) rescued by *DBP2* cDNA plasmid (pIG70). The tetrad dissection was performed by Derek Jamieson.

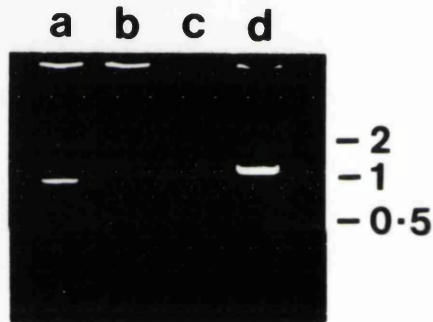


Figure 7.2 Identification of transformants containing disrupted *DBP2* genes. a & b, *S.cerevisiae*; c & d, *S.pombe*; b & c, parental diploids; a & d, transformants containing one disrupted allele. Whole yeast were amplified directly by PCR using *ura* and p68 primers. Markers in kb.

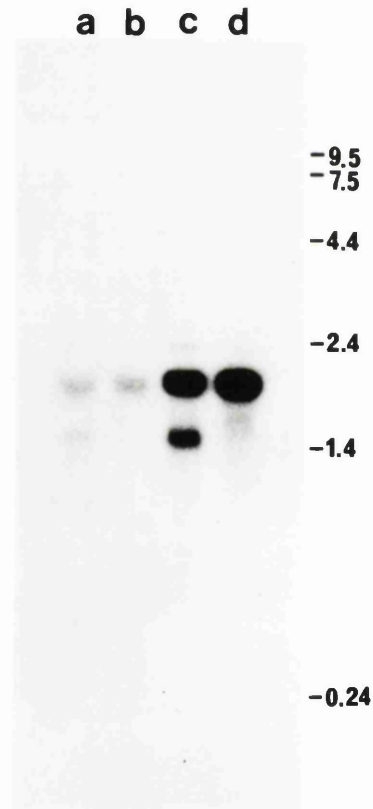


Figure 7.3 Northern blot. Total RNA (lanes a and b) and oligo-dT selected RNA (lanes c and d) from parental (lanes b and d) and disrupted (*DBP2/DBP2::URA3*, lanes a and c) strains was hybridised with an antisense *DBP2* cDNA probe. Note the presence of an extra band in the disrupted strain. RNA markers in kb.

that all of the surviving spores were *ura*⁻. The *ura*⁺ spores germinated on schedule but arrested after about five rounds of cell division. Microscopic examination of the dying cells gave no clue to the function of *DBP2* or *dbp2*. Specifically, the cells did not show a cell cycle arrest phenotype (S. MacNeill, personal communication).

Viability is restored by expression of a *DBP2* cDNA

A good test of whether lethality results directly from disruption of a particular gene is to show rescue of the inviable cells by adding back the gene. A plasmid (pIG70, table 2) was constructed which expresses a *DBP2* cDNA under the control of the *GAL1* promoter. Western blotting with MaD1 shows that the plasmid gives good induction of DBP2 when yeast are transferred to medium containing galactose, which activates the promoter (figure 7.4). Only two bands are seen on blots from protease deficient cells: the lower band is DBP2, the upper band is probably DED1. The plasmid was transformed into cells heterozygous for the *DBP2* disruption, and the tetrad dissection was repeated (figure 7.1 c). In tetrads where the plasmid segregated with the disrupted *DBP2* allele all four spores were viable, confirming that the inviability of *DBP2::URA3* spores is due to the disruption and that *DBP2* is an essential gene. Northern blotting of the rescued *S.cerevisiae* strain (yIG4) using a *DBP2* probe showed that the plasmid gave good induction at the RNA level (figure 7.5). Western blotting of two different rescued strains (yIG4 and yIG5) showed that the plasmid also gave good induction at the protein level (figure 7.6). Even on overdeveloped blots DBP2 was barely detectable in uninduced haploid cells containing the *DBP2* plasmid and the disrupted *DBP2* allele (figure 7.6 d & f). Despite the absence of DBP2, cells were still able to grow very slowly in medium containing glucose (which represses the *GAL1* promoter). The doubling time of wild type cells is 160 minutes in glucose and 195 minutes in galactose; that of the disrupted/rescued cells is 260 minutes in glucose and 360 minutes in galactose. This shows that the plasmid copy of *DBP2* does not restore normal growth. There are several possible explanations. The disrupted allele could be toxic

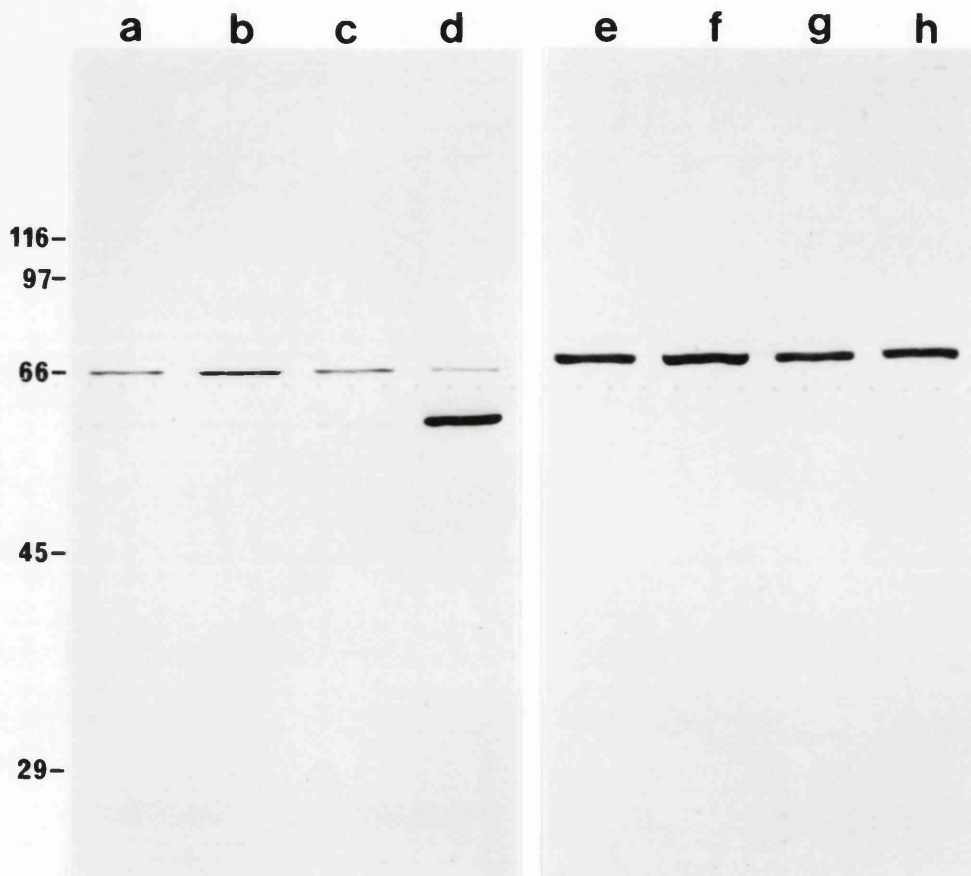


Figure 7.4 Western blot of protease deficient (PY26) cells probed with MaD1 (a, b, c & d) or the anti-hsp 70 monoclonal antibody 7.10 (e, f, g & h). a, b, e & f, parental cells; c, d, g & h, cells transformed with pIG70, a plasmid containing a *DBP2* cDNA under the control of the *GAL1* promoter. a, c, e & g, glucose medium; b, d, f & h, galactose medium. DBP2 is the lower of the two bands in a - d ; the upper band is probably DED1. Markers in kD. 7.10 was included as a loading control.

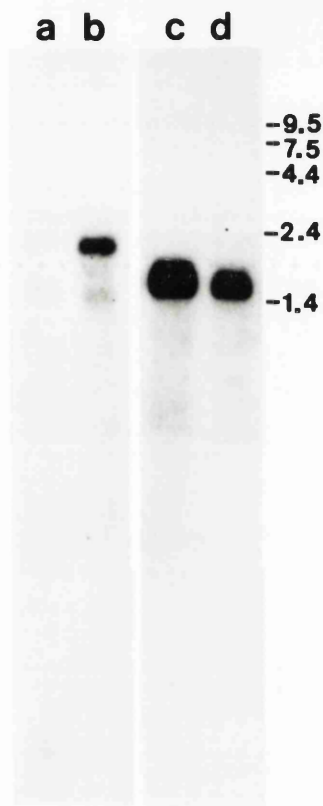


Figure 7.5 Northern blot of total RNA from yIG4 cells (haploid *DBP2::URA3* cells rescued with pIG70). a and c, glucose medium; b and d, galactose medium; a and b, DBP2 probe; c and d, actin probe. RNA markers in kb.

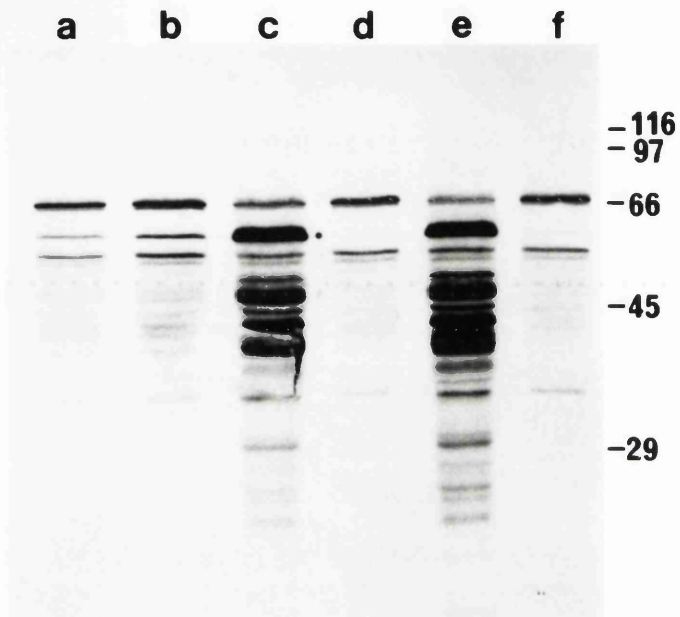


Figure 7.6 Western blot of yIG4 and yIG5 cells (haploid *DBP2::URA3* cells rescued with pIG70) and yIG6 cells (wild type haploid cells from the same dissection) probed with MaD1. a and b, yIG6; c and d, yIG5; e and f, yIG4; b, d and f, glucose medium; a, c and e, galactose medium. DBP2 is the second band from the top of the blot. The top band (probably DED1) acts as a loading control. Markers in kD.

in the rescued strains, although it is not toxic in heterozygotes. It is possible that the level of DBP2 needs to be carefully titrated to achieve normal growth: the level of *DBP2* in the rescued cells is very low in glucose medium and grossly elevated when the *GAL1* promoter is induced with galactose. Interestingly, one rescued strain did appear to grow only in 2% galactose/0.2% glucose, suggesting that the absolute level of DBP2 is important. However, that strain was morphologically grossly abnormal, with giant malformed buds and fragmented nuclear DNA.

The trivial explanation for the slow growth of the rescued cells is that there is a mutation in the cDNA clone. The Western blots in figures 7.4 and 7.6 show that the open reading frame is intact. However, the cloning strategy (chapter three) inevitably resulted in the production of a composite cDNA clone: the plasmid has an adaptor followed by a genomic PCR product then a cDNA PCR product and a genomic λ gt11 clone. At each step in the construction the sequence of the DNA inserted was checked. Nevertheless, the possibility that a point mutation is present should not be dismissed.

The conserved intron may regulate *DBP2* expression

One obvious difference between wild type and rescued strains is the absence from the latter of the large intron in helicase motif V. Fickett plots (chapter three) are unable to predict whether regions of a few hundred base pairs are likely to encode proteins. It would not be surprising if the intron encoded a small protein or even a small RNA molecule such as a small nuclear RNA. To address these possibilities a Northern blot of *S.cerevisiae* RNA was hybridised with a *DBP2* intron probe under stringent conditions (figure 7.7). There is a band at 500 bp which is strongly enriched in oligo(dT)-selected RNA. Lane a contains RNA from cells (yIG4) which contain no intron DNA. If the 500 bp band is a polyadenylated transcript encoded by the intron there should be no band in this lane. That there is indeed no band is encouraging but the experiment needs to be repeated with oligo(dT)-selected RNA because the signal from total RNA is too weak to be interpreted with confidence.

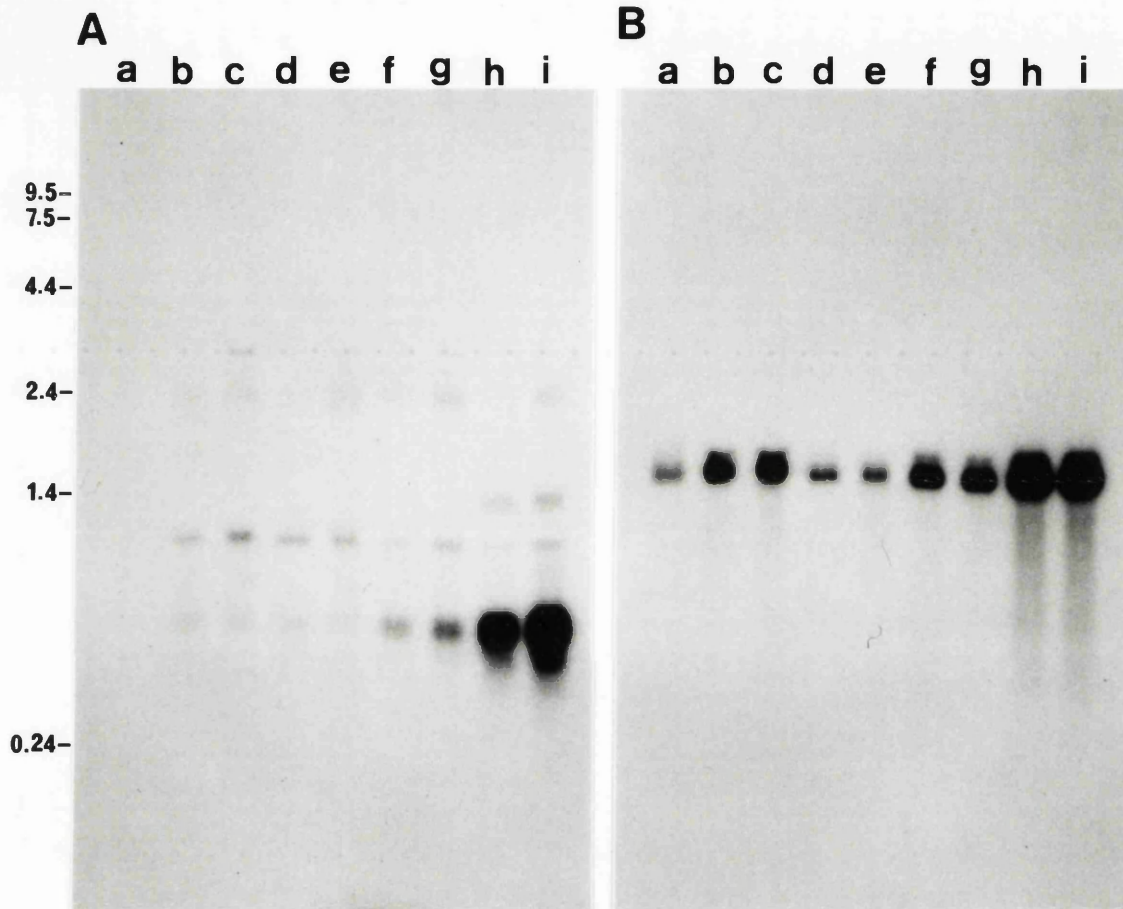
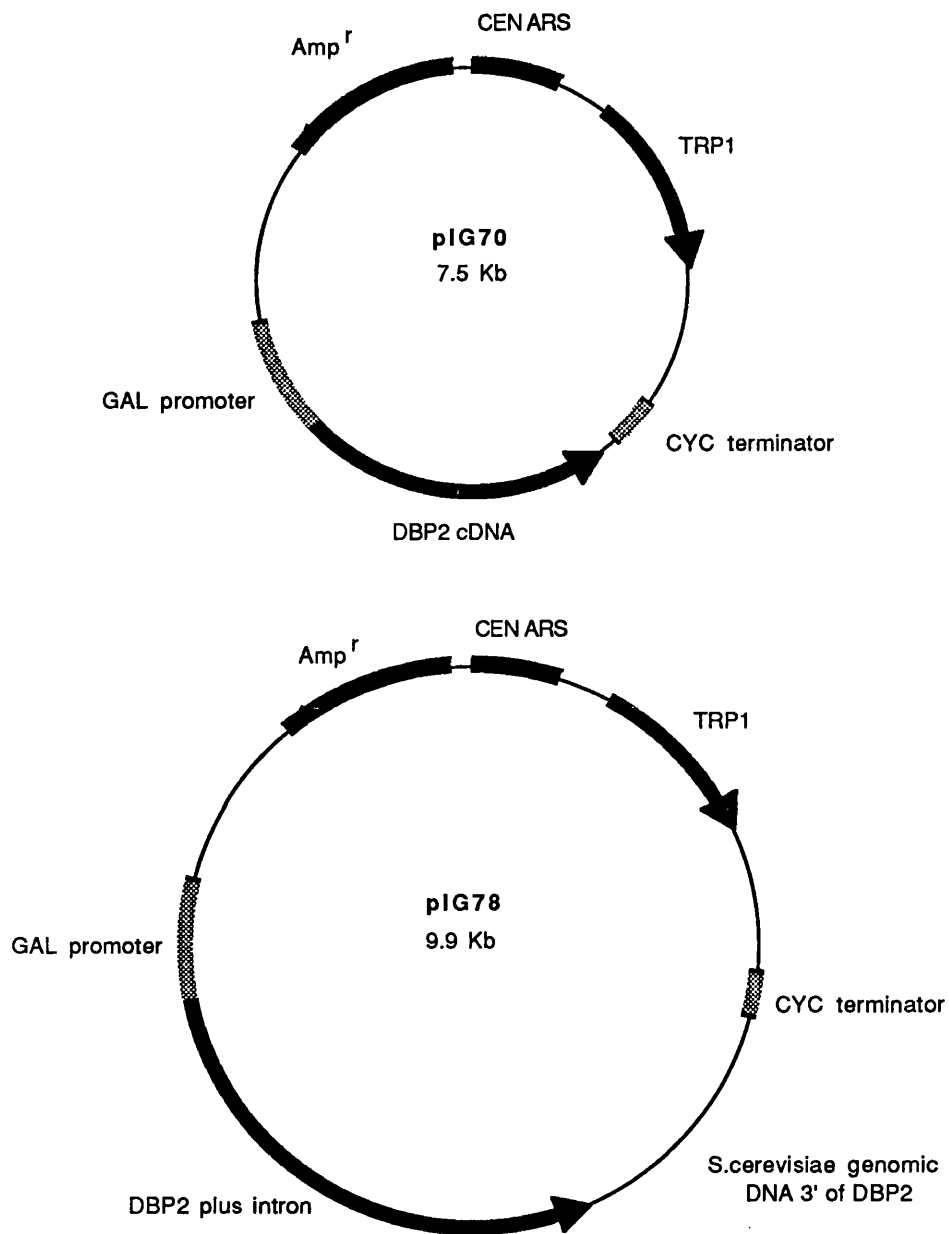


Figure 7.7 Northern blot hybridised with an antisense intron probe (A) and an actin probe (B). a - g, total RNA; h & i, oligo-dT selected RNA. a, yIG4 (haploid *DBP2::URA3* cells rescued with pIG70); b & c, DJY98 (*DBP2/DBP2::URA3*) plus pIG70; d & e, DJY84 (wild type) plus pIG70; f & h, DJY98; g & i, DJY84. pIG70 contains a *DBP2* cDNA under the control of the *GAL1* promoter. The promoter was induced with galactose in lanes a, b and d. RNA markers in kb.

If the intron encodes a protein or RNA it should be possible to rescue the sick phenotype of cells lacking the intron by supplying the intron in *trans*. No wild type intron clones were obtained by PCR subcloning. However, it was possible to use the available inverse PCR clones and genomic clones to construct a plasmid (pIG74) which was successfully used to obtain a wild type genomic *DBP2* clone by gap repair (pIG75, see chapter two for details). Gap repair is a high fidelity cloning technique because there is no *in vitro* polymerase step. The wild type EcoRV-BglII fragment encompassing the intron was subcloned from pIG75 into a centromeric plasmid (pRS313, [Sikorski and Hieter, 1989]) which was transformed into the haploid *DBP2::URA3* cells which had been rescued with the *DBP2* cDNA. Growth of these cells requires a triple selection (histidine for the intron plasmid, tryptophan for the *DBP2* cDNA plasmid and uracil for the disruption) which means growing the cells in minimal medium plus galactose and amino acids (rather than minimal plus galactose and casamino acids if histidine selection is not required). The growth rate of cells containing the parental plasmid was the same as that of cells containing pIG76, the plasmid with the intron cloned into the polylinker. However, it is difficult to be confident of this result because of the very slow growth rate in galactose minimal medium.

An alternative function for the intron would be to provide a target for regulation of *DBP2* expression. In its simplest form this might involve inhibition of splicing by *DBP2* itself. To begin to address this problem constructs expressing *DBP2* with and without the intron were engineered. Preliminary results show that, unlike cells containing the cDNA plasmid described above, cells transformed with an analogous plasmid containing the intron (pIG78) do not overexpress *DBP2*. RNA analysis (figure 7.8) shows that unspliced forms do not accumulate when the *GAL1* promoter is switched on. This does not entirely eliminate a model in which splicing is inhibited, but one would have to postulate that the unspliced forms were rapidly degraded. The RNA loading was markedly uneven in this experiment, which makes it difficult to interpret the results with confidence. The size difference between the pIG70 RNA and the wild type RNA



pIG70 and pIG78 place DBP2 under the control of the GAL1 promoter. pIG70 contains a composite DBP2 cDNA clone and uses the CYC terminator and polyadenylation signal. pIG78 contains a genomic DBP2 clone and uses its own terminator and polyadenylation signal. The plasmids were used to study the effect of the DBP2 intron on DBP2 expression.

} gluc.
 + cDNA }
 + gene }
 + cDNA } gal.
 + gene }
 } PA⁺

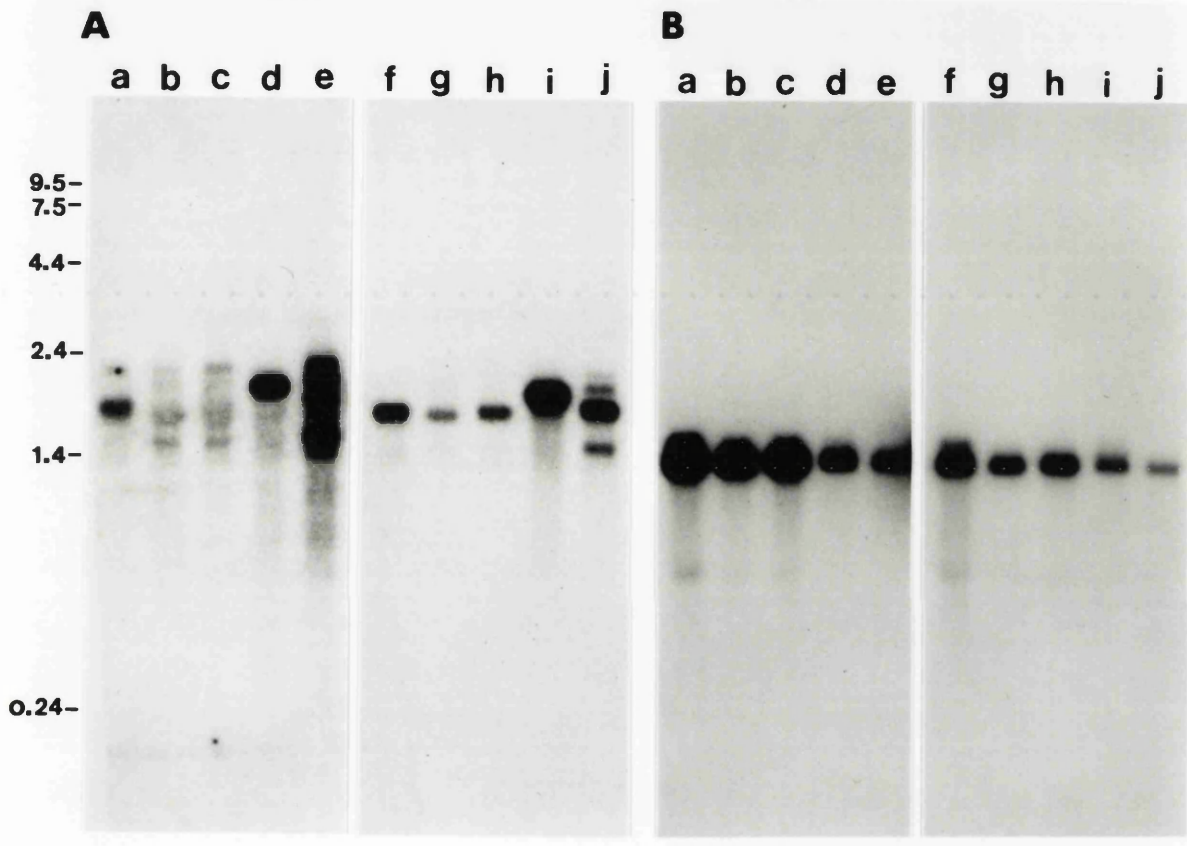


Figure 7.8 Northern blot showing the effect of the intron on *DBP2* expression.
 A, *DBP2* probe, B, actin probe. a-e, total RNA; f-j, oligo-dT selected RNA.
 a & f, DJY84; b, d, g & i, DJY84 plus pIG70; c, e, h & i, DJY84 plus pIG78.
 a, b, c, f, g & h, glucose medium; d, e, i & j, galactose medium.
 The *DBP2* locus is intact in these cells. pIG70 and pIG78 contain *DBP2* genes under the control of the *GAL1* promoter; the main difference between the two constructs is that pIG78 contains the *DBP2* intron.

band probably reflects a difference in polyadenylation site. The absence of the wild type band in lane i suggests that negative feedback on expression of the endogenous RNA does occur; the problem is to identify the level at which this control is exerted. The cDNA construct has a *CYC* terminator immediately downstream of the stop codon; the genomic construct has an extra 1.6 kb of non-coding genomic DNA between the stop codon and the terminator. This means that the genomic construct uses the normal terminator, which is unfortunate because it is impossible to be sure that the presumptive *DBP2* messenger RNA band in lane j is derived from the transfected gene. Some of the bands in lane e are absent from poly(A) selected RNA (lane j). The identity of these bands is unknown but they may be prematurely terminated transcripts, splicing intermediates or degradation products.

Discussion

The identification of p68-like genes in yeast has made possible entirely new approaches to determine the biological function of p68. The most important benefit from using yeast is the ability to use genetics. This should permit the identification of genes which interact with *DBP2*. It is essential to keep the limitations of the approach in perspective. The key point is that mechanisms are much better conserved than signals: the same messenger can carry different messages. The relevance of this to the general point made in the introduction about the functions of DEAD proteins is that yeast can provide powerful insight into higher eukaryotic "housekeeping" functions but potentially misleading insight into higher eukaryotic "gene control" functions. That *DBP2* is essential is good news in this regard since it biases the analysis towards the housekeeping argument. However, death is not a helpful phenotype. To exploit yeast genetics it will be necessary to generate conditional mutants. As soon as genetic targets are identified it should be possible to return to mammalian systems to ask whether p68 interacts with the same set of genes.

Although the reason for moving to yeast was to exploit yeast genetics, the incidental findings about the yeast introns

may permit rapid progress using classical molecular biological techniques. If the conserved intron in motif V autoregulates *DBP2* expression it should be possible by deletion mapping to identify one RNA target for *DBP2*. Needless to say, comparison with the human and *S.pombe* introns should allow rapid characterisation of the target. This will be immensely valuable in hunting for molecules with which *DBP2* and p68 interact.

The results presented here on the role of the *S.cerevisiae* *DBP2* intron are preliminary observations which require confirmation. However, it appears that the model to test has changed to one where *DBP2* exposes a nuclease site in the intron by modifying the secondary structure of the RNA. Naturally, in yeast one would be disappointed if the intron did not encode the nuclease. Whether the 500 bp band detected by intron probes on Northern blots of *S.cerevisiae* RNA is an intron encoded transcript can be sorted out by examining oligo(dT)-selected RNA from strains lacking the intron, mapping the domain in the intron responsible for the signal with strand specific probes, and PCR cloning the transcript using an oligo(dT) anchor (Frohman, et al., 1988). It would be worthwhile making a strain lacking the intron. It should be relatively easy to do this by transformation of *DBP2::URA3* cells with a *DBP2* cDNA on a linear fragment in the presence of 5-fluoro-orotic acid and uracil. Transformation with the *DBP2* cDNA will result in a low frequency of gene conversion of the cDNA into the chromosomal *DBP2* locus by homologous recombination. In the disruption *URA3* replaces the *DBP2* intron, hence gene conversion of a *DBP2* cDNA into the *DBP2* locus in these cells will result in loss of *URA3*. Although this will be a rare event, non-recombinants can be killed with 5-fluoro-orotic acid, which is converted to 5-fluoro-uracil by orotidine-5'-phosphate decarboxylase, the *URA3* gene product (Winston, et al., 1984).

Tackling the question of autoregulation requires more sophisticated constructs than are presently available. It is important to be able to distinguish endogenous from transfected *DBP2* and it would be helpful to know if translation of the message is required. These goals can be achieved by epitope tagging the transfected gene and putting stop codons at the start of the message. For analysis at the protein level it would be

much better to work in protease deficient strains than in DJY98 (compare figure 7.4 with figure 7.6). Working in a *DBP2::URA3* background has the advantage that the short transcript from the disrupted gene acts as a reporter for feedback on the promoter, thus it would be sensible to repeat the disruption in protease deficient strains. If these experiments give unequivocal evidence of autoregulation the next step would be to transfer the *DBP2* intron into a splicing reporter gene and show that expression of the reporter was inhibited by expression of *DBP2*. Although this is the most decisive experiment it is also the most likely to fail for spurious reasons; hence the attraction of looking initially at the intron in its normal context.

Chapter 8 Discussion

When the work described in this thesis began the only members of the DEAD family which had been identified were p68 and eIF4A. At the time it was not clear whether p68 was a DNA or an RNA helicase, or even if it was a helicase at all. That situation has been transformed over the last three years by work from a large number of laboratories. The results presented in the preceding five chapters have contributed materially not only to our understanding of p68 itself but also to our knowledge of the workings of the DEAD family as a whole. p68 is now one of the best understood DEAD box proteins at a biochemical level. The key outstanding problem is the identification of the biological process in which p68 is involved. The most rigorous way to solve this problem is by genetic analysis of the yeast p68-like genes described in chapter three. The least useful approach has been cell biology, although the striking cell cycle changes in nuclear location described in chapter five will impose demanding constraints on any final understanding of the function of p68. Ironically, the starting point of the work, the link between p68 and SV40 large T antigen, remains as enigmatic and ill-understood as ever. Specific aspects of the results presented in the preceding chapters have already been discussed at some length. This final discussion will, therefore, concentrate more on outstanding problems and ways in which they may be solved.

In the long term the yeast system is probably the most tractable, always bearing in mind that results obtained in yeast with the *DBP2* gene will need to be confirmed in mammalian cells with p68. Indeed, if a great deal more work is to be done on *DBP2* it would be reasonable to attempt rescue of *DBP2* mutants with p68 to clarify whether the two are genetically homologous. If full length p68 did not complement *DBP2* mutants it might still be possible to show a more limited form of complementation by transferring only the central domain from p68 into *DBP2*. To pursue the function of *DBP2* in yeast requires conditional mutants. These can take two forms, expression mutants and (primary) structural mutants. In principle, the *GALI* expression

system can be used to study what happens to cells when *DBP2* is removed. For this to work the endogenous *DBP2* allele must be disrupted. With this system it is possible to grow large quantities of cells in galactose (which activates the *GALI* promoter), transfer the cells to medium containing glucose (which represses the *GALI* promoter) and then look for a phenotype, such as the presence of ribosomal RNA precursors or unspliced pre-messenger RNA. Preliminary work suggests that neither phenotype is seen with the cDNA construct described in chapter seven (note, for example, that the actin control in figure 7.5 is normally spliced). Of course, it is possible that the slow growth in glucose of cells rescued with the cDNA construct reflects the presence of a severe RNA processing defect which can only be detected using a specific assay. Since the cells were able to grow when the promoter was switched off it is clear that yeast require very little *DBP2* to survive. It may be that the RNA processing defect would be more obvious in cells containing even less *DBP2*. One way to achieve this would be to repeat the *DBP2::URA3* rescue experiments using *GALI::DBP2* expression constructs containing the *DBP2* intron.

If it is not possible to switch off expression completely by modifying the *GALI* constructs it will be necessary to create structural mutants. A large number of mutations in other DEAD genes have now been sequenced. Since the mutations all appear to change conserved residues it might be possible to engineer a temperature sensitive or cold sensitive mutation directly. Alternatively, it would be possible to mutagenise plasmid non-specifically and ask for conditional rescue of *DBP2::URA3* cells at different temperatures. A class of mutants which would be particularly useful would be transdominant mutants. These could be identified by mutagenising the *GALI* expression construct *in vitro*, transforming it into wild type cells and looking for growth arrest in galactose. Structural mutants could be tested in the same way as expression mutants. However, this search might suffer from the same problem as was alluded to above, namely that a gross defect can be missed without a specific assay. The only remaining option would be to use classical yeast genetics to isolate allele-specific suppressors.

Comparison of the human p68 gene with the *S.pombe dbp2* and *S.cerevisiae DBP2* genes revealed the presence of a highly conserved intron. It is only the position and not the sequence of the intron that is conserved. There are no other examples of conservation of gene structure over such vast expanses of evolutionary time. Furthermore, the specific features of the intron in *S.cerevisiae* make it most unlikely that the conservation is purely accidental. The *S.cerevisiae DBP2* intron is twice as big as any other known intron in *S.cerevisiae*, and its location towards the 3'-end of the gene is very unusual. The possibility that the conserved intron encodes a short stable RNA or messenger RNA deserves careful scrutiny. Although sequence analysis suggests that the intron is non-coding the software available can not reliably detect short exons. The data presented in chapter seven indicate that there is a ~500 base pair polyadenylated transcript which hybridises with *S.cerevisiae DBP2* intron probes on Northern blots of *S.cerevisiae* RNA. Whether the band is encoded by the intron can be determined by examining messenger RNA from strains lacking the intron. The domain in the intron responsible for the signal can then be mapped with strand specific probes. Finally, having established the polarity and boundaries of the sequence the poly(A) tail can be used for anchor PCR cloning (Frohman, et al., 1988).

Faced with conservation of non-coding DNA the molecular biologist naturally considers the possibility that the conserved DNA functions at the nucleic acid level. This problem is made more intriguing by the fact that *S.cerevisiae* has apparently chosen not to scatter transcriptional control elements in introns the way higher eukaryotes do. This makes excision of junk DNA much less risky and is a key prerequisite for the Fink reverse transcription/homologous recombination model (Fink, 1987). Unless the intron is an aggressive molecular parasite there must be either a mechanistic obstacle to its removal or a high price to pay for its removal. There may well be sequences which can not be removed by reverse transcription and homologous recombination for purely technical reasons, but that is a very tenuous explanation for the conservation of the *DBP2* intron. Deletion of a *DBP2* intron encoded transcript would, by contrast,

carry an obvious penalty. Nevertheless, retention of the intron logically demands more than just the presence of specific coding information, it requires that placement of that information in the *DBP2* intron is essential for the proper functioning of either the intron encoded gene or the *DBP2* gene. The most apposite model, autoregulation of bacterial and yeast ribosomal protein gene expression (Dabeva, et al., 1986), predicts that *DBP2*'s primary RNA target is mimicked by the *DBP2* intron. When the primary target sites are saturated *DBP2* binds to its own message and prevents its own expression. It would obviously be quite simple for an RNA helicase to autoregulate its expression in this way. Since *DBP2* is the only DEAD gene which is known to contain an intron in *S.cerevisiae* it follows that there must be something special about the RNA target to which *DBP2* binds which makes conventional transcriptional feedback unattractive. This would be particularly remarkable if it were also the basis for the conservation of the intron in human p68, since higher organisms are adept at highly complex transcriptional control. If the model is correct it will not only provide an extremely elegant means for an RNA helicase to regulate its own expression but also identify at least one RNA target to which *DBP2* can bind. If so, it should be relatively straightforward to define the RNA domain responsible for binding by deletion mapping. Once a minimal domain had been defined biologically it would be reasonable to look for *in vitro* binding of intron transcripts to *DBP2* (and p68), for example by nitrocellulose filter binding. The identification of one RNA which binds to *DBP2* with high affinity opens up the possibility of hunting for similar sequences, either in libraries or by inverse PCR. The multiple benefits from defining the role of the intron dictate that this should be the most important field of enquiry in the immediate future.

Without other clues the only way to identify p68's RNA target is to look for p68-specific RNA. The observation in chapter six that immunoprecipitated p68 has endogenous ATPase activity which is abolished by treatment with ribonuclease A suggests that RNA copurifies with p68. The failure to label any coprecipitating RNA using pCp and T4 RNA ligase may indicate that p68-bound RNA has a specific structure or modification at

its 3'-end. If there is any RNA present at all it should be possible to label it by growing cells in $^{32}\text{P}_i$. This RNA could be extracted and examined by ribonuclease fingerprinting, which should give an idea of its complexity. If large amounts of labelled RNA were available they could be used for Northern blotting and library screening. Another obvious way to identify the RNA would be to use some form of cDNA cloning. This could be done using oligo(dT)-primed or random-primed cDNA synthesis followed by amplification either *in vitro* by PCR or *in vivo* using bacteria. It would be possible to increase the RNA yield by loading cellular RNA onto p68 or *DBP2* columns. The fashionable way to identify RNA ligands is by affinity PCR (Tuerk and Gold, 1990; Ellington and Szostak, 1990), but there are two difficulties. The first is knowing where to begin, because it is obviously better to randomise a candidate sequence than to start from scratch. The second is that the technique may be powerful enough to generate extraneous ligands, although this could be mitigated by using specific elution of product in the early cycles with, say, poly(A). A particularly elegant approach would be to use affinity PCR to explore the binding requirements of randomised derivatives of the *DBP2* intron.

The issues addressed above are specific to the p68/*DBP2* RNA helicase family. There are general problems of a more biochemical nature which apply to all DEAD proteins. These are questions of kinetics, catalysis and structure. There is no compelling reason to study these problems purely in the context of p68 or *DBP2*. Indeed, it could be argued that SrmB is a better candidate for biochemical work. This is because soluble SrmB is available in milligram quantities using the T7 expression constructs created by Nishi, et al. (1988). SrmB has the additional advantage that it is one of the smallest DEAD proteins, yet it retains RNA-dependent ATPase activity. That makes it a better candidate for structural analysis than eIF4A, which probably lacks a specific RNA binding domain. If it is not possible to grow SrmB crystals it would be worth finding out whether the 40 kD *E.coli* protein recognised by MaD1 (figure 3.29) is an authentic DEAD box protein because 40 kD is the theoretical upper limit for structural analysis using four-

dimensional nuclear magnetic resonance techniques (Kay, et al., 1990). The large number of conserved residues in DEAD proteins means that solution of one structure will immediately suggest structures for the other members of the family. More mundane biochemical problems include the need to demonstrate enzyme turnover rather than just stoichiometric binding, and functional definition of the central domain as the catalytic unit. If the terminal domains provide RNA binding specificity it should be possible to demonstrate this *in vitro*, preferably using specific RNA ligands. The immunoprecipitation assay should provide a rather elegant way to distinguish stoichiometric binding from turnover. Since all p68 remains attached to the beads, any single-stranded RNA present in the supernatant after pelleting the beads must have been unwound and then released by the helicase. This sort of assay should also make it simpler to study problems like polarity and RNA target sites. In the case of *PRP28*, for example, it might only be possible to recover the U4 small nuclear RNA in the supernatant, which would indicate that the *PRP28* remained bound to U6.

Biochemical techniques can be used to ask what p68 and *DBP2* do in the cell. This requires the development of specific inhibitors. The main difficulty with this approach is the likelihood that closely related members of the DEAD family will also be affected. The ideal inhibitor would be a small molecule, such as an ATP analogue, but even if a suitable compound already exists it is unlikely to be recognised as such without assistance from, say, the pharmaceutical industry. The alternative would be to use a biological ligand. One of the reasons for making a monoclonal antibody against the DEAD peptide (chapter four) was to produce a reagent which would inhibit p68's ATPase activity. That MaD1 is not inhibitory is not really very surprising because the hybridoma fusion was not screened for inhibition. Devising a suitable screen would be time-consuming but not difficult. For example, one could develop a microtitre plate screen using the activated charcoal assay described in chapter six. The key requirement is that the antibody should recognise native protein, hence it might be possible to use an even simpler screen, such as staining cells

arrested in telophase. A more attractive option would be to use phage antibody (McCafferty, et al., 1990) or affinity PCR technology (Tuerk and Gold, 1990; Ellington and Szostak, 1990). The antibody approach would require two rounds of selection. In the first round single chain variable region phage antibodies derived from mice immunised with the DEAD peptide (see chapter four) would be selected on p68 columns. p68 immunoprecipitated on PAb204 columns would be ideal because it is still active in the ATPase assay (see chapter six). Low affinity antibodies against sites which undergo allosteric changes on ATP hydrolysis could be selected by eluting the phages with ATP. Conversion of these phage antibodies to high affinity ligands by *in vitro* mutagenesis followed by reselection on p68 columns (with elution by p68 instead of ATP) would stand a reasonable chance of creating inhibitory antibodies. Since the changes in ras on ATP hydrolysis are most striking in regions outside the ATP binding site (Schlichting, et al., 1990) there is in fact no reason to start with spleen RNA from animals immunised with the DEAD peptide rather than full length p68. An analogous procedure could be tried using PCR ligands, although in this case the starting material would be a specific RNA, such as the putative target in the *DBP2* intron. Inhibitors of RNA binding could be produced using an analogous approach involving elution of phage antibodies or PCR ligands with RNA. Affinity PCR ligands which recognised the RNA binding domain in p68 or *DBP2* would be analogous to transdominant mutants in the RNA target. Inhibitors which worked *in vivo* could be used to look for an RNA phenotype in treated cells. If, as seems more likely, inhibition was only effective *in vitro* it would still be possible to screen candidate assays. Assays which might be tested include *in vitro* translation, splicing, transcription and ribosomal RNA processing.

Another biochemical approach which might yield clues to the function of p68 would be some form of protein coprecipitation technique. That no proteins appear to be coimmunoprecipitated with p68 by PAb204 certainly does not rule out this approach. Cellular proteins which bound to the PAb204 epitope would never be seen in this assay. One way to

pursue this approach would be to make new antibodies against p68. However, an alternative would be to use an epitope tag or even to use a non-immunological system such as the glutathione-S-transferase fusion protein/glutathione column system (Smith and Johnson, 1988).

The cell biological data presented in chapter five is fascinating. During interphase p68 is dispersed throughout the nucleoplasm but excluded from the nucleoli. Since the majority of nuclear antigens give granular nucleoplasmic staining it is not possible to deduce much from this pattern. The striking prenucleolar body staining which occurs in telophase is much more informative. It can be interpreted either in functional or in structural terms. The functional argument poses the question: what RNA processing event is specific to the nucleolus in telophase? One can postulate a role for p68 in reassembly of the nucleolus after mitosis, but this presupposes a degree of nucleolar complexity which may not be justified. Identification of a specific population of RNA molecules, small nucleolar RNAs perhaps, by *in situ* hybridisation with 2'O-methyl RNA probes (Carmo-Fonseca, et al., 1991) might suggest the existence of a specific RNA processing event in prenucleolar bodies. The alternative would be to use cell fractionation techniques, but purification of nucleolar RNA from cells arrested in telophase does not seem a very promising task. Interpretation of the localisation data in structural rather than functional terms dictates that the problem should be investigated using cell biological techniques. The first step is to map p68's nuclear transport signal by transfection of mutant p68 genes. The key question is whether nuclear transport and nucleolar transport are linked. If they are not it should be possible to confer nucleolar localisation on reporter genes by transfer of p68's nucleolar localisation signal. The reason to suspect that they are not linked is that the most conspicuous homology between p68 and nucleolar proteins is not the lysine-rich putative nuclear transport signal between amino acids 32 and 45 but the glycine-rich domain at the extreme amino-terminus. Identification of a sequence capable of conferring cell cycle specific nucleolar targeting would be a very exciting result. However, there are

two sides to any localisation problem and identification of the prenucleolar anchor which interacts with p68 is a prerequisite for a proper understanding of the prenucleolar staining. Investigation of the cell biology of the yeast p68-like genes would be particularly interesting because they appear to lack the lysine-rich nuclear transport signal but have much more extensive glycine-rich "nucleolar" domains. The attraction of the yeast genes is that it is possible to ask whether the domains responsible for localisation are also necessary for viability, and to look for RNA processing defects in cells containing only DBP2 molecules defective in nuclear transport. However, two points should be borne in mind: first, that yeast and mammalian cells have very different nucleolar architectures (Potashkin, et al., 1990), and second, that the *DBP2* genes may not be functionally equivalent to p68. The post-mitotic nucleolar staining seen when PAb204 is injected into the cytoplasm is intriguing. The most interesting feature is the mitotic requirement. The general question which arises is how proteins segregate at the end of mitosis. It would obviously be more efficient to form the nuclear membrane around regions of the cell already enriched in nuclear components. This could be achieved quite simply by sticking nuclear proteins to DNA at the end of mitosis. In the case of p68 and PAb204 this would involve an interaction (possibly indirect) with the nucleolar organiser regions which contain the ribosomal RNA genes.

The most contentious issue is the link between p68 and SV40 large T antigen. What is required is an *in vivo* experiment showing that p68 and T antigen are functionally related. A good example would be a cell staining experiment showing dramatic changes in p68 staining following SV40 infection. One might imagine that T antigen would replace p68 in prenucleolar bodies, but in practice this does not occur for the simple reason that SV40 infected cells do not divide. Preliminary results with the affinity purified anti-p68 rabbit serum give no hint of changes in the granular interphase p68 staining pattern after infection, but it is doubtful whether anything other than gross changes would be detectable in what is by definition an ill-defined staining pattern. The underlying barrier to understanding the

link between T antigen and p68 is that neither p68's role in cellular RNA processing nor T antigen's role in viral RNA processing is known. T antigen mutants defective in RNA processing, or transdominant p68 mutants would be very useful. Gannon and Lane (1987) showed that the PAb204 epitope is obscured when either DNA polymerase α or the p53 oncogene binds to T antigen. If the PAb204 epitope were central to this interaction one might expect that the corresponding region in p68 would be able to bind to DNA polymerase α and p53. A role for p53 in RNA processing is hinted at by the discovery that a carboxy-terminal serine residue in p53 is covalently attached to RNA (Samad and Carroll, 1991). A role for DNA polymerase α in RNA processing would presumably be incidental to its role in DNA replication. For example, the primase subunit of DNA polymerase α might, in special circumstances, have some use for an RNA helicase. DNA polymerase α presumably also encounters nascent transcripts when it replicates actively transcribed genes. That p53 and DNA polymerase α are not seen in PAb204 immunoprecipitates certainly does not rule out an interaction between p53, DNA polymerase α and p68; indeed, it is the expected result. The reverse experiment, looking for p68 in p53 or DNA polymerase α immunoprecipitates, has not been done.

There are clearly a great many interesting experiments on p68 waiting to be done. Whatever the final verdict, the next few years should be a very exciting time for those who study DEAD proteins.

References

- Abramson, R. D., Dever, T. E., Lawson, T. G., Ray, B. K., Thach, R. E. and Merrick, W. C. (1987) The ATP-dependent interaction of eukaryotic initiation factors with mRNA, *J Biol Chem*, **262**, 3826-3832.
- Adam, S. A., Lobl, T. J., Mitchell, M. A. and Gerace, L. (1989) Identification of specific binding proteins for a nuclear location sequence, *Nature*, **337**, 276-279.
- Altmann, M., Handschin, C. and Trachsel, H. (1987) mRNA cap-binding protein: cloning of the gene encoding protein synthesis initiation factor eIF-4E from *Saccharomyces cerevisiae*, *Mol Cell Biol*, **7**, 998-1003.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1987) Current Protocols in Molecular Biology, Wiley-Interscience, New York.
- Bai, Y., Sheer, D., Hiorns, L., Knowles, R. W. and Tunnacliffe, A. (1982) A monoclonal antibody recognizing a cell surface antigen coded for by a gene on human chromosome 17, *Ann Hum Genet*, **46**, 337-347.
- Beach, D. and Nurse, P. (1981) High frequency transformation of the fission yeast *Schizosaccharomyces pombe*, *Nature*, **290**, 140-142.
- Bell, L. R., Maine, E. M., Schedl, P. and Cline, T. W. (1988) Sex-lethal, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins, *Cell*, **55**, 1037-1046.
- Blow, J. J. and Laskey, R. A. (1988) A role for the nuclear envelope in controlling DNA replication within the cell cycle, *Nature*, **332**, 546-548.
- Borer, R. A., Lehner, C. F., Eppenberger, H. M. and Nigg, E. A. (1989) Major nucleolar proteins shuttle between the nucleus and cytoplasm, *Cell*, **56**, 379-390.
- Brandziulis, R. J., Swanson, M. S. and Dreyfuss, G. (1989) RNA binding proteins as developmental regulators, *Genes Devel*, **3**, 431-437.
- Bravo, R., Fey, S. J., Bellatin, J., Larsen, P. M., Arevalo, J. and Celis, J. E. (1981) Identification of a nuclear and of a cytoplasmic

polypeptide whose relative proportion are sensitive to changes in the rate of cell proliferation, *Exp Cell Res*, **136**, 311-319.

Brinkmann, U., Mattes, R. E. and Buckel, P. (1989) High-level expression of recombinant genes in *E.coli* is dependent on the availability of the *dnaY* gene product, *Gene*, **85**, 109-114.

Bugler, B., Bourbon, H., Lapeyre, B., Wallace, M. O., Chang, J.-H., Amalric, F. and Olson, M. O. J. (1987) RNA binding fragments from nucleolin contain the ribonucleoprotein consensus sequence, *J Biol Chem*, **262**, 10922-10925.

Buluwela, L., Forster, A., Boehm, T. and Rabbitts, T. H. (1989) A rapid procedure for colony screening using nylon filters, *Nucl Acids Res*, **17**, 452.

Burgess, S., Couto, J. R. and Guthrie, C. (1990) A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing, *Cell*, **60**, 705-717.

Carmo-Fonseca, M., Tollervey, D., Pepperkok, R., Barabino, S. M. L., Merdes, A., Brunner, C., Zamore, P. D., Green, M. R., Hurt, E. and Lamond, A. I. (1991) Mammalian nuclei contain foci which are highly enriched in components of the pre-mRNA splicing machinery, *EMBO J*, **10**, 195-206.

Carr, A. M., MacNeill, S. A., Hayles, J. and Nurse, P. (1989) Molecular cloning and sequence analysis of mutant alleles of the fission yeast *cdc2* protein kinase gene: implications for *cdc2+* protein structure and function, *Mol Gen Genet*, **218**, 41-49.

Carroll, R. B., Samad, A., Mann, A., Harper, J. and Anderson, C. W. (1988) RNA is covalently linked to SV40 large T antigen, *Oncogene*, **2**, 437-444.

Cashel, M. and Rudd, K. E. (1987) The stringent response, in F. C. Neidhardt (ed.), *Cellular and Molecular Biology of E.coli and S.typhi*, **2**, ASM, pp. 1410-1438.

Chang, T.-H., Arenas, J. and Abelson, J. (1990) Identification of five putative yeast RNA helicase genes, *Proc Natl Acad Sci USA*, **87**, 1571-1575.

Chen, J. W. and Lin, R. J. (1990) The yeast PRP2 protein, a putative RNA-dependent ATPase, shares extensive sequence homology with two other pre-mRNA splicing factors, *Nucl Acids Res*, **18**, 6447.

- Cheng, J., Fogel-Petrovic, M. and Maquat, L. E. (1990) Translation to near the distal end of the penultimate exon is required for normal levels of spliced triosephosphate isomerase mRNA, *Mol Cell Biol*, **10**, 5215-5225.
- Clark, R., Lane, D. P. and Tjian, R. (1981) Use of monoclonal antibodies as probes of simian virus 40 T antigen ATPase activity, *J Biol Chem*, **256**, 11854-11858.
- Cobianchi, F., Sengupta, D. N., Zmudzka, B. Z. and Wilson, S. H. (1986) Structure of the rodent helix-destabilizing protein revealed by cDNA cloning, *J Biol Chem*, **261**, 3536-3543.
- Coles, C., Thompson, A. M., Elder, P. A., Cohen, B. B., Mackenzie, I. M., Cranston, G., Chetty, U., Mackay, J., MacDonald, M., Nakamura, Y., Hoyheim, B. and Steel, C. M. (1990) Evidence implicating at least two genes on chromosome 17p in breast carcinogenesis, *Lancet*, **336**, 761-763.
- Company, M., Arenas, J. and Abelson, J. (1991) Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes, *Nature*, **349**, 487-493.
- Crawford, L., Leppard, K., Lane, D. and Harlow, E. (1982) Cellular proteins reactive with monoclonal antibodies directed against simian virus 40 T-antigen, *J Virol*, **42**, 612-620.
- Daar, I. O. and Manquat, L. E. (1988) Premature translation termination mediates triosephosphate isomerase mRNA degradation, *Mol Cell Biol*, **8**, 802-813.
- Dabeva, M. D., Post, B. M. and Warner, J. R. (1986) Autogenous regulation of splicing of the transcript of a yeast ribosomal protein gene, *Proc Natl Acad Sci USA*, **83**, 5854-5857.
- Dalbadie-McFarland, G. and Abelson, J. (1990) PRP5: A helicase-like protein required for mRNA splicing in yeast, *Proc Natl Acad Sci USA*, **87**, 4236-4240.
- Dang, C. V. and Lee, W. M. F. (1989) Nuclear and nucleolar targeting sequences of c-erb-A, c-myb, N-myc, p53, HSP70 and HIV tat proteins, *J. Biol. Chem.*, **264**, 18019-18023.
- Darlix, J. L., Khandjian, E. W. and Weil, R. (1984) Nature and origin of the RNA associated with simian virus 40 large tumor antigen, *Proc Natl Acad Sci USA*, **81**, 5425-5429.
- DeCaprio, J. A., Ludlow, J. W., Figge, J., Shew, J.-Y., Huang, C.-M., Lee, W.-H., Marsillo, E., Paucha, E. and Livingston, D. M. (1988)

SV40 large tumor antigen forms a specific complex with the product of the retinoblastoma susceptibility gene, *Cell*, **54**, 275-283.

Delahodde, A., Goguel, V., Becam, A. M., Creusot, F., Perea, J., Banroques, J. and Jacq, C. (1989) Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria, *Cell*, **56**, 431-441.

Delsert, C., Morin, N. and Klessig, D. F. (1989) cis-acting elements and a trans-acting factor affecting alternative splicing of adenovirus L1 transcripts, *Mol Cell Biol*, **9**, 4364-4371.

Demple, B. and Halbrook, J. (1983) Inducible repair of oxidative DNA damage in *Escherichia coli*, *Nature*, **304**, 466-468.

Dever, T. E., Glynias, M. J. and Merrick, W. C. (1987) GTP-binding domain: three consensus sequence elements with distinct spacing, *Proc Natl Acad Sci USA*, **84**, 1814-1818.

Devereux, J., Haeberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX, *Nucl Acids Res*, **12**, 387-395.

Dingwall, C. (1990) Protein structure: plugging the nuclear pore, *Nature*, **346**, 512-514.

Dorer, D. R., Christensen, A. C. and Johnson, D. H. (1990) A novel RNA helicase gene tightly linked to the *triplo-lethal* locus of *Drosophila*, *Nucl Acids Res*, **18**, 5489-5494.

Doudna, J. A. and Szostak, J. W. (1989) RNA-catalysed synthesis of complementary-strand RNA, *Nature*, **339**, 519-522.

Duesberg, P. H. (1987) Latent cellular oncogenes: the paradox dissolves, *J Cell Sci Suppl*, **7**, 169-87.

Dyson, N., Buchkovich, K., Whyte, P. and Harlow, E. (1989) The cellular 107K protein that binds to adenovirus E1A also associates with the large T antigens of SV40 and JC virus, *Cell*, **58**, 249-55.

Eccles, D. M., Cranston, G., Steel, C. M., Nakamura, Y. and Leonard, R. C. F. (1990) Allele losses on chromosome 17 in human epithelial ovarian carcinoma, *Oncogene*, **5**, 1599-1601.

Ellington, A. D. and Szostak, J. W. (1990) *In vitro* selection of RNA molecules that bind specific ligands, *Nature*, **346**, 818-822.

England, T. E., Bruce, A. G. and Uhlenbeck, O. C. (1980) Specific labeling of 3' termini of RNA with T4 RNA ligase, *Meth Enzymol*, **65**, 65-74.

Ewen, M. E., Ludlow, J. W., Marsilio, E., DeCaprio, J. A., Millikan, R. C., Cheng, S. H., Paucha, E. and Livingston, D. M. (1989) An N-terminal transformation-governing sequence of SV40 large T antigen contributes to the binding of both p110Rb and a second cellular protein, p120, *Cell*, **58**, 257-267.

Fickett, J. W. (1982) Recognition of protein coding regions in DNA sequences, *Nucl Acids Res*, **10**, 5303-5318.

Fink, G. R. (1987) Pseudogenes in yeast?, *Cell*, **49**, 5-6.

Ford, M. J., Anton, I. A. and Lane, D. P. (1988) Nuclear protein with sequence homology to translation initiation factor eIF4A, *Nature*, **332**, 736-738.

Frohman, R. A., Dush, M. K. and Martin, G. R. (1988) Rapid production of full length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer, *Proc Natl Acad Sci USA*, **85**, 8998-9002.

Fry, D. C., Kuby, S. A. and Mildvan, A. S. (1986) ATP-binding site of adenylate kinase: mechanistic implications of its homology with ras-encoded p21, F1-ATPase, and other nucleotide-binding proteins, *Proc Natl Acad Sci USA*, **83**, 907-911.

Fu, L., Ye, R., Browder, L. W. and Johnston, R. N. (1991) Translational potentiation of messenger RNA with secondary structure in *Xenopus*, *Science*, **251**, 807-810.

Gannon, J. V., Greaves, R., Iggo, R. and Lane, D. P. (1990) Activating mutations in p53 produce a common conformational effect. A monoclonal antibody specific for the mutant form, *EMBO J*, **9**, 1595-1602.

Gannon, J. V. and Lane, D. P. (1987) p53 and DNA polymerase α compete for binding to SV40 T antigen, *Nature*, **329**, 456-458.

Garcia, G. M., Mar, P. K., Mullin, D. A., Walker, J. R. and Prather, N. E. (1986) The *E.coli dnaY* gene encodes an arginine transfer RNA, *Cell*, **45**, 453-459.

Gateman, K. B., Hoffman, A., Rosenberg, G. H. and Kaufer, N. F. (1989) Introduction of functional artificial introns into the naturally intronless *ura4* gene of *Schizosaccharomyces pombe*, *Mol Cell Biol*, **9**, 1526-1535.

- Gilbert, W., Marchionni, M. and McKnight, G. (1986) On the antiquity of introns, *Cell*, **46**, 151-153.
- Gorbalenya, A. E. and Koonin, E. V. (1989) Viral proteins containing the purine NTP-binding sequence pattern, *Nucl Acids Res*, **17**, 8413-8440.
- Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. and Blinov, V. M. (1988) A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination, *Nature*, **333**, 22.
- Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. and Blinov, V. M. (1989) Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes, *Nucl Acids Res*, **17**, 4713-4730.
- Gorbalenya, A. E., Koonin, E. V. and Wolf, Y. I. (1990) A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses, *Febs Lett*, **262**, 145-148.
- Gorman, C., Gies, D. R. and McCray, G. (1990) Transient production of proteins using an adenovirus transformed cell line, *DNA Prot Eng Tech*, **2**, 3-10.
- Gott, J. M., Zeeh, A., Bell, P. D., Ehrenman, K., Belfort, M. and Shub, D. A. (1988) Genes within genes: independent expression of phage T4 intron open reading frames and the genes in which they reside, *Genes Dev*, **2**, 1791-1799.
- Graham, F. L., Smiley, J., Russell, W. C. and Nairn, R. (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5, *J Gen Virol*, **36**, 59-74.
- Grifo, J. A., Abramson, R. D., Satler, C. A. and Merrick, W. C. (1984) RNA-stimulated ATPase activity of eukaryotic initiation factors, *J Biol Chem*, **259**, 8648-8654.
- Gururajan, R., Perry-O'Keefe, H., Melton, D. A. and Weeks, D. L. (1991) The *Xenopus* localized messenger RNA An3 may encode an ATP-dependent RNA helicase, *Nature*, **349**, 717-719.
- Gyllensten, U. B. and Erlich, H. A. (1988) Generation of single stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus, *Proc Natl Acad Sci USA*, **85**, 7652-7656.

Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B. and King, M. (1990) Linkage of early-onset familial breast cancer to chromosome 17q21, *Science*, **250**, 1684-1689.

Harlow, E. and Lane, D. P. (1988) *Antibodies: A laboratory manual*, Cold Spring Harbor Laboratory Press, New York.

Hartwell, L. H. (1967) Macromolecule synthesis in temperature-sensitive mutants of yeast, *J Bacteriol*, **93**, 1662-1670.

Hawkins, J. D. (1988) A survey on intron and exon lengths, *Nucl Acids Res*, **16**, 9893-9908.

Hay, B., Ackerman, L., Barbel, S., Jan, L. Y. and Jan, Y. N. (1988a) Identification of a component of *Drosophila* polar granules, *Development*, **103**, 625-640.

Hay, B., Jan, L. Y. and Jan, Y. N. (1988b) A protein component of *Drosophila* polar granules is encoded by vasa and has extensive sequence similarity to ATP-dependent helicases, *Cell*, **55**, 577-587.

Hay, B., Jan, L. Y. and Jan, Y. N. (1990) Localization of vasa, a component of *Drosophila* polar granules, in maternal-effect mutants that alter embryonic anteroposterior polarity, *Development*, **109**, 425-433.

Hay, N., Skolnik, D. H. and Aloni, Y. (1982) Attenuation in the control of SV40 gene expression, *Cell*, **29**, 183-193.

Herschlag, D. and Cech, T. R. (1990) DNA cleavage catalysed by the ribozyme from Tetrahymena, *Nature*, **344**, 405-409.

Hickson, I. D., Robson, C. N., Atkinson, K. E., Hutton, L. and Emmerson, P. T. (1985) Reconstitution of RecBC DNase activity from purified *Escherichia coli* RecB and RecC proteins, *J Biol Chem*, **260**, 1224-1229.

Hirling, H., Scheffner, M., Restle, T. and Stahl, H. (1989) RNA helicase activity associated with the human p68 protein, *Nature*, **339**, 562-564.

Hloch, P., Schneider, G. and Stahl, H. (1990) p68 cDNA sequence, *Nucl Acids Res*, **18**, 3045.

Hodgman, T. C. (1988) A new superfamily of replicative proteins, *Nature*, **333**, 22-23.

Hunt, T. (1989) Cytoplasmic anchoring proteins and the control of nuclear localization, **59**, 949-951, *Cell*.

Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (1990) PCR protocols, Academic Press, San Diego.

- Huynh, T. V., Young, R. A. and Davis, R. W. (1985) Constructing and screening cDNA libraries in λ gt11 and λ gt10, in D. M. Glover (ed.), DNA Cloning I, IRL Press, pp. 49-78.
- Iggo, R. D. and Lane, D. P. (1989) Nuclear protein p68 is an RNA-dependent ATPase, *EMBO J*, **8**, 1827-1831.
- Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983) Transformation of intact yeast cells with alkali cations, *J Bacteriol*, **153**, 163-168.
- Jamieson, D. J. and Beggs, J. D. (1991) A suppressor of yeast *spp81/ded1* mutations encodes a very similar putative ATP-dependent RNA helicase, *Molec Microbiol*, in press.
- Jamieson, D. J., Rahe, B., Pringle, J. and Beggs, J. D. (1991) A suppressor of a yeast splicing mutation (*prp8-1*) encodes a putative ATP-dependent RNA helicase, *Nature*, **349**, 715-717.
- Jan, L. Y. and Jan, Y. N. (1982) Antibodies to horseradish peroxidase as specific neuronal markers in *Drosophila* and in grasshopper embryos, *Proc Natl Acad Sci USA*, **72**, 2700-2704.
- Jensen, F.C., Girardi, A.J., Gilden, R.V., Koprowski, H. (1964) Infection of human and simian tissue cultures with Rous sarcoma virus, *Proc Natl Acad Sci USA*, **52**, 53-59.
- Jaramillo, M., Browning, K., Dever, T. E., Blum, S., Trachsel, H., Merrick, W. C., Ravel, J. M. and Sonenberg, N. (1990) Translation initiation factors that function as RNA helicases from mammals, plants and yeast, *Biochim Biophys Acta*, **1050**, 134-139.
- Jones, E. A., Goodfellow, P., Kennett, R. H. and Bodmer, W. F. (1976) Independent expression of HLA and beta2 microglobulin on human-mouse hybrids, *Cytogenet Cell Genet*, **16**, 161-163.
- Jong, A., Clark, M. W., Gilbert, M., Oehm, A. and Campbell, J. L. (1987) *Saccharomyces cerevisiae* SSB1 protein and its relationship to nucleolar RNA binding proteins, *Mol Cell Biol*, **7**, 2947-2955.
- Joshi, B. S., Rychlik, W. and Rhoads, R. E. (1990) Alteration of the major phosphorylation site of eukaryotic protein synthesis initiation factor 4E prevents its association with the 48S initiation complex, *J Biol Chem*, **265**, 2979-2983.
- Kay, L. E., Clore, G. M., Bax, A. and Gronenborn, A. M. (1990) Four-dimensional heteronuclear triple-resonance NMR

spectroscopy of interleukin-1 beta in solution, *Science*, **249**, 411-414.

Khandjian, E. W., Loche, M., Darlix, J. L., Cramer, R., Turler, H. and Weil, R. (1982) Simian virus 40 large tumor antigen: a "RNA binding protein"?, *Proc Natl Acad Sci USA*, **79**, 1139-1143.

Kohara, Y., Akiyama, K. and Isono, K. (1987) The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library, *Cell*, **50**, 495-508.

Kornberg, A., Scott, J. F. and Bertsch, L. L. (1978) ATP utilization by rep protein in the catalytic separation of DNA strands at a replicating fork, *J Biol Chem*, **253**, 3298-3304.

Kumura, K. and Sekiguchi, M. (1984) Identification of the *uvrD* gene product of *Escherichia coli* as DNA helicase II and its induction by DNA-damaging agents, *J Biol Chem*, **259**, 1560-1565.

Lachmann, P. J., Strangeways, L., Vyakarnam, A. and Evan, G. (1986) Raising antibodies by coupling peptides to PPD and immunising BCG-sensitised animals, in *Ciba Symposium 119, Synthetic peptides as antigens*, Wiley, pp. 25-57.

Lain, S., Martin, M. T., Riechmann, J. L. and Garcia, J. A. (1991) Novel catalytic activity associated with positive-strand RNA virus infection: Nucleic acid-stimulated ATPase activity of the plum pox potyvirus helicase-like protein, *J Virol*, **65**, 1-6.

Lane, D. P. and Crawford, L. V. (1979) T-antigen is bound to a host protein in SV40-transformed cells, *Nature*, **278**, 261-263.

Lane, D. P. and Hoeffler, W. K. (1980) SV40 large T shares an antigenic determinant with a cellular protein of molecular weight 68,000, *Nature*, **288**, 167-170.

Lapeyre, B., Bourbon, H. and Amalric, F. (1987) Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence, *Proc Natl Acad Sci USA*, **84**, 1472-1476.

Lasko, P. F. and Ashburner, M. (1988) The product of the *Drosophila* gene *vasa* is very similar to eukaryotic initiation factor-4A, *Nature*, **335**, 611-617.

- Lasko, P. F. and Ashburner, M. (1990) Posterior localization of vasa protein correlates with, but is not sufficient for, pole cell development, *Genes Dev*, **4**, 905-921.
- Lazaris, K. A., Montine, K. S. and Sonenberg, N. (1990) Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap, *Nature*, **345**, 544-7.
- Lerner, E. A., Lerner, M. R., Janeway, C. A. and Steitz, J. A. (1981) Monoclonal antibodies to nucleic acid-containing cellular constituents: Probes for molecular biology and autoimmune diseases, *Proc Natl Acad Sci USA*, **78**, 2737-2741.
- Leroy, P., Alzari, P., Sassoon, D., Wolgemuth, D. and Fellous, M. (1989) The protein encoded by a murine male germ cell specific transcript is a putative ATP-dependent RNA helicase, *Cell*, **57**, 549-559.
- Lin, R. J., Lustig, A. J. and Abelson, J. (1987) Splicing of yeast nuclear pre-mRNA *in vitro* requires a functional 40S spliceosome and several extrinsic factors, *Genes Dev*, **1**, 7-18.
- Linder, P., Lasko, P. F., Ashburner, M., Leroy, P., Nielsen, P. J., Nishi, K., Schnier, J. and Slonimski, P. P. (1989) Birth of the D-E-A-D box, *Nature*, **337**, 121-122.
- Linder, P. and Slonimski. (1989) An essential yeast protein, encoded by duplicated genes *TIF1* and *TIF2* and homologous to the mammalian translation initiation factor eIF-4A, can suppress a mitochondrial missense mutation, *Proc Natl Acad Sci USA*, **86**, 2286-2290.
- Linder, P. and Slonimski, P. P. (1988) Sequence of the genes *TIF1* and *TIF2* from *Saccharomyces cerevisiae* coding for a translation initiation factor, *Nucl Acids Res*, **16**, 10359.
- Linzer, D. I. H. and Levine, A. J. (1979) Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40 transformed cells and uninfected embryonal carcinoma cells, *Cell*, **17**, 43-52.
- Lossky, M., Anderson, G. J., Jackson, S. P. and Beggs, J. (1987) Identification of a yeast snRNP protein and detection of snRNP-snRNP interactions, *Cell*, **51**, 1019-1026.
- Martin, D., Tucker, D. F., Gorman, P., Sheer, D., Spurr, N. K. and Trowsdale, J. (1987) The human placental alkaline phosphatase gene and related sequences map to chromosome 2 band q37, *Ann Hum Genet*, **51**, 145-152.

- Mastrangelo, I. A., Hough, P. V., Wall, J. S., Dodson, M., Dean, F. B. and Hurwitz, J. (1989) ATP-dependent assembly of double hexamers of SV40 T antigen at the viral origin of DNA replication, *Nature*, **338**, 658-662.
- McCafferty, J., Griffiths, A. D., Winter, G. and Chiswell, D. J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains, *Nature*, **348**, 552-554.
- McCrae, M. A. and Joklik, W. K. (1978) The nature of the polypeptides encoded by each of the 10 double-stranded RNA segments of reovirus type 3, *Virology*, **89**, 578-593.
- Merrick, W. C., Dever, T. E., Kinzy, T. G., Conroy, S. C., Cavallius, J. and Owens, C. L. (1990) Characterization of protein synthesis factors from rabbit reticulocytes, *Biochim Biophys Acta*, **1050**, 235-240.
- Michalovitz, D., Halvey, O. and Oren, M. (1990) Conditional inhibition of transformation and of cell proliferation by a temperature-sensitive mutant of p53, *Cell*, **62**, 671-680.
- Michel, F. and Lang, B. F. (1985) Mitochondrial type II introns encode proteins related to the reverse transcriptases of retroviruses, *Nature*, **316**, 641-643.
- Michel, M. R. and Schwyzer, M. (1982) Messenger ribonucleoproteins of cells infected by simian virus 40 contain large T-antigen, *Eur J Biochem*, **129**, 25-32.
- Milburn, S. C., Hershey, J. W., Davies, M. V., Kelleher, K. and Kaufman, R. J. (1990) Cloning and expression of eukaryotic initiation factor 4B cDNA: sequence determination identifies a common RNA recognition motif, *EMBO J*, **9**, 2783-2790.
- Miller, A. M. (1984) The yeast *MATa1* gene contains two introns, *EMBO J*, **3**, 1061-1065.
- Miller, M. A., Korn, D. and Wang, T. S.-F. (1988) The evolutionary conservation of DNA polymerase α , *Nucl Acids Res*, **16**, 7961-7973.
- Mitchell, P. J., Wang, C. and Tjian, R. (1987) Positive and negative regulation of transcription *in vitro*: enhancer-binding protein AP-2 is inhibited by SV40 T antigen, *Cell*, **50**, 847-861.
- Molenaar, C. M. T., Woudt, L. P., Jansen, A. E. M., Mager, W. H., Planta, R. J., Donovan, D. M. and Pearson, N. J. (1984) Structure

- and organisation of two linked ribosomal protein genes in yeast, *Nucl Acids Res*, **12**, 7345-7358.
- Moreno, S., Klar, A. and Nurse, P. (1990) An introduction to molecular genetic analysis of the fission yeast, *Meth Enzymol*, **194**, 793-823.
- Nielsen, P. J., McMaster, G. K. and Trachsel, H. (1985) Cloning of eukaryotic protein synthesis initiation factor genes: isolation and characterization of cDNA clones encoding factor eIF-4A, *Nucl Acids Res*, **13**, 6867-6880.
- Nielsen, P. J. and Trachsel, H. (1988) The mouse protein synthesis initiation factor 4A gene family includes two related functional genes which are differentially expressed, *EMBO J*, **7**, 2097-2105.
- Nishi, K., Morel-Deville, F., Hershey, J. W. B., Leighton, T. and Schnier, J. (1988) An eIF4A-like protein is a suppressor of an *Escherichia coli* mutant defective in 50S ribosomal subunit assembly, *Nature*, **336**, 496-498.
- Nunn, W. D. (1987) Two-carbon compounds and fatty acids as carbon sources, in F. C. Neidhardt (ed.), *Cellular and Molecular Biology of E.coli and S.typhi*, **1**, ASM, pp. 285-301.
- Ochs, R., Lischwe, M., O'Leary, P. and Busch, H. (1983) Localisation of nucleolar phosphoproteins B23 and C23 during mitosis, *Exp Cell Res*, **146**, 139-149.
- Ochs, R. L., Lischwe, M. A., Shen, E., Carroll, R. E. and Busch, H. (1985) Nucleologenesis: Composition and fate of prenucleolar bodies, *Chromosoma*, **92**, 330-336.
- Polisky, B. (1988) ColE1 replication control circuitry: sense from antisense, *Cell*, **55**, 929-932.
- Potashkin, J. A., Derby, R. J. and Spector, D. L. (1990) Differential distribution of factors involved in pre-mRNA processing in the yeast cell nucleus, *Mol Cell Biol*, **10**, 3524-3534.
- Povey, S., Jeremiah, S. J., Barker, R. F., Hopkinson, D. A., Robson, E. B., Cook, P. J., Solomon, E., Bobrow, M., Carritt, B. and Buckton, K. E. (1980) Assignment of the human locus determining phosphoglycolate phosphatase (PGP) to chromosome 16, *Ann Hum Genet*, **43**, 241-248.
- Prat, A., Schmid, S. R., Buser, P., Blum, S., Trachsel, H., Nielsen, P. J. and Linder, P. (1990) Expression of translation initiation factor

4A from yeast and mouse in *Saccharomyces cerevisiae*, *Biochim Biophys Acta*, **1050**, 140-145.

Prives, C. (1990) The replication functions of SV40 T antigen are regulated by phosphorylation, *Cell*, **61**, 735-738.

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) ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors, *J Biol Chem*, **260**, 7651-7658.

Rebagliati, M. R., Weeks, D. L., Harvey, R. P. and Melton, D. A. (1985) Identification and cloning of localized maternal RNAs from *Xenopus* eggs, *Cell*, **42**, 769-777.

Rhoads, R. E. (1988) Cap recognition and the entry of mRNA into the protein synthesis initiation cycle, *Trends Biochem Sci*, **13**, 52-56.

Robbins, J., Dilworth, S. M., Laskey, R. A. and Dingwall, C. (1991) Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: Identification of a class of bipartite nuclear targeting sequence, *Cell*, **64**, 615-623.

Rothstein, R. J. (1983) One step gene disruption in yeast, *Meth Enzymol*, **101**, 202-211.

Rozen, F., Edery, I., Meerovitch, K., Dever, T. E., Merrick, W. C. and Sonenberg, N. (1990) Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F, *Mol Cell Biol*, **10**, 1134-1144.

Russell, P. and Nurse, P. (1986) *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*: A look at yeasts divided, *Cell*, **45**, 781-782.

Russell, S. E. H., Hickey, G. I., Lowry, W. S., White, P. and Atkinson, R. J. (1990) Allele loss from chromosome 17 in ovarian cancer, *Oncogene*, **5**, 1581-1583.

Ruther, U. and Muller, H. B. (1983) Easy identification of cDNA clones, *EMBO J*, **2**, 1791-1794.

Rychlik, W., Domier, L. L., Gardner, P. R., Hellmann, G. M. and Rhoads, R. E. (1987) Amino acid sequence of the mRNA cap-binding protein from human tissues, *Proc Natl Acad Sci USA*, **84**, 945-949.

Sachs, A. B. and Davis, R. W. (1989) The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation, *Cell*, **58**, 857-867.

Sachs, A. B. and Davis, R. W. (1990) Translation initiation and ribosomal biogenesis: involvement of a putative rRNA helicase and RPL46, *Science*, **247**, 1077-1079.

Sachs, A. B., Davis, R. W. and Kornberg, R. D. (1987) A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability, *Mol Cell Biol*, **7**, 3268-3276.

Saiki, R. K., Gelfand, D. H., Stoffel, B., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988) Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase, *Science*, **239**, 487-490.

Samad, A. and Carroll, R. B. (1991) The tumour suppressor p53 is bound to RNA by a stable covalent linkage, *Mol Cell Biol*, **11**, 1598-1606.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press, New York.

Scheffner, M., Knippers, R. and Stahl, H. (1989) RNA unwinding activity of SV40 large T antigen, *Cell*, **57**, 955-963.

Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J. and Howley, P. M. (1991) The E6 oncoprotein encoded by the human papillomavirus types 16 and 18 promotes degradation of p53, *Cell*, **63**, 1129-1136.

Schimmang, T., Tollervey, D., Kern, H., Frank, R. and Hurt, E. C. (1989) A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability, *EMBO J*, **8**, 4015-4024.

Schlichting, I., Almo, S. C., Rapp, G., Wilson, K., Petratos, K., Lentfer, A., Wittinghofer, A., Kabsch, W., Pai, E. F., Petsko, G. A. and Goody, R. S. (1990) Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis, *Nature*, **345**, 309-315.

Schmitt, M. E., Brown, T. A. and Trumppower, B. L. (1990) A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*, *Nucl Acids Res*, **18**, 3091-3092.

Smith, D.B. and Johnson, K.S. (1988) Single-step purification of polypeptides expressed in *E.coli* as fusions with glutathione S-transferase, *Gene*, **67**, 31-40.

- Schwer, B. and Guthrie, C. (1991) PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome, *Nature*, **349**, 494-499.
- Seraphin, B., Simon, M., Boulet, A. and Faye, G. (1989) Mitochondrial splicing requires a protein from a novel helicase family, *Nature*, **337**, 84-87.
- Sherman, F., Fink, G. R. and Hicks, J. B. (1982) *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory Press, New York.
- Shub, D. A., Xu, M. Q., Gott, J. M., Zeeh, A. and Wilson, L. D. (1987) A family of autocatalytic group I introns in bacteriophage T4, *Cold Spring Harbor Symp Quant Biol*, **52**, 193-200.
- Sikorski, R. S. and Hieter, P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*, *Genetics*, **122**, 19-27.
- Simanis, V. and Lane, D. P. (1985) An immunoaffinity purification procedure for SV40 large T antigen, *Virology*, **144**, 88-100.
- Siomi, H., Shida, H., Nam, S. H., Nosaka, T., Maki, M. and Hanataka, M. (1988) Sequence requirements for nucleolar localisation of human T cell leukaemia virus type I pX protein, which regulates viral RNA processing, *Cell*, **55**, 197-209.
- Smale, S. T. and Tjian, R. (1986) T-antigen-DNA polymerase α complex implicated in simian virus 40 replication, *Mol Cell Biol*, **6**, 4077-4087.
- Solomon, E., Bobrow, M., Goodfellow, P. N., Bodmer, W. F., Swallow, D. M., Povey, S. and Noel, B. (1976) Human gene mapping using an X/autosome translocation, *Somatic Cell Genet*, **2**, 125-140.
- Solomon, E., Swallow, D., Burgess, S. and Evans, L. (1979) Assignment of the human acid alpha-glucosidase gene (alphaGLU) to chromosome 17 using somatic cell hybrids, *Ann Hum Genet*, **42**, 273-281.
- Sonenberg, N. (1988) Cap-binding proteins of eukaryotic messenger RNA: functions in initiation and control of translation, *Progr Nucl Acids Res Mol Biol*, **35**, 173-207.
- Spector, D. L., Ochs, R. L. and Busch, H. (1984) Silver staining, immunofluorescence and immunoelectron microscopic

localization of nucleolar phosphoproteins B23 and C23, *Chromosoma*, **90**, 139-148.

Spurr, N. K., Durbin, H., Sheer, D., Parkar, M., Bobrow, L. and Bodmer, W. F. (1986) Characterization and chromosomal assignment of a human cell surface antigen defined by the monoclonal antibody AUAI, *Int J Cancer*, **38**, 631-636.

Stahl, H., Droge, P. and Knippers, R. (1986) DNA helicase activity of SV40 large tumor antigen, *EMBO J*, **5**, 1939-1944.

Stoker, N. G., Grant, K. A., Dockrell, H. M., Howard, C. R., Jouy, N. F. and McAdam, K. P. (1989) High level expression of genes cloned in phage lambda gt11, *Gene*, **78**, 93-99.

Strauss, E. J. and Guthrie, C. (1991) A cold-sensitive mRNA splicing mutant is a member of the RNA helicase gene family, *Genes Dev*, in press.

Struhl, K. (1985) Nucleotide sequence and transcriptional mapping of the yeast *pet56-his3-ded1* gene region, *Nucl Acids Res*, **13**, 8587-8601.

Sturzbecher, H.-W., Brain, R., Maimets, T., Addison, C., Rudge, K. and Jenkins, J. R. (1988) Mouse p53 blocks SV40 DNA replication in vitro and downregulates T antigen DNA helicase activity, *Oncogene*, **3**, 405-413.

Sung, P., Prakash, L., Matson, S. W. and Prakash, S. (1987) RAD3 protein of *Saccharomyces cerevisiae* is a DNA helicase, *Proc Natl Acad Sci USA*, **84**, 8951-8955.

Swallow, D. M., Solomon, E. and Pajunen, L. (1977) Immunochemical analysis of the N-acetyl hexosaminidases in human-mouse hybrids made using a double selective system, *Cytogenet Cell Genet*, **18**, 136-148.

Tabor, S. and Richardson, C. C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes, *Proc Natl Acad Sci USA*, **82**, 1074-1078.

Tack, L. C., Wright, J. H., Deb, S. P. and Tegtmeyer, P. (1989) The p53 complex from monkey cells modulates the biochemical activities of simian virus 40 large T antigen, *J Virol*, **63**, 1310-1317.

Triglia, T., Peterson, M. G. and Kemp, D. J. (1988) A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences, *Nucl Acids Res*, **16**, 8186.

- Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, *Science*, **249**, 505-510.
- Van Heyningen, V., Bobrow, M., Bodmer, W. F., Gardiner, S. E., Povey, S. and Hopkinson, D. A. (1975) Chromosome assignment of some human enzyme loci: mitochondrial malate dehydrogenase to 7, mannosephosphate isomerase and pyruvate kinase to 15 and probably, esterase D to 13, *Ann Hum Genet*, **38**, 295-303.
- Vijayraghavan, U., Company, M. and Abelson, J. (1989) Isolation and characterization of pre-mRNA splicing mutants of *Saccharomyces cerevisiae*, *Genes Dev*, **3**, 1206-1216.
- Voss, R., Lerer, I., Povey, S., Solomon, E. and Bobrow, M. (1980) Confirmation and further regional assignment of aminoacylase 1 (acy-1) on human chromosome 3 using a simplified detection method, *Ann Hum Genet*, **44**, 1-9.
- Walker, J. E., Saraste, M. J., Runswick, J. J. and Gay, N. J. (1982) Distantly related sequences in the alpha and beta subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold, *EMBO J*, **1**, 945-951.
- Wang, E. H., Friedman, P. N. and Prives, C. (1989) The murine p53 protein blocks replication of SV40 DNA in vitro by inhibiting the initiation functions of SV40 large T antigen, *Cell*, **57**, 379-392.
- Warner, J. R. (1989) Synthesis of ribosomes in *Saccharomyces cerevisiae*, *Microbiol Reviews*, **53**, 256-271.
- Warner, J. R. (1990) The nucleolus and ribosome formation, *Current Opinion in Cell Biology*, **2**, 521-527.
- Waseem, N. H. and Lane, D. P. (1990) Monoclonal antibody analysis of the proliferating cell nuclear antigen (PCNA). Structural conservation and the detection of a nucleolar form, *J. Cell Sci.*, **96**, 121-129.
- Wassarman, D. A. and Steitz, J. A. (1991) RNA Splicing: Alive with DEAD proteins, *Nature*, **349**, 463-464.
- Williams, J. G. and Mason, P. J. (1985) Hybridisation in RNA analysis, in D. B. Hames and S. J. Higgins (eds), *Nucleic Acid hybridisation: A practical approach*, IRL press, pp. 152-155.

- Winston, F., Chaleff, D. T., Valent, B. and Fink, G. R. (1984) Mutations affecting Ty-mediated expression of the HIS4 gene of *Saccharomyces cerevisiae*, *Genetics*, **107**, 179-197.
- Wood, E. R. and Matson, S. W. (1989) The molecular cloning of the gene encoding the *Escherichia coli* 75-kDa helicase and the determination of its nucleotide sequence and genetic map position, *J Biol Chem*, **264**, 8297-8303.
- Xu, W. M., Gorman, P. A., Rider, S. H., Hedge, P. J., Moore, G., Prichard, C., Sheer, D. and Solomon, E. (1988) Construction of a genetic map of human chromosome 17 by use of chromosome-mediated gene transfer, *Proc Natl Acad Sci USA*, **85**, 8563-8567.
- Yamasaki, L., Kanda, P. and Lanford, R. E. (1989) Identification of four nuclear transport signal-binding proteins that interact with diverse transport signals, *Mol Cell Biol*, **9**, 3028-3036.
- Yanofsky, C. (1981) Attenuation in the control of expression of bacterial operons, *Nature*, **289**, 751-758.
- Yung, B. Y.-M., Busch, R. K., Busch, H., Mauger, A. B. and Chan, P. K. (1985) Effects of actinomycin D analogs on nuclear phosphoprotein B23 (37 kD/pI 5.1), *Biochem Pharmacol*, **34**, 4059-4063.
- Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule, *Science*, **244**, 48-52.

Publications relevant to this thesis
RD Iggo 1989-1991

Nuclear protein p68 is an RNA-dependent ATPase

R.D.Iggo and D.P.Lane

Molecular Immunology Laboratory, ICRF, Clare Hall Laboratories, South Mimms, Herts EN6 3LD, UK

Communicated by D.P.Lane

The human nuclear antigen p68 cross reacts with a monoclonal antibody to SV40 large-T antigen. Its deduced amino acid sequence contains short motifs which place it in a large superfamily of proteins of known or putative helicase activity. Recently, a p68 subfamily (DEAD box proteins) which share more extensive regions of homology has been identified in mouse, *Drosophila*, *Saccharomyces cerevisiae* and *Escherichia coli*. These proteins are involved in translation, ribosome assembly, mitochondrial splicing, spermatogenesis and embryogenesis. We show here that immunopurified human p68 has RNA dependent ATPase activity. In addition, we show that the protein undergoes dramatic changes in cellular location during the cell cycle.

Key words: ATPase/DEAD box/helicase/p68/T antigen

Introduction

SV40 large-T antigen is a complex multi-functional protein that is absolutely required for viral DNA replication and can transform a wide range of host cells (Rigby and Lane, 1983; Livingstone and Bradley, 1987). It interacts with key host proteins to divert them from their normal function for the benefit of the virus. It interacts directly with two anti-oncogenes, p53 (Lane and Crawford, 1979) and p105 RB (DeCaprio *et al.*, 1988), to create a cellular environment favouring viral DNA replication. It binds to the viral origin of replication (Tjian, 1978) and through its intrinsic DNA helicase activity (Stahl *et al.*, 1986) unwinds the viral template (Dean *et al.*, 1987; Wold *et al.*, 1987). It forms a complex with DNA polymerase α (Smale and Tjian, 1986; Gannon and Lane, 1987) which probably serves to introduce the polymerase to the viral origin. It binds to the transcription factor AP2 (Mitchell *et al.*, 1987) and reduces *in vitro* AP2 dependent transcription. There is strong evidence that it has dramatic effects on steady state levels of both viral and cellular transcripts (Schutzbach *et al.*, 1982; Scott *et al.*, 1983; reviewed in Acheson, 1981) although the reason for this is unclear.

There must exist cellular proteins whose normal role it is to interact with p53, p105 RB, polymerase α and AP2. It is likely that these 'T-equivalent proteins' will play an important role in cell regulation. One potential way to identify them is to look for T-related antigens in the cell. We (Lane and Hoeffler, 1980; Mole *et al.*, 1987) and others (Gurney *et al.*, 1980; Harlow *et al.*, 1981; Ball *et al.*, 1984)

have generated large libraries of monoclonal antibodies against large T and used them to search for cross-reactive host proteins. One such cross reaction has been studied in detail (Lane and Hoeffler, 1980). The monoclonal antibody PAb 204 binds to a site that has been mapped with synthetic peptides to amino acids 453–469 on T antigen (Mole *et al.*, 1987). The epitope lies close to the consensus nucleotide binding motif. Antibody binding inhibits the ATPase (Clark *et al.*, 1981), DNA helicase (Stahl *et al.*, 1986) and replicative functions of large T (Smale and Tjian, 1986; Gough *et al.*, 1988). The epitope is obscured when T is bound to either DNA polymerase α or p53 and under some conditions PAb 204 may displace p53 from T (Lane and Gannon, 1986; Gannon and Lane, 1987). The antibody cross reacts very specifically with a 68 000-dalton host protein, p68, that is present in the nuclei of growing cells from a wide range of vertebrate species (Lane and Hoeffler, 1980). The antibody was used to screen λ gt11 cDNA libraries (Ford *et al.*, 1988). A cDNA was isolated whose deduced amino acid sequence is related to that of eIF4-A, a eukaryotic translation factor. Both proteins belong to a subset of the superfamily of proteins with known or suspected helicase activity described by Hodgman (1988), Lane (1988) and Gorbalenya *et al.* (1988). Recently several new members of the eIF4A subfamily have been discovered (Linder *et al.*, 1989). While genetic evidence suggests that these proteins play distinct cellular roles, their sequence homology suggests that they share a core biochemical activity as RNA helicases. We report here the isolation of p68 from human cells and demonstrate that it is an RNA dependent ATPase. Finally, we note striking changes in the cellular location of p68 during the cell cycle.

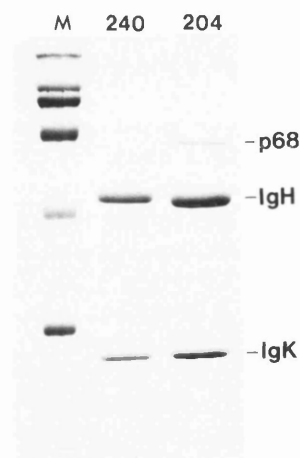


Fig. 1. SDS-polyacrylamide gel stained with Coomassie blue showing p68 immunoprecipitated from HeLa cells by monoclonal antibody PAb 204. IgH, immunoglobulin heavy chain; IgK, immunoglobulin kappa light chain; PAb 240, negative control antibody; M, relative mol. wt markers in kd: 205, 116, 97, 66, 45 and 29.

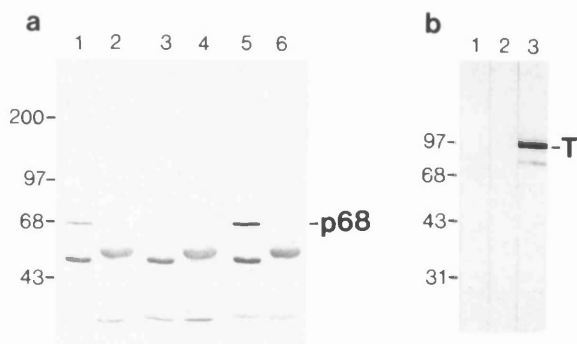


Fig. 2. (a) Immunoblot of immunoprecipitates from 293 cells probed with MF7 anti-p68 serum (lanes 1 and 2), BG2 plus normal mouse serum (lanes 3 and 4, negative controls) or PAb 204 (lanes 5 and 6). Lanes 1, 3 and 5 contain p68 immunoprecipitated by PAb 204. Lanes 2, 4 and 6 are BG2 immunoprecipitates (negative control). Immunoglobulin heavy and light chains are visible below the p68 band because alkaline phosphatase-conjugated rabbit anti-mouse immunoglobulin was used to detect the first antibody. (b) Immunoblot of SV40 large-T antigen probed with MF7 (lane 1), BG2 plus normal mouse serum (lane 2) or PAb 204 (lane 3). Relative mol. wt is indicated in kd.

Results

Immunopurification of p68

The p68 protein can be detected in a wide range of mammalian cells by immunoblotting with PAb 204 (data not shown). Initial attempts to purify p68 by immunoaffinity chromatography were unsuccessful because we lysed the cells at physiological salt concentration, which extracts much less p68 than 0.5 M salt (data not shown). In addition, PAb 204 is an IgG1 kappa mouse monoclonal antibody with a low affinity for protein A. The hybridoma was therefore grown in serum free medium, from which antibody could easily be purified either bound to protein G-Sepharose beads or by conventional means (Harlow and Lane, 1989). We immunoprecipitated p68 from concentrated HeLa and 293 cell extract with the antibody on protein G beads. Coomassie blue staining of HeLa material run on a polyacrylamide gel (Figure 1) reveals a single band at 68 kd which is absent from the control immunoprecipitate performed with a monoclonal antibody of identical isotype grown under identical conditions. The other two bands seen are immunoglobulin heavy and light chain.

Antisera to a p68 fusion protein confirm the identity of our human p68 cDNA clone

To confirm that our human cDNA clone (Ford *et al.*, 1988) is authentic p68 rather than another host protein which cross reacts with PAb 204, we immunized mice with the cloned p68 expressed as a fusion protein in *E. coli*. Antisera to the fusion protein detect a 68 kd band on immunoblots of authentic p68 immunoprecipitated from 293 cells (Figure 2a). Because the fusion protein contains the PAb 204 epitope, it is formally possible that the serum is directed only against this epitope. We exclude this possibility by showing that the serum does not recognize SV40 large-T antigen (Figure 2b).

p68 is an RNA dependent ATPase

p68 is homologous to known RNA and DNA dependent ATPases (Hodgman, 1988; Lane, 1988; Gorbalenya *et al.*, 1988) and antigenically related to T antigen, a DNA

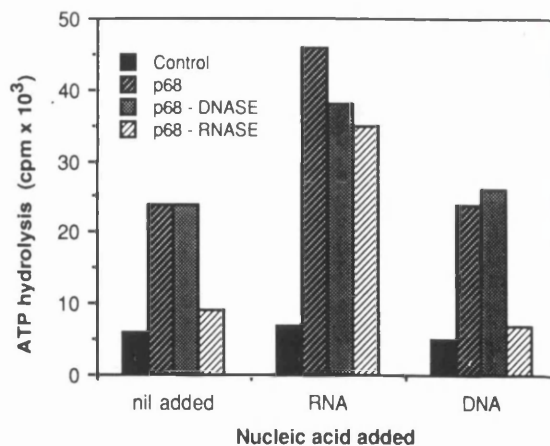


Fig. 3. ATPase assays showing that p68 is stimulated by RNA but not DNA. Immunoprecipitates with PAb 204 (p68) or PAb 240 (control) were incubated with 50 μ M γ -³²P-labelled ATP. ATP hydrolysis was measured as release of free ³²P_i (10³ c.p.m. is ~1 fmol/s). For details see text.

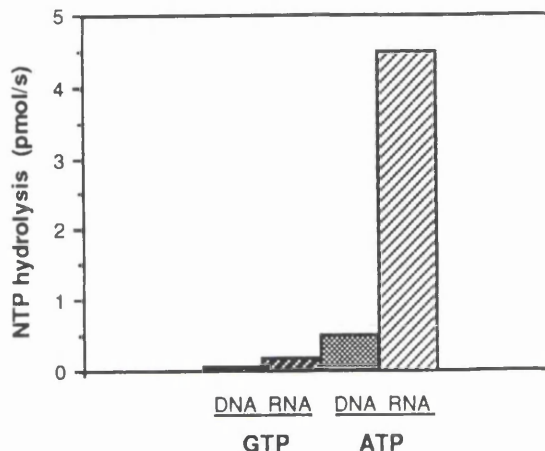


Fig. 4. ATPase and GTPase assays showing that p68 does not hydrolyse GTP. Immunoprecipitated p68 was incubated with 1 mM γ -³²P-labelled ATP or GTP and ³²P_i release measured in the presence of 0.2 mg/ml HeLa cell RNA or salmon testis DNA. The p68 was prepared from cells lysed in high salt and RNase A. Background counts due to spontaneous NTP hydrolysis have been deducted.

dependent ATPase. We have tested HeLa cell p68 for nucleic acid dependent ATPase activity. Because p68 is difficult to purify we chose to examine its activity while still bound to PAb 204 on protein G beads. The PAb 204 binding site on p68 lies outside the region of homology with eIF4-A and distant from the ATPase and helicase motifs in the primary sequence (Ford *et al.*, 1988). We used the assay of Clark *et al.* (1981) in which ATP hydrolysis is measured as release of ³²P_i from γ -³²P-labelled ATP.

When cells are lysed at physiological salt concentration, we find that immunoprecipitated p68 has substantial ATPase activity which is stimulated by added RNA but not by DNA (Figure 3). We used HeLa cell RNA and boiled sonicated salmon testis DNA (0.2 mg/ml) to stimulate the ATPase because these should contain a wide spectrum of secondary structures. The presence of significant ATPase activity even without added RNA made us consider the possibility that there was endogenous nucleic acid in our immunoprecipitates. To address this question, we pretreated immunoprecipitates with either DNase I (100 U/ml) or RNase A

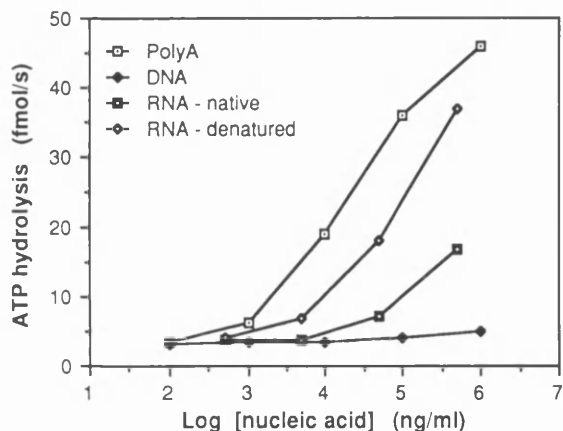


Fig. 5. ATPase assays showing that p68 is stimulated more effectively by single-stranded than double-stranded RNA. Immunoprecipitated p68 was incubated with 10 μ M ATP and 32 P_i release measured in the presence of increasing amounts of poly(A), salmon testis DNA (DNA), native rotavirus genomic RNA (RNA-native) or DMSO-denatured rotavirus genomic RNA (RNA-denatured). The immunoprecipitates were RNase treated to reduce endogenous activity.

(0.2 mg/ml) and then washed them thoroughly to remove the nuclease. RNase treatment reduces ATPase activity to near control levels. Only RNA can restore activity to RNased beads. DNA has no effect, even on DNase treated p68. We conclude that with cell lysis at 150 mM NaCl there is probably endogenous RNA which co-purifies with p68.

We have obtained a K_m value for ATP of 10^{-3} to 10^{-4} M. This is one order of magnitude higher than that of eIF4-A (Grifo *et al.*, 1984), although still below the cellular ATP level (10^{-2} to 10^{-3} M). We therefore tested p68 for GTPase activity (Figure 4). We see no GTPase activity with either RNA or DNA, which is consistent with the lack of the NKXD/W GTPase motif (Dever *et al.*, 1987) in p68.

We have not seen significant differences in ATPase activity with different species of cellular RNA or poly(A). The ATPase activity of eIF4-A is reduced by secondary structure in the RNA template (Abramson *et al.*, 1987), which is one of the hallmarks of a helicase. We therefore tested p68 with double-stranded rotavirus genomic RNA. The viral RNA was divided into equal aliquots, one of which was denatured using the dimethyl sulphoxide (DMSO) technique, which minimizes RNA degradation (McCrae and Joklik, 1978). In its double stranded form the RNA is less effective at stimulating ATP hydrolysis than it is after DMSO treatment (Figure 5).

Cell cycle variation in p68 distribution

Immunohistochemical examination of mammalian cells with PAb 204 reveals a weak granular staining pattern which is localized exclusively in the cell nucleus but spares the nucleoli. The extent of granularity depends on the individual cell type. PTK-1 cells characteristically give an exceptionally punctate pattern (Lane and Hoeffler, 1980). In mitotic cells PAb 204 does not stain chromosomes on the metaphase plate although weak peri-chromosomal staining is observed. The most striking phenomenon is the presence of a few intensely staining bodies in the nuclei of cells that have just divided. The granules usually number two or three arranged in a symmetrical pattern in sister cell nuclei (Figure 6). The

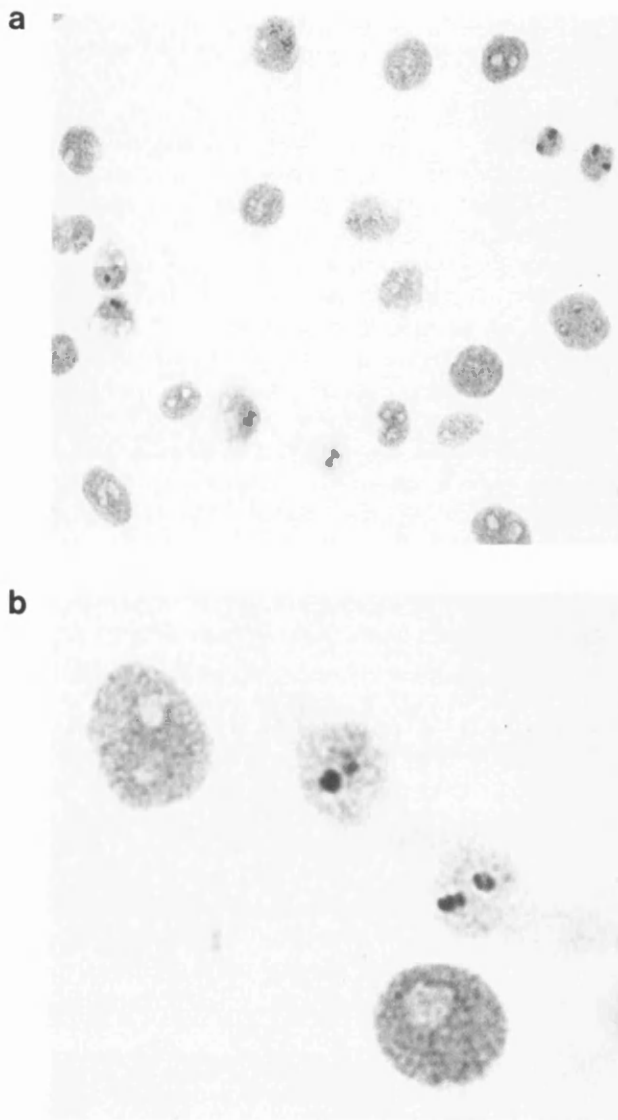


Fig. 6. Immunoperoxidase staining of p68 in CV1 cells using PAb 204, rabbit anti-mouse HRP and diaminobenzidine. (a) Low power, (b) high power.

pattern is present in all the cell lines we have examined to date.

Discussion

When we first reported the homology between p68 and eIF4-A we speculated that it might define a new gene family. The recent identification of four diverse new genes related to p68 confirms this idea. Linder *et al.* (1989) proposed the term 'DEAD box' proteins to describe the family, because the sequence aspartate, glutamate, alanine, aspartate (part of the ATP binding site) uniquely identifies these proteins in the database. The sequence data strongly suggests that they share a core biochemical activity.

We report here that p68 is an RNA dependent ATPase which is highly stimulated by a range of natural and synthetic RNAs. It is stimulated weakly by dsRNA and not at all by DNA. Two lines of evidence lead us to suspect that p68 is bound to nuclear RNA: first, p68 extracted in conditions which favour RNA-protein interactions has endogenous ATPase activity which can be abolished by RNase treat-

ment; and second, the PAb 204 cell staining pattern is abolished by RNase A treatment of fixed cells (data not shown).

Three DEAD box proteins (eIF-4A, Grifo *et al.*, 1984; srmB, Nishi *et al.*, 1988; and p68) have now been shown to be RNA dependent ATPases and it is reasonable to assume that all members of the family possess this activity. We speculate that these proteins bring the power to manipulate RNA secondary structure to many different molecular organelles in the cell. Only eIF4-A has a clearly defined role: by melting secondary structure near the mRNA cap it allows the ribosome to scan along the message looking for the start codon. Cap dependence requires that eIF4-A interact with several other initiation factors (reviewed in Sonenberg, 1988). The strong inference for other members of the family is that they will be subunits of multimeric complexes that act on specific RNAs. For example, MSS116 (Seraphin *et al.*, 1989) is probably a component of the mitochondrial splicing apparatus, which presumably requires the ability to modify secondary structure in pre-mRNA. Vasa protein is a component of *Drosophila* polar granules, complex ribonucleoprotein particles which are thought to govern the expression of some posterior group maternal effect genes (Hay *et al.*, 1988; Lasko and Ashburner, 1988).

This concept may hold the key to understanding the biological significance of the cross reaction between p68 and SV40 T antigen. T is known to bind, in part covalently, to cellular RNA (Darlix *et al.*, 1984; Carroll *et al.*, 1988). It may replace p68 in ribonucleoprotein particles by interacting with p68's normal ligand via the PAb204 epitope. T antigen thus mimics p68 in order to recruit the particle for the benefit of the virus. Conservation of the PAb 204 epitope on T and p68 is thus analogous to conservation of the p105 RB binding site on T and adenovirus E1A (Moran, 1988).

The p68 protein, like many other nuclear antigens, has a granular distribution in the interphase cell nucleus and appears to be excluded from the nucleoli. Granular staining is associated with a variety of functional entities including sites of DNA replication (Bravo and Macdonald-Bravo, 1987) and ribonucleoprotein particles (Spector *et al.*, 1987). The antigen is either absent or inaccessible to PAb 204 on condensed chromosomes but as sister cells emerge from mitosis the antigen is strikingly clustered. At this stage in the cell cycle nucleoli are being assembled at the chromosomal nucleolar organizers and it is possible that the protein may be associated with the formation of nucleoli. Interestingly another DEAD box protein, srmB (Nishi *et al.*, 1988), appears to be involved in ribosome assembly. Previous studies on mitosis (Albrecht-Buehler, 1977) drew attention to mirror symmetry in post-mitotic sister cells. The results with p68 extend this observation to the cell nucleus and imply that despite complete disruption at mitosis, elements of nuclear architecture are recalled after cell division.

Materials and methods

Cell lines, reagents and buffers

Mouse monoclonal hybridomas PAb 204 (Lane and Hoeffler, 1980), PAb 240 and BG2 were produced in this laboratory. CV1 cells were provided by ICRF cell production. 293 cells were provided by Dr P.Gallimore.

Protein G-Sepharose was purchased from Pharmacia. RNase-free DNase I was purchased from Stratagene. Rabbit anti-mouse immunoglobulins conjugated to horseradish peroxidase (HRP), rabbit anti-mouse immunoglobulins conjugated to alkaline phosphatase (ALP) and normal mouse serum

were purchased from Dako. Serum free hybridoma culture medium (HB101) was purchased from Dupont. Rotavirus genomic RNA was kindly provided by Dr Malcolm McCrae. Total cellular RNA was made from HeLa cells by the guanidinium method (Ausubel *et al.*, 1987). SV40 large-T antigen was prepared by the method of Simanis and Lane (1985). ³²P-labelled ATP and GTP were purchased from Amersham. Unless stated otherwise all other reagents were purchased from Sigma.

Lysis buffer is 150 mM NaCl, 50 mM Tris pH 8.0, 5 mM EDTA, 1% Nonidet P-40 (NP-40), 1 mM dithiothreitol (DTT), 1 mM phenylmethylsulphonyl fluoride (PMSF). ATPase buffer is 150 mM NaCl, 5 mM KCl, 1.5 mM MgCl₂, 20 mM Hepes pH 7.0, 1 mM DTT, 1 mM PMSF. DNase buffer is 150 mM NaCl, 5 mM KCl, 10 mM MgCl₂, 20 mM Pipes pH 7.0, 1 mM DTT.

Immunoprecipitation

3 × 10⁹ HeLa or 293 cells were incubated in lysis buffer for 30 min on ice. The cell extract was centrifuged at 100 000 g for 30 min and the pellet discarded. Monoclonal antibody was added either in solution (Figures 1, 2 and 3) or pre-bound to protein G-Sepharose beads (Figures 4 and 5) and incubated on a rotating wheel overnight at 4°C. Protein G-Sepharose beads were added (Figures 1, 2 and 3) and the incubation continued for 1 h. The beads were then washed in lysis buffer.

The lysis conditions were varied as follows in individual experiments: 500 mM NaCl and sonication (Figures 4 and 5); RNase A 0.1 mg/ml (Figures 1, 4 and 5). After washing, the beads for Figure 5 were incubated in lysis buffer containing 0.1 mg/ml RNase A for 1 h. Finally, the beads were washed in ATPase buffer (Figures 1, 2, 4 and 5), or DNase buffer (Figure 3). Subsequent manipulations are described in the text.

Electrophoresis and immunoblotting

Denaturing polyacrylamide gel electrophoresis and immunoblotting were performed as described in Harlow and Lane (1989) using BioRad minigel apparatus. Protein from 10 µl of beads immunoprecipitated as above was loaded per track (Figures 1 and 2a). Pure T antigen was loaded in a single slot, blotted and strips cut from the blot with ~10 ng of T antigen per strip. Blots were blocked in 2% dried milk in phosphate buffered saline (PBS) and washed in 1% NP-40 in PBS. Antibodies were suspended in blocking buffer plus 0.1% Tween: PAb 204 and BG2 diluted 1:2, MF7, normal mouse serum and rabbit anti-mouse ALP diluted 1:1000. Blots were developed with bromo-chloro-indolyl phosphate and nitro blue tetrazolium. MF7 is pooled serum from mice immunized with a β-galactosidase fusion protein made in *E. coli* by expressing a PUR290 plasmid with the p68 cDNA (Ford *et al.*, 1988) cloned into the *EcoRI* site. The fusion protein was purified as described in Harlow and Lane (1989).

ATPase assays

ATPase and GTPase activity is expressed as P_i released from γ-³²P-labelled ATP or GTP. Assays were performed in 30 µl ATPase buffer containing ~10 µl of protein G-Sepharose beads bound to PAb 204 or PAb 240. From Coomassie staining of polyacrylamide gels we estimate that there was ~50 ng p68 per assay. Nucleic acid and nucleotide triphosphate were included as indicated in the text. The assays were incubated for 30 min at 37°C. 300 µl of 7% activated charcoal in 50 mM HCl, 5 mM H₃PO₄ was then added and the mixture centrifuged at 10 000 g for 10 min to remove unreacted ATP. The supernatant (containing free ³²P_i) was mixed with Aquasol liquid scintillation cocktail and counted in a Packard scintillation counter on the ³²P channel.

Cell staining

CV1 cells grown on cover slips in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum (FBS) were fixed in cold 50% acetone in methanol for 5 min and dried. They were then incubated for 3 h with PAb 204 hybridoma supernatant, rinsed in PBS, incubated for 3 h with rabbit anti-mouse HRP diluted 1:100 in 10% FBS in PBS, rinsed in PBS and developed with diaminobenzidine.

Acknowledgements

We thank Martin Ford for preparation of the MF7 serum, Lu Xin for help with the cell staining and Malcolm McCrae for providing the rotavirus RNA and for helpful discussions.

References

Acheson, N.H. (1981) In Tooze, J. (ed.), *DNA Tumour Viruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 125–204.

- Albrecht-Buehler, G. (1977) *J. Cell. Biol.*, **72**, 595–603.
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) *Current Protocols in Molecular Biology*. Wiley Interscience, pp. 421–425.
- Abramson, R.D., Dever, T.E., Lawson, T.G., Ray, B.K., Thach, R.E. and Merrick, W.C. (1987) *J. Biol. Chem.*, **262**, 3826–3832.
- Ball, R.K., Siegl, B., Quellhorst, S., Brandner, G. and Braun, D.G. (1984) *EMBO J.*, **3**, 1485–1491.
- Bravo, R. and Macdonald-Bravo, H. (1987) *J. Cell. Biol.*, **105**, 1549–1554.
- Clark, R., Lane, D.P. and Tjian, R. (1981) *J. Biol. Chem.*, **56**, 11854–11858.
- Carroll, R.B., Samad, A., Mann, A., Harper, J. and Anderson, C.W. (1988) *Oncogene*, **2**, 437–444.
- Darlix, J., Khandjian, E.W. and Weil, R. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5425–5429.
- DeCaprio, J.A., Ludlow, J.W., Figge, J., Shew, J.-H., Huang, C.-M., Lee, W.-H., Marsilio, E., Paucha, E. and Livingston, D.M. (1988) *Cell*, **54**, 275–283.
- Dean, F.B., Bullock, P., Murakami, Y., Wobbe, C.R., Weissbach, L. and Hurwitz, J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 16–20.
- Dever, T.E., Glyniadis, M.J. and Merrick, W.C. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1814–1818.
- Ford, M.J., Anton, I.A. and Lane, D.P. (1988) *Nature*, **332**, 736–738.
- Gannon, J.V. and Lane, D.P. (1987) *Nature*, **329**, 456–458.
- Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. and Blinov, V.M. (1988) *Nature*, **333**, 22.
- Gough, G., Gannon, J.V. and Lane, D.P. (1988) *Cancer Cells*, **6**, 153–158.
- Grifo, J.A., Abramson, R.D., Satler, C.A. and Merrick, W.C. (1984) *J. Biol. Chem.*, **259**, 8648–8654.
- Gurney, E.G., Harrison, R.O. and Fenno, J. (1980) *J. Virol.*, **34**, 752–763.
- Harlow, E. and Lane, D.P. (1989) *Antibodies: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Harlow, E., Crawford, L.C., Pim, D.C. and Williamson, N.M. (1981) *J. Virol.*, **39**, 861–869.
- Hay, B., Jan, L.Y. and Jan, Y.N. (1988) *Cell*, **55**, 577–587.
- Hodgman, T.C. (1988) *Nature*, **333**, 22–23, 578.
- Lane, D.P. (1988) *Nature*, **334**, 478.
- Lane, D.P. and Crawford, L.V. (1979) *Nature*, **278**, 261–263.
- Lane, D.P. and Gannon, J. (1986) *Cancer Cells*, **4**, 387–393.
- Lane, D.P. and Hoefler, W.K. (1980) *Nature*, **288**, 167–170.
- Lasko, P.F. and Ashburner, M. (1988) *Nature*, **335**, 611–617.
- 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.
- Livingstone, D.M. and Bradley, M.K. (1987) *Mol. Biol. Med.*, **4**, 63–80.
- McCrae, M. and Joklik, W.K. (1978) *Virology*, **89**, 578–593.
- Mitchell, P.J., Wang, C. and Tjian, R. (1987) *Cell*, **50**, 847–861.
- Moran, E. (1988) *Nature*, **334**, 168–170.
- Mole, S., Gannon, J., Ford, M. and Lane, D.P. (1987) *Phil. Trans. R. Soc. Lond. B*, **317**, 445–469.
- Nishi, K., Morel-Deville, F., Hershey, J.W.B., Leighton, T. and Schnier, J. (1988) *Nature*, **336**, 496–498.
- Rigby, P.W.J. and Lane, D.P. (1983) In Klein, G. (ed.), *Advances in Viral Oncology*. Raven Press, New York, Vol. 3, pp. 31–57.
- Schutzbank, T., Robinson, R., Oren, M. and Levine, A.J. (1982) *Cell*, **30**, 481–490.
- Scott, M.R.D., Westphal, K.H. and Rigby, P.W.J. (1983) *Cell*, **34**, 557–567.
- Seraphin, B., Simon, M., Boulet, A. and Faye, G. (1989) *Nature*, **337**, 84–87.
- Simanis, V. and Lane, D.P. (1985) *Virology*, **144**, 88–100.
- Smale, S.T. and Tjian, R. (1986) *Mol. Cell Biol.*, **6**, 4077–4087.
- Sonenberg, N. (1988) *Progr. Nucleic Acids Res. Mol. Biol.*, **35**, 173–207.
- Spector, D., Watt, R.A. and Sullivan, N.F. (1987) *Oncogenes*, **1**, 5–12.
- Stahl, H., Droge, P. and Knippers, R. (1986) *EMBO J.*, **5**, 1939–1944.
- Tjian, R. (1978) *Cell*, **13**, 165–179.
- Wold, M., Li, J. and Kelly, T. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 3643–3647.

Received on February 22, 1989

Chromosome mapping of the human gene encoding the 68-kDa nuclear antigen (p68) by using the polymerase chain reaction

(chromosome-mediated gene transfer/simian virus 40 tumor antigen/helicase)

R. IGGO*, A. GOUGH†, W. XU‡§, D. P. LANE*, AND N. K. SPURR†¶

Departments of *Molecular Immunochimistry and †Human Genetic Resources, Imperial Cancer Research Fund Clare Hall Laboratories, South Mimms, Herts, EN6 3LD, United Kingdom; and ‡Somatic Cell Genetics Laboratory, Imperial Cancer Research Fund, London, WC2A 3PX, United Kingdom

Communicated by Walter Bodmer, May 4, 1989 (received for review February 21, 1989)

ABSTRACT Chromosome mapping by Southern analysis of DNA from somatic cell hybrids is often unsuccessful when only cDNA probes are available. p68 is a putative RNA helicase that is antigenically related to the simian virus 40 large tumor antigen. By using the polymerase chain reaction and oligodeoxynucleotide primers based on the cDNA sequence, we have identified introns in the p68 gene. Comparison of human and mouse DNA fragments amplified with these primers revealed length differences that allowed us to identify the human gene. Application of this technique to DNA from human-mouse somatic cell hybrids and cell lines derived from them by chromosome-mediated gene transfer allowed us to map p68 to the distal part of the long arm of chromosome 17.

The rapid chromosomal assignment of protein-encoding genes is a major goal of human molecular genetics. The standard approach is Southern blot analysis of DNA from hybrid cell lines of known cytogenetic composition. This technique depends upon the ready identification of species-specific restriction fragments of the gene of interest. cDNA probes detect pseudogenes and may fail to distinguish between related genes within a species or between highly conserved genes in different species. Intron probes overcome all of these problems but classically can only be obtained by screening a genomic library.

The p68 protein is a proliferation-associated nuclear antigen first identified through its highly specific cross reaction with the simian virus 40 tumor antigen (1). Sequencing of cDNA clones encoding human p68 showed the protein to be homologous to the eukaryotic translation initiation factor (eIF) 4A (2). eIF-4A has two closely related forms that show tissue-specificity and has multiple pseudogenes (3). Subsequently other members of this gene family have been identified including a yeast gene involved in the splicing of mitochondrial transcripts (4), a *Drosophila* maternal-effect gene *vasa* (5), and an *Escherichia coli* gene involved in ribosomal assembly (6). These proteins are about 30% identical at the amino acid sequence level and contain long blocks of absolutely conserved sequence. Some of the conserved blocks are characteristic of a large superfamily of proteins with putative helicase activity (4, 7-9).

We were unable to map p68 by Southern analysis of somatic cell hybrids using cDNA probes. We present an alternative approach exploiting the polymerase chain reaction (PCR; ref. 10). We guessed the position of introns in p68 by examination of the gene structure of eIF-4AII (3) and *vasa* (5) in the conserved regions. By using primers around these putative splice sites (Fig. 1A), we identified species-specific differences in PCR products presumably due to intron length

variation and used these in combination with somatic cell hybrids and cell lines derived from them by chromosome-mediated gene transfer to localize p68 to the distal part of the long arm of chromosome 17 in the region 17q23-q25.

METHODS

Derivations of most of the hybrid cell lines used have been given (see Table 1). TRID62, PJT2/A1, KLT-8, PLT-6B, and PLT-8 were described in Xu *et al.* (11). P7A2.7 was described in Voss *et al.* (12). PCR was performed on a Perkin-Elmer/Cetus PCR machine using Cetus *Thermus aquaticus* polymerase according to the manufacturer's instructions (Perkin-Elmer/Cetus) with the following modifications: each dNTP at 300 μ M, 200 ng of DNA, and 10% (vol/vol) dimethyl sulfoxide. The following primers were used: primer 1, GGC-CAAAAGAAGTAAGACAGC; primer 2, TTCTGG-TAAGCTCATCACATC; primer 3, CAACAAGAGCGT-GACTGGGTT; and primer 4, ATCCTCTGAGGAGT-TAGGGTA (Fig. 1A). All of the hybrids were tested with both pairs of primers. The cycles used were 92°C for 15 s, 55°C for 30 s, and 72°C for 60 s (primers 1 and 2) or 72°C for 120 s (primers 3 and 4) for 25 cycles. For data in Fig. 2, 1 ng of primer 2 end-labeled with ³²P was included in the reaction mix. The PCR products were extracted with phenol/chloroform, ethanol-precipitated, and examined by either denaturing polyacrylamide gel electrophoresis and autoradiography (see Fig. 2) or agarose gel electrophoresis and ethidium bromide staining (see Fig. 3). For DNA sequencing, the PCR products were extracted with phenol/chloroform, passed through a Sephadex G-50 column equilibrated with 10 mM Tris-HCl/1 mM EDTA, pH 8, ethanol-precipitated, and resuspended in 10 μ l of water. The PCR product (5 μ l) was mixed with Sequenase buffer (United States Biochemical Sequenase kit) and 30 ng of ³²P-labeled sequencing primer (AAAGACTATATTCATAT), which lies between primers 1 and 2, boiled for 5 min, and placed on ice. Thereafter, sequencing was carried out with Sequenase 2 according to the manufacturer's instructions (United States Biochemical).

RESULTS

In the initial hybrid screen with primers 1 and 2 (Fig. 2) the PCR product was end-labeled with ³²P, electrophoresed on a denaturing polyacrylamide gel, and visualized by autoradiography. Lane a shows the PCR product with mouse DNA

Abbreviations: PCR, polymerase chain reaction; eIF, eukaryotic translation initiation factor.

§Present address: Shanghai Institute of Cell Biology, Academia Sinica, 320 Yo-Yang Road, Shanghai, China.

¶To whom reprint requests should be addressed.

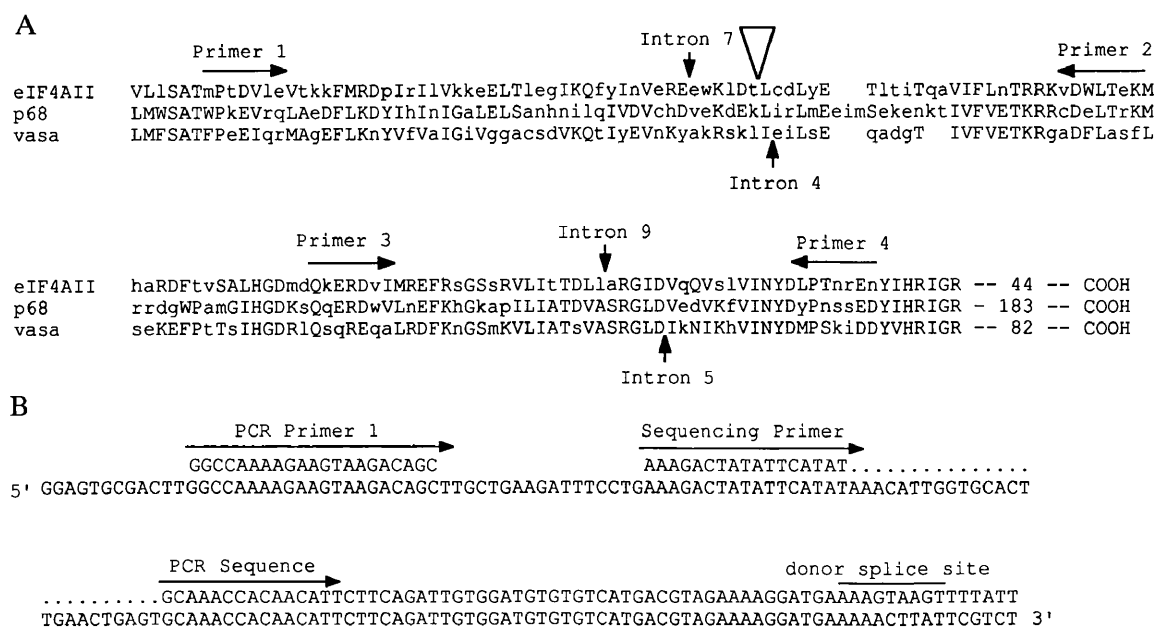


FIG. 1. PCR mapping strategy. (A) Part of the deduced amino acid sequences of eIF-4AII, vasa, and p68 are shown using the single-letter code. Closely related amino acids (A and G; S and T; Q and N; E and D; K and R; F, Y, and W; V, L, I, and M) shared by two sequences are shown in capital letters. Vertical arrows indicate the position of introns in eIF-4AII and vasa. The open triangle shows the splice site we identified in p68. (B) The lower line shows part of the cloned cDNA sequence of p68. The upper line shows the sequence we obtained from the PCR product (the first 25 nucleotides after the sequencing primer were unreadable).

alone; lane b shows the PCR product with human DNA alone. There is a band in lane a (mouse DNA) at 234 nucleotides, which is the size we would expect for an intronless pseudogene. With both mouse and human DNAs, there is a band at ≈600 nucleotides: the human p68 product is ≈15 base pairs (bp) longer than the mouse product. The mouse PCR products act as positive controls for the PCR process. We have sequenced part of the human PCR product (Fig. 1B) and find complete agreement with our cDNA sequence for 59 nucleotides up to the expected splice site, where we find a consensus splice donor sequence. We are thus confident that the band seen in lane b is indeed derived from the human p68 gene, with an intron of ≈370 bp at the predicted site. Lanes c-o are from mouse-human somatic cell hybrids. Even on this overexposed film the human (upper) band is clearly absent in lanes c, i, and m. Analysis of the results (Table 1; see Fig. 4) suggests that p68 is on chromosome 17. Lane o is from hybrid PCTBA1.8, whose human DNA is solely derived from chromosome 17, and confirms that p68 is on chromosome 17.

Fig. 3 shows the results with primers 3 and 4. We have not

sequenced the human product in this case, but from the size we conclude that there is an intron of 1.3 kilobases. We obtained identical mapping results with these primers, including the presence of a presumed pseudogene in the mouse DNA (data not shown). Lanes e-j are from cell lines containing only fragments of human chromosome 17 (see Fig. 4); lanes e-g are from somatic cell hybrids with translocations of chromosome 17; lanes h-j are from cell lines developed using chromosome-mediated gene transfer (11). The presence of the human (lower) band in lanes e, f, g, and i localizes p68 to the distal part of the long arm of chromosome 17 in the region 17q23-q25 (Fig. 4).

DISCUSSION

We attempted to map the chromosomal location of p68 by using Southern blot analysis and human p68 cDNA probes. A complex pattern was seen with both human and rodent DNA (Fig. 5). The restriction patterns we obtained did not

Table 1. Chromosomal assignment of p68 by PCR

Hybrid	Human chromosome																						p68	Ref.	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22			X
CTP34B4	+	+	+	-	+	+	+	+	-	-	-	+	-	+	-	+	+	+	-	-	-	-	+	+	13
CTP41A2	-	+	+	-	+	+	+	-	-	-	-	-	-	+	-	+	+	+	-	-	-	-	+	+	13
DUR4.3	-	-	+	-	+	-	-	-	-	+	+	+	+	+	+	-	+	+	-	+	+	+	+	+	14
DT1.2R	-	-	+	-	-	-	-	-	-	+	+	-	+	-	+	-	+	+	-	+	+	+	-	+	15
DT1.2.4	-	-	+	-	-	-	-	-	-	+	-	-	-	+	-	+	+	+	-	+	+	-	+	+	16
PLTI.S	+	-	+	-	+	+	+	-	-	+	+	-	+	-	+	+	+	+	-	+	+	+	-	+	16
PCTBA1.8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+	17
LSR34S49	+	+	+	-	+	+	-	-	-	+	-	-	+	+	+	+	+	+	+	+	+	-	+	-	18
MOG34A4	+	-	+	+	+	+	+	+	-	+	+	+	+	+	-	+	-	+	+	-	+	-	+	-	19
MOG13/10	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	+	-	-	20
HORP9.5	-	-	-	-	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-	-	-	+	+	-	21
HORP27RC14	-	-	-	+	-	-	+	-	-	-	+	+	-	+	+	-	-	-	-	-	+	-	-	-	21

Whether (+) or not (-) a hybrid contains the listed human chromosome (blank = unassigned) and a hybrid yields the human p68 PCR product is indicated. The concordant chromosome is chromosome 17 (boxed). Descriptions of the hybrids are found in the references listed.

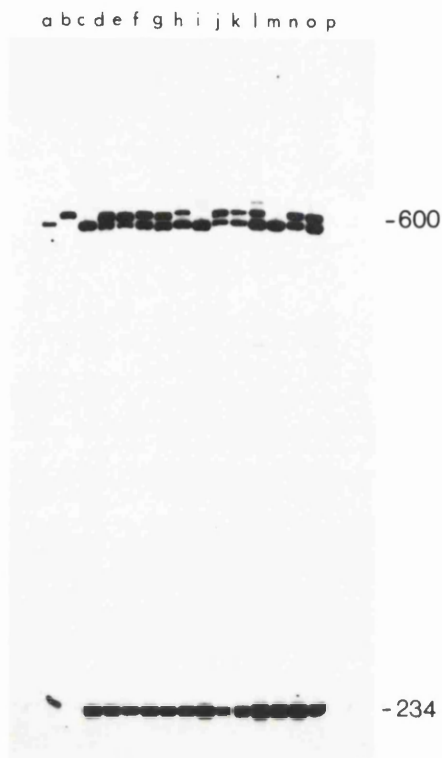


FIG. 2. PCR products using primers 1 and 2. Lanes: a, mouse DNA; b, human DNA; c, MOG13/10; d, P7A2.7; e, TRID62; f, PJT2/A1; g, CTP41A2; h, PLT1.S; i, MOG34A4; j, CTP34B4; k, DUR4.3; l, DT1.2R; m, LSR34S49; n, DT1.2.4; o, PCTBA1.8; p, no DNA. Sizes of the DNA fragments (in nucleotides) shown were measured on an M13 sequencing ladder.

allow us to identify the human p68 gene, presumably because of conservation of the p68 gene in man and mouse. We, therefore, tried a different approach using pairs of p68 cDNA primers to identify interspecies variation in PCR fragment length. Two sets of primers were chosen that we predicted would lie either side of splice sites in p68 since introns occur at these sites in eIF-4AII (3) and *vasa* (5) (Fig. 1A).

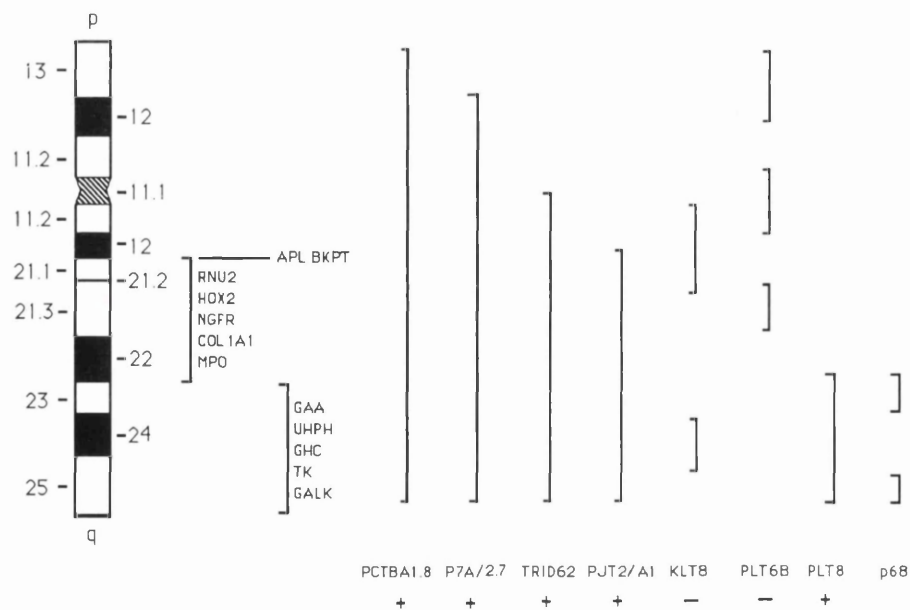


FIG. 4. Localization of p68 to the region 17q23-25. Whether plus (+) or not (-) a cell line yielded the human p68 PCR product is indicated. The gene order is taken from Xu *et al.* (11). RNU2, U2 small nuclear RNA; HOX2, homeobox region 2; NGFR, nerve growth factor receptor; COL1A1, collagen α 1[1]; MPO, myeloperoxidase; GAA, acid α -glucosidase; UMPH, uridine-5'-monophosphate phosphohydrolase; GHC, growth hormone complex; TK, thymidine kinase; GALK, galactokinase; APL BKPT, acute promyelocytic leukemia breakpoint.

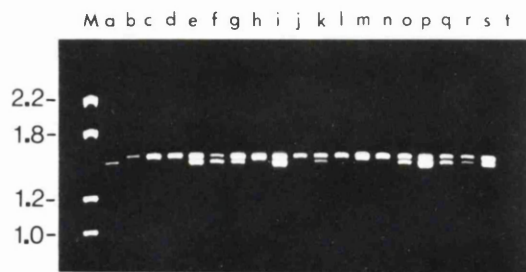


FIG. 3. PCR products using primers 3 and 4. Lanes: a, human DNA; b, mouse DNA; c, HORP9.5; d, HORP27RC14; e, P7A2.7; f, TRID62; g, PJT2/A1; h, KLT8; i, PLT8; j, PLT6B; k, PLT1.S; l, MOG34A4; m, MOG13/10; n, LSR34S49; o, PCTBA1.8; p, CTP34B4; q, DUR4.3; r, DT1.2R; s, CTP41A2; t, no DNA; M, DNA markers. Sizes in kilobases are shown.

Chromosome mapping by PCR is extremely rapid, can be carried out on very small samples of impure DNA, and avoids the use of radioactivity. There is no known evolutionary constraint on absolute intron length. Although we were fortunate in being able to guess the position of introns in p68, a survey of intron and exon size (22) suggests that our strategy should be successful without this information. The mean length of vertebrate internal exons is 137 bp. For vertebrate internal introns the modal intron length is 80–99 bp, with only 19% >1600 bp. Therefore, any pair of PCR primers separated by several hundred base pairs should amplify an intron and still remain within the size range that can easily be amplified by PCR, provided that the primers lie within the coding sequence. Long introns are unlikely to be a problem if several pairs of primers are used. This suggests that this method of mapping should be widely applicable and indeed we have identified species-specific differences in PCR products in another gene (unpublished data).

We thank Iain Goldsmith for oligonucleotide synthesis, Ellen Solomon and Sue Rider for samples of DNA from the chromosome-mediated gene transfer derived cell lines, and Mark Meuth for helpful discussions on the PCR.

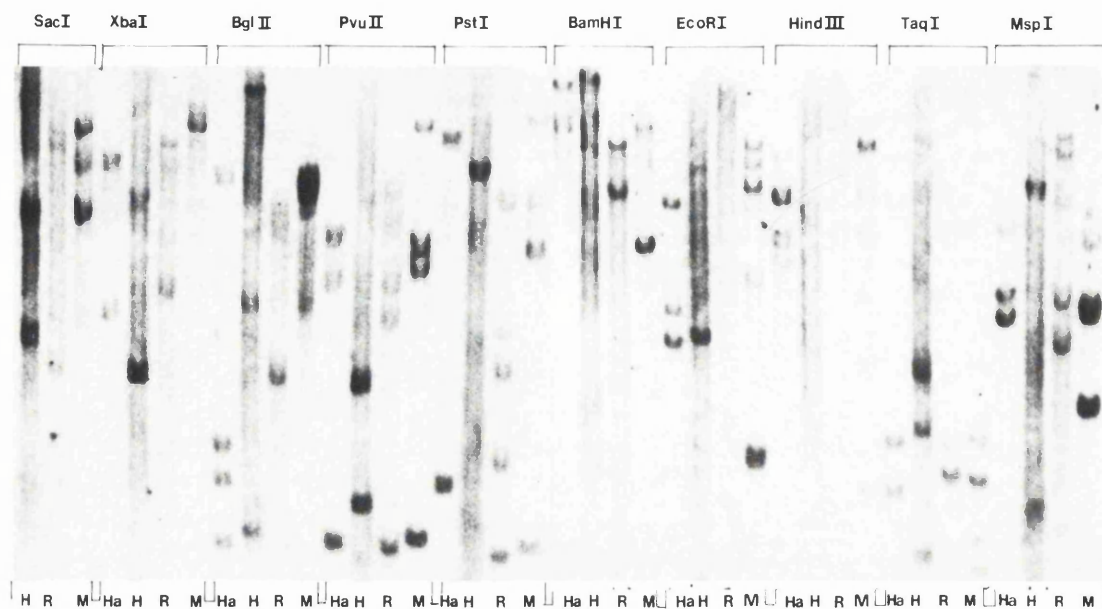


FIG. 5. Southern blot of genomic DNA probed with a 1-kilobase fragment spanning the central region of the p68 cDNA (nucleotides 532–1599). Human (lanes H), mouse (lanes M), rat (lanes R), and hamster (lanes Ha) DNA was digested with the restriction enzymes listed above the relevant lanes.

- Lane, D. P. & Hoeffler, W. K. (1980) *Nature (London)* **288**, 167–170.
- Ford, M. J., Anton, I. A. & Lane, D. P. (1988) *Nature (London)* **332**, 736–738.
- Nielsen, P. J. & Trachsel, H. (1988) *EMBO J.* **7**, 2097–2105.
- Seraphin, B., Simon, M., Boulet, A. & Faye, G. (1989) *Nature (London)* **337**, 84–87.
- Lasko, P. F. & Asburner, M. (1988) *Nature (London)* **335**, 611–617.
- Nishi, K., Morel-Deville, F., Hershey, J. W. B., Leighton, T. & Schnier, J. (1988) *Nature (London)* **336**, 496–498.
- Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1988) *Nature (London)* **333**, 22.
- Hodgman, T. C. (1988) *Nature (London)* **333**, 22–23, and correction (1988) **333**, 578.
- Lane, D. (1988) *Nature (London)* **334**, 478.
- Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985) *Science* **230**, 1350–1354.
- Xu, W., Gorman, P. A., Ridler, S. H., Hedge, P. J., Moore, G., Prichard, C., Sheer, D. & Solomon, E. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8563–8567.
- Voss, R., Lerer, I., Povey, S., Solomon, E. & Bobrow, M. (1980) *Ann. Hum. Genet.* **44**, 1–9.
- Jones, E. A., Goodfellow, P. N., Kennet, R. H. & Bodmer, W. F. (1976) *Somatic Cell Genet.* **2**, 483–496.
- Solomon, E., Bobrow, M., Goodfellow, P. N., Bodmer, W. F., Swallow, D. M., Povey, S. & Noel, B. (1976) *Somatic Cell Genet.* **2**, 125–140.
- Swallow, D. M., Solomon, E. & Pajunen, L. (1977) *Cytogenet. Cell Genet.* **18**, 136–148.
- Martin, D., Tucker, D. F., Gorman, P., Sheer, D., Spurr, N. K. & Trowsdale, J. (1987) *Ann. Hum. Genet.* **51**, 145–152.
- Bai, Y., Sheer, Y., Hiorns, L., Knowles, R. W. & Tunnacliffe, A. (1982) *Ann. Hum. Genet.* **46**, 337–347.
- Spurr, N. K., Durbin, H., Sheer, D., Parkar, M., Bobrow, L. & Bodmer, W. F. (1986) *Int. J. Cancer* **38**, 631–636.
- Solomon, E., Swallow, E., Burgess, S. & Evans, L. (1979) *Ann. Hum. Genet.* **42**, 273–281.
- Povey, S., Jeremiah, S. J., Barker, R. F., Hopkinson, D. A., Robson, E. B., Cook, P. J. L., Solomon, E. & Bobrow, M. (1980) *Ann. Hum. Genet.* **43**, 241–248.
- Van Heyningen, V., Bobrow, M., Bodmer, W. F., Gardiner, S. E., Povey, S. & Hopkinson, D. A. (1975) *Ann. Hum. Genet.* **38**, 295–303.
- Hawkins, J. D. (1988) *Nucleic Acids Res.* **16**, 9893–9905.

Identification of a putative RNA helicase in *E.coli*

R.Iggo, S.Picksley, J.Southgate, J.McPheat and D.P.Lane*

Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Herts EN6 3LD, UK

Received June 26, 1990; Revised and Accepted July 23, 1990

EMBL accession no. X52647

ABSTRACT

The human p68 protein, an SV40 large T related antigen, is an RNA dependent ATPase and RNA helicase. It belongs to a new large and highly conserved gene family, the DEAD box proteins, whose members are involved in a variety of processes requiring manipulation of RNA secondary structure such as translation and splicing. Multiple DEAD box genes are present in *S.cerevisiae*, but only one has previously been described in *E.coli*. Low stringency screening of an *E.coli* genomic library with a p68 cDNA probe led to the identification of *dbpA*, a new *E.coli* DEAD box gene located at 29.6 minutes on the W3110 chromosome. We report here the nucleotide and deduced amino acid sequences of the gene. We have overexpressed *dbpA* from its own promoter on a high copy number plasmid and identified the gene product as a ~50 kD protein by immunoblotting with an anti-DEAD antibody.

INTRODUCTION

The term 'DEAD box protein' was coined by Linder *et al* (1) to describe a family of RNA helicases which has recently been identified in bacteria, yeast, *drosophila*, mouse and man. Overall, the deduced amino acid sequences share about 30% homology, but there are multiple short runs of absolute conservation, including the LDEADXXL motif which gives the family its name. The general helicase motifs identified by Hodgman (2) and Gorbalenya *et al* (3) in viruses, bacteria and yeast lie within the DEAD box motifs, but the DEAD box motifs are longer and clearly identify a subfamily. Biochemical evidence that the proteins are RNA helicases only exists for human p68 (4,5) and the mouse translation initiation factor eIF4A (6), although the bacterial protein SrmB has been shown to possess RNA dependent ATPase activity (7). Several other members of the family are strongly implicated in RNA processing: MSS116 is required for yeast mitochondrial splicing (8); SPP8-1 mutants suppress mutations in PRP8, a yeast nuclear splicing gene (J. Beggs personal communication); prp16-1 suppresses a yeast lariat branchpoint mutation (9); and spb4-1 suppresses the inhibition of translation following deletion of yeast polyA binding protein (10). There are at least 12 DEAD box genes in *S.cerevisiae* (TIF1&2, MSS116, PRP16, SPB4, SPP8-1, DED1, p68, CA3,4,5/6 and 8 (11) and unpublished data) and in most cases

tested (TIF1 plus TIF2 (12), DED1 (13), p68 unpublished data) they are essential. For yeast to have evolved such a diverse array of RNA helicases implies that control of RNA secondary structure is a major preoccupation for the cell.

We are interested in p68 (14), the SV40 large T related antigen in which the DEAD box motifs were first identified (15). p68 is one of the best characterised DEAD box proteins *in vitro* (4,5), but crucial biological questions such as the nature of its RNA target remain to be answered. To allow us to pursue the function of p68 genetically we cloned it from yeast (Iggo and Lane, in preparation). When we looked for it in *E.coli* using an *S.pombe* p68 probe we cloned instead *dbpA*, the bacterial DEAD box gene which we describe in this paper.

MATERIALS AND METHODS

For details of bacterial strains used see Table 1.

Library screening

An *E.coli* library containing partial *Sau3A1* fragments of genomic DNA in λ PPE11 (16) was screened by the method of Huynh *et al* (17) using the low stringency hybridisation conditions of Miller *et al* (18) with a 1.2 kb fragment of the *S.pombe* p68 cDNA. (The human, *S.pombe* and *S.cerevisiae* p68 sequences have been given the accession numbers X15729, X52648 and X52649 respectively in the EMBL database.)

Hfr mapping

Selected positive clones from the λ library were used to create λ lysogens in strain AB1157 as described in Silhavy *et al* (19).

Table 1 *E.coli* strains

Strain	Genotype
AB1157	F ⁻ <i>thi-1, his-4, Δ(gpt-proA2)62, argE3, thr-1, leuB6, kgdK51, rtdD1, ara-14, lacY1, galK2, xyl-5, mtl-1, supE44, tsx-33, rpsL31. Rac⁺, λ⁻.</i>
KL99	Hfr <i>thi-1, rel-1. λ⁻.</i>
KL208	Hfr <i>Rac⁺, λ⁻.</i>
W3110	F ⁻ <i>IN(rrnD-rrnE). Rac⁺, λ⁻.</i>
DH5α	F ⁻ <i>endA1, hsdR17 (r_k⁻, m_k⁺), thi-1, recA1, gyrA96, relA1, Δ(argF-lacZYA)U169, supE44, #80dlacZΔM15. λ⁻.</i>

* To whom correspondence should be addressed

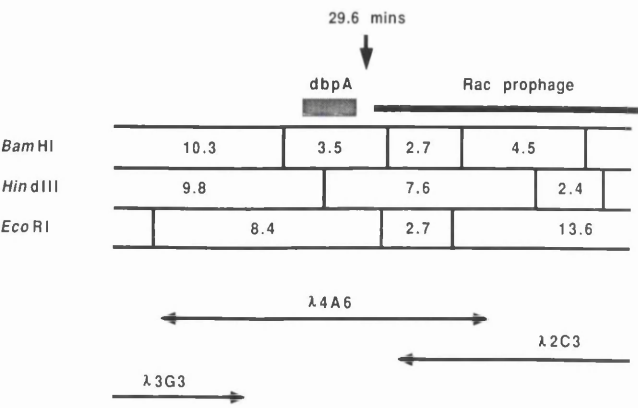


Figure 1. Physical map of the W3110 chromosome showing the location of *dbpA*, the Rac prophage and the three phages supplied by Dr Kohara (23,24,25). Restriction fragment sizes are in kb.

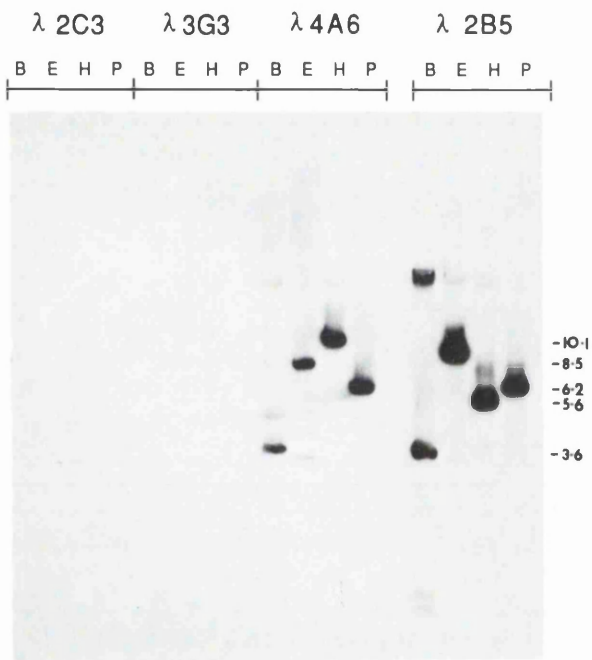


Figure 2. Southern blot of DNA from λ2B5 and the Kohara phages probed with the λ2B5 *Bam*HI fragment containing *dbpA*. B, *Bam*HI; E, *Eco*RI; H, *Hind*III; P, *Pst*I. Restriction fragment sizes are in kb.

Hfr mapping was performed essentially as described in Miller (20). Lysogens and non-lysogens were identified by cross-streaking with λ*vir* and λ*cintH80*.

DNA Sequencing

DNA subcloned in Bluescript (Stratagene) was sequenced on both strands using a Sequenase2 kit (United States Biochemical). Templates spanning the entire gene were prepared by making deletions in the 3.5 kb *Bam*HI fragment using restriction sites and sequenced with Bluescript primers or *dbpA* specific primers. The 5' and 3' ends of the open reading frame were checked by nested asymmetric PCR sequencing of genomic DNA essentially as described in Gyllensten and Erlich (21).

Sequence analysis

Fickett and diagon plots were performed using the University of Wisconsin Genetics Computer Group programs 'Testcode',

```

GCA AAT TCC TGT CCC TGA ACC GAT TCC TGG CGA TCC TGT TCC AGT GCC CGA TCC
GAT CCC TCG CCC GCA ACC CAT GCC TGA CCC ACC ACC CGA TGA AGA ACC GAT TAA
ATT GTC GCA TCG TGA GCG TAG ATC TGC GAG GAT ACG CGC CTG CTA ACT TTG CGT
CGA TGA CCA CGA GAA TAG ATT GTG ACC GCT TTT TCT ACC CTG AAT GTT TTG CCT
CCC GCC CAA CTC ACG AAC CTT AAT GAG TTG GGT TAT TTA ACC ATG ACG CCG GTG
M T P V

CAG GCC GCC GCG CTT CCG GCG ATC CTT GCC GGA AAA GAT GTT CGC GTG CAG GCG
Q A A A L P A I L A G K D V R V Q A

AAA ACC GGC AGC GGC AAA ACG GCG GCT TTT GGC CTC GGC TTG TTA CAG CAA ATT
K T G S G K T A A F G L G L L Q Q I

GAT GCG TCG CTA TTT CAA ACC CAG GCT TTA GTG CTG TGT CCT ACG CGT GAA CTG
D A S L F Q T Q A L V L C P T R E L

GCG GAT CAG GTG GCA GGT GAA TTG CGT CGG CTG GCG CGT TTT CTG CCA AAT ACC
A D Q V A G E L R R L A R F L P N T

AAA ATT TTG ACG TTG TGC GGT GGT CAA CCG TTC GGT ATG CAG CGT GAT TCG TTG
K I L T L C G G Q P F G M Q R D S L

CAA CAT GCG CCG CAT ATT ATC GTG GCA ACG CCG GGG CGT TTG CTG GAT CAC CTG
Q H A P H I I V A T P G R L L D H L

CAA AAA GGC ACG GTA TCA CTG GAT GCG TTG AAT ACG CTG GTG ATG GAT GAG GCC
Q K G T V S L D A L N T L V M D E A

GAC GCG ATG CTG GAT ATG GGA TTT AGC GAT GCC ATT GAT GAT GTC ATC CGT TTT
D R M L D M G F S D A I D D V I R F

GCG CCT GCA TCT CGA CAG ACG CTT CTG TTT TCG GCA ACC TGG CCG GAA GCC ATC
A P A S R Q T L L F S A T W P E A I

GCT GCA ATC AGC GGA CGA GTG CAA CGC GAT CCT TTG GCG ATT GAA ATT GAC TCA
A A I S G R V Q R D P L A I E A D S

ACA GAT GCT TTG CCA CCC ATT GAA CAA CAA TTT TAT GAG ACA TCC AGC AAA GGC
T D A L P P I E Q Q F Y E T S S K G

AAA ATT CCT CTG TTG CAA CCG TTA TTA AGC TTG CAT CAG CCA TCC TCT TGC GTG
K I P L L Q R L L S L H Q P S S C V

GTG TTT TGC AAT ACC AAA AAA GAT TGC CAG GCT GTC TGC GAC GCG CTG AAT GAA
V F C N T K K D C Q A V C D A L N E

GTA GGG CAA AGT GCA TTG TCA TTA CAC GGC GAT TTG GAG CAA CGC GAT CGC GAT
V G Q S A L S L H G D L E Q R D R D

CAG ACC CTG GTA CGT TTT GCT AAC GGT AGC GCC CGT GTA CTG GTC GCG ACT GAT
Q T L V R F A N G S A R V L V A T D

GTT GCT GCG CGT GGT CTG GAT ATT AAA TCG CTT GAG CTG GTG GTG AAC TTT GAG
V A A R G L D I K S L E V G N F E

CTG GCG TGG GAC CCT GAA GTT CAT GTA CAT CGC ATC GGT CGT ACA GCT CGT GCA
L A W D P E V H V H R I G R T A R A

GGA AAT AGC GGT CTG GCG ATC AGT TTC TGT GCT CCG GAA GAA GCA CAG CCG GCC
G N S G L A I S F C A P E E A Q R A

AAT ATC ATT TCT GAC ATG TTG CAG ATA AAA CTT AAC TGG CAA ACG CCG CCA GCT
N I I S D M L Q I K L N W Q T P A

AAT AGT TCC ATT GCG ACG CTG GAA GCA GAA ATG GCA ACG TTG TGT ATC GAT GGC
N S S I A T L E A E M A T L C I D G

GGG AAA AAA GCC AAA ATG CGC CCG GGT GAT GTA TTA GGT GCA CTG ACA GGA GAT
G K K A K M R P G D V L G A L T G D

ATC GGG CTT GAT GGC GCA GAT ATT GGC AAA ATC GCC GTG CAT CCG GCG CAT GTC
I G L D G A D I G K I A V H P A H V

TAT GTC GCG GTC CGT CAG GCT GTT GCT CAT AAA GCA TGG AAA CAG TTA CAG GGC
Y V A V R Q A V A H K A W K Q L Q G

GGG AAG ATT AAA GGA AAA ACG TGC CCG GTG CCG TTA TTA AAA TAA TGA AAT GTT
G K I K G K T C R V R L L K

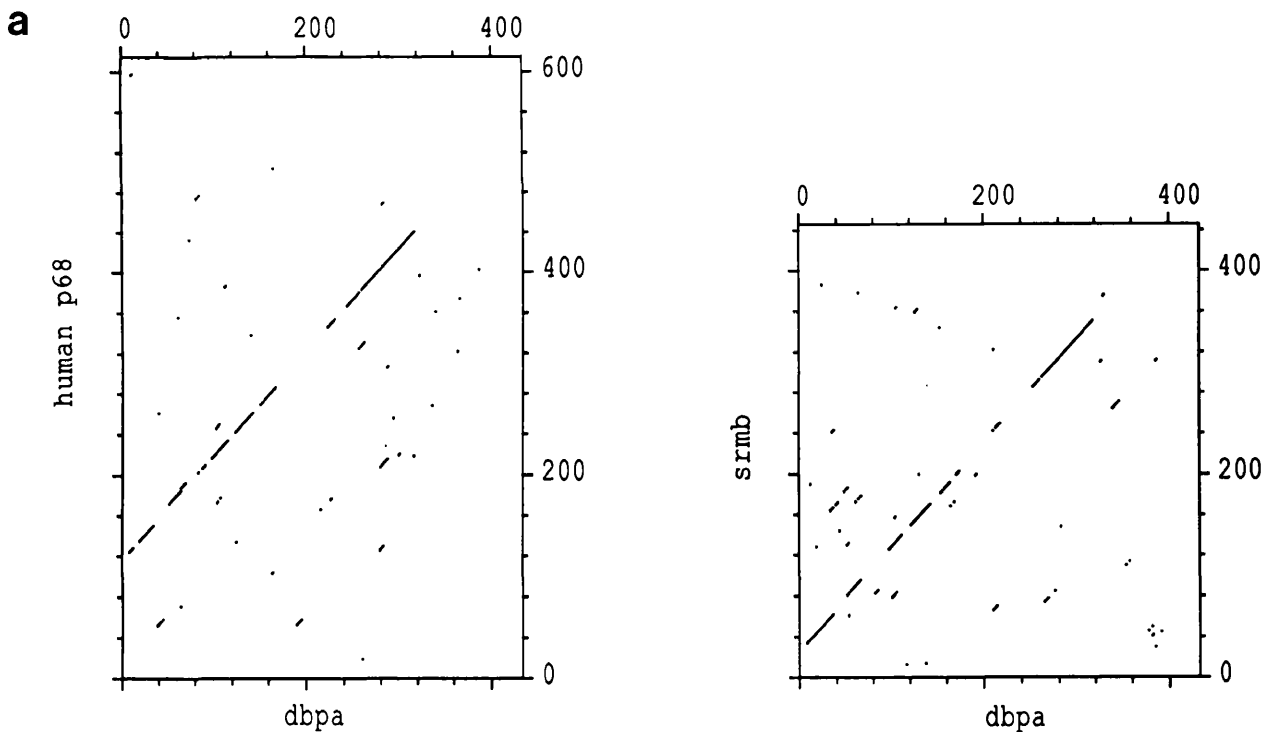
GAA TTG CCG GGT GCA AGA GTA AAC ATC TTA TTC GGG ATT GCC GGA TGC GAC GCT
GGC CGC GTC TTA TCC GGC CTC CAT AAG AGT AGC CCG ATA CGC TTG CGC
    
```

Figure 3. Nucleotide and deduced amino acid sequence of *dbpA*. Nucleotides matching the ribosome binding site consensus are underlined. The sequence has been deposited in the EMBL database under the accession number X52647.

'Compare' and 'Dotplot'. 'Compare' was used with a stringency of 15 and a window of 10.

Gene expression

DH5α cells grown on LB plates were lysed directly in SDS sample buffer and loaded immediately on a 12% polyacrylamide gel for Western blotting as described in Harlow and Lane (22). Nitrocellulose blots were blocked for 30 minutes in 2% dried milk/0.1% NonidetP40 in phosphate buffered saline (PBS),



b

```

MSGYSSDRDRGRDRGFGAPRFGGSRAGPLSGKKFGNPGEKLVKKKWNLDLDPKFEKNFYQEHPDLARRTAQEVETYYRRSKE - p68

ITVRGHNCPKPVLNFYEANFPANVMDVIARQNFTEPTAIQAQGWVVALSGLDMVGVVAQTGSGKTLSYLLPAIVHINHQPFFLERGDGPIC - p68
MTPVQAAALPAILAGKDVVRVQAKTGSKTAAFGLGLLQQIDASLFQTA - dbpa
MTVTTTFSELELDESLEALQDKGFRPTAIQAAAIPPALDGRDVLGSAPTGTGKTAAYLLPALQHLLDFPRKKSPPRI - srmB

..v.f.e.....ft.pTaiQAaa.P.aL.G.Dv.g.A.TGsGKTAaylLpalqhi...pf...g...i - cons
motif I

LVLAPTRELAQQVQVAAEYCRACRLKSTCIY GGAPKGPQIRDLEGRVEICIAATPGRLLIDFLECGKTNLRRTTYLVLDEADRMLDMGF - p68
LVLCPRELAQVAGELRRLARFLPNTKILTLCGGQPFQMGQRDSLQHAPHIIVATPGRLLDHLQKGTVSLDALNTLVMDADRMLDMGF - dbpa
LILTPRELAQVSDHARELAKHHLDIATIT GGVAYMNHAEVFSENQDIVVATTGRLLQYIKEENFDCRAVETLILDEADRMLDMGF - srmB

LvL.PTRELA.QV...arelar...l...i..GG.p.g.q...l...I.vATpGRLLd.l.g...lra..tLvLDEADRMLDMGF - cons
motif Ia motif II

EPQIRKIVDQIRPDRQTLMWSATWPKEVRQLAEDFLKDYIHINIGALELSANHNILQIVDVCHDVEKDEKLIRLMEEIMSEKENKTIVF - p68
SDAIDDVIRFAPASRQTLLFSATWPEAI AASGRVQRDPLAIEIDSTDALPPIEQQFYETSSKGKIPLLQRLLSLHQFPSSC VVF - dbpa
AQDIEHIAGETWRKQTLFSATLEGDAIQDFAERLLEDPVEVSANPSTRERKKIHQWYYRADDLEHKTALLVHLLKQPEATRS IVF - srmB

...I.i...r.rQTLlfsATwp...q.a...l.....i...st.a...I.Q..y...d.ek...L.rll.....iVF - cons
motif III motif IV

VETKRRCDLTKRMRDQWPMAGIHGDKSQQERDQVWLNDFKHKAPILIAATDVASRGLDVEDVKFVINYDYNPNSSEDIYHRIGRTARST - p68
CNTKKDCQAVCDALNEVGQSALS LHGDLEQRDRDQTLVRFANGSARVLVATDVAARGLDKSLELVNFELAWDPEVHVHRIGRTARAG - dbpa
VRKRERVHELANWLREAGINNCYLEGEMVQKRNIAIKRLTEGRVNVLVATDVAARGIDIPDVSHVFNFDMPRSGDTYLHRIGRTARAG - srmB

v.tk.rc.el...lre.G.a..lhGd..Q..Rd..l.rf..G.a.vLvATDVAARGLDi.dv..V.Nfd.p.s.e.y.HRIGRTARag - cons
motif V motif VI

KTGTAYTFFTPNNIKQVSDLISVLRANQAINPKLLQLVEDRGSRSRGRGGMKDDRRDRYSAGKRGGFNTFRDRENYDRGYSSLLKRD - p68
NSGLAISFCAPEEAQRANIISDMLQIKLNWQTPPANSSIATLEAEMATLIDGGKKAKMRPGDVLGALTGDIGLDGADIGKIAVHPAHV - dbpa
RKGTAISLVEAHDHLLGKVGGRYIEEPKARVIDELRPKTRAPSEKQTKGPKKVLAKRAEKKKAKEKEKPRVKKRHRDTKNIKGRKRP - srmB

..G.Aisf..p.....l.e...a.p.l.....se..tg...k.ak.r.....d.k..... - cons

FGAKTQNGVYSAANYTNGSFGSNFVSAGIQTSFRITGNPTGTQNGYDSTQQYGSNVNPMHNGMNQQAAYAPATAAAPMIGYPMPGTYSQ - p68
YVAVRQAVAHKAWKQLQGGKIKGKTCRVRLK - dbpa
SGTGVPPTTEE - srmB

.ga..q....a....g..... - cons
    
```

Figure 4. Comparison of *dbpA* with human p68 and *srmB*. a, diagonal plots, b, deduced amino acid sequences in single letter code. In the consensus line amino acids identical in all three proteins are shown in capitals and those shared by only two sequences are in lower case. Sequences matching the Hodgman/Gorbalenya motifs are indicated; motif II is the DEAD box (2,3).

probed overnight with hybridoma supernatant plus 0.1% NonidetP40, washed in PBS, probed for 3 hours with rabbit anti-mouse immunoglobulin conjugated to alkaline phosphatase (Dako) diluted 1/1000 in blocking buffer, washed in PBS and developed with nitroblue tetrazolium/bromochloroindolyl phosphate. Preparation of the monoclonal anti-DEAD antibody will be described elsewhere.

RESULTS

We screened an *E. coli* genomic library by low stringency DNA hybridisation with an *S. pombe* p68 probe and obtained four λ clones (2A3, 2B2, 2B3 and 2B5). Hfr crossing of lysogens showed that λ 2A3, 2B2 and 2B5 map between the origins of Hfr KL99 and KL208, ie between 24 and 31 minutes. The location of λ 2B3 was not determined.

The *S. pombe* p68 probe hybridized with a 3.5 kb *Bam*HI fragment of *E. coli* genomic DNA and an apparently identical fragment of phages 2A3, 2B2 and 2B5 (data not shown). We subcloned this *Bam*HI fragment from λ 2B5 and located it at 29.6 minutes on the physical map (23) of the *E. coli* chromosome (figure 1) by genomic Southern blotting (data not shown). Dr Kohara kindly supplied us with three of the λ phage used to create this region of the map (2C3, 3G3 and 4A6), which allowed us to confirm the deduced position (figure 2).

We identified an open reading frame, which we call *dbpA*, within the 3.5 kb *Bam*HI fragment (figure 3). The deduced amino acid sequence is 432 amino acids long and has a predicted molecular weight of 46.5 kD. Figure 4 shows a comparison of *dbpA* with human p68 and *srnB*. In the region of overlap 29% of amino acids are identical, a value similar to that for other distinct members of the DEAD box family. *dbpA* contains all the previously identified DEAD box motifs (1,11). Despite the fact that it was cloned by hybridisation with p68 DNA, *dbpA* is no more closely related to p68 than it is to any other member of the family. Overall, the DEAD box proteins have a conserved core region and divergent amino and carboxy termini. *dbpA* has the shortest amino terminal region of any DEAD box protein, with only 25 amino acids before the G-GKT motif (part of the predicted nucleotide binding site). We are convinced that we have correctly identified the amino terminus because Fickett (third codon preference) plots strongly predict that the region 5' of the deduced methionine is non-coding; there is an 8 out of 11 match to the consensus Shine Dalgarno sequence ending at position -12 relative to the methionine (underlined in figure 2); and finally we rechecked the sequence by direct asymmetric PCR sequencing from genomic DNA. We searched the PIR database with the carboxy terminal sequence of *dbpA* using PROSRCH (JF Collins and AFW Coulson, University of Edinburgh) on an Active Memory Technology Distributed Array Processor and found no significant homology to other proteins.

We have generated a monoclonal antibody against the p68 peptide LVLDEADRMLDMGFEPQ. Preliminary results show that it reacts strongly with Western blots of bacterially overexpressed p68 and *SrmB*, and is specifically blocked by peptide (data not shown). Western blotting of *E. coli* whole cell extract (figure 5, lane 4) shows that endogenous DbpA and *SrmB* are below the limit of detection with this antibody. Since the antibody readily detects human p68 in Hela cell extracts this suggests that both *dbpA* and *srnB* are normally expressed at very low levels. When we transformed the same *E. coli* strain with a high copy number plasmid (Bluescript) containing the 3.5 kb

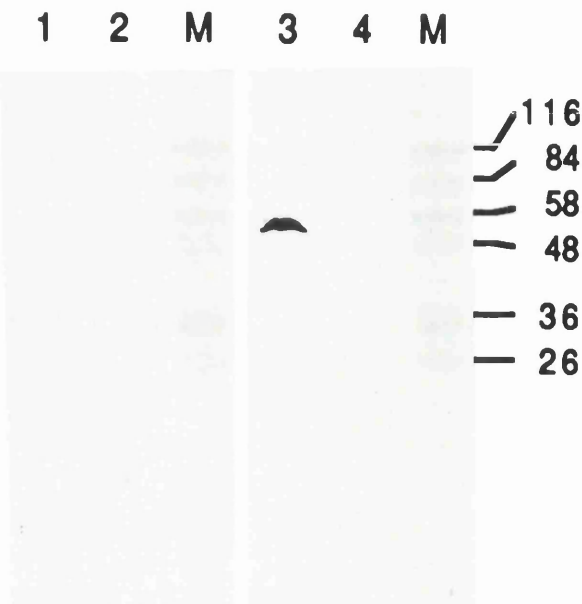


Figure 5. Western blot of bacteria overexpressing DbpA (lanes 1 and 3) or containing a control plasmid (lanes 2 and 4) probed with an anti-DEAD peptide monoclonal antibody, plus (lanes 1 and 2) or minus (lanes 3 and 4) 100 μ g/ml DEAD peptide. Sizes of prestained molecular weight markers are given in kD.

*Bam*HI fragment expressing *dbpA* from its own promoter we readily detected DbpA by Western blotting (figure 5, lane 3). The complete abolition of binding of the antibody to the protein by the LVLDEADRMLDMGFEPQ peptide confirms the specificity of the reaction (figure 5, lane 1).

DISCUSSION

We have cloned, sequenced, mapped and overexpressed a new gene encoding a novel DEAD box protein in *E. coli*. We call the gene *dbpA*, for *dead box protein A*. It is remarkable that it was cloned by DNA cross hybridisation from a eukaryotic to a prokaryotic organism. The gene product is a ~50,000 Dalton protein. Our identification of the DbpA protein was dependent upon the use of a new anti-peptide monoclonal antibody to the conserved motif LVLDEADRMLDMGFEPQ. Anti-DEAD antibodies should prove extremely useful in identifying and characterising new members of the DEAD box family.

Hfr mapping localised *dbpA* to between 24 and 31 minutes on the *E. coli* chromosome. By comparison of the restriction map of the *dbpA* locus with that of the chromosome we physically mapped the gene to 29.6 minutes and were able to confirm this directly using the λ clones from which the map was originally created. This is a relatively silent area of the chromosome, close to the termination site for DNA replication. It has been noted before that genes from this region are only expressed at low levels and this is consistent with our immunochemical analysis of *dbpA* expression. The *dbpA* gene maps near the site of integration of the *Rac* prophage. Our mapping and sequencing data clearly indicate that *dbpA* is not part of the prophage. Furthermore, we have shown by Southern blotting that *dbpA* is present in strain AB1157, which lacks the *Rac* prophage (data not shown). Examination of the most recent map of the *E. coli* chromosome and consultation with Dr B. Bachmann (Curator, *E. coli* Genetic Stock Center) establish that *dbpA* is a new gene.

The striking sequence homology of *dbpA* to other DEAD box proteins permits a confident prediction that the protein will be involved in altering the secondary structure of RNA molecules. Possible roles for the protein include ribosome assembly, translational control, transcriptional control or tRNA processing. Overexpression of the only other known bacterial DEAD box gene, *srnB*, suppresses a defect in ribosome assembly at the non-permissive temperature in cells harboring a temperature sensitive mutation in the ribosomal protein L24 (7). However, the normal function of *srnB* is unknown. Since *E. coli* contains at least two divergent members of the DEAD box family we speculate that there may be a multigene family of these proteins in all living organisms. The very early emergence of an ancestral DEAD box protein which was able to catalyse the allosteric conversion of RNA molecules is consistent with current speculation that life evolved in an RNA world. The ability to replicate nucleic acid depends critically on the ability to separate template from copied strands, hence it is reasonable to expect that helicases should be as ancient as polymerases.

Since *dbpA* was obtained by 'reverse genetics' we do not yet know the phenotype of *dbpA* mutants. Central to our understanding of the function of *dbpA* will be the identification of its RNA substrate.

ACKNOWLEDGEMENTS

We thank Dr P.T. Emmerson for the *E. coli* library, Iain Goldsmith for oligonucleotide synthesis and Michelle Ginsberg for help with computing.

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.
- Hodgman, T. C. (1988) *Nature* **333**, 22–23.
- Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. and Blinov, V. M. (1989) *Nucl. Acids. Res.* **17**, 4713–4730.
- Hirling, H., Scheffner, M., Restle, T. and Stahl, H. (1989) *Nature* **339**, 562–564.
- Iggo, R. D. and Lane, D. P. (1989) *EMBO J.* **8**, 1827–1831.
- Rozen, F., Edery, I., Meerovitch, K., Dever, T. E., Merrick, W. C. and Sonenberg, N. (1990) *Mol. Cell. Biol.* **10**, 1134–1144.
- Nishi, K., Morel-Deville, F., Hershey, J. W. B., Leighton, T. and Schnier, J. (1988) *Nature* **336**, 496–498.
- Seraphin, B., Simon, M., Boulet, A. and Faye, G. (1989) *Nature* **337**, 84–87.
- Burgess, S., Couto, J. R. and Guthrie, C. (1990) *Cell* **60**, 705–717.
- Sachs, A. B. and Davis, R. W. (1990) *Science* **247**, 1077–1079.
- Chang, T.-H., Arenas, J. and Abelson, J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1571–1575.
- Linder, P. and Slonimski, P. P. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2286–2290.
- Struhl, K. (1985) *Nucl. Acids Res.* **13**, 8587–8600.
- Lane, D. P. and Hoeffler, W. K. (1980) *Nature* **288**, 167–170.
- Ford, M. J., Anton, I. A. and Lane, D. P. (1988) *Nature* **332**, 736–738.
- Arthur, H. M., Bramhill, D., Eastlake, P. B. and Emmerson, P. T. (1982) *Gene* **19**, 285–295.
- Huynh, T. V., Young, R. A. and Davis, R. W. (1988) In Glover, D. M. (ed) *DNA Cloning—A Practical Approach*, IRL Press, Oxford Vol. 1, pp49–78.
- Miller, M. A., Korn, D. and Wang, T. S.-F. (1988) *Nucl. Acids Res.* **16**, 7961–7973.
- Silhavy, T. J., Berman, M. L., Enquist, L. W. (1984) *Experiments with gene fusions*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Miller, J. H. *Experiments in molecular genetics* (1972) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Gyllenstein, U. B. and Erlich, H. A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7652–7656.
- Harlow, E. and Lane, D. P. (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Kohara, Y., Akiyama, K. and Isono, K. (1987) *Cell* **50**, 495–508.
- Kaiser, K. and Murray, N. E. (1979) *Molec. Gen. Genet.* **175**, 159–174.
- Willis, D. K., Satin, L. H. and Clark, A. J. (1985) *J. Bacteriol.* **162**, 1166–1172.

p68 RNA Helicase: Identification of a Nucleolar Form and Cloning of Related Genes Containing a Conserved Intron in Yeasts

R. D. IGGO,¹ D. J. JAMIESON,² S. A. MACNEILL,³ J. SOUTHGATE,¹ J. MCPHEAT,¹ AND D. P. LANE^{1*}

Molecular Immunochimistry Laboratory, Imperial Cancer Research Fund Clare Hall Laboratories, South Mimms, Herts, EN6 3LD,¹ Department of Molecular Biology, University of Edinburgh, Edinburgh, EH9 3JR,² and Imperial Cancer Research Fund Cell Cycle Group, Department of Biochemistry, University of Oxford, Oxford, OX1 3QU,³ United Kingdom

Received 1 October 1990/Accepted 4 December 1990

The human p68 protein is an RNA-dependent ATPase and RNA helicase which was first identified because of its immunological cross-reaction with a viral RNA helicase, simian virus 40 large T antigen. It belongs to a recently discovered family of proteins (DEAD box proteins) that share extensive regions of amino acid sequence homology, are ubiquitous in living organisms, and are involved in many aspects of RNA metabolism, including splicing, translation, and ribosome assembly. We have shown by immunofluorescent microscopy that mammalian p68, which is excluded from the nucleoli during interphase, translocates to prenucleolar bodies during telophase. We have cloned 55% identical genes from both *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* and shown that they are essential in both yeasts. The human and yeast genes contain a large intron whose position has been precisely conserved. In *S. cerevisiae*, the intron is unusual both because of its size and because of its location near the 3' end of the gene. We discuss possible functional roles for such an unusual intron in an RNA helicase gene.

Manipulation of RNA secondary structure is essential for the proper execution of a large number of processes in the cell. The importance that the cell attaches to this problem has only recently come to light with the discovery of a large family of proteins with putative ATP-dependent RNA helicase activity. These proteins are readily identified because they all possess a core region of highly conserved motifs in their primary amino acid sequences. One of the most conserved motifs is the DEAD box (LDEADxxL), which gives the family its name (34). DEAD box proteins form a subset of a more loosely defined family, identified by Hodgman (27) and Gorbalenya et al. (21), of proteins with known or putative helicase activity.

RNA helicase activity has been demonstrated in vitro for only two DEAD box proteins: p68 (26), in which the DEAD motifs were first identified (18), and eIF4A (44), a translation initiation factor. Although the biological processes in which the other members are involved may prove very diverse, it seems reasonable to expect that all of them will show RNA helicase activity, at least in strand displacement assays in vitro. The family appears to be very large, since DEAD box proteins have been found in all prokaryotes and eukaryotes examined, and even the relatively small genome of *Saccharomyces cerevisiae* contains genes encoding at least 12 different members (*TIF1* and -2 [35], *MSS116* [49], *SPB4* [46], *PRP5* [11], *PRP28* [52a], *DED1* [42a], *DBP1* [29a], *DBP2* [this report], and *CA3*, *CA4*, and *CA5/6* [7]).

We are interested in p68 because of its specific immunological cross-reaction with the simian virus 40 large T antigen (31). We have taken two approaches to determine the function of p68. First, we have used immunofluorescent microscopy and antibody microinjection to study the cell biology of p68 in mammalian cells. Using antibodies to

nucleolar antigens, we show that p68 is transiently associated with the nucleoli during late telophase, at the time when prenucleolar bodies are condensing after mitosis. Antibodies to p68 microinjected into the cell cytoplasm are specifically transported to the nucleoli, although they do not grossly disrupt cell growth. Second, by cloning related genes from *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*, we show that both the deduced amino acid sequence and the gene structure of p68 have been highly conserved in evolution. Gene disruption demonstrates that the yeast genes are essential for viability in both yeast species.

MATERIALS AND METHODS

Cell staining. C6 cells were provided by M. Oren, 293 cells were provided by P. Gallimore, and CV1, 3T3, and HFF cells were provided by Imperial Cancer Research Fund cell production. Antinucleophosmin monoclonal antibody hybridoma supernatant (41) was provided by P. K. Chan. PAb204 (anti-p68), BG2 (anti- β -galactosidase), and CAT2 (anti-chloramphenicol acetyltransferase) monoclonal antibodies were prepared in this laboratory. Human antinuclear antibody (nucleolar pattern: ANA-N) was purchased from Sigma. For immunofluorescent microscopy, cells were fixed for 5 min in acetone-methanol. Primary antibodies were visualized with commercial fluorescein- and rhodamine-conjugated anti-immunoglobulin sera except for the cells shown in Fig. 1C (biotinylated PAb204 plus streptavidin-fluorescein), 2E (biotinylated PAb204 plus streptavidin-Texas red), and 2F (directly conjugated CAT2-fluorescein).

Affinity-purified anti-p68 immunoglobulin was prepared by acid elution from a p68 column of serum from a rabbit (MF6) immunized with a gel-purified bacterially expressed p68/ β -galactosidase fusion protein. The p68 column was made by incubating HeLa cell extract with PAb204 (28) and immobilizing the immune complexes on a rabbit anti-mouse immunoglobulin column by coupling with dimethyl pimelidate. The rabbit anti-mouse immunoglobulin column was prepared

* Corresponding author.

† Present address: Department of Biochemistry, The University, Dundee. DD1 4HN, United Kingdom.

by reacting Dako rabbit anti-mouse immunoglobulins with Pharmacia cyanogen bromide-activated Sepharose 4B. PAb204 leakage from the column can be ruled out because a peptide corresponding to the PAb204/p68 epitope (18) abolishes immunofluorescent cell staining with PAb204 but has no effect on affinity-purified MF6 staining.

Microinjection. PAb204 was purified from serum-free hybridoma supernatant by chromatography on protein A-Sepharose in the presence of 3.3 M NaCl (23). BG2 was purified from monoclonal ascites fluid by chromatography on a rabbit anti-mouse immunoglobulin column (23). Antibody at 10 mg/ml was injected in buffer containing 57 mM KCl, 7.5 mM NaCl, 2.75 mM MgCl₂, and 5 mM Tris (pH 7.0). Cells in *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES)-buffered Dulbecco modified Eagle medium maintained at 35°C on a heated microscope stage were injected under phase contrast, using a Zeiss model MR micromanipulator and an Eppendorf model 5242 microinjector.

Transfection. 293 cells were transfected with pCIS5p68, which contains a full-length human p68 cDNA clone under the control of the cytomegalovirus major immediate-early promoter, using the method of Gorman et al. (22). To distinguish endogenous from transfected p68, cells were cotransfected with pCIS-CAT (which contains the chloramphenicol acetyltransferase gene under the control of the immediate-early promoter) and identified by double labeling with CAT2 and PAb204. Control transfections with pCIS-CAT alone showed that expression of chloramphenicol acetyltransferase does not alter the pattern of endogenous p68 staining. pCIS5 and pCIS-CAT (22) were kindly supplied by C. Gorman.

DNA cloning. General cloning techniques were performed as described by Sambrook et al. (47). The Fickett analysis in Fig. 5 was done by using the University of Wisconsin Genetics Computer Group program Testcode with a window of 300 bp (14, 16). To map *DBP2*, *S. cerevisiae* chromosomes were resolved by using a CHEF-DRII gel system (Bio-Rad), Southern blotted, and hybridized to a *DBP2* probe spanning nucleotides 928 to 2242.

Disruption. Deletion of the *S. pombe dbp2* and *S. cerevisiae DBP2* genes was accomplished by one-step gene replacement (43). The *S. cerevisiae URA3* gene was cloned into the *EcoRV* and *BglII* sites of *DBP2* (between amino acids 375 and 497). A *BalI* fragment (from nucleotide 963 to a *BalI* site 0.7 kb downstream of the *BglII* site) was gel purified and transformed into DJY84 (a *MAT α ura3-52/ura3-52 his3- Δ 1/his3- Δ 200 ADE2/ade2-101 LYS2/lys2-801 trp1-289/trp1- Δ 1 leu2-3,112/LEU2* strain) by the lithium acetate technique (29). Stable diploid Ura⁺ colonies were picked and analyzed by polymerase chain reaction (PCR) (from a *DBP2* primer [CTCAACCATTATTGGCTCCAG] to a *URA3* primer [GGTAGAGGGTGAACGTTACAG]) and Southern blotting to identify transformants in which one copy of *DBP2* had been disrupted. Cells from the strain (DJY98) containing the disruption were sporulated, and tetrads were dissected and allowed to germinate on YPD plates (50). To demonstrate rescue of the inviable spores by *DBP2*, a full-length *DBP2* cDNA (from nucleotides 61 to 2776) under the control of the *GAL1* promoter on a *CEN/ARS/TRP1* plasmid was transformed into diploid DJY98 cells. This strain was then sporulated, and tetrads were dissected as described above, analyzed for uracil prototrophy (indicating the presence of a disrupted chromosomal copy of *DBP2*) and tryptophan prototrophy (indicating the presence of the plasmid copy of the *DBP2* cDNA), and

subjected to PCR to confirm the integrity of the recombinant DNA molecules in the recovered strains.

The *S. pombe dbp2* gene (nucleotides 187 to 2648) with the *S. pombe ura4* gene inserted between the *MluI* and *AflII* sites (from amino acid 205 to a position within the intron) was cloned into the *EcoRV* site of Bluescript (Stratagene); ~10 μ g of insert DNA (from the *BamHI* to *SalI* sites in Bluescript) was gel purified and transformed by the method of Beach and Nurse (1) into a *leu1-32/leu1-32 ura4-D18/ura4-D18 ade6-M210/ade6-M216 h⁺lh⁻* strain (40). Following 7 days of growth on minimal medium plus leucine, stable diploid Ura⁺ transformant colonies were picked and analyzed by PCR (from a *dbp2* primer [AACCTTGTAACC TCAAATTACC] to a *ura4* primer [CCACAAATTCATGT TTGGACT]) to identify transformants in which *dbp2* had been partially deleted as intended. Diploids containing the deletion were sporulated on malt extract medium, and tetrads were dissected by using a micromanipulator (40).

Nucleotide sequence accession numbers. The nucleic acid sequences (including complete sequences for the intron shown in Fig. 4) have been deposited in the EMBL data base under accession numbers X15729 (human), X52648 (*S. pombe*), and X52649 (*S. cerevisiae*).

RESULTS

p68 is present in prenucleolar bodies. We have previously reported that mammalian p68 shows dramatic changes in nuclear location during the cell cycle (28). Double-label immunofluorescent cell staining with anti-p68 and antinucleolar antibodies (Fig. 1) shows that the densely staining structures seen with PAb204 during telophase are prenucleolar bodies. This is in contrast with the distribution during interphase, when p68 is present in the nucleoplasm and excluded from the nucleoli. Nucleophosmin is an abundant protein in the granular component of the interphase nucleolus (52). Although nucleophosmin is detectable in prenucleolar bodies by electron microscopy (42), it is not seen by light microscopy until early G₁ (41). Double labeling of acetone-methanol-fixed cells shows that there is very little overlap between the p68 and nucleophosmin staining patterns: nucleolar PAb204 staining is at its most intense when the majority of nucleophosmin is in the cytoplasm (Fig. 1C and D).

The proliferating cell nuclear antigen gives transient S-phase nucleolar staining (5), apparently as a result of masking of the dominant epitope on the proliferating cell nuclear antigen in the nucleolus during the rest of the cell cycle (55). To establish whether the cell cycle variation in p68 staining reflects epitope masking or bulk movement of p68, we repeated the staining with affinity-purified anti-p68 serum (MF6) depleted of antibodies to the PAb204 epitope. Affinity-purified MF6 stains prenucleolar bodies in an identical manner to PAb204 (Fig. 1E to H), suggesting that the cell cycle changes are not simply a result of epitope masking.

Microinjection of PAb204. The strong conservation of the PAb204 epitope, both across species and between p68 and simian virus 40 T antigen, suggests that it may be part of a protein- or RNA-binding site (28). According to this model, microinjection of PAb204 antibody should significantly interfere with p68's normal function. In practice, 3T3, HFF, C6, and CV1 cells all continue to grow and divide apparently normally following injection (Fig. 2A and C). The antibody was injected into the cytoplasm but should have gained free access to nuclear p68 molecules following nuclear envelope breakdown at mitosis. Even in postmitotic cells, antibody

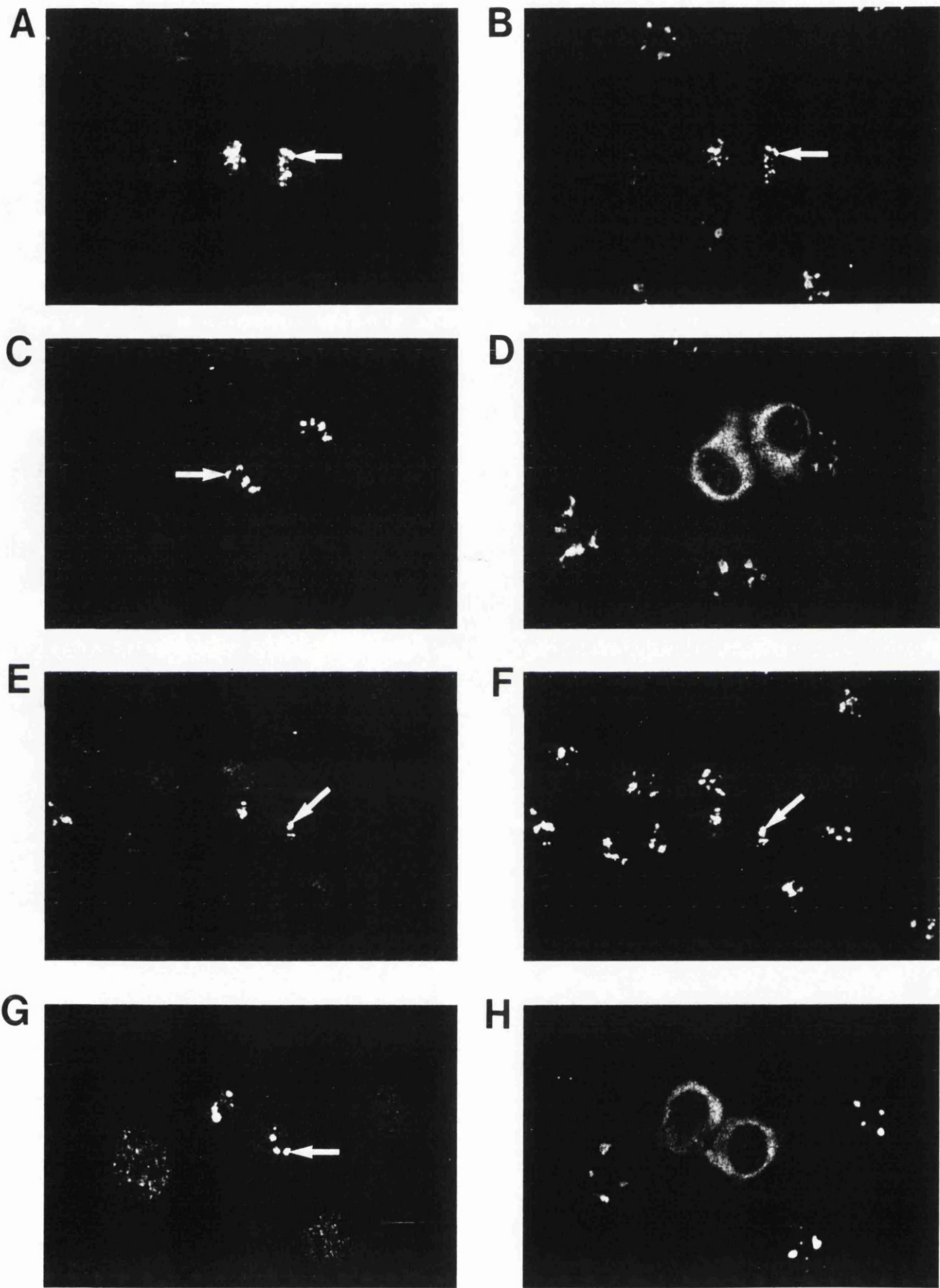


FIG. 1. Double-label immunofluorescence showing anti-p68 (A, C, E, and G) and antinucleolar (B, D, F, and H) staining of C6 cells. (A and C) PAb204; (E and G) MF6; (B and F) ANA-N; (D and H) antinucleophosmin. Prenucleolar bodies are marked with arrows. Magnifications: A, B, E, and F, $\times 375$; C, D, G, and H, $\times 590$.

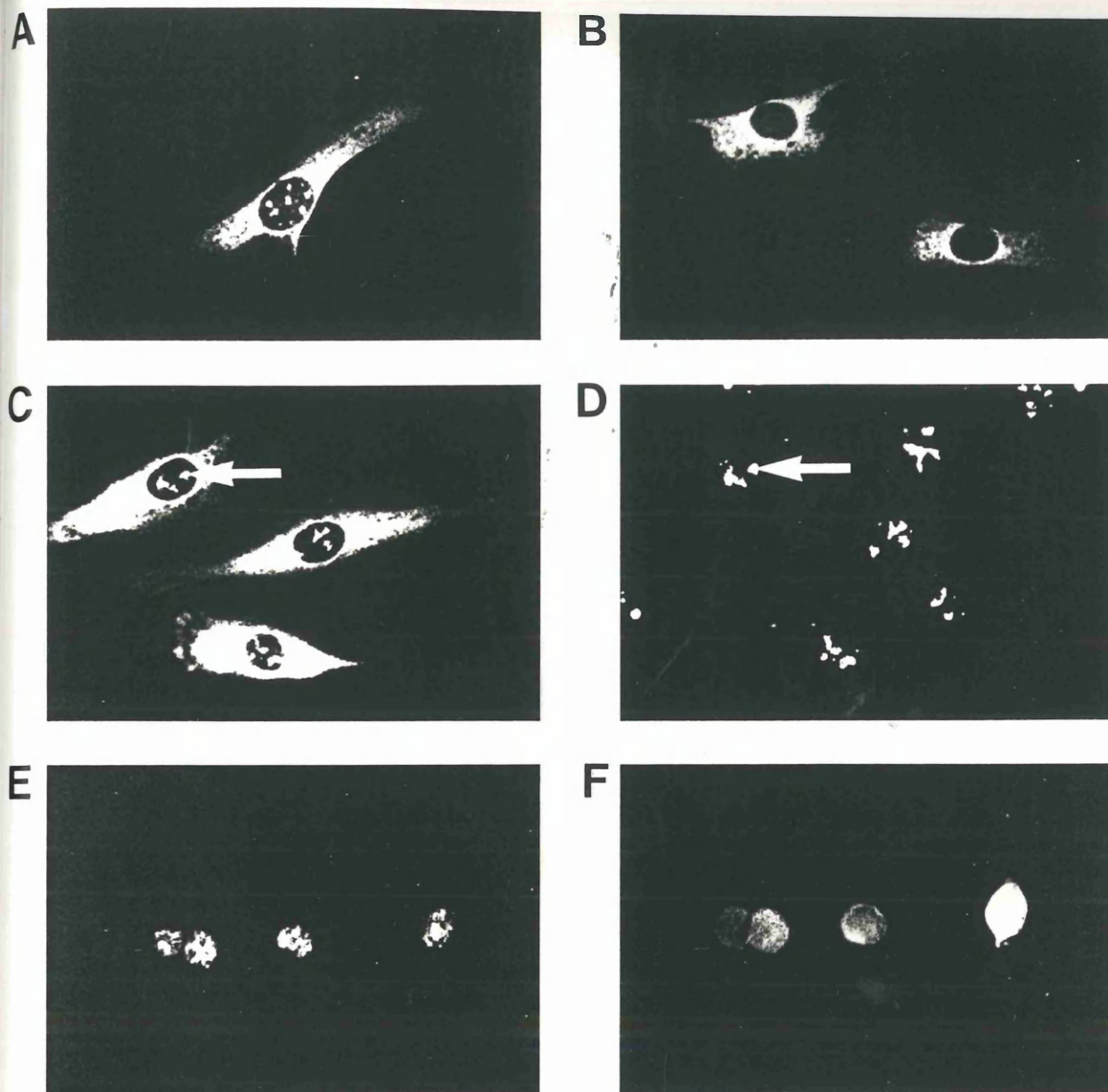


FIG. 2. Microinjection of PAB204 (A, C, and D) and BG2 (B) into 3T3 (A and B) and C6 (C and D) cells fixed after 24 h and stained with rhodamine-conjugated anti-mouse immunoglobulin (A to C) and ANA-N (D; same field as in panel C). Nucleolar staining in panels C and D is marked with arrows. Double label of 293 cells transfected with a p68 overexpression construct: (E) PAB204; (F) CAT2 (a marker for transfected cells). Magnifications: A, B, E, and F, $\times 330$; C and D, $\times 375$.

never accumulated to high levels in the nucleoplasm, although p68 could still be detected there by double labeling (data not shown). In the three fibroblast cell lines, PAB204 was transported to the nucleoli following injection into the cytoplasm (Fig. 2C and D). This is not an artifact because it could be seen in living cells injected with directly labeled antibody (data not shown) and was not seen with control antibodies (Fig. 2B). The nucleolar staining was only seen in cells that had divided. One possible explanation is that p68 enters the nucleus only at mitosis. This is unlikely because transfected p68 expressed from an exogenous promoter readily enters the nucleus, with no cytoplasmic accumulation in interphase cells (Fig. 2E and F).

Cloning of yeast genes related to p68. To permit genetic analysis of p68's function, we sought yeast homologs by low-stringency screening of yeast λ gt11 libraries (38). We obtained two identical *S. cerevisiae* genomic clones, both of which start within a 1-kb intron, and four *S. pombe* cDNA clones, of which three are identical and the fourth encodes the same open reading frame interrupted by a 700-bp intron. Full-length clones were obtained by a combination of library and PCR techniques and shown to be single-copy genes by Southern blotting of genomic DNA (data not shown). We call the genes *dbp2* (*S. pombe*) and *DBP2* (*S. cerevisiae*), for DEAD box protein 2. Southern blots of *S. cerevisiae* chromosomes resolved by pulsed-field gel electrophoresis

MTYGGRDQQYNKTKYKSRGGDFRGGNRSDRNSYNDRPQGGNYRGGFGGSRSNYNQPOELIKPNWDEELPKLPTFEKNFYVEHESVRDRS - *S. cerevisiae*
 MSYRDNEYSNGYNGKEDGYNSRGRYGGYRNYSRGGGRGGFNDGASYGYDQRGQGRNFYESDGPANLVKDKWNETLIPFQKDFYKEHENVRNRS - *S. pombe*
 MSGYSSDRDRGRDRGFAPRFGSRRAGPLSGKKFGNPGEKLVKKKWNLDLDPKFEKNFYQEHPLARRT - Human

.....sr...r...ndga..g...rg.g.....pg..lvkkdw.l..Lp.FeKnFY.EHe.vr.Rs - cons

DSEIAQFRKENEMTISGHDIPKPIITTFDEAGFPDYVLNEVKAEGFDKPTGIQQGWPMALSGRDMVGIATGSGKTLSCYCLPGIVHINAQPLLAPGD - *S. cerevisiae*
 DAEVTEYRKEKEIVVHGLNVPKPVITTFEAGFPNYVLKEVKQLGFEAPTPIQQQAWPMAMSGRDMVGIATGSGKTLSCYCLPAIVHINAQPLLSPGD - *S. pombe*
 AQEVETYSRKEITVRGHNCPPKLVNFYEANFPANVMDVIARQNFTEPTAIOAQGWVVALSGLDMVGAQTGSGKTLSCYLLPAIVHINHQPFLERGD - Human

d.Ev..yRkekEitv.Ghn.PKPVtTF.EAgFP.yvL.evk..gF..PT.IQ.QgWPMAlSGrDMVGIaaTGSGKTLSCYcLPaIVHINAQpLL.pGD - cons
 motif I

GPIVLVLAPTRELAVQIQTECSKFGHSSRIKNTCVYGGVPKSGQIRDLRSRSEIVIATPGRLLDMLDIEIGKTNLKRVTYLVLEADRMLDMGFEPQIR - *S. cerevisiae*
 GPIVLVLAPTRELAVQIQTECKFKGSSRIKNTCVYGGVPLGPQILDLRGVEICIAITPGRLLDMLDSNKTNLRRVTYLVLEADRMLDMGFEPQIR - *S. pombe*
 GPICLVLAPTRELAVQIQVAAEYCRACRLKSTCIYGGAPKGPQIRDLERLVEICIAITPGRLLDMLDIEIGKTNLRRVTYLVLEADRMLDMGFEPQIR - Human

GPIVLVLAPTRELAVQIQec.kfg.ssRirntCvYGGVPkqPQIRDL.RGvEicIATPGRLLDmLe.gKTNLrRvTYLVLEADRMLDMGFEPQIR - cons
 motif Ia motif II

KIVDQIRPDRQTLMSATWPKEVKQLAADYLNQVQVGSLELSASHNITQIVVWVSDFEKDRNLNKYLETASQDNEYKTLIFASTKRMCDITKY - *S. cerevisiae*
 KIVDQIRPDRQTVMSATWPKEVRLARDYLDNDYIQVTVGSLDLAASHNIKQIVVVDNADKRRALGKIDIEVLKDRDNKVLIFTGTRVADDITRF - *S. pombe*
 KIVDQIRPDRQTLMSATWPKEVRLAEDFLKDYIHNIGALELSANHNILQIVDVCHDVEKDEKLIRLMEEIMSEKENKTIVFVETKRRCEDELTRK - Human

KIVDQIRPDRQTLMSATWPKEV.qLA.DyLnDyIqv.vGsLeLsAshNI.QIVeVv.d.eKr.rL.k..Ee...d.enK.liF..TKR.cDdiTr. - cons
 motif III motif IV

LREDGWPALAIHGDKQDRERDQVLEFRNGRSPIMVATDVAARGIDVKGINYYINYDMPGNIEDYVHRIGRTGRAGATGTAFSFFTEQNKGLGAKLI - *S. cerevisiae*
 LRQDGPALAIHGDKAQDERDQVLEFRNGRSPIMVATDVAARGIDVKGITHVFNDFPNTEDYVHRIGRTGRAGATGTAFSFFTEQNKGLGAKLI - *S. pombe*
 MRRDGPAMPAGIHGDKSQQERDQVLEFRNGRSPIMVATDVAARGIDVKGIVINYDMPGNIEDYVHRIGRTGRAGATGTAFSFFTEQNKGLGAKLI - Human

1R.DGWPALAIHGDK.Q.ERDQVLEFr.GksPImvATDVAARGIDVkgi..ViNYD.Pgn.EDYVHRIGRTgRagatGTAYtfft..N.kq...Li - cons
 motif V motif VI

SIMREANQNIPELLKYDRRSYGGGHPRYGGGRGGYGRGGYGGGGRGGYGNRQRDGGWGNRGRSNY - *S. cerevisiae*
 SILSEAKQIDPKLEEMARYSSGGGRGGYRRGGYGRGGFRGGYGNRNRGFTGSNSAPLARSFW - *S. pombe*
 SVLREANQAINPKLLQLVEDRGSRGRGGMKDRDRDRYSAGKRGGFNTFRDRENYDRGYSLLKRDGAKTQNGVYSAANYTNGSFGSNFVSAGI - Human

SilrEanQ.I.PkLl...r.s.gGr...yggg..gRgg..r.GgyGg.n.g..g.n...g..s.. - cons

QTSFRTGNPTGTYQNGYDSTQQYGSNVPMNHGMNQAYAPATAAAPMIGYPMPTGYSQ - Human

FIG. 3. Amino acid alignment of the human and yeast genes. Amino acids identical in all three sequences are shown in uppercase in the consensus (cons) line; those identical in only two sequences are in lowercase. Sequences matching the helicase motifs of Hodgman (27) and Gorbalenya et al. (21) are marked.

indicate that *DBP2* lies on chromosome XIII (data not shown). The yeast genes contain predicted open reading frames encoding 61-kDa proteins that are homologous throughout their lengths, but the human gene differs from the yeast genes at both ends (Fig. 3). This analysis showed that 296 amino acids (54%) of the human and *S. pombe* sequences, 301 amino acids (55%) of the human and *S. cerevisiae* sequences, and 358 amino acids (63%) of the *S. pombe* and *S. cerevisiae* sequences are identical, compared with ~35% for the other members of the DEAD box family. The deduced carboxy termini of the yeast proteins contain the sequences YGGGHPRYGGGRRGGYGRGGYGGGRGGYGNR (*DBP2*) and GGRGGNYRRGGYGRGGFR

RGGYGNRNRGF (*dbp2*). PIR data base searches using the program PROSRCH (J. F. Collins and A. F. W. Coulson, University of Edinburgh) identified several RNA-binding proteins with related glycine-rich domains, including the A1 heterogeneous nuclear ribonucleoprotein (hnRNP1) (9) and the nucleolar proteins nucleolin (32), fibrillarlin (48), and yeast SSB1 (30). Among the DEAD box proteins, the *Drosophila* gene vasa (25, 33) contains a glycine-rich fivefold heptad repeat, F/SRGGG/QGG. Nucleolin, hnRNP1, and SSB1 contain the RNP1 and RNP2 consensus RNA-binding domains (4), which are not present in p68, *DBP2*, or *dbp2*. Northwestern blotting (i.e., using RNA to probe a Western blot) of nucleolin fragments suggests that the RNP1 and

human	GCCTCCAGAGGGCTAG	GTTAGT--1155bp---CTAAC---63bp--GCAACACCTTACAG	ATGTGGAAGATGTGAAAT
	A S R G L D		V E D V K F
<i>S. cerevisiae</i>	GCCGCCAGAGGTATCG	GTATGT---952bp--TACTAAC--28bp--CTTTATTTTTCAG	ATGTCAAAGGTATCAATT
	A A R G I D		V K G I N Y
<i>S. pombe</i>	GCTTCTAGAGGAATAG	GTAAGT---685bp-----CTAACTTTTATAG	ATGTTAAGGGTATTACAC
	A S R G I D		V K G I T H

FIG. 4. Sequence comparison showing that human p68, *S. pombe dbp2*, and *S. cerevisiae DBP2* contain an intron at the same site in helicase motif V. The predicted human branchpoint sequence was identified by homology to the higher eukaryotic branchpoint consensus (YNRAY).

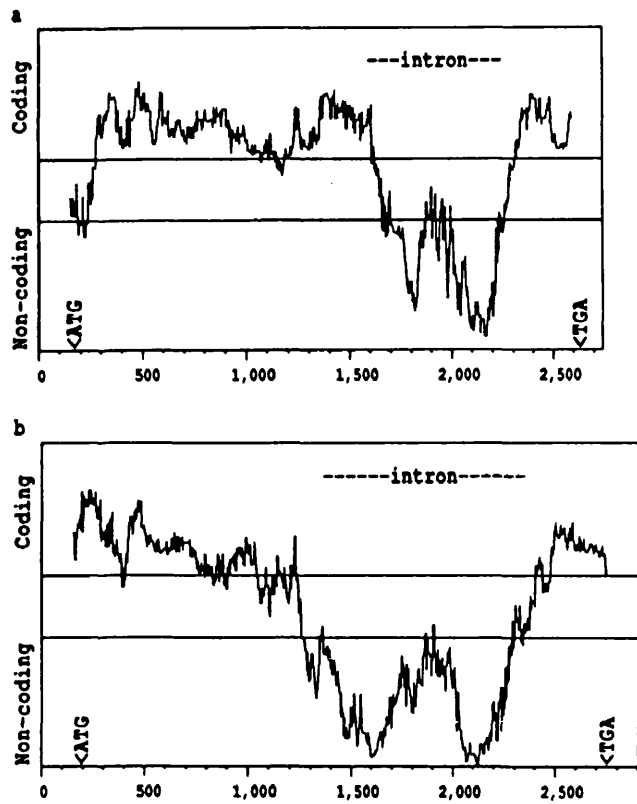


FIG. 5. Third-codon preference plots. (a) *S. pombe* *dbp2*; (b) *S. cerevisiae* *DBP2*. The assignment of coding regions for values of the test code statistic between the two horizontal lines is uncertain. Units on the x axis are in base pairs.

RNP2 domains bind RNA but the glycine-rich domain does not (6). However, the high arginine and aromatic amino acid content of the glycine-rich domains may permit electrostatic and base stacking interactions which modify the specificity of nucleic acid binding (30).

Sequence analysis suggests that the PAb204 epitope (FR DRENYDRG in p68; PLDLRLNFELGVAIDQFL in T antigen [18; unpublished data]) is not present in *DBP2* or *dbp2*. We have confirmed the absence of the epitope from *dbp2* by immunoblotting of a β -galactosidase/*dbp2* fusion protein with PAb204 (data not shown).

p68, *DBP2*, and *dbp2* contain introns conserved from yeasts to humans. The p68, *DBP2*, and *dbp2* genes all contain an intron at exactly the same position in motif V (the ARGID motif; Fig. 3 and 4). The intron is 700 bp long in *S. pombe*, 1 kb long in *S. cerevisiae*, and 1.2 kb long in humans. We identified an alternately spliced form of the human intron in one clone obtained from an Okayama-Berg library. The sequence of this cDNA is identical with that of the genomic DNA up to nucleotide 2428, where an alternate 5' splice site is used (CT/GTAAGG). This introduces a stop codon into the human sequence immediately after the prototypic 5' splice site (nucleotide 1476). Translation of this variant would produce a truncated form of the p68 protein which would almost certainly be nonfunctional because it lacked part of motif V and all of motif VI. Many *S. pombe* genes contain small introns, although we identified none in the data base that contain an intron as large as that in *dbp2*. We have identified no *S. cerevisiae* genes in the GenBank or EMBL data base that contain introns longer than the 551-bp intron in ribosomal protein S16A (24, 39) and only one intron near

the 3' end of a gene (the 52-bp second intron in *MATa1* [37]). The intron in *DBP2* is thus remarkable both because of its size and because of its location near the 3' end of the gene. One reason for conservation of an intron is the existence of an intron-encoded protein. However, the nucleotide sequences of the p68, *DBP2*, and *dbp2* introns are not closely related, their deduced amino acid sequences contain no long open reading frames, and data base searches identify no closely related sequences (data not shown). Third-codon preference plots (16) strongly suggest that they are noncoding (Fig. 5).

***DBP2* and *dbp2* are essential genes.** We disrupted the yeast genes by insertion of *ura4* into *dbp2* and *URA3* into *DBP2*. Tetrad analysis of the disrupted strains revealed that only two out of four spores were viable and that all the surviving spores were *Ura*⁻. The *Ura*⁺ spores germinated on schedule but arrested after about five rounds of cell division. Microscopic examination of the dying cells gave no clue to the function of *DBP2* or *dbp2*. Specifically, the cells did not show a cell cycle arrest phenotype. To confirm that the effect in *S. cerevisiae* was due to disruption of the *DBP2* gene, we transformed the diploid *DBP2/DBP2::URA3* strain with a *CENARS* plasmid encoding the *DBP2* cDNA under the control of the *GAL1* promoter and repeated the tetrad analysis. In tetrads where the plasmid segregated with the disrupted *DBP2* allele, all four spores were viable, confirming that the inviability of *DBP2::URA3* spores was due to the disruption and that *DBP2* is an essential gene.

DISCUSSION

We have demonstrated that p68 undergoes dramatic changes in nuclear location during the cell cycle, lying in the nucleoplasm during interphase and transiently entering the nucleoli during telophase. When the nucleolus re-forms at the end of mitosis, the first structures to appear are pre-nucleolar bodies, which contain mainly dense fibrillar components (54). Although these components are classically associated with rRNA processing (54), it is difficult to envisage p68 being directly involved in rRNA processing because anti-p68 antibodies do not stain the nucleoli of interphase cells. Perhaps p68 is required for an early step in nucleolar assembly.

There are many reasons why microinjected antibody might have no obvious effect: PAb204 may not bind to an essential site (it does not inhibit *in vitro* ATPase [28] or helicase [26] reactions), it may require many rounds of cell division to deplete the cell of p68's target, or the epitope may be masked on the majority of p68 molecules in living cells. The migration of antibody from the cytoplasm to the nucleoli is intriguing. Several models could account for this behavior: nucleolar shuttling (3), which is unlikely given the requirement for cell division; mitotic loading (2), which is unlikely given the failure of p68 to accumulate during interphase in the cytoplasm of transfected cells; or cell cycle-regulated synthesis of p68. The nuclear transport signal in p68 has not been identified experimentally, but the region between amino acids 32 and 45 (KK...KKK) is a likely candidate. Nucleolar transport signals (12, 51) resemble extended but otherwise unremarkable nuclear transport signals. Unless all nuclear proteins transit the nucleolus, for example to allow recycling of a nuclear transporter on the ribosomes, there seems no *a priori* reason why nucleolar targeting could not be achieved by entirely separate nuclear transport and nucleolar binding signals.

p68 is more closely related to *DBP2* and *dbp2* than it is to

any other known DEAD box gene in yeasts (*TIF1* and 2 [35], *MSS116* [49], *SPB4* [46], *PRP5* [11], *PRP28* [52a], *DED1* [42a], *DBP1* [29a], and CA3, CA4, and CA5/6 [7]). The major disagreement between p68 and the yeast *DBP2* and *dbp2* genes is at the carboxy terminus, which is the site of the PAb204 epitope (18). Since PAb204 does not recognize any proteins on Western blots of yeast whole-cell extract, it would have been surprising if the yeast genes had been homologous with p68 in this region. The absence of the PAb204 epitope from the yeast sequences and the presence of an extra 86 amino acids at the carboxy terminus of the human sequence argue for caution in equating p68 with the yeast genes that we cloned, which is why we have chosen to call them *DBP2* and *dbp2*.

Of the 12 *S. cerevisiae* DEAD box genes that have been identified so far (*TIF1* and 2 [35], *MSS116* [49], *SPB4* [46], *PRP5* [11], *PRP28* [52a], *DED1* [42a], *DBP1* [29a], *DBP2* (this report), and CA3, CA4, and CA5/6 [7]), *DBP2* is the only member of the family that contains an intron in motif V. We recently learned of the existence of a *Drosophila* p68 homolog called RM62 (15) which also contains an intron in motif V (29b). The presence of an intron at the same site in the *Drosophila* gene *vasa* indicates that it is not unique to p68.

The evolutionary distance between *S. cerevisiae* and *S. pombe* is as great as the distance between yeasts and humans (45). We are aware of no other example of conservation of an intron in *S. cerevisiae*, *S. pombe* and humans. It implies that an ancestral intron was present at least 1.2 billion years ago, as has been postulated for the triosephosphate isomerase (TPI) introns in *Aspergillus* spp., chickens, and humans (20). Interestingly, the TPI introns may autoregulate TPI expression, since Manquat et al. (8, 10) have shown that premature termination of translation in exon 6 of human TPI mRNA reduces mRNA levels to 20% of normal and have speculated that translation past codon 189 facilitates splicing of TPI pre-mRNA. Few *S. cerevisiae* genes have introns, and in several cases those introns have specific functions: mitochondrial introns encode reverse transcriptases (36) and maturases and endonucleases (13); L32 prevents splicing of its own message by promoting base pairing between the 5' end of the transcript and the 5' splice site (53); and *MATa1* contains two introns that are inefficiently spliced and may play a role in the gene conversion events during mating-type switching (17). *S. pombe* introns over 180 bp are known to abolish splicing in a model system (19), but the existence of a 700-bp intron in *S. pombe* p68, which is an essential gene, indicates that this size limit does not operate in vivo. Although it is perhaps not surprising that one of the *S. pombe* clones was polyadenylated but unspliced, the fact that one of the human cDNA clones was alternately spliced suggests that splicing of the intron is subject to some form of physiological regulation. The unusual nature of the introns hints at the possibility of an intriguing form of autoregulation whereby an RNA helicase may control its own expression by modifying the secondary structure of its mRNA, a model that we are now in the process of testing experimentally in *S. cerevisiae*.

ACKNOWLEDGMENTS

We thank Iain Goldsmith for synthesis of oligonucleotides, Michelle Ginsburg for help with computing, Mike Goman for help with the CHEF blot, Alan Christensen, Evi Strauss, and John Pringle for communication of results prior to publication, and Paul Nurse and Jean Beggs for useful discussions.

D.J.J. is a Beit Memorial Fellow.

REFERENCES

1. Beach, D., and P. Nurse. 1981. High frequency transformation of the fission yeast *Schizosaccharomyces pombe*. *Nature (London)* **290**:140-142.
2. Blow, J. J., and R. A. Laskey. 1988. A role for the nuclear envelope in controlling DNA replication within the cell cycle. *Nature (London)* **332**:546-548.
3. Borer, R. A., C. F. Lehner, H. M. Eppenberger, and E. A. Nigg. 1989. Major nucleolar proteins shuttle between the nucleus and cytoplasm. *Cell* **56**:379-390.
4. Brandziulis, R. J., M. S. Swanson, and G. Dreyfuss. 1989. RNA binding proteins as developmental regulators. *Genes Dev.* **3**:431-437.
5. Bravo, R., S. J. Fey, J. Bellatin, P. M. Larsen, J. Arevalo, and J. E. Celis. 1981. Identification of a nuclear and of a cytoplasmic polypeptide whose relative proportions are sensitive to changes in the rate of cell proliferation. *Exp. Cell Res.* **136**:311-319.
6. Bugler, B., H. Bourbon, B. Lapeyre, M. O. Wallace, J.-H. Chang, F. Amalric, and M. O. J. Olson. 1987. RNA binding fragments from nucleolin contain the ribonucleoprotein consensus sequence. *J. Biol. Chem.* **262**:10922-10925.
7. Chang, T.-H., J. Arenas, and J. Abelson. 1990. Identification of five putative yeast RNA helicase genes. *Proc. Natl. Acad. Sci. USA* **87**:1571-1575.
8. Cheng, J., M. Fogel-Petrovic, and L. E. Manquat. 1990. Translation to near the distal end of the penultimate exon is required for normal levels of spliced triosephosphate isomerase mRNA. *Mol. Cell. Biol.* **10**:5215-5225.
9. Cobiainchi, F., D. N. Sengupta, B. Z. Zmudzka, and S. H. Wilson. 1986. Structure of the rodent helix-destabilizing protein revealed by cDNA cloning. *J. Biol. Chem.* **261**:3536-3543.
10. Daar, I. O., and L. E. Manquat. 1988. Premature translation termination mediates triosephosphate isomerase mRNA degradation. *Mol. Cell. Biol.* **8**:802-813.
11. Dalbadie-McFarland, G., and J. Abelson. 1990. PRP5: a helicase-like protein required for mRNA splicing in yeast. *Proc. Natl. Acad. Sci. USA* **87**:4236-4240.
12. Dang, C. V., and W. M. F. Lee. 1989. Nuclear and nucleolar targeting sequences of c-erb-A, c-myc, N-myc, p53, HSP70 and HIV tat proteins. *J. Biol. Chem.* **264**:18019-18023.
13. Delahodde, A., V. Goguel, A. M. Becam, F. Creusot, J. Perea, J. Banroques, and C. Jacq. 1989. Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. *Cell* **56**:431-441.
14. Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**:387-395.
15. Dorer, D. R., A. C. Christensen, and D. H. Johnson. 1990. A novel RNA helicase gene tightly linked to the triplo-lethal locus of *Drosophila*. *Nucleic Acids Res.* **18**:5489-5494.
16. Fickett, J. W. 1982. Recognition of protein coding regions in DNA sequences. *Nucleic Acids Res.* **10**:5303-5318.
17. Ford, G. R. 1987. Pseudogenes in yeast? *Cell* **49**:5-6.
18. Ford, M. J., I. A. Anton, and D. P. Lane. 1988. Nuclear protein with sequence homology to translation initiation factor e1F4-A. *Nature (London)* **332**:736-738.
19. Gaterman, K. B., A. Hoffman, G. H. Rosenberg, and N. F. Kaufer. 1989. Introduction of functional artificial introns into the naturally intronless *ura4* gene of *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* **9**:1526-1535.
20. Gilbert, W., M. Marchionni, and G. McKnight. 1986. On the antiquity of introns. *Cell* **46**:151-153.
21. Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov. 1989. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* **17**:4713-4730.
22. Gorman, C., D. R. Gies, and G. McCray. 1990. Transient production of proteins using an adenovirus transformed cell line. *DNA Protein Eng. Tech.* **2**:3-10.
23. Harlow, E. E., and D. P. Lane. 1988. *Antibodies: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. Hawkins, J. D. 1988. A survey on intron and exon lengths.

25. Hay, B., L. Y. Jan, and Y. N. Jan. 1988. A protein component of *Drosophila* polar granules is encoded by vasa and has extensive sequence similarity to ATP-dependent helicases. *Cell* 55:577-587.
26. Hirling, H., M. Scheffner, T. Restle, and H. Stahl. 1989. RNA helicase activity associated with the human p68 protein. *Nature* (London) 339:562-564.
27. Hodgman, T. C. 1988. A new superfamily of replicative proteins. *Nature* (London) 333:22-23.
28. Iggo, R. D., and D. P. Lane. 1989. Nuclear protein p68 is an RNA-dependent ATPase. *EMBO J.* 8:1827-1831.
29. Ito, H., Y. Fukuda, K. Murata, and A. Kimura. 1983. Transformation of intact yeast cells with alkali cations. *J. Bacteriol.* 153:163-168.
- 29a. Jamieson, D. Unpublished data.
- 29b. Johnson, D. H., and A. C. Christensen. Personal communication.
30. Jong, A., M. W. Clark, M. Gilbert, A. Oehm, and J. L. Campbell. 1987. *Saccharomyces cerevisiae* SSB1 protein and its relationship to nucleolar RNA binding proteins. *Mol. Cell. Biol.* 7:2947-2955.
31. Lane, D. P., and W. K. Hoeffler. 1980. SV40 large T shares an antigenic determinant with a cellular protein of molecular weight 68,000. *Nature* (London) 288:167-170.
32. Lapeyre, B., H. Bourbon, and F. Amalric. 1987. Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence. *Proc. Natl. Acad. Sci. USA* 84:1472-1476.
33. Lasko, P. F., and M. Ashburner. 1988. The product of the *Drosophila* gene vasa is very similar to eukaryotic initiation factor-4A. *Nature* (London) 335:611-617.
34. Linder, P., P. F. Lasko, M. Ashburner, P. Leroy, P. J. Nielsen, K. Nishi, J. Schnier, and P. P. Slonimski. 1989. Birth of the D-E-A-D box. *Nature* (London) 337:121-122.
35. Linder, P., and P. P. Slonimski. 1989. An essential yeast protein, encoded by duplicated genes TIF1 and TIF2 and homologous to the mammalian translation initiation factor eIF-4A, can suppress a mitochondrial missense mutation. *Proc. Natl. Acad. Sci. USA* 86:2286-2290.
36. Michel, F., and B. F. Lang. 1985. Mitochondrial type II introns encode proteins related to the reverse transcriptases of retroviruses. *Nature* (London) 316:641-643.
37. Miller, A. M. 1984. The yeast MATa1 gene contains two introns. *EMBO J.* 3:1061-1065.
38. Miller, M. A., D. Korn, and T. S.-F. Wang. 1988. The evolutionary conservation of DNA polymerase α . *Nucleic Acids Res.* 16:7961-7973.
39. Molenaar, C. M. T., L. P. Woudt, A. E. M. Jansen, W. H. Mager, R. J. Planta, D. M. Donovan, and N. J. Pearson. 1984. Structure and organisation of two linked ribosomal protein genes in yeast. *Nucleic Acids Res.* 12:7345-7358.
40. Moreno, S., A. Klar, and P. Nurse. 1990. An introduction to molecular genetic analysis of the fission yeast. *Methods Enzymol.* 194:793-823.
41. Ochs, R., M. Lischwe, P. O'Leary, and H. Busch. 1983. Localisation of nucleolar phosphoproteins B23 and C23 during mitosis. *Exp. Cell Res.* 146:139-149.
42. Ochs, R. L., M. A. Lischwe, E. Shen, R. E. Carroll, and H. Busch. 1985. Nucleologenesis: composition and fate of pre-nucleolar bodies. *Chromosoma* 92:330-336.
- 42a. Pringle, J. Personal communication.
43. Rothstein, R. J. 1983. One step gene disruption in yeast. *Methods Enzymol.* 101:202-211.
44. Rozen, F., I. Edery, K. Meerovitch, T. E. Dever, W. C. Merrick, and N. Sonenberg. 1990. Bidirectional RNA helicase activity of eukaryotic translation initiation factors 4A and 4F. *Mol. Cell. Biol.* 10:1134-1144.
45. Russell, P., and P. Nurse. 1986. *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*: a look at yeasts divided. *Cell* 45:781-782.
46. Sachs, A. B., and R. W. Davis. 1990. Translation initiation and ribosomal biogenesis: involvement of a putative rRNA helicase and RPL46. *Science* 247:1077-1079.
47. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
48. Schimmang, T., D. Tollervey, H. Kern, R. Frank, and E. C. Hurt. 1989. A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability. *EMBO J.* 8:4015-4024.
49. Seraphin, B., M. Simon, A. Boulet, and G. Faye. 1989. Mitochondrial splicing requires a protein from a novel helicase family. *Nature* (London) 337:84-87.
50. Sherman, F., G. R. Fink, and J. B. Hicks. 1982. *Methods in yeast genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
51. Siomi, H., H. Shida, S. H. Nam, T. Nosaka, M. Maki, and M. Hanataka. 1988. Sequence requirements for nucleolar localisation of human T cell leukaemia virus type I pX protein, which regulates viral RNA processing. *Cell* 55:197-209.
52. Spector, D. L., R. L. Ochs, and H. Busch. 1984. Silver staining, immunofluorescence and immunoelectron microscopic localisation of nucleolar phosphoproteins B23 and C23. *Chromosoma* 90:139-148.
- 52a. Strauss, E. Personal communication.
53. Warner, J. R. 1989. Synthesis of ribosomes in *Saccharomyces cerevisiae*. *Microbiol. Rev.* 53:256-271.
54. Warner, J. R. 1990. The nucleolus and ribosome formation. *Curr. Opin. Cell Biol.* 2:521-527.
55. Waseem, N. H., and D. P. Lane. 1990. Monoclonal antibody analysis of the proliferating cell nuclear antigen (PCNA). Structural conservation and the detection of a nucleolar form. *J. Cell Sci.* 96:121-129.