

**A study into the conserved nature of yeast and  
mammalian transcription factors**

Robert Hugh Jones  
Department of Cell Biology  
University College London

A thesis submitted for the degree of  
Doctor of Philosophy  
April 1990

Gene Regulation Laboratory  
Imperial Cancer Research Fund  
PO Box no. 123  
Lincoln's Inn Fields  
London WC2A 3PX.

ProQuest Number: 10609306

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 10609306

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

**THIS THESIS IS DEDICATED TO MY PARENTS.**

## **ACKNOWLEDGEMENTS**

I would like to thank all friends and colleagues who have advised and helped me throughout my thesis. In particular special thanks should go Mike Fried, Richard Treisman, Steve Goodbourne and the members of the gene regulation laboratory; Helen Hurst, Gunvanti Patel, Tom Dooley, Julie Jordan, Anders Follin, John Schneider, Gos Micklem, Alex Eccleston, Doris Bembrook, Shuske Kuge and Kate Flint, and also Claire Middlemiss for her secretarial help.

I acknowledge the financial support of the ICRF and also the services of the photography department and the efforts of Iain Goldsmith in supplying me with oligonucleotides.

Most of all however I would like to acknowledge the support and guidance of my supervisor Nic Jones who provided a constant source of ideas and lager.

# **A study into the conserved nature of yeast and mammalian transcription factors.**

## **ABSTRACT.**

The study of transcriptional regulation in mammalian cells has led to the identification of a large number of DNA binding transcription factors. Two such activities have been called ATF and AP1. The DNA sequence to which ATF binds to has been shown to mediate transcriptional activation by E1A and cAMP. On the other hand the AP1 binding site has been shown to mediate transcriptional activation by phorbol esters such as TPA. The consensus sequences of these two functional elements differ by only a single nucleotide but this difference is critical in determining specific transcription factor interaction.

Previous studies have also defined sequence specific DNA binding proteins which can mediate transcriptional activation in S.cerevisiae. In this thesis I show that there are transcription factors in two yeast systems, that of S.cerevisiae and S.pombe, which have identical DNA binding specificities to the mammalian ATF and AP1 proteins. I have designated these proteins yATF and yAP1. In addition I show that viral promoters which contain binding sites for these transcription factors are expressed in S.pombe utilising transcriptional start sites also used in mammalian cells.

By placing ATF and AP1 binding sites upstream of basal promoters I have demonstrated that S.pombe yAP1 and S.cerevisiae yATF are most likely transcriptional activators. I have also shown that both the afore mentioned proteins are sensitive to phosphatase treatment indicating they are themselves phosphoproteins, a property which they would share with their mammalian counterparts.

I have purified S.cerevisiae yAP1 and identified it as a major species of 85kd. In addition I have obtained a clone for a fission yeast protein which preferentially recognises an ATF site to an AP1 site by using a  $\lambda$ gt11 expression library binding site screen.

This work has provided the first demonstration of a fission yeast transcription factor. In addition it has shown a remarkable conservation in the DNA binding specificities of yeast and mammalian transcription factors. Together with the parallel developments which have taken place in this field this indicates that certain fundamental properties of the transcriptional machinery have been conserved throughout eukaryotes

<b>CONTENTS</b>	<b>Page</b>
<b>ABSTRACT</b>	1
<b>CONTENTS</b>	2
<b>ABBREVIATIONS</b>	11
<b>CHAPTER 1; INTRODUCTION;</b>	14
<b>1. YEAST AS AN EXPERIMENTAL SYSTEM</b>	15
<b>2. THE EUKARYOTIC PROMOTER</b>	16
2.1. Summary.	16
2.2. Organisation of regulatory sequences in Eukaryotes	16
2.3. Sequences important for transcriptional initiation; the TATA box	17
2.4 Other considerations in transcriptional initiation.	18
<b>3. INITIATION FACTORS</b>	19
3.1. RNA polymerases	19
3.1.1. RNA polymerase II	19
3.2. Other factors required for transcriptional initiation	20
<b>4. STUDIES ON REGULATORY RESPONSES</b>	21
4.1. Regulatory Studies in yeast	22
4.1.1. Amino acid Starvation response	22
4.1.1.1. The isolation of <u>S.cerevisiae GCN4</u>	22
4.1.1.2. GCN4 binds as a dimer;	22
4.1.1.3. Regulation of GCN4.	23
4.1.2. Galactose inducible expression	24
4.1.3. HAP1 and induction by heme	24
4.1.4. Heat shock induction.	24
4.1.5. Mating type determination.	24
4.2. Studies in mammalian systems	25

4.2.1.	Transcriptional activation in response to TPA	25
4.2.1.1	Characterisation of TRE	
4.2.1.2	Cloning and characterisation of <u>v-jun</u>	26
4.2.2.	Regulation by cAMP and E1A	27
4.2.2.1	cAMP regulation of the somatostatin gene	27
4.2.2.2.	Adenovirus E1A activation of transcription	28
4.2.2.3.	The ATF site and CRE are very similar	29
5.	GENERAL PRINCIPLES OF TRANSCRIPTION FACTORS.	29
5.1.	Seperation of DNA binding and Activation domains	30
5.2.	DNA-binding motifs	30
5.2.1.	The helix turn helix	31
5.2.2.	Zinc finger	31
5.2.3.	Leucine zipper	32
5.2.4.	Helix-Loop-Helix	33
5.3.	Activation Domains	34
5.4.	Interaction of an activator protein with the TATA factor	36
6.	OTHER MECHANISIMS OF TRANSCRIPTIONAL REGULATION	37
6.1.	Chromatin Structure	37
6.2.	DNA methylation	38
7.	SUMMARY; REASONS FOR CARRYING OUT MY PROJECT	38
CHAPTER 2;	MATERIALS AND METHODS	40
1.	DNA/RNA PREPARATION	41

1.a.	Plasmids	41
1.a.i.	Small scale	41
1.a.i.i.	Large scale	41
1.b.	Buffers	41
1.c.	Phenol and Phenol /chloroform extractions	41
1.d.	Isolation of <u>S.pombe</u> RNA	42
1.e	Isolation of <u>S.pombe</u> DNA	42
2.	DNA MANIPULATIONS	42
2.a.	General	42
2.b.	Restriction Enzyme Digestion	43
2.c.	Phosphatase treatment	43
2.d.	Fragment purification	43
2.e.	Ligation	43
2.f.	Nested Deletions	43
3.	DNA SEQUENCING	44
3.a.	Denaturing the template	44
3.b.	Annealing	45
3.c.	Labelling	45
3.d.	Termination	45
3.e.	Buffers	45
4.	CONSTRUCTS	45
4.a.	E2-CAT constructs	45
4.b.	Construction of plasmids containing oligonucleotide AP-1 sites	46
4.c.	SV40-CAT constructs	46
4.d.	E1A expression plasmids	46
4.e.	Budding yeast plasmids	47

<b>5.</b>	<b>GEL ELECTROPHORESIS</b>	<b>47</b>
<b>5.a.</b>	Loading buffers	47
<b>5.b.</b>	Agarose gels	47
<b>5.c.</b>	Buffer Gradient Gels	47
<b>5.d.</b>	Footprinting	48
<b>5.e.</b>	Gel shifts	48
<b>5.f.</b>	Protein gel electrophoresis	48
<b>6.</b>	<b>BLOTTING</b>	<b>49</b>
<b>6.a.</b>	Western Blot Analysis	49
<b>6.b.</b>	Colloidal gold staining	50
<b>6.c.</b>	Southwestern analysis	50
<b>7.</b>	<b>DNA LABELLING</b>	<b>50</b>
<b>7.1.</b>	Filling in a 5' overhang	50
<b>7.2.</b>	Phosphorylation of 5' ends	50
<b>7.3.</b>	Random priming	51
<b>7.4.</b>	Preparation of G-50 sephadex column	51
<b>8.</b>	<b>STRAINS</b>	<b>52</b>
<b>8.a.</b>	Bacterial	52
<b>8.b.</b>	Budding yeast	52
<b>8.c.</b>	Fission yeast	52
<b>9.</b>	<b>MEDIA</b>	<b>52</b>
<b>9.a.</b>	Bacterial	52
<b>9.b.</b>	Budding yeast	52
<b>9.c.</b>	Fission yeast	53
<b>10.</b>	<b>TRANSFORMATIONS</b>	<b>54</b>
<b>10.a.</b>	Budding yeast	54
<b>10.b.</b>	fission yeast	54



10.c.	Bacterial	55
11.	FISSION YEAST EXTRACTS	56
12.	BUDDING YEAST EXTRACTS	56
12.a.	General	56
12.b.	Cell culture and breakage	56
12.c.	Protein purification	57
12.d.	Preparation of collumns	58
12.d.i.	Heparin Agarose	58
12.d.ii.	DNA sepharose	58
12.d.iii.	Oligonucleotide affinity collumns	58
12.e.	Buffers	59
12.f.	Small scale budding yeast preparations	60
13.	$\lambda$ gt11 EXPRESSION SCREEN	61
13.a.	Plating	61
13.b.	Probing	61
13.c.	Selection and amplification of positive clones	61
13.d.	Preparation of high titre virus stocks	62
13.e.	Isolation of recombinant phage lysogens	62
13.f.	Buffers	62
14.	$\beta$ -GALACTOSIDASE ASSAYS	63
15.	PRIMER EXTENSIONS	63
16.	CAT ASSAYS	63
17.	DNase 1 FOOTPRINTING	64
17.a.	DNase 1 reactions	64
17.b.	A+G marker lane	64
18.	GEL SHIFT ASSAYS	65

18.a.	General	65
18.b.	Specific competition	65
18.c.	Phosphatase reactions	65
18.d.	Kinase Reactions	65
19.	OLIGONUCLEOTIDES	66
19.a.	Preparation of single stranded oligonucleotides	66
19.b.	Oligonucleotides used	66
<b>CHAPTER 3;</b>	<b>RESULTS IDENTIFICATION AND CHARACTERISATION OF AN AP1 LIKE ACTIVITY IN YEAST.</b>	<b>69</b>
1.	SUMMARY	70
2.	THE SV40 PROMOTER FUNCTIONS EFFICIENTLY IN FISSION YEAST	71
3.	FISSION YEAST CONTAINS AN AP1 LIKE BINDING ACTIVITY	72
4.	BUDDING YEAST CONTAINS TWO AP1 LIKE FACTORS	72
5.	THE AP1 BINDING SITE CAN SERVE AS A UAS IN BOTH YEAST SYSTEMS	74
6.	THE AP1-LIKE FACTOR OF S.POMBE IS SENSITIVE TO PHOSPHATASE TREATMENT	74
<b>CHAPTER 4;</b>	<b>RESULTS; IDENTIFICATION AND CHARACTERISATION OF A YEAST ATF-LIKE BINDING ACTIVITY</b>	<b>76</b>
1.	SUMMARY	77

2.	EXPRESSION OF THE ADENOVIRUS E2A PROMOTER IN FISSION YEAST	77
3.	FISSION YEAST FACTORS INTERACTING WITH THE E2A PROMOTER	78
4.	YEAST ATF CAN INTERACT WITH THE ADENOVIRUS E4 PROMOTER	79
5.	<u>S.CEREVISIAE</u> ALSO CONTAIN AN ATF LIKE BINDING ACTIVITY	79
6.	YEAST ATF AND YEAST AP1 ARE DISTINGUISHABLE	80
7.	THE YATF BINDING SITE CAN ACT AS AN EFFICIENT UAS	81
<b>CHAPTER 5;</b>	<b>RESULTS; PURIFICATION OF BUDDING YEAST AP1 AND ATF.</b>	<b>82</b>
1.	INTRODUCTION AND SUMMARY	83
2.	PURIFICATION OF YAP1	84
2.a.	Heparin Agarose Chromotography	84
2.b.	DNA-sepharose	85
2.c.	Oligonucleotide affinity chromatography	86
2.d.	Colloidal gel staining	86
2.e.	Southwestern analysis	87
2.f.	Peptide sequencing	87
3.	PURIFICATION OF $\gamma$ ATF	89
3.a.	Summary	89
3.b.	Crude extracts and Heparin Agarose purification	89
3.c.	DNA sepharose purification	90
3.d.	Affinity chromatography	90

3.e.	Phosphorylation of $\gamma$ ATF	91
<b>CHAPTER 6;</b>	<b>RESULTS; USING <math>\lambda</math> GT11 SCREENS TO CLONE YEAST DNA BINDING PROTEINS.</b>	<b>93</b>
1.	INTRODUCTION AND SUMMARY	94
2.	PICKING AND ENRICHING POSITIVE SIGNALS	94
3.	CLONES A AND B PREFERENTIALLY BIND AN ATF SITE	94
4.	GEL SHIFT ANALYSIS ON LYSOGEN DERIVED EXTRACTS	96
5.	SEQUENCE ANALYSIS OF CLONE B	97
<b>CHAPTER 7;</b>	<b>DISCUSSION</b>	<b>98</b>
1.	EXPRESSION OF VIRAL PROMOTERS IN FISSION YEAST	99
1.1.	The Adenovirus E2A promoter	99
1.2.	The SV40 promoter/enhancer	100
2.	CONSERVATION IN DNA BINDING SPECIFICITY BETWEEN YEAST AND MAMMALIAN PROTEINS	102
2.1.	Multiple proteins binding to the TRE	102
2.2.	Many factors can bind CRE/ATF sites	104
2.3.	Is Clone B a member of the CREB/ATF transcription factor family	105
2.3.1.	Homology in the DNA binding domain	105
2.3.2.	Leucine Zipper	106
2.3.3.	Perspectives	107

2.4.	Yeast homologues to another family of transcription factors	108
2.4.1.	CCAAT box proteins	108
2.4.2.	SRF and MCM1	109
2.4.3.	Heat Shock Factor	109
2.4.4.	TATA factor	110
3.	<b>CONSERVATION IN ACTIVATION AND REGULATORY PROPERTIES</b>	111
3.1.	AP1 and ATF binding sites can act as UASs in yeast	111
3.2.	Regulatory Considerations	111
3.2.1.	AP1	112
3.2.2.	ATF/CREB	113
3.3.	Why have multiple proteins recognising the same site	114
3.3.1.	Molecular Regulatory Interplay	115
3.3.2.	Developmental Considerations	115
3.4.	Conservation of activation in general	116
4.	<b>SUMMARY AND PERSPECTIVES</b>	117
<b>CHAPTER 8;</b>	<b>REFERENCES</b>	120

**ABBREVIATIONS:**

AP1	Activating protein 1
AP2	Activating protein 2
APS	Ammonium persulphate
ARS	autonomously replicating sequence
ATF	Activating transcription factor
ATP	Adenosine triphosphate
$\alpha$	alpha
bp	base pair
BSA	bovine serum albumen
C	degrees centigrade
cAMP	3', 5' cyclic adenosine monophosphate
CAT	chloramphenicol acetyl transferase
CEN	centromere
Ci	curie
Cl	chloride
cps	counts per second
dATP	3' deoxyadenosine 5' triphosphate
dCTP	3' deoxycytosine 5' triphosphate
dGTP	3' deoxyguanosine 5' triphosphate
dTTP	3' deoxythymidine 5' triphosphate
dNTP	3' deoxynucleotide 5' triphosphate
ddATP	2', 3'-dideoxyadenosine 5' triphosphate
ddCTP	2', 3'-dideoxycytosine 5' triphosphate
ddGTP	2', 3'-dideoxyguanosine 5' triphosphate
ddTTP	2', 3'-dideoxythymidine 5' triphosphate
ddNTP	2', 3'-dideoxynucleoside triphosphate
DNA	deoxyribonucleic acid
DNAse	deoxyribonucleic acid endonuclease
dNTP	2' deoxyribonucleoside triphosphate
DTT	dithiothreitol
EDTA	Ethylenediamine tetra acetic acid
Fig.	figure
g	gramme
gDW	glass distilled water.
GCN.	General control of nitrogen
$\gamma$	gamma

HCl	hydrochloric acid
HEPES	N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulphonic acid]
kb	kilobase pair
kg	kilogramme
kdal	kilodaltons
l	litre
$\lambda$	bacteriophage lamda
M	Molar
mM	milli Molar
mA	milli ampere
ml	milli litre
mg	milli gramme
$\mu$ g	micro gramme
$\mu$ l	micro litre
MOPS	3-[N-Morpholino] propanesulphonic acid
mRNA	messenger RNA
MW	molecular weight
Na	Sodium
nm	nanometres
nt	nucleotide
NTP	ribonucleoside triphosphate
OAc	acetate
OD	optical density
oligo	synthetic DNA oligonucleotide
ONPG	O-nitrophenol-b-D-galactopyranoside
ORF	Open reading frame
PMSF	Phenylmethyl sulphonyl fluoride
PNK	Polynucleotide Kinase
pol	polymerase
rDNA	ribosomal DNA
RNA	ribonucleic acid
rpm	revolutions per minute
SDS	Sodium dodecylsulphate
T,C,G,A	Thymine, Cytosine, Guanine, Adenine
TE	10mM tris/Cl, 1mM Na <sub>2</sub> EDTA
TEMED	N N N' N'-tetramethyl-1,2-diaminothane
tris	Tris (hydroxymethyl) aminomethane
tRNA	transfer RNA

u	unit of enzyme activity.
UAS	upstream activating sequence
V	Volt
W	Watt
μ	micro
Δ	deletion

Single letter amino acid code;

A	Alanine
C	Cysteine
D	Aspartic acid
E	Glutamic acid
F	Phenylalanine
G	Glycine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
N	Asparagine
P	Proline
Q	Glutamine
R	Arginine
S	Serine
T	Threonine
V	Valine
Y	Tyrosine



## **CHAPTER 1; INTRODUCTION**

The control of cellular growth and differentiation is of fundamental importance to all eukaryotic organisms. Consequently it is hardly surprising that there are many different regulatory mechanisms which can operate in determining a cell's potential in a particular environment. For the most part it is the particular genes which a cell expresses which is the predominant determinant of behaviour. However the regulation of gene expression itself is influenced at many different levels, the first and probably most important of which is the initiation of RNA transcription. The work I describe in this thesis is directed at obtaining additional information about this process and for most of the time I have used yeast as my experimental system. In the introduction I shall review most the relevant background to my work and also outline why I initiated the experimental approach which my results chapters describe.

## **1. Yeast as an experimental system.**

The work in this thesis is predominantly focussed on features of the transcriptional biology in two yeast systems; *Saccharomyces cerevisiae* (budding yeast) and *Schizosaccharomyces pombe* (fission yeast) . It therefore seems appropriate that I should outline some features of their general biology which makes them an appropriate experimental system.

The major reason yeast has been such a popular system to study eukaryotic cell biology is because it is so amenable to approaches in both classical and molecular genetics (See Moreno et al 1989 and Struhl 1986 for reviews). Historically it has been their tractability for classical genetics that has been more widely exploited. Both yeasts can be grown as a haploid of one of two mating types. A haploid of one mating type can be fused to that of another thus enabling growth in either a haploid or diploid state. In this way mutations, which are easily isolated in haploids, can be tested for their dominance and placed into a particular complementation group simply by fusing the mutant in question with a haploid of a different strain.

The advent of recombinant DNA technology has facilitated the extension of this analysis to the molecular level. The gene responsible for a particular mutation can be isolated by complementing mutant strains with an expression library. As well as identifying the mutated product this method frequently leads to the isolation of other gene products which can act as suppressors of the mutant phenotype when overexpressed in the cell. In addition it is possible to replace the endogenous copy of a gene with a copy of it which has been artificially manipulated in vitro. This gene replacement technology is a powerful tool in studying how a selected alteration of a gene in its endogenous locus affects the phenotype of the cell.

This technology was originally developed in S.cerevisiae but much of it can now be applied to S.pombe.

In addition to their genetic tractability budding yeast in particular also offers a number of advantages for biochemical studies. This centres around the ability to grow large volumes of culture and therefore obtain large amounts of yeast material fairly easily and cheaply. It thus provides a convenient system to isolate reasonable quantities of low abundance proteins, such as transcription factors, using protein purification procedures. One example where this advantage has been put to good practical use involves the purification and subsequent cloning of the gene encoding yeast heat shock transcription factor (Sorger and Pelham 1988, Wiederrecht and Parker 1988).

## **2. The eukaryotic promoter.**

### **2.1. Summary;**

The control of gene expression at the level of RNA synthesis is determined by cis-acting DNA regulatory sequences. Such regulatory sequences are normally located upstream (5') of the transcriptional start site although occasionally they can be found downstream (3'). These DNA elements mediate their effect by binding protein factors (transcription factors) in a sequence specific manner. A number of such protein binding elements can occur in any given promoter and it is the specific combination of transcription factors that a particular promoter binds which determines its own regulatory properties. The variable presence of such regulatory elements is the basis for differential gene expression.

Although the exact means of regulation can be specific to individual genes there are a number of underlying principles which appear to be common to the regulation of eukaryotic gene expression as a whole. In the first part of the Introduction I shall discuss some of these underlying principles. Later I shall review studies addressing a small number of specific regulatory responses which serve to illustrate approaches which have been taken to understand the field as a whole and in themselves are also of specific relevance to the studies in my thesis.

### **2.2. Organisation of regulatory sequences in Eukaryotes;**

Traditionally the mammalian promoter has been separated<sup>a</sup> into three types of element: The TATA box, critical for transcriptional initiation; proximal regulatory elements, promoter

sequences which only function when in a single orientation and when placed close to and 5' of the TATA; and enhancer elements, regulatory sequences which can function either 5' or 3' of the initiation site and in an essentially position and orientation independent manner. For now I will let this type of definition stand but, as will become apparent in the discussion, the manner in which these functional elements can be distinguished has recently become somewhat blurred. Regulatory elements can take the form of either activators or repressors of transcription with the exact property of any particular element dependent on both its actual sequence and the context within a promoter that it is found (See Goodbourne 1989 for review and Chiu et al 1989 for a specific example).

A similar type of characterisation has also taken place in budding yeast. Here activating elements are placed into a general category called upstream activating sequences (UASs) and negatively acting elements are called operators (OP) (see Struhl 1989 for review). Both types are normally relatively short, 10-30bp, and the specific sequence within the element determines its regulatory properties. UASs are similar to mammalian enhancer sequences in that they can function in either orientation and at variable distances from the transcriptional start site (100-1500 bps). They cannot however activate transcription when placed downstream of the start site (Struhl 1989). Operator sequences like UASs can function bidirectionally and at variable distances from the TATA. In general they are maximally effective at repressing the effect of a UAS when placed between the TATA box and the UAS (see Brent 1985 for review). However in the case of the mating type silencer transcription can be efficiently repressed when placed at distances of up to 2kb either upstream or downstream of the transcriptional start site (Herskowitz 1989).

In addition to studies in the budding yeast system there has been a limited investigation into promoters in fission yeast *S.pombe*. This work has demonstrated that TATA box sequences exist 25-30 bp upstream of transcriptional start sites indicating a similar type of transcriptional start mechanism to mammalian cells (Russell 1983). However at the point I started my thesis no fission yeast transcription factors had been identified.

### **2.3. Sequences important for transcriptional initiation; the TATA Box;**

This element consists of an AT rich region for which a TATAA consensus has been derived. It has therefore been called the TATA box and is located 30bp upstream of the transcriptional start site. A large number of experiments have shown that the TATA box is very important in determining the site of transcriptional initiation; modest changes in the position of the TATA box almost always causes similar changes in the site of transcriptional initiation. In addition the introduction of point mutations which disrupt the TATA sequence

often causes spurious initiation of transcription (for review see Breathnach and Chambon 1981).

The situation is very similar in budding yeast where most promoters also contain TATA boxes (Struhl 1987a). However unlike higher eukaryotes the distance between the TATA element and the actual start site of transcription in any given promoter can vary between 40-120 bp. Specific mutagenesis experiments on a number of yeast promoters has shown that the TATA element can be moved around within these distances without changing the transcriptional start site. This has led to the identification of the initiator element which occurs very close to the site of initiation itself (Hahn et al 1985, Nagawa and Fink, 1985, Chen, W. and K.Struhl 1985). As long as the initiator element occurs between 40-120 bp of the TATA element it determines a preferred start site but beyond these distances it has no power and transcripts start heterogenously about 60 bp downstream of the TATA box. There is no good sequence consensus which has been worked out for the initiator element as yet and it is still unclear exactly how its effect is mediated. (Hahn et al 1985, Nagawa and Fink, 1985, Chen, W. and K.Struhl 1985). Interestingly an 'initiator' element has also recently been characterised in a mammalian promoter (Smale and Baltimore 1989). Like the yeast element it mediates transcriptional start actually in the position it is placed rather than a fixed distance downstream, but unlike yeast, the mammalian initiator can apparently function in the absence of a TATA box.

#### **2.4. Other considerations in transcriptional initiation;**

The introduction above has implied that the TATA box represents a consensus element which is highly conserved among different promoters in many different eukaryotes. It has also implied that for higher eukaryotes the TATA box is the primary if not only functional element required to determine the start site of transcription. Although this is essentially true it is also an oversimplification. Therefore it should be noted that;

1. Approximately 20% of eukaryotic promoters lack sequences resembling the classic TATAAA motif (reviewed in Harbury and Struhl 1989). These promoters appear to recruit other fairly general transcription factors to help pol II mediate transcriptional initiation.
2. Recent analysis in yeast and mammalian systems has indicated that functionally distinct TATA box binding factors may exist which interact and respond differently to different regulatory factors and responses (Harbury and Struhl 1989, Simon et al 1988).

3. In certain TATA box promoters transcriptional start also depends on sequences around the actual start site itself (Breathnach and Chambon 1981, Jones, K.A. et al 1988).

4. In certain cases the TATA box can be mutated without significantly affecting the position of transcriptional start (Jones, K.A. et al 1988).

5. Recently Smale and Baltimore (1989) have identified another control element, the initiator, in a lymphocyte specific promoter which can determine the transcriptional start site. This element can act independently of the TATA and also respond to regulatory factors.

### **3. Initiation factors.**

#### **3.1. RNA polymerases;**

RNA Polymerase is the enzymatic activity that actually produces RNA from DNA. Eukaryotic cells contain three types of RNA polymerase, RNA Polymerase I, II and III. These were separated chromatographically from mammalian cell extracts and have been shown differentially sensitive to  $\alpha$ -amanitin (Roeder and Rutter 1969, reviewed by Chambon ). In the normal cellular environment polymerase I makes large ribosomal RNAs and polymerase III makes a variety of very small stable RNAs including the small 5S ribosomal RNA and tRNAs. RNA Polymerase II transcribes those genes whose RNA becomes translated into protein and is therefore the polymerase of interest in most gene regulation studies.

##### **3.1.1. RNA polymerase II;**

RNA polymerase II is composed of 8-10 subunits ranging in molecular mass from 240 to 10 kD. The gene for the largest subunit has been isolated from a number of eukaryotic systems including budding yeast, Drosophila and humans. Sequence comparison has revealed a 65% homology between the yeast and Drosophila proteins and a 76% homology between the Drosophila and human proteins (for review see Saltzman and Weinman, 1989). An interesting feature of the large subunit Pol II clone is a repeated seven amino acid sequence located at the carboxyl-end. This occurs 52 times in the mouse and 26 times in S.cerevisiae. Removal of this domain is a lethal effect in yeast but the cell can be rescued by

adding back the equivalent structure from a mouse clone (Allison et al 1988). This indicates that the structure is both important and its function has been conserved between yeast and mammalian cells. It has been suggested that the repeat may make direct contacts with transcriptional activator proteins. If this is the case it would indicate that the actual mechanism of transcriptional activation may be highly conserved among eukaryotes (see discussion).

### **3.2. Other factors required for transcriptional initiation; identification of TFII-A, -B, -D, and -E;**

RNA Polymerase is therefore the central enzymatic activity responsible for the synthesis of RNA. However experimental analysis quickly revealed that purified Pol II could not direct synthesis from the correct start site of cellular promoters in vitro. The inability to observe accurately initiated transcription in vitro was eventually overcome by simply including crude S-100 cellular extracts, or nuclear extracts, in the transcription reaction (Dingham and Roeder (1983) Weil et al (1979) Manley et al (1980). Although there was endogenous Polymerase activity within these extracts the transcriptional activity could be increased 10-20 fold by adding purified RNA polymerase II. Thus it transpired that in addition to Pol II and the TATA box certain accessory factors were also required for correct transcriptional initiation. This then led to attempts by a number of laboratories to characterise and purify these accessory factors.

Reconstitution assays have resulted in the identification and partial purification of four additional activities designated TFIIA, -B, -D and -E which together with pol II can accurately initiate from a minimal promoter containing a TATA element and start site. Studies on these factors and their role in the initiation process has led to the recognition of various functional stages involved in transcriptional initiation which reflect the combination and state of the above mentioned factors in the protein DNA complex (Sawadogo and Roeder 1985, Van Dyke et al 1988, Fire et al 1984, Buratowski et al 1989, Reinberg et al 1987). It has been possible to visualise the formation of the protein DNA interactions by footprinting and gel shift assays. TFIIID can bind the TATA element alone but does so much better in the presence of TFIIA to form a minimal complex. This is the initial event which subsequently facilitates the interaction of all the other initiation factors to form what is known as template committed complex (Hawley and Roeder 1987). The order of interaction in vivo is not clear but a hierarchy of complexes indicating the sequential addition of TFIIIB, Pol II, and TFIIIE has been demonstrated in vitro (Buratowski et al 1989). Subsequent addition of ATP or dATP and other NTPs allows transcriptional initiation to occur.

## **4. Studies on regulatory responses.**

The processes which have been addressed in detail so far have concerned general transcription factors which mediate the initiation process. However there have been an enormous number of studies concerning the characterisation of regulatory factors which bind upstream of the TATA element. By their very nature regulatory factors should have features of individuality so that they can mediate their appropriate responses to specific regulatory stimuli. Despite this fact it is becoming increasingly clear that certain fundamental features are conserved among many different transcription factors. In a later section I shall discuss some of these general observations.

However I shall first review some more specific studies which have attempted to identify transcription factors responsible for two mammalian responses, the induction of gene expression by cAMP and by phorbol esters, and a yeast transcription factor responsible for transcriptional induction in response to aminoacid starvation.

### **4.1. Regulatory studies in yeast;**

The study of transcriptional regulatory pathways in yeast has been largely, but not entirely, based on defining regulatory loci genetically. Therefore before a product is defined at the molecular level it has already been assigned a specific position in a genetic hierarchy of a particular physiological response. Thus the determination of the molecular basis of a regulatory pathway has a more informed starting point than in most other eukaryotic systems. An obvious consequence of this approach is that a cloned product for a transcription factor is isolated before it is established that the protein can itself bind DNA. In this sense therefore the approach is somewhat the reverse of that applied to the study of mammalian transcriptional responses.

There have been a number of systems which have been productively studied. The better characterised include the regulation of galactose metabolism and the aminoacid starvation response. Other studies have included mating type determination, heat shock response, and induction of cytochrome c genes by heme. In this section I shall discuss points from all these studies but will focus primarily on the transcriptional regulatory response to aminoacid starvation.



#### 4.1.1. Amino acid Starvation response.

##### 4.1.1.1. The isolation of S.cerevisiae GCN4, a positive regulator in general amino acid control.

The discovery of GCN4 came from studies of the biosynthetic pathways relating to amino acid synthesis in S.cerevisiae. In this system genes encoding enzymes involved in different amino acid biosynthetic pathways become coordinately derepressed (activated) when any single amino acid becomes limiting. For example starvation for histidine leads to activation of enzymes in histidine, arginine, tryptophan and lysine pathways (Wolfner 1975). The regulation of activity was shown to reflect the steady state mRNA levels of the enzymes in question indicating that regulation could be occurring at the level of transcription. (Struhl and Davis 1981, Aebi et al 1984, Zalkin and Yanofsky 1982, Donahue et al 1983). Subsequent analysis proved this idea correct. A study on the promoter region of one of the regulated genes, HIS3, indicated a region of between -86 and -99 as critical for correct regulation. This region was shown to contain the sequence 5'-TGACTC-3' which was repeated in other regions of this promoter and was also found upstream of other coregulated genes (Struhl 1982, Hinnenbusch et al 1983b). Conclusively it was shown that this element could confer starvation inducibility on an otherwise uninducible promoter. (Hinnenbusch et al 1985, Donahue et al 1983). It therefore became pertinent to identify a DNA binding protein which interacted with this element to mediate the response.

Investigations into the components of the regulatory pathway resulted in the isolation of a number of mutants which are unable to derepress under conditions of amino acid limitation (Wolfner 1975, <sup>et al</sup> Donahue et al 1983, Hinnenbusch et al 1983b). Classical and molecular genetic complementations constructed a regulatory hierarchy for these products and led to the identification of GCN4 as a positive regulator of amino acid biosynthesis which acted downstream of the other mutations (Hinnenbusch et al 1983b). This suggested that GCN4 protein could be a transcriptional activator which interacted with the 5'-TGACTC-3' positive promoter element. Hope and Struhl (1985) subsequently demonstrated that GCN4 protein synthesised in vitro did indeed bind to this regulatory sequence of the HIS3 promoter and other regulated genes. Other experiments demonstrated that once bound to the DNA GCN4 activates transcription (Hope and Struhl 1986).

A detailed study of presumptive binding sites in other regulated promoter led to the formation of a consensus palindromic GCN4 binding site 5'-ATGA(C/G)TCAT-3'. Saturation mutagenesis of the HIS3 regulatory site demonstrated that this consensus was the

optimal binding site for in vitro produced GCN4 and also acted as the best activating sequence when tested in vivo. Interestingly, however, although this was both the optimal and consensus GCN4 binding site, it never actually appeared in any of the regulated genes (Hill et al 1986). A possible reason for this apparent anomaly is discussed later.

#### 4.1.1.2. GCN4 binds as a dimer;

The palindromic nature of the GCN4 binding site raised the possibility that the GCN4 protein might bind as a dimer. Experiments addressing this theory were carried out by Hope and Struhl (1987). Deletion analysis of GCN4 had established that the carboxyterminal 60 amino acids of the protein were sufficient for specific DNA binding and produced a high mobility complex (relative to the full length product) in a gel shift assay (Hope and Struhl 1986). By mixing the truncated and full length products in the same reaction Hope and Struhl observed not two but three different mobility complexes. They concluded that the middle band was a heterodimer of the full length and truncated forms, the other two being homodimers.

#### 4.1.1.3. Regulation of GCN4;

The role GCN4 plays in derepression is therefore apparently clear. However as a regulatory protein it must itself be differentially active. Subsequent experiments have demonstrated that GCN4 expression is itself derepressed upon amino acid starvation through translational regulation of its message (Thieros et al . 1984; Hinnenbusch, 1984). Genetic analysis of the derepression pathway as a whole has shown that other mutations can either constitutively derepress or stop regulated derepression of GCN4 mRNA translation. One gene required to derepress GCN4 levels is GCN2, another locus isolated in the genetic dissection of this pathway. Interestingly GCN2 encodes a protein kinase whose levels are in turn transcriptionally regulated by GCN4 (Hinnenbusch 1985, Roussou et al 1988, Tzamaris et al 1989). The molecular mechanisms involved in this network are presently being studied.

#### 4.1.2. Galactose inducible expression;

Genes encoding enzymes involved in the metabolism of galactose are subject to regulation according to the carbon source they are grown in. Basal level gene expression can be seen in a glycerol based medium but this can be activated greatly on addition of galactose. This activation is in turn repressed on addition of glucose. Deletion analysis has identified a short DNA element, the UASG, which can confer the response to heterologous promoters. Genes involved in the regulation of galactose induced gene expression have, like GCN4 and GCN2, been isolated genetically. Two particularly well characterised examples are the GAL4 and GAL80 gene products. GAL4 is an activator in the galactose induction of gene expression and GAL80 has been genetically defined to repress activation through the GAL4

product (for review see Johnson 1987b). Molecular analysis has shown that GAL4 is a DNA binding protein which associates with the UASG whether inducible expression is at a basal (in the presence of glycerol) or activated (in the presence of galactose) level. Therefore activation does not appear to occur by altering DNA binding properties of GAL4 (Giniger et al 1985, Bram and Kornberg 1985). In addition GAL4 contains a well characterised transcriptional activator domain (see activation domains) and can act as a constitutive activator in gal80 cells whether they are grown in glycerol or galactose. It appears that constitutive activation is prevented in wildtype cells by the binding of GAL80 to a region in the 30 carboxyterminal aminoacids of GAL4. This is very close to where the predominant activation domain of GAL4 has been mapped to and it is possible that GAL80 masks these domains to prevent activation. Mutants of GAL4 lacking their carboxyterminus aminoacids are no longer subject to repression by GAL80. Additionally, over expression of this region of GAL4, which does not contain the DNA binding domain, selectively relieves the repression elicited on a wildtype GAL4 product presumably, by titrating out the endogenous GAL80 product (Ma and Ptashne 1987b). In the normal situation it is believed that galactose induction causes dissociation of GAL80 from GAL4 thereby initialising the activation response. Such genetic theorising awaits a biochemical test.

In contrast the presence of glucose appears to prevent GAL4 associating with DNA in the first place which explains its dominant activity in this whole response (Giniger et al 1985). Thus regulation in this response can occur via two separable mechanisms.

#### **4.1.3. HAP1 and induction by heme;**

Another example of regulation by modulation of DNA binding activity can be seen with the HAP1 activator. This gene was once again isolated genetically and subsequently shown to bind UAS's of two cytochrome c genes, CYC1 and CYC7 which are activated in response to elevated levels of intracellular heme. The DNA binding activity of HAP1 is itself increased in response to elevated levels of heme, a property which can be reproduced in vitro. The manner in which this regulation occurs is not entirely clear but heme appears to mediate the unmasking of the DNA binding domain (Pfeifer et al 1987).

#### **4.1.4. Heat shock induction;**

In contrast the heat shock transcription factor (HSTF), which associates with the heatshock element of heat inducible promoters, associates with its UAS in conditions of either basal or induced activity. Its transcriptional activation in response to heatshock is correlated rather with an increase in its phosphorylation state although it is not yet clear how the increased phosphorylation in itself results in increased activity. (Sorger and Pelham 1988).

#### **4.1.5. Mating type determination;**

The regulation of gene expression in yeast mating type determination provides an interesting example of gene regulation by the combinatorial action of transcription factors. S.cerevisiae haploid cells can be one of two mating types known as **a** and  $\alpha$ . The understanding of how this decision is made has resulted from a vast number of genetic and biochemical studies which have demonstrated the importance of many types of regulatory process, including rearrangements in the DNA (for review see Herskowitz, 1989). I <sup>do not</sup> have room to discuss all this background and so I will concentrate only on one salient feature. This is the ability of a single protein, MCM1, to form part of a DNA binding complex that can either activate or repress gene expression from different promoters according to the accessory protein it associates with. In  $\alpha$  yeast cells it can associate with the  $\alpha$ 1 product to stimulate transcription of  $\alpha$  cell specific genes such as STE3, and in the same cells it can repress expression of **a** specific genes, such as STE6, by forming a complex on their promoter's with  $\alpha$ 2. On the other hand in **a** cells, where the  $\alpha$ 2 and  $\alpha$ 1 products are not expressed, MCM1 can apparently act independently to activate **a** gene expression. Thus the MCM1 product which is expressed in both cell types can act to determine cell type gene expression by interacting with cell type specific factors, whose own expression is regulated by a gene rearrangement mechanism (Herskowitz 1989).

Thus from looking at only these few examples it is apparent that there are many ways in which transcription factor activity can be regulated in yeast. As will become clearer later, mammalian cells share many of these basic properties.

#### **4.2. Studies in mammalian systems;**

The classic route for studying regulatory transcription factors in higher eukaryotes is to choose a gene which is regulated by the response of interest and then determine the promoter sequences and subsequently the transcription factors involved. The apparent progress made on each response has varied considerably and has largely depended on the complexity of the promoter elements involved. Certain salient features concerning a number of investigations are addressed in the discussion. However at this stage it is pointless to catalogue all the pathways and factors which have received experimental interest. Therefore I shall describe in detail the experimentation which has taken place on only two. These factors are of particular relevance as they represent the mammalian versions of the yeast factors I have investigated in my thesis.

##### **4.2.1. Transcriptional activation in response to TPA;**

#### **4.2.1.1. Characterisation of a promoter element and transcription factor involved in response to TPA;**

Phorbol esters such as 12-O-tetradecanoyl-phorbol-13-acetate (TPA) are potent tumour promoters, capable of potentiating the effect of a subcarcinogenic dose of an initiating carcinogen (reviewed in Nishizuka 1984). Differential screens of TPA treated and untreated cells have revealed a number of genes, including certain proto-oncogenes such as c-myc and c-fos, whose expression is induced with TPA (for example see Greenberg and Ziff 1984). These observations have led to the hypothesis that TPA may cause tumour promoting activity by transcriptionally activating the expression of certain growth stimulating gene products. The dysregulation itself is thought to be mediated through protein kinase C which is a major cellular target for TPA and other tumour promoters (Nishizuka 1984). The manner in which kinase C may affect cellular gene expression is unknown. In an attempt to identify transcription factors which may be involved in this response Lee et al (1987a) studied the DNA binding proteins which interacted with two TPA inducible promoters, the SV40 promoter/enhancer and the human metallothionein IIA promoter. This study revealed two transcription factors, SP1 and AP1, which bound to both promoters. The role of AP-1 in the induction process was further implicated when it was noted that its binding site was present in a number of other TPA inducible promoters such as the collagenase gene (Angel et al 1987). Formal proof of AP1's activity was gained when a multimerised AP1 binding site was shown to confer TPA inducibility on a heterologous promoter. A point mutated multimerised site, which could not bind AP1, was unable to confer this inducibility. (Lee et al 1987b, Angel et al 1987). AP1 was purified as a 43-47 kd species and the next obvious step was to gain peptide sequence information from the purified protein in order to clone its gene. While this was in progress a series of independent experiments were indicating that the AP1 gene had already been cloned.

#### **4.2.1.2. Cloning and characterisation of v-jun;**

The v-jun gene was characterised as a cell-derived genetic insert in the avian sarcoma virus-17 (Maki et al 1989). When its predicted aminoacid sequence was compared to those of other proteins the best homology with the budding yeast protein GCN4 (Vogt et al 1987). The carboxy terminus of the predicted jun protein showed a 45% aminoacid identity to the 66 C-terminal aminoacids of GCN4. This region of GCN4 had previously been mapped as its DNA binding domain (Hope and Struhl 1986) and so the possibility arose that the v-jun protein was a DNA binding protein itself, perhaps able to interact with the GCN4 recognition sequence. This idea was tested by Struhl (1987) when he swapped the DNA binding domain of GCN4 with the homologous region of v-jun. The resultant protein was

indeed able to bind the GCN4 site and so it appeared that, in terms of DNA binding specificity, the v-jun protein was a mammalian equivalent of yeast GCN4.

The connection with AP1 came when the GCN4 and AP1 recognition sequences were compared. The optimal GCN4 site 5'-TGA(C/G)TCA-3' was identical to the AP1 consensus recognition site 5'-TGA(C/G)TCA-3'. It was therefore believed that the cellular equivalent of v-jun would encode AP1 and so this gene was isolated and sequenced. Peptide sequence information obtained from purified AP1 confirmed that the C-JUN protein was indeed in these preparations (Bohmann et al 1987), although other peptides which were present in these preparations did not match the predicted aminoacid sequence. It was also demonstrated that C-JUN recognised the AP1 site with the same specificity as purified AP1 and that the V-JUN protein could stimulate transcription from an AP1 site in vivo (Bohmann et al 1987, Angel et al 1988). Thus the DNA binding factor mediating TPA responsiveness appeared to have been isolated. However the DNA binding activity associating with the AP1 site proved to be a good deal more complex.

#### **4.2.2. Regulation by E1A and cAMP;**

##### **4.2.2.1. cAMP regulation of the somatostatin gene;**

Somatostatin is a 14 aminoacid peptide that modulates the secretion of other regulatory peptides in the pituitary, pancreatic islets and gastrointestinal tract. In this case previous work had established that cAMP regulated somatostatin biosynthesis in certain cell types. PC12 cells were one such cell type and were thus chosen as a cell line for experimentation (for review see M.R.Montminy and L.M.Bilezikjian 1987). The somatostatin promoter was fused to the reporter gene CAT which resulted in cAMP inducible CAT expression. Deletion analysis of the somatostatin promoter implicated a region between -71 and -48 as important in mediating the cAMP regulation. In addition a promoter fragment containing sequences -29 to -60 fused to a heterologous promoter was shown to confer a 15-fold stimulation of expression from the promoter in response to cAMP. Careful analysis revealed that within this region there was an 8 bp palindrome 5'-TGACGTCA-3' which was also present in other promoters regulated by cAMP (Montminy et al 1986). A protein shown to bind to this sequence was subsequently purified by affinity chromatography. The CRE binding protein (CREB) was shown to have a molecular weight of 43kD and it was also shown that the phosphorylation state of CREB increased 3-4 fold when PC12 cells were treated with forskolin, an activator of adenylate cyclase in these cells (M.R.Montminy and L.M.Bilezikjian 1987). In addition it was shown that purified CREB could be phosphorylated in vitro by the catalytic subunit of cAMP-dependent kinase, a cellular enzyme known to play a key role in mediating cellular responses to cAMP (M.R.Montminy and L.M.Bilezikjian 1987, Yamamoto et al 1988). Thus it appeared that the CRE and CREB

were specific and autonomous for the cAMP response. However a number of complications have since become apparent which arose, in part, from the realisation that this motif was also implicated in another regulatory pathway (see below).

#### **4.2.2.2. Adenovirus E1A- dependent trans-activation of transcription;**

Regulation of viral transcriptional units have proved a popular model to investigate the mechanisms of gene regulation and Adenovirus is a classic example. Upon viral infection efficient early gene expression is dependent upon the synthesis of products from the E1A gene. The two major products which have received most experimental attention, are polypeptides of 289 and 243 aminoacids. These are identical apart from a 46 aminoacid stretch unique to the larger polypeptide. These 46 aminoacids appear to play a critical, but as yet poorly understood, role in the activation process as a number of studies have indicated that the 289, but not the 243, aminoacid product can mediate transactivation of other early viral genes.(See Berk 1986 and Nevins 1989 for reviews)

Since this regulation was noted there have been a number of studies investigating the promoter regions of the early genes important for E1A transactivation in an attempt to identify transcription factors and/or mechanisms through which this regulation may occur. The early promoters investigated, which use RNA polymerase II, include those of the E1A, E1B, E2, E3, E4 and Major late genes (See Berk 1986 and Nevins 1989 for reviews). Two promoters of particular relevance to my thesis are the E4 and E2AE promoter.

##### **E4 Promoter.**

In the case of the E4 promoter two elements, apart from the TATA box, were characterised as important (Lee and Green 1987). Close sequence analysis of these elements demonstrated that they shared a core homology. The distal element contains two copies of this core and the proximal, only one. Binding assays showed that each element could bind the same transcription factor which was subsequently called activating transcription factor (ATF) (K.Lee and Green 1987, K.Lee et al 1987). The importance of this factor in E1A activation was further implicated when it was shown that the distal E4 promoter element could act as an E1A inducible enhancer when placed on a heterologous promoter (K.Lee and Green 1987). In addition it was shown that ATF sites were present and functionally important in a number of the other Adenovirus early gene promoters (K.Lee et al 1987). One such promoter was that of the E2A gene.

##### **E2AE promoter.**

The basic structure of the E2 promoter is outlined in fig. 4.1.1. E2 is slightly unusual in the fact that it has two sites of transcriptional initiation. Both these sites have ' TATA like ' boxes at appropriate distances upstream, but the downstream site accounts for around 80%

of message whether the promoter is transactivated by E1A or not (Elkaim et al 1983, Boeuf et al 1987). Transcriptional activation from the downstream site has therefore received the greatest attention although most mutations appear to affect both similarly (Boeuf et al 1987). Linker scanner analysis, conventional deletion analysis, and specific mutational analysis of the promoter has implicated three regions of major functional significance (Murthy et al 1985, Boeuf et al 1987, Imperiale and Nevins 1984, Loeken and Brady 1989). These are outlined in the figure as one ATF binding site and two binding sites for the transcription factor E2F. DNA binding assays have complemented these findings showing that sequence specific transcription factors interact with the functionally important regions (SivaRaman et al 1986, Kovesdi et al 1987, Boeuf et al 1987). However there have also been reports which demonstrate protein-DNA interactions in other regions of the promoter for which no function has been ascribed (Boeuf et al 1987).

These two promoters therefore represent examples where ATF binding sites have been implicated as important for mediating, or helping to mediate, responsiveness to E1A. However it is important to note that ATF is by no means the only factor which has been implicated in E1A induction and some inducible promoters do not contain any ATF binding sites whatsoever. Other sequences such as the E2F site and the TATA element itself are probably equally important (Nevins 1990).

#### **4.2.2.3. The ATF site and CRE are very similar;**

Through these studies it rapidly became apparent that the consensus ATF binding site and the consensus CRE were identical 5'-TGACGTCA-3'. Subsequent detailed mutational studies indicated that factors binding the CRE and the ATF site had identical DNA binding specificity and very similar apparent molecular weights (43kd) (Lee et al 1988, Montminy and Bilezikjian 1987, Hardy and Shenk 1988). In addition certain E1A inducible promoters could also be stimulated by elevated levels of cAMP (Sassone-Corsi 1988). It was therefore widely believed that CREB and ATF were the products of a single gene and thus the same transcription factor protein was the converging point for both E1A and cAMP transcriptional induction. However there was also additional evidence at this time of other transcription factors which could recognise certain ATF/CRE sites but not others. For example Cortes et al (1988) reported a 65-72 kd factor, which they called EivF, that was capable of recognising an ATF site in the E4 promoter and a CRE in the somatostatin promoter but was unable to bind efficiently to an ATF site derived from the E3 promoter. Similarly Raychaudhuri et al (1987) identified a factor, E4F, which interacted with ATF sites derived from the E4 promoter but not with ATF sites from any of the other Adenovirus early promoters. The significance of these observations was not immediately clear, but as will



become apparent in the Discussion, they probably represent very real and pertinent phenomena.

## **5. General principles of transcription factors.**

The previous section has described the progress made towards an understanding of the nuclear proteins mediating two transcriptional responses. Similar studies on other systems has resulted in the isolation of a number of transcription factor clones and these clones have themselves been the subject of intense experimentation. Deletion analysis has been used to elucidate the regions of the protein required for DNA binding and activation of transcription. These experiments have shown that the domain required for DNA binding is separable from that responsible for activation ie there are mutant proteins which can bind DNA but cannot activate transcription.

In addition sequence comparisons have identified certain structures which are shared in the binding domains of different factors and this has resulted in the formulation of four classes of binding motifs. A similar comparison and classification has taken place for activation domains. In this section I shall review in varying detail important features which have resulted from this analysis.

### **5.1. Separation of DNA binding and Activation domains;**

Genetic experiments using phage lambda repressor were the first to suggest that DNA binding and activation of transcription are separable functions on a transcription factor (Guarente et al, 1982) and since the cloning of transcription factor genes in other systems this observation has been extended to a number of eukaryotes (See Ptashne 1988 for review).

The underlying message from all these experiments is that the binding of an activator protein to its cognate site is necessary but not sufficient for stimulation of transcription. Deletion studies using the yeast transcriptional activators GCN4 and GAL4 have clearly defined forms which can bind to DNA without activating transcription. Conversely when small regions of these proteins not required for DNA binding were fused to the binding domain of a bacterial transcription factor, Lex A, it was possible to isolate regions of both GAL4 and GCN4 which could activate transcription through the bacterial binding site (Brent and Ptashne 1985, Keegan et al 1986, Hope and Struhl 1986). This type of experimental approach has now been extended to transcription factor clones isolated from higher eukaryotic systems and a very similar pattern is emerging.

### **5.2. DNA-binding motifs;**

In elucidating the manner in which a DNA binding protein recognises its cognate site X-ray crystallographic studies provide the definitive data. In most cases studied this stage of analysis has not been reached due to the inability to generate or purify sufficient amounts of these low abundance proteins. Instead predictions on what structure a particular transcription factor sequence will adopt has been based on existing structural data obtained from other proteins. Such analysis has created defined a number of general motifs into which a number of different factors may be placed. I will point out the most important features from each group but will only consider one motif, the leucine zipper, in great detail.

### 5.2.1. The helix turn helix;

This was the first motif to be defined and is also the only one for which firm X-ray crystallographic data has been obtained. This data has been obtained from prokaryotic repressor (CRO and CI) and activator (CAP) factors which were originally isolated from a genetic screen. The structure essentially consists of two alpha helices separated from each other by a  $\beta$ -turn with some aminoacids from one of these helices (the recognition helix) making contacts with the bases exposed on the major groove of the DNA (Anderson et al 1987, Jordan and Pabo 1988, Otwiniski et al 1988).

From these studies it became possible to correlate certain features of primary aminoacid sequence with the structure which was formed. Thus it was predicted that the yeast mating type loci MAT $\alpha$ 1 and MAT $\alpha$ 2 could form the same structure (Shepherd et al 1984). Both these factors were subsequently shown to bind DNA (Johnson and Herskowitz 1985). It was also predicted that the homeobox region, found initially in a number of Drosophila genes of developmental significance and then also in a number of mammalian genes, could also form a similar structure (Laughton and Scott 1984). Recent in vitro footprint analysis using a number of these homeobox proteins has also shown them binding DNA (see Biggin and Tjian 1989 for review).

### 5.2.2. Zinc finger;

This structure was proposed after studies on the transcription factor TFIIIA which is required for the expression of 5S ribosomal RNA genes by Pol III (Miller et al 1985). The sequence of TFIIIA revealed a repetitive motif consisting of an ordered pattern of cysteine and histidine residues which was critical for the factor's ability to bind DNA and to complex with zinc. The motif consisted of cys-n2-cys-n12-his-n3-his (where n is any residue) and was repeated nine times in the TFIIIA factor. The model suggested that the 12 residues between his and cys would loop out to allow the cys and his pairs to coordinate a single zinc ion (Miller et al 1985). Since then zinc finger motifs have been identified as repeated structures in a number of Pol II transcription factors and in many eukaryotic systems. In

some cases such as steroid hormone receptors and the yeast GAL4 protein a slightly different structure which consists of two pairs of cysteines has been observed. These motifs can nevertheless bind zinc and are required for DNA binding (Freedman et al 1988, Johnson 1987a, Johnson and McKnight 1989). No crystallographic study has yet been performed for a zinc finger protein and so it is unclear exactly how this structure interacts with DNA. The original model predicts that the 12 intervening residues between cys and his loop out to contact DNA and also provide specificity. In anticipation of the structural studies, finger swap experiments may provide a genetic test.

### 5.2.3. Leucine zipper;

The concept of the leucine zipper as a structural motif was first introduced by Landschultz et al (1988a) when they noticed a homology in the primary amino acid sequence between the nuclear proto-oncogenes, C-MYC and C-FOS, and the DNA binding proteins GCN4, C-JUN (See relevant sections for details of these proteins) and C-EBP. C-EBP had been characterised as a rat liver nuclear protein capable of binding certain viral promoter sequences (Johnson et al 1987). The common feature between these proteins is that in a region which can be predicted to form an alpha helical structure a number of leucine residues, generally four or five, appear as a heptad repeat and align on one side of this idealised alpha helix (Landschultz et al 1988b). In the case of GCN4, JUN, and C-EBP this structure appeared in a region of the protein known to be important for DNA binding. At the time this model was proposed GCN4 was the only one of these proteins for which a dimeric DNA binding form had been well characterised but it was widely predicted C-EBP and jun would do likewise as both bound to palindromic sequences (Hope and Struhl 1987). On the basis of this information Landschultz et al (1988a) predicted that the alpha helices of two monomeric units could align in an antiparallel fashion with the leucine residues of one helix interdigitating with the other, in a manner akin to a zipper, to provide a structural basis for dimer formation. As such the zipper provided only a means of obtaining protein-protein interactions but not protein-DNA interactions.

However it was also noticed that a stretch of positively charged basic residues appeared immediately N-terminal of the zipper in a region also necessary for DNA binding of JUN, GCN4, and C-EBP. The zipper model predicted that the zipper would bring the basic residues of each monomer into very close proximity so together they could form specific interactions with the DNA. It was also suggested that zipper regions of different monomers may interact to produce heterodimers harbouring potentially different properties from the homodimer (Landschultz et al 1988b). Subsequent experimentation has indicated many of the predictions are correct although the interdigitating nature of the leucine residues has been somewhat questioned (Turner and Tjian 1989, Landschultz et al 1989, Kouzarides and Ziff

1988, Gentz et al 1989, O'Shea et al 1989). In the case of GCN4 it has been shown that the helices align in a parallel, not antiparallel, fashion and a structure more akin to the previously described coiled coil has been proposed (O'Shea et al 1989). This coiled coil model, originally formulated by Crick (1952) on the basis of X-ray data, has been shown to apply to a number of fibrous proteins such as myosin and keratin (See Cohen and Parry 1986 for review). Like the leucine zipper coiled coils have a heptad repeat of hydrophilic residues but in this case there are two such repeats interspersed within the same helix. The two heptad repeats align on the same side of the helix with the leucine repeat of one helix interacting with the second hydrophilic repeat of the other helix. Such an interaction could theoretically occur for any of the leucine zipper proteins discussed. However structural studies will be needed to firmly establish the exact nature of these interactions.

In this respect the zipper model has become somewhat altered, however in a number of other criteria it appears essentially correct and a number of interesting developments have occurred.

One prediction of the zipper model is that the zipper itself provides a means of bringing together the basic regions of the protein to confer a DNA binding specificity. As such DNA binding requires the presence of a functional zipper but dimerisation should not require the binding domain. Recent mutational analysis has indeed shown this. Alteration of leucine residues in the zipper results in both loss of binding and dimerisation capabilities, whereas mutations in the basic region which abolish DNA binding do not abolish dimerisation (Landschultz et al 1989a, Gentz et al 1989, Turner and Tjian 1989). Recent experiments have indicated that leucine zipper proteins can form functional heterodimers, with the best example involving the association of C-JUN with C-FOS. Heterodimerisation appears to be quite a specific event however as GCN4 cannot associate with either C-JUN or C-FOS (Kouzarides and Ziff 1989). The significance of these interactions, particularly relating to JUN and FOS are addressed in the Discussion.

#### **5.2.4. Helix-Loop-Helix;**

This DNA binding/ dimerisation motif is the most recent to be defined. It was initially discovered in two proteins, E12 and E47, that bind to a specific DNA sequence, κE2, found in the immunoglobulin enhancer (Murre et al 1989a). Because it is the most recent motif described, its structural characterisation is rather preliminary. The predicted structure essentially consists of two short amphipathic helices separated by a 20-25 aminoacid loop region that contains residues characteristic of β-turn and a secondary loop structure. This intervening region is therefore longer than the short turns of the Helix-turn-Helix motif. Sequence comparison<sup>n</sup> within the HLH domain revealed this structure was also apparent in a number of other proteins including MyoD and myogenin, which are products which play critical roles in myogenesis, and also certain Drosophila gene products which regulate the

development of the peripheral nervous system. There are a number of residues within the two helices which are highly conserved among all the different family members indicating the alignment was not simply fortuitous. Functional analysis addressing the role of HLH region itself has shown it to mediate dimerisation, rather like the zipper motif of proteins such as c-jun and c-fos. Also like the zipper proteins the dimerisation motif is immediately preceded by a short stretch of basic amino acids which, in the case of MyoD at least, have been shown to make direct contacts with DNA and would therefore be predicted to mediate the specificity of binding (Davies et al 1990).

An obvious question is whether different members of the HLH family can heterodimerise with each other. Very recent work has demonstrated that they can and the heterodimers that are formed can bind to the  $\kappa$ E2 sequence that was originally used to isolate the E12 and E47 (Murre et al 1989b). It is interesting to consider the regulatory consequences for such heterodimerisation processes. When MyoD is expressed in CH310T<sup>1/2</sup> cells it can induce their differentiation into myoblasts (Davis et al 1987). Thus one might predict that when myoblast differentiation is induced by other effectors one may see a concurrent increase in MyoD expression. This is not the case and thus invokes a model whereby MyoD activity during differentiation may be potentiated by differential heterodimerisation processes (see Discussion).

All motifs described in this section are simply structures which facilitate an interaction with DNA in a general sense. There is presently little insight into the molecular associations which determine the specificity of what protein can bind which nucleotide sequence. Once again it is studies on prokaryotic proteins which have provided the little knowledge we have (Ptashne 1988).

In addition there are also a number of transcription factors such as SRF, AP2, and NF1 whose DNA binding domains can not be modelled into any of the motifs thus far described. It is therefore likely that new structures will become apparent in the future (see Mitchell and Tjian 1989 for review).

### **5.3. Activation Domains;**

Functional deletion studies have delineated domains of eukaryotic transcription factors required for either DNA binding or activation. In both studies yeast factors have been the first analysed and certainly in the case of transcriptional activation yeast has provided the most pertinent information on general mechanisms.

Activation domains per se were first defined in the GAL4 protein where deletion analysis demonstrated two short amino acid stretches, one at the carboxyterminal end and the other close to the aminoterminal end of the protein which together could confer almost wild-type activation potential, even when a large internal section of the protein was deleted (Ma and Ptashne 1987a). A similar type of approach also identified a single short stretch of aminoacids critical for activation in the GCN4 protein (Hope and Struhl 1986). The only shared property between these activating domains was their negative charge, due to the presence of a high proportion of acidic aminoacid residues. The importance of this acidity was demonstrated in the point mutational analysis of one of the GAL4 activating domains. Mutants with increased negative charge of the region were seen to increase activating potential but the opposite was true when a mutation decreased the negative charge (Gill and Ptashne 1987). It should be noted however that not all down mutations were a result of a reduction of overall acidity, indicating other considerations may be important.

In an attempt to identify a comprehensive series of activator sequences Ma and Ptashne (1987c) cloned random bacterial DNA fragments upstream of the GAL4 DNA binding domain and isolated fusion genes which could activate transcription from a GAL4 binding site. Analysis of the bacterial sequences isolated demonstrated that all had an overall negative charge reinforcing the apparent importance of acidic residues for activation. However an additional feature of these sequences was that most could form amphipathic alpha helices, ie an alpha helix with charged residues on one side and hydrophobic ones on another. The GCN4 activator region could also form such a structure and so the amphipathic helix became a subject of experimentation. Giniger and Ptashne (1987) demonstrated that a synthesised 15 aminoacid peptide with the potential of forming an amphipathic alpha helix could indeed confer activation properties on a GAL4 DNA binding domain at about 20% efficiency of the wildtype protein. A similar peptide containing the same aminoacid residues but unable to form this amphipathic helix was unable to activate. Thus a structural as well as a charged criteria was realised and in any one case both may play important roles (Hope et al 1988). An interesting development on this work was the demonstration that such acidic regions could also confer activation properties in mammalian cells. Thus if a promoter bearing a GAL4 binding site is introduced into mammalian cells it can be activated if GAL4 is co-expressed. However if a GAL4 deletion containing only the binding site is expressed then no activation occurs (Kakadini and Ptashne 1988, Webster et al 1988). Fusion gene experiments have also been extended to viral transactivators such as VP16 (Sadowski et al 1988). VP16 is a virion protein of the Herpes Simplex virus which activates immediate early gene expression by associating to viral promoters via an attachment to a cellular transcription factor (see Goding and O'Hare for review). When a highly acidic portion of VP16 is fused to the GAL4 binding domain it activates transcription from a GAL4 binding

site about 100-fold more efficiently as GAL4 and, as such, it is the strongest activator yet tested in mammalian cells (Sadowski et al 1988).

However, despite their obvious ability to function in mammalian cells, the presence of acidic domains in cloned mammalian transcription factors has not been highly documented. JUN and the human glucocorticoid receptor are notable exceptions (Bohmann et al 1987, Hollenberg and Evans 1988).

The study of mammalian transcription factors has led to the identification of new motifs. Deletion analysis has demonstrated that two activator regions of the transcription factor SP1 contain approximately 25% glutamine (Courey and Tjian 1988). Such glutamine stretches are also found in other transcription factors although their functional significance has not yet been determined (Williams et al 1988, Sturm et al 1988).

Another apparent activating motif consists of a proline rich domain. Such a region is found in a number of transcription factors including CTF/NF-1, AP-2, SRF, and JUN (see Mitchell and Tjian 1989 for review). Removal of this region from CTF/NF1 results in a transcription factor capable of binding to DNA but unable to activate transcription. In addition the 100 aminoacid proline rich stretch can confer activation properties onto the otherwise inactive DNA binding domain of SP1 (Mermod et al 1989).

Although I have described these three activation motifs separately, more than one can be observed on any one transcription factor. For example within a 52 aminoacid activator region of JUN which is itself rich in glutamine and proline there is a particularly important 22 amino acid stretch which has overall charge of -6 (Bohmann and Tjian 1989). However it is not clear whether all activation motifs are equally as active in all eukaryotic systems.

#### **5.4. Interaction of an activator protein with the TATA factor;**

In the Introduction I have described work characterising initiation factors and regulatory factors separately. However it is clear that these two classes of proteins must interact either directly or indirectly to facilitate activation of transcription. Therefore a number of studies, using *in vitro* transcription and binding assays, have investigated how upstream factors effect the initiation process (Sawadogo and Roeder 1985, Horikoshi et al 1988a, Hai et al 1988a).

One example of this analysis involves the transcriptional activator ATF (Horikoshi et al 1988a, Hai et al 1988a). It is clear that the addition of a number of ATF sites to a basal TATA box containing promoter can increase transcription 800-fold in nuclear extracts. By careful dissection of the initiation process itself it can be seen that ATF appears to be needed only during the formation of the template committed complex but that its presence during this stage is critical. In this way it is believed ATF

interacts both with TFIID and certain other components of the initiation machinery to form an active preinitiation complex. However once formed this complex no longer requires the presence of ATF as assayed by footprinting and in vitro transcription. Although this study involves only a single activator protein the results are probably of general significance.

Some experiments have demonstrated that small variations in the TATA sequence can influence the ability of certain activators to mediate their effect. For example Harbury and Struhl (1989) have identified variants in the TATA box of the HIS3 promoter that are capable of mediating induction by GCN4 but not GAL4. This has led to the idea that there may be more than a single TATA box binding protein .

In other cases however activators appear to be able to mediate transcription in the absence of a TATA box. For example a GCN4 site has been shown to be capable of mediating polymerase initiation from the correct site of the HIS3 gene when all TATA sequences were removed (Chen and Struhl 1989). In this experiment the acidic activating domain of GCN4 was shown to be essential for function, in agreement with the idea that the heptad repeat of Polymerase may be capable of interacting with acidic helices of transcription factors. Another study by Brandel et al (1989) has demonstrated a direct interaction of purified yeast Pol II and the yeast activator GCN4 in vitro. Paradoxically the in vitro analysis has shown that the DNA binding domain, in the absence of the activating region, is necessary and sufficient for direct interactions between these two factors (Brandel et al 1989). A clear understanding of the interactions required for activation will require far more experimentation both in vivo and in vitro.

## **6. Other ways in which transcription is regulated;**

### **6.1. Chromatin structure;**

A prerequisite for the regulation of gene expression by transcriptional activators is that the activators themselves can gain access to the DNA. The interaction of DNA with histones and the consequential packing of nucleosomes to form chromatin fibres obviously applies certain constraints of access. However a large body of work has demonstrated that regions of chromatin which are being, or poised to be, transcribed exist in an ' active or open form ' (for reviews see Gross and Garrard 1988, and Svaren and Chalkley 1989) . Active chromatin is characterised by having an increased sensitivity to DNase1; this in itself demonstrates an increased accessibility of the DNA to proteins reflecting a less packed chromatin structure. A closer look within regions of active chromatin has shown certain



short stretches (50-400 bp) more sensitive to DNase 1 than the rest (see Nahon et al 1987 for example). In some cases such hypersensitive regions have been shown to be histone free (Han and Grunstein 1988). In addition some experiments have shown the position of the hypersensitive region correlates with the region of transcriptional control. Based on where and when they are hypersensitive to DNase 1 sites become differentially classified as either constitutive, inducible, developmental or tissue specific (Gross and Garrard 1988). Although it is now widely apparent that hypersensitive sites occur around transcriptionally active genes it has also been shown that the presence of a hypersensitive site in itself is not sufficient to conclude transcription is occurring (Weintraub et al 1982).

## 6.2. DNA methylation;

Methylation is another manner in which the expression of certain genes can be regulated. By methylating CpG residues on cytosine it is possible to create a small structural change in the DNA which can be propagated to further cell generations. Thus methylation in a promoter region, where CpG residues are clustered, provides a way in which protein DNA interactions may be altered. The overwhelming consensus of a number of experiments is that methylation of CpG residues in promoter regions serves to repress gene expression (see Bird 1987, and Cedar 1988 for reviews). One simple way this has been demonstrated is by in vitro manipulation of DNA sequences.

Studies on the  $\gamma$  globin gene shows that in fibroblasts the endogenous allele is not expressed and is also heavily methylated. Expression is seen however if an unmethylated gene is introduced into the same cells, a phenomena which is reversed if the new DNA is methylated before introduction (Buslinger et al 1983). The exact manner in which methylation influences gene expression in vivo is not entirely clear but it probably involves influencing protein DNA interactions in some way. Some data indicates that methylation of CpGs in transcription factor binding sites can inhibit protein binding (Kovesdi et al 1987, Watt and Molloy 1988) but this phenomena is by no means universal (Holler et al 1988). Indeed in the case of the  $\gamma$  globin promoter it appears that methylation need not be site specific but simply region specific which indicates a mechanism which does not simply involve the exclusion of a particular transcription factor (Murray and Grosveld 1987). Recent experimentation has, in fact, supported an idea of methylation facilitating the binding of a methylation specific DNA binding protein which could in turn exclude transcriptional activators, a mechanism which could account for the globin promoter data (Antequera et al 1989, Meehan et al 1989).

## 7. Summary; Reasons for carrying out my project.

The Introduction has outlined a number of features of transcriptional regulation. On the one hand studies in mammalian cells, due to the very fact they address mammalian regulatory pathways, are of intrinsic interest and importance to an understanding of human biology as a whole. On the other hand the genetic tractability of the yeast system, which is the envy of the mammalian biologist, has resulted in the identification of a large number of regulatory loci thereby paving the way for a molecular understanding of certain regulatory interactions. The experiments in my thesis are aimed at addressing whether the yeast and mammalian transcriptional machinery are conserved. The long term aim of this work was to determine whether a study exploiting the genetic tractability of yeast might be instructive in understanding certain fundamental concepts of higher eukaryotic transcriptional biology. A number of other laboratories have also been addressing this question and consequently many new developments have occurred since I started my thesis. These Developments will be analysed in the discussion. At this point I shall simply summarise the information that was available to me at the beginning of my thesis which, even then, indicated my work may provide an instructive line of investigation.

1. The structure of yeast and mamalian promoters show some striking similarities. I have discussed this information in an earlier section.
2. Genes encoding the largest subunit of RNA polymerase II from a number of eukaryotes indicated that this enzyme had been quite well conserved during evolution (see basic transcription factor section).
3. It was known that the SV40 promoter/enhancer could function in fission yeast. This had been demonstrated by the studies of Lee and Nurse (1987) where a fission yeast mutation was rescued by a human cDNA expressed from this promoter/enhancer.

## **CHAPTER 2; MATERIALS AND METHODS**

## 1. DNA/RNA PREPARATION;

### 1.a. Plasmid;

Both large scale and small scale preparation of DNA employed the alkaline lysis method of DNA preparation essentially as described in Maniatis et al (1982)

**1.a.i. Small scale;** 1.5mls of an overnight culture was pelleted and the bacteria resuspended in 100 $\mu$ l of solution 1. Cells were lysed by addition of 200 $\mu$ l solution 2, placed on ice for 5 minutes after which 150 $\mu$ l of solution 3 was added. This was then kept on ice for a further 5 minutes and then spun at 4 C for 5 minutes. The supernatant was decanted, phenol extracted, and ethanol precipitated. The pellet was washed in 70% ethanol, dried and resuspended in 20 $\mu$ l TE buffer.

**1.a.ii. Large scale;** 400mls of overnight bacterial culture was pelleted by centrifuging 5 K (unless otherwise stated K in this context refers to x 1000 revolutions per minute) for 10 minutes in a Sorval GS3 rotor. Pellets were resuspended in 10 mls solution 1 (plus 5mg/ml lysozyme (sigma)). Cells were lysed in 20mls solution 2 and chromosomal DNA was precipitated by the addition of 15mls of solution 3. This was centrifuged in the sorval at 18K for 20 minutes in an SS-34 rotor. To the supernatant was added 20mls of isopropanol and the DNA/RNA was pelleted by centrifugation at 10K for 10 minutes in a Sorval GSA rotor. The pellet was resuspended in 4.5 mls TE, 5.0g of caesium chloride and 50 $\mu$ l (10mg/ml) of ethidium bromide. This was loaded into 5ml Beckman polyallomer tubes which were sealed and loaded into a Beckman vti65 rotor. Centrifugation lasted for 4 hrs at 64K or overnight at 55K. The resulting bands gained from the caesium chloride gradient were removed with a syringe and needle. 2.5 mls of dDW was added and the ethidium bromide was extracted with repeated washes with isobutanol. DNA was isopropanol precipitated, washed in 70% ethanol, dried and resuspended in TE. DNA concentration was determined by 260:280 spectrophotometry.

**1.b. Buffers;** solution 1; 1% glucose, 25mM Tris pH 7.4, 10mM EDTA

solution 2; 200mM NaOH, 1% SDS

solution 3; 60ml 5M KOAc, 11.5mls glacial acetic acid, 28.5mls dDW

**1.c. Phenol and Phenol /chloroform extractions;** Phenol was supplied by BRL in a redistilled form and it was prepared in batches of 400mls. 0.4g of the antioxidant hydroxyquinoline was added to 400mls of phenol. This was then

equilibrated to pH. 7.8-8.0. by mixing and decanting an equal volume of 0.5M and subsequently 0.1M Tris.Cl pH 8.0. Once the phenol phase had reached a pH of over 7.8 it was divided into aliquots of 40 mls and 10 mls of 0.1M Tris.Cl pH 8.0 containing 0.2% B- mercaptoethanol was added to each. This was then stored at 4 C. Calibrated phenol was also added to chloroform and isoamyl alcohol in a ratio of 25:24:1. 10mls of 0.1M Tris.Cl pH 8.0 was also added to 40ml aliquots of this mixture and stored at 4 C.

A standard extraction involved adding to the sample an equal volume of Phenol: chloroform : isoamyl alcohol (25:24:1), vortexing and centrifuging in a microfuge for 2 minutes. The upper aqueous phase was then removed and the process repeated if necessary.

**1.d. Isolation of S.pombe RNA;** RNA was prepared essentially as Kaufer et al (1985). Throughout the protocol particular caution was taken to keep all reagents, gilson tips and tubes absolutely sterile.

A 200 ml overnight culture was grown to an OD600 of 0.5. The cells were centrifuged at 2K for 5 minutes in a benchtop centrifuge and washed in ddW. The pellet was resuspended in 500µl HEN buffer with an equal volume of glass beads (acid washed and baked). Tubes were then vortexed for 2 minutes and then supernatant removed into an <sup>Eppendorf</sup> tube. The beads were then washed a further 2 times with 0.5ml HE buffer pooling the washes with original lysate. This pool was then microfuged for 1 minute and the supernatant mixed with an equal volume of 2x PK buffer and incubated at 37 C for 20 mins. This was then phenol extracted twice and ethanol precipitated. RNA was quantified by spectrophotometry.

#### Buffers

HEN buffer; 50mM Tris pH 8.0, 10mM EDTA 150mM NaCl

HE buffer; 50mM Tris pH 8.0, 10mM EDTA

2X PK buffer; 100mM Tris.HCL (pH 7.5), 2% SDS, 20 mM EDTA, 400µg/ml proteinase K

**1.e. Isolation of S.pombe DNA;** This was prepared by S.Moreno as described in R.Jones et al (1988)

## 2. DNA MANIPULATIONS;

**2.a. General;** BSA and DTT were usually added separately to each reaction, and where necessary RNase A to 2µg/ml final.), which was supplemented to 1mM ATP

and 20mM DTT final concentrations for ligations. Enzymes quantities used were approximately as suggested by the manufacturer.

**2.b. Restriction Enzyme Digestion;** 1-5 $\mu$ l of miniprep DNA or 1-3  $\mu$ g of maxiprep DNA was digested at 37 C in a 20 $\mu$ l volume for 1 to 2 hours before adding loading dye and loading onto agarose gels. All digests were carried out in reaction buffers recommended by and supplied by restriction enzyme manufacturers.

**2.c. Phosphatase treatment;** The 5' terminal phosphate of vector fragments was removed using calf intestinal phosphatase (CIP : Boehringer Mannheim) to prevent vector recircularisation on ligation. Restricted DNA was phenol extracted, precipitated and resuspended in 40 $\mu$ l of phosphatase buffer (0.1mM EDTA, 50mM Tris.HCl pH 8.0, 0.1mM ZnCl<sub>2</sub>, 1mM spermidine), and incubated with 2- 20 units of CIP (Boehringer Mann.) at 37 C for 30 minutes. The appropriate fragment was then gel purified.

**2.d. Fragment purification;** DNA fragments were purified on 150ml low melt 1-2% Tris acetate agarose gels with ethidium bromide. Gels were run at 50-80 mA in a tank from which light was excluded after which the gel was removed and visualised on a long wave UV light box. The appropriate fragment was excised placed into an eppendorf and heated to 65 C for 10 minutes to melt the agarose. The tube was then vortexed and placed at 37C in preparation for ligation step.

**2.e. Ligation;** Ligations were generally carried out in a 20 $\mu$ l volume containing 2 $\mu$ l 10x ligase buffer (0.66M Tris pH 7.5, 50mM MgCl<sub>2</sub>, 50mM DTT and 10mM ATP) 100ng of purified vector and 400ng of purified fragment and 10U T4 DNA ligase (Biolabs). When fragments or vectors were in LM agarose the final agarose concentration did not exceed 0.4%. Ligations were left 4-16 hrs after which they were transformed into E.coli.

**2.f. Nested Deletions;** 2 $\mu$ g of DNA per 160bp was digested with appropriate enzymes to leave a 5' overhang at the end to be deleted and a 3' overhang at the end to be saved. If a convenient 3' site was unavailable then a 5' overhang site was cut and then subsequently filled in by with deoxy-thioderivatives (Stratagene), resistant to Exo III digestion. After phenol/ chloroform extraction and precipitation each pellet was resuspended in 36 $\mu$ l of ExoIII buffer per 2 $\mu$ g of DNA, and prewarmed to 37 C. After taking a zero timepoint, a fresh mixture of 20U of ExoIII in 4 $\mu$ l of ExoIII buffer was added per 2 $\mu$ g DNA, and 9.0  $\mu$ l timepoints collected every 30 seconds. Adjacent sets

of 4 timepoints were added to the same aliquot of stop mix (4 $\mu$ l of 10X ExoVII buffer) on ice. Each pool covers about 150-200bp of deletion. At the end of the timecourse, 2 $\mu$ l of ExoVII buffer containing 0.3U ExoVII was added to each pool of 4 timepoints, incubated at 37 C for 45 mins, and then the enzyme inactivated at 75 C for 15 mins. (10 $\mu$ l of every pool was run on a 1% agarose gel to confirm the extent and tightness of the deletions.)

20  $\mu$ l of each pool was added to 4 $\mu$ l of 75mM MgCl<sub>2</sub> and 2mM of each of the four dNTPs. To this was added 4 $\mu$ l of Klenow/T4 mix in order to repair the termini, which are not blunted by ExoVII. After leaving at room temperature for 30-40 mins, the reactions were again heated to 75 C for 15 mins and then 5 $\mu$ l of the reaction ligated overnight at room temperature in a 20 $\mu$ l total volume. Half of this was then transformed into competent TG1 bacteria and a selection of resulting colonies tested for the presence of suitably sized fragments by appropriate restriction analysis. Chosen miniprep colonies were then grown for large scale production of DNA. These DNA preparations were then sequenced.

Buffers;

ExoIII buffer; 50mM Tris.HCl pH 8.0, 5mM MgCl<sub>2</sub>, 1mM DTT.

10X ExoVII buffer; 670mM Potassium phosphate pH 7.9, 83mM EDTA, 100mM  $\beta$ -mercaptoethanol

Klenow/T4 mix; 5U Klenow (Boehringer), 3U T4 DNA polymerase (Biolabs), 6.4 $\mu$ l 10X TA buffer, dDW to 64 $\mu$ l.

10X TA buffer; 330mM Tris.HCL pH 7.9, 660mM KOAc, 100mM MgOAc, 5mM DTT, 1mg/ml BSA.

### 3. DNA SEQUENCING.

DNA sequencing was carried out using sequenase version 2.0 (United States Biochemical Corporation). All DNA sequenced was double stranded plasmid template that had been purified on a caesium chloride gradient and all primers were 20-30bp oligonucleotides synthesised by Ian Goldsmith.

**3.a. Denaturing the template;** 5 $\mu$ l (5 $\mu$ g) of DNA was added to 20 $\mu$ l of 0.25M NaOH and left 5 minutes at room temperature. To this was added 10 $\mu$ l 1.5M NH<sub>4</sub>OAc and 90 $\mu$ l ethanol. After centrifugation the pellet was washed in 70% ethanol and dried and resuspended 25 $\mu$ l of dDW.

**3.b. Annealing;** 5 $\mu$ l of this DNA was added to 1 $\mu$ l (10 ng) of primer, 2 $\mu$ l reaction buffer, and 2 $\mu$ l ddW. This provides a molar stoichiometry approximately of 1:1 primer:template. The reaction was warmed to 65 C in a 500ml beaker within a water bath for 2 minutes and then the beaker removed to allow the reaction to cool slowly to 30 C.

**3.c. Labelling;** To the 10 $\mu$ l of annealed DNA was added the following; 1 $\mu$ l DTT, 2 $\mu$ l labelling mix (diluted 1:5 in H<sub>2</sub>O), 0.5 $\mu$ l <sup>35</sup>S dATP, and 2 $\mu$ l sequenase (diluted 1:8 in enzyme dilution buffer). This was mixed and incubated for 5 minutes at room temperature.

**3.d. Termination;** 4 tubes, each containing 2.5 $\mu$ l of one of ddATP, ddCTP, ddTTP or ddGTP were prewarmed at 37 C. To each of these tubes was added 3.5 $\mu$ l of the labelling reaction. The termination reactions were incubated for 5 minutes at 37 C after which 4 $\mu$ l of stop mix was added to each tube. The samples were then heated to 80 C for 2 minutes after which they were loaded on to a denaturing sequencing gradient gel and run for 2-6 hrs at 33mA. Gels were fixed in 10% acetic acid, vacuum dried and exposed overnight on to film.

**3.e. Buffers;** Reaction buffer; 200mM Tris.HCL pH 7.5, 100mM MgCl<sub>2</sub>, 250mM NaCl.

Labelling mix; 7.5 $\mu$ M each of dGTP, dCTP and dTTP.

Termination mixes; each mix contained 50mM NaCl and 80 $\mu$ M of three dNTPs. One dNTP was missing from each of the four and replaced by ddNTP. For example ddA termination mix had 80 $\mu$ M dGTP, dCTP and dTTP and 8 $\mu$ M ddATP.

Stop mix; 95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.5% xylene cyanol.

#### 4. CONSTRUCTS;

**4.a. E2-CAT constructs;** The plasmid PURA4 was a gift from P.Russell and contains the URA4 gene as a selectable marker as well the ars6 sequence in a pBR322 background. Unique cloning sites include Bam H1, Sal1, Sph1 and Cla1. pE2-E-CAT contains sequences -284 to -62 of the adenovirus type 5 E2A early transcription unit fused to the CAT coding sequences derived from pSVO-CAT which also contains the splice and poly A signals of SV40 (Murthy et al 1985). This whole fusion unit was



excised from pE2A-E-CAT as a 1.95 kb Xho1-Xba1 fragment and was ligated into the Sal1 and Xba1 sites of pUC19 (Stratagene). PU4E2-CAT was constructed by subcloning a Bam HI-SphI fragment from this intermediate into the Bam HI-SphI sites of pURA4.

**4.b. Construction of plasmids containing oligonucleotide AP-1 sites;** The plasmids pYT3, pYT3CP and pYT3E3 were constructed by inserting specific oligonucleotide sequences upstream of the CAT coding sequences in the plasmid pU4E2CAT. pYT3 was constructed by removing the E2 promoter sequences present in pU4E2CAT with a Sal1 BamHI cut and replaced with oligonucleotide E355-TATA containing the adenovirus E3 TATA region +10 to -37 (pYT3) linked to sequences -79 to -103 of the E3 promoter bearing the same sites. An Sph1 and Bgl2 restriction site was included in between the TATA and ATF site so that the E3 site could be replaced by other promoter fragments. pYT3CP was created by removing the E3 ATF site with a Sal1 and Sph1 cut and replacing with a Col API oligo. Similarly the E3 promoter AP-1 site was cloned in the same way. The TATA construct with no upstream promoter elements was created by cutting pE355TATA with Bgl 2 and Sph1, blunting the resulting overhangs, and ligating the large gel purified fragment. In all cases the upstream promoter region was separated from the TATA box by 32-34bp (see oligonucleotide section).

**4.c. SV40-CAT constructs;** Sergio Moreno constructed the plasmid pSABCAT by cloning the HindIII-BamHI fragment from pSV2CAT (Gorman et al., 1982) that contains the CAT coding sequence and SV40 3' regulatory sequences into the plasmid pSAB1 (Durkacz et al., 1985). pSMCAT was constructed by cloning the 323 bp PvuII-HindIII fragment containing the SV40 promoter-enhancer region upstream of CAT in pSABCAT. SMd11CAT was prepared by cloning the Bgl 2 -Hin dIII fragment of pA10CAT2 (Gorman et al., 1982) into pSABCAT; this fragment contains the sequences of the SV40 promoter region to be the Sph1 site at -198. SMd12CAT was prepared by cloning into SABCAT the HincII-HindIII fragment from the plasmid p21Tg (a kind gift of C.Goding); this fragment contains the promoter sequences up to position -150. SMd13CAT was prepared by cloning into SABCAT the Nco1-HindIII fragment from p21Tg which contains promoter sequences up to position -107.

**4.d. E1A expression plasmids;** The plasmids pPB13s and pPB12s were constructed by restricting the E1A 13s and 12s coding sequences from the plasmids pm13s and pm12s (a gift from Colin Goding) with Hind III and Sal 1 and cloning these fragments into into pSM1 (a gift from Sergio Moreno) restricted with the same

enzymes. This placed the 12s and 13s sequences immediately downstream of the SV40 early promoter and enhancer in a yeast episomal plasmid which contained a LEU selectable marker. Plasmids were transformed into the S.pombe mutant strain (leu1-32, his3-237, ura4-d18).

**4.e. Budding yeast plasmids;** The plasmid, pLG559D-312 (A255) (Guarente and Mason 1983) was the parent of plasmids used in this study. This is a 2 $\mu$  origin based, CYC1/LacZ fusion plasmid activated by the yeast activator, HAP1, which binds CYC1 UAS (see introduction). It also contains a URA3 selectable marker. The Sma1-Xho1 fragment which contains the CYC1 wild type UAS was replaced with an oligonucleotide linker containing sites for the restriction enzymes Sph1 and Sal1. Oligonucleotides containing wild type or mutant ATF and AP-1 binding sites were subsequently inserted between these two sites of the linker so that the ATF or AP1 sites were in the position originally occupied by the CYC1 UAS. (see oligonucleotide section for sequences). Details of the sites used are outlined in fig. 4.1.4.

## **5. GEL ELECTROPHORESIS;**

All radiographs exposures of radioactives gels which are shown in the thesis were generated using Kodak XOMAT film.

### **5.a. Loading buffers;**

**Denaturing;** 95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.5% xylene cyanol.

**Non denaturing;** 0.25% bromophenol blue, 0.25% xylene cyanol, and 30% glycerol in dDW.

**5.b. Agarose gels;** 0.6-1.0% 150mls agarose gels contained 1 x TBE , 5 $\mu$ g/ml ethidium bromide and varying amounts of either ultrapure agarose (BRL) or sea plaque agarose (low gelling temperature, FMC). They were poured into taped horizontal agarose gel trays and run in 1x TBE in horizontal gel tanks (BRL).

10X TBE buffer per liter: 108g tris base, 55g boric acid, 9.3g Na<sub>2</sub>EDTA

**5.c. Buffer Gradient Gels;** Buffer gradient gels were poured using 5X gel mix and 0.5X gel mix . The mixes were degassed before use. To a 7ml aliquot of 5X mix 12.5 $\mu$ l of 25% APS and 12.5 $\mu$ l of TEMED were added to initiate polymerisation, and

to a 40ml aliquot of 0.5X mix 75 $\mu$ l of APS and 75 $\mu$ l of TEMED were added. 6ml of 0.5X mix was taken up in a 10ml pipette followed by all of the 5X mix. 2-3 bubbles were taken up to form a rough gradient, and the mix was poured down the open top (20cm) edge of two taped 20cm by 50cm degreased glass plates separated with 0.35mm spacers. This was followed by the rest of the 0.5X mix, and then a 44 slot comb clamped in position with two bulldog clips. After being allowed to set for 30 mins to 1 hr the gel was run in a vertical apparatus with 1X TBE in top and bottom buffer tanks. After running, the gel was fixed in 10% glacial acetic acid for 15 mins while still attached to one of the gel plates, before being transferred to 3MM paper (Whatman), covered with saran wrap, and dried at 80 C on a flatbed vacuum drier. The saran wrap was removed before exposure with X-Ray film.

#### Buffers;

5X gel mix (30ml of 40% stock acrylamide (38:2 acrylamide : NN-methylene bisacrylamide in dDW), 100ml 10X TBE, 92g urea (B.R.L. ultrapure), ~10mg bromophenol blue, made up to 200ml with dDW)

0.5X gel mix (as above but no bromophenol blue, and 0.5X TBE final)

TBE; (10X, per liter): 108g Tris base, 55g boric acid, 9.3g Na<sub>2</sub>EDTA

Samples were heated to 90 C for 4 minutes in denaturing sample buffer just prior to loading

**5.d. Footprinting;** the procedure was essentially the same; a 40 ml total volume mixture containing 8mls 40% acrylamide mix (38:2 acrylamide : NN-methylene bisacrylamide in dDW), 20g urea (BRL) and 4 mls 10X TBE was prepared. Just prior to pouring 40 $\mu$ l TEMED and 250 $\mu$ l 10% AMPS was added.

**5.e. Gel shifts;** A 40ml mixture was prepared consisting of 7.2 mls acrylamide (44% acrylamide; 0.8% bis acrylamide) 2mls 10X TBE, 30.6 mls dDW, 40 $\mu$ l TEMED and 250 $\mu$ l 10% AMPS and poured into a 18cm x 20cm x 0.8mm gel apparatus. This was allowed to set for 1 hr after which samples were loaded in a vertical gel apparatus and gels run at room temperature in 0.5X TBE for 90 mins. Gels were then transferred to Whatman 3MM paper, dried under vacuum and exposed, generally overnight at -70 C.

**5.f. Protein gel electrophoresis;** Protein gel dimensions were typically 18cm x 20 cm x 0.8 mm. Gels consisted of an acrylamide plug, on top of which was poured a

30ml separating gel mix . Once this had set a stacking gel mix was poured on top. An equal volume of sample and 2X loading buffer were boiled for 2 minutes, loaded on to the gel and the gel run for 14 hrs in running buffer. The reagents for this protocol were kept sterile and all handling procedures used rubber gloves.

#### **5.f.i. Gel Mixes;**

Acrylamide plug; (2mls of 30% acrylamide: 0.8% bisacrylamide, 25 $\mu$ l AMPS, and 25 $\mu$ l 8.4% TEMED)

Separating gel mix; (15 mls of 2x separating buffer pH 8.8, 9 mls of 30%: 0.8% acrylamide:bis, 5.3mls of dDW, 150 $\mu$ l of 8.4% TEMED, 300 $\mu$ l of 10% SDS, and 250 $\mu$ l of 10% AMPS)

Stacking gel mix; 6mls of 2x stacking gel buffer 2mls of 30:0.8 acryl:bis 3.7 mls dDW, 120 $\mu$ l 10% SDS, 120 $\mu$ l 8.4% TEMED and 120 $\mu$ l 10% AMPS

running buffer (6g Tris, 28.8g glycine and 1g SDS made to 1litre)

Loading mix (Laemmli buffer); 2% SDS, 10% glycerol, 100mM DTT, 60mM Tris.HCL pH 6.8, and 0.001% bromophenol blue.

Samples were heated to 90 C for 5 minutes prior to loading.

## **6. BLOTTING**

Protein gels were blotted on to nitrocellulose using a Hoefer Scientific Instruments electronic transfer apparatus in 5 litres of transfer buffer (72.1g glycine, 15.0g Tris, 1 litre methanol (plus dDW to 5 litres). Gels were laid out onto Whatmann 3MM paper, and two sheets of nitrocellulose followed by a further two sheets of Whatmann 3MM paper layed on top. All paper was appropriately cut to size and presoaked in transfer buffer. Transfer at full power took around 100 mins after which the nitrocellulose was carefully removed for further treatment.

#### **6.a. Western Blot Analysis;**

The gel was run and blotted as described in the southwestern protocol. However once the protein had been transferred to nitrocellulose it was blocked overnight at 4 C or for 30 minutes at room temperature in TBS with 7% BSA. The filter was then washed for fifteen minutes in TBS after which 40 $\mu$ l of a Rabbit polyclonal E1A antibody preparation was added to 20mls of RIPA buffer and 1% BSA and incubated for 60 minutes. The antibody was then removed and the filter washed three times for 10 minutes each in fresh 50ml aliquots of RIPA buffer. 60 $\mu$ l of biotinylated antirabbit secondary antibody (Amersham) was then added in 25mls of TBS and incubated for 60 minutes at room temperature. The filter was then washed 3 times as above after which

60 ml of the streptavidin peroxidase complex (Amersham) was added in 25mls of TBS and incubated with the filter for 30 minutes at room temperature. The filter was subsequently washed again as above and the conjugate visualised with a hydrogen peroxide and 4-chloro-1-naphthol based substrate reaction as described in the Amersham protocol leaflet.

Buffers;

TBS; 50mM Tris.HCl pH 7.5, 200mM NaCl

RIPA buffer; 150 mM NaCl, 1% NP-40, 0.5% deoxycholate, 0.1% SDS, 50mM Tris.HCl, pH 8.0

**6.b. Colloidal gold staining;** the blot was then washed three times for 20 minutes each in 100mls of TBS and then for a further three times for 2 minutes each in 100mls of H<sub>2</sub>O. The blot was then covered with colloidal gold staining solution (Biorad) and gently agitated for 20 minutes. Protein bands then appeared as a deep red colour. The blot was then gently dried on whatman 3MM paper and carefully stored.

**6.c. Southwestern analysis;** This was carried out as described for probing of filters in the expression screen

## 7. DNA LABELLING

**7.a. Filling in a 5' overhang;** This labelling reaction was used for preparing DNA probes for footprinting and gel shift reactions. labelling oligonucleotides 100ng (1µl) of oligonucleotide, 1µl 10X super reverse transcriptase buffer (see primer extension protocol for details), 1µl cold dTTP (10mM), 1µl cold dGTP (10mM), 1.5µl α <sup>32</sup>P dATP and 1.5µl ap<sup>32</sup> dCTP, 3.5µl H<sub>2</sub>O and 40 units of super reverse transcriptase (Anglian) were mixed and incubated at 37 C for 30 mins. The reaction was then made up to 50µl with dDW and the unincorporated counts were separated on a G50 sephadex column.

Labelling probes for footprinting was carried out in essentially the same way. Plasmid DNA containing the sequences of interest was linearised by restriction enzyme digestion and the DNA phenol extracted and ethanol precipitated. This DNA fragment was then end labelled and the unincorporated counts separated on a G50 column. The plasmid was then cut with a second restriction enzyme and the reaction run on a 6% non denaturing polyacrylamide gel for 1-2hrs. The probe was then visualized by exposing a piece of XOMAT film on the gel for 1 minute. A piece of acrylamide

containing the DNA fragment was then excised and the DNA extracted by electroelution in dialysis tubing in 0.5X TBE for 1 hr. The electroeluted DNA was then ethanol precipitated as described for the preparation of single stranded oligonucleotides.

**7.b. Phosphorylation of 5' ends;** This procedure was carried out in the preparation of probes for use in  $\lambda$ gt11 expression screen and southwestern analysis of blotted protein gels. 200 ng of single stranded oligonucleotide was mixed with 2 $\mu$ l of 10X kinase/ligase buffer (0.5M Tris pH 7.5 and 0.1M MgCl<sub>2</sub>, 2 $\mu$ l 0.1M DTT, 1 $\mu$ l 10mM spermidine, 3 $\mu$ l dDW, 30 units (3 $\mu$ l) T4 polynucleotide kinase and 5 $\mu$ l  $\gamma$ 32P ATP. This was incubated at 37 C for 1 hr after which 1 $\mu$ l of cold 10mM ATP was added and incubated a further 10 minutes. Unincorporated counts were then separated using a G50 sephadex column. The 2 complementary strands were then annealed as previously described, made up to 60  $\mu$ l with 6  $\mu$ l 0.1M DTT, 3 $\mu$ l 10mM spermidine, 6 $\mu$ l 10X kinase/ligase buffer 4 $\mu$ l dDW and 1 $\mu$ l T4 DNA ligase. This was then incubated overnight at 15 C.

**7.c. Random priming;** for this method a 'Random primed DNA labelling kit' (Boehringer) was used. DNA excised in a low melt agarose gel fragment was heated to 95 C for 10 minutes and then cooled to 37 C for 5 minutes. Approximately 50ng of this was mixed with 3 $\mu$ l of dATP, dGTP, dGTP mixture (each one at 16mM), 2 $\mu$ l reaction mixture, 1 $\mu$ l klenow (2Units) and made up to 20 $\mu$ l with dDW. This was incubated for 45 minutes at 37 C, heated at 65 C for 5 minutes and passed over a G-50 column to separate unincorporated counts.

**7.d. Preparation of G-50 sephadex column;** 5g of Sephadex G-50 (Pharmacia) was added to 40mls of dDW and left to stand overnight to let the beads hydrate. The hydrated beads were then autoclaved. A 1ml syringe (sabre) was plugged with polymer wool (Interpet) and filled with hydrated G-50. The column was placed in a 15ml falcon<sup>tube</sup> and spun in a benchtop centrifuge at 1.5 K for 4 minutes. Fluid in the falcon was discarded and an eppendorf tube placed at the bottom of the falcon. The syringe was replaced and the labelling reaction placed into the syringe. This was then spun at 1.5K for 4 minutes. The labelled DNA (now in the eppendorf) was stored at -20C until required.

## 8. STRAINS;

### 8.a. Bacterial;

**HB101;** F<sup>-</sup>, hsdS20 (r<sub>B</sub><sup>-</sup>, m<sub>B</sub><sup>-</sup>), recA13, ara-14, proA2, lacY1, galK2, rpsL20 (Sm<sup>r</sup>), xyl-5, mtl-1, supE44, λ<sup>-</sup> (Boliva and Backman 1979)

**TG1;** K12, Δ(lac-pro), supE, thi, hsdD5/F' traD36, proA+B+, lacq, lacZΔM15 (Gibson 1984)

**Y1089;** ΔlacU169 proA<sup>+</sup> Δ lon araD139 strA hflA150 [chr Tn10] (pmc9)

**Y1090;** ΔlacU169 proA<sup>+</sup> Δ lon araD139 strA supF [trpC22 Tn10] (pmc9)

### 8.b. Budding Yeast;

**BJ21168** (leu2, trp1, ura3-52, prb1-1122, pep4-3, prc1-407, gal2). This strain was used for extract preparations. It is mutated at three protease loci; PEP4, PRC1, and PRB1. (Sorger and Pelham 1987)

**W303** (ade2-1, trp1, can1-100, leu2-3, -112, his-11, 14, ura3; Shore and Naysmyth 1987). This was used for all β-galactosidase assays.

### 8.c. Fission Yeast;

The parent strain for all transformation experiments was donated by V.Simanis (leu1-32, his3-237, ura4-d18).

**E2C1**, and **E2C2** are two individual strains which have both been stably transformed with the plasmid pU4E2CAT.

**E2C113a**, and **E2C112a** are the above **E2C1** strain but additionally transformed stably with pSM113s and pSM112s respectively.

## 9. MEDIA;

### 9.a. Bacterial;

**LB** (Luria-Bertani) medium; (per litre)

10g Bacto-tryptone, 5g Bacto-yeast extract (Difco), and 10g NaCl. The pH is adjusted to 7.5 with sodium hydroxide. For plates 15g of Bacto Agar was also added.

For selection purposes ampicillin (Sigma) was added at 100 μg/ml.

### 9.b. Budding yeast;

Recipes are given for liquid medium, to which supplements and glucose are added before use. For plates, 22g Bacto Agar (Difco) was added and media was always autoclaved. This was cooled to 50 C-60 C before adding 0.1 volume of 20% glucose and any other supplements, stirring and then pouring.

**Rich Medium (YEPD);** 11g Yeast extract (Difco), 22g Bacto Peptone (Difco), 55mg adenine sulphate, dDW to 1 litre, autoclave. 0.1 volume of 20% glucose was added before use.

**Minimal Medium (SD):** 8g Difco nitrogen base (without amino acids), dDW to 1 litre. 0.1 volume of 20% glucose was added before use.

**Selective Media:** As minimal medium, and all contain 55mg / litre each of adenine sulphate, uracil, and tyrosine, unless stated otherwise.

**-URA:** no uracil, plus 11g/ litre casamino-acids (Vitamin assay : Difco), supplemented with 10ml/litre 0.5% tryptophan and 20ml/litre 0.5% leucine.

**Top Agar :** per litre, 200g sorbitol, 8g yeast nitrogen base (Difco), 55mg Adenine sulphate, 55mg tyrosine sulphate, 33g agar (Difco). After autoclaving in a 2.5 litre flask and cooling to 50-60 C, 100ml of 20% Glucose was added. After mixing, the agar was dispensed into 100ml aliquots for storage at room temperature. When used, the agar was melted, held at ~65 C, and the appropriate supplements were added.

**Supplements :** For Ura transformation, 6.7ml of a 20% Casamino acids (Difco) 0.4% leucine solution and 0.25% tryptophan.

Cultures were grown at 30 C unless stated, either on solid medium or shaking in liquid. The generation time of wildtype yeast at 30 C in liquid medium is about 90 minutes. At 25 C it is about 150 minutes.

### 9.c. Fission yeast;

**Salt stock (x50)** 1 litre; 21.4g MgCl<sub>2</sub>, 0.74g CaCl<sub>2</sub>, 50g KCl, 2g Na<sub>2</sub>SO<sub>4</sub>.

**Vitamin stock (x 1000)** 1 litre; Na Pantothenate 1g, Nicotinic acid 10g, Inositol 10g, Biotin 10mg

**Minerals (x 10000)** 1 litre; 5g H<sub>3</sub>BO<sub>3</sub>, 4g MnSO<sub>4</sub>, 4g ZnSO<sub>4</sub>, 2g FeCl<sub>3</sub>, 0.4g H<sub>2</sub>MoO<sub>4</sub>

Chemicals obtained from BDH and FSA.

**Minimal media (1 litre)** KH pthallate 3g, Na<sub>2</sub>HPO<sub>4</sub> 2.2g, NH<sub>4</sub>Cl 5g, Glucose 20g, Salts stock 20ml, vitamins stock 1ml, Mineral stock 0.1ml.



Minimal media (plates) 1 litre; as above except additionally; agar 20g, 1M NaOH 1ml

Minimal Variations; 1 litre

+ 1.2M sorbitol; as above + 218.6g sorbitol

Low glucose; as above but 5g glucose

**Supplements;** Stock solutions of adenine, histidine and leucine were made up at 7.5g/litre and 12mls of each were added to minimal medium when required. Uracil was made up at 3.75g/litre and 24 mls were added to 1 litre of minimal medium when required.

## 10. TRANSFORMATIONS

**10.a. S.cerevisiae.** A 50ml yeast culture was grown overnight in YEP-glucose and the cells harvested in a bench top centrifuge at 2k for 5 minutes when the culture reached an OD<sub>600</sub> of 0.5-1.0. Cells were washed in 20mls dDW and 20mls 1M sorbitol before being resuspended in 20mls SCEM and 1000U lyticase (sigma) and incubated 30 minutes at 30 C. Spheroplast formation was monitored by seeing if 100µl of suspension cleared when a drop of 1% SDS was added. Following spheroplast formation the tube was centrifuged 3-4 minutes at 1K. The spheroplasts were gently washed in 20mls STC, repelleted and resuspended in 2mls of STC buffer. 100µl aliquots of this was mixed with 5µg of carrier salmon sperm DNA and 1µg of plasmid DNA. These were then left at room temperature for 10 minutes when 1ml PEG buffer was added and left for a further 10 minutes and pelleted. The pellet was resuspended in 150µl SOS and relevant supplements and left at 30 minutes at 30 C. 8mls of TOP (1M sorbitol, 2.55 agar in SD medium) was added, the tube inverted quickly to mix the contents, and then poured out onto SORB plates (SD plates containing 0.9M sorbitol and 3% glucose) and incubated at 30 C for 3-4 days.

Buffers and media;

SCEM; 1M sorbitol, 0.1M sodium citrate pH 5.8, 10mM EDTA, 30mM β-mercaptoethanol

PEG buffer; 10mM Tris pH 7.5, 10mM CaCl<sub>2</sub>, 20% polyethylene glycol 800

STC buffer; 1M sorbitol, 10mM Tris pH 7.5, 10mM CaCl<sub>2</sub>

SOS buffer; 1M sorbitol, 6.5 mM CaCl<sub>2</sub>, 0.25% yeast extract, 0.5% bactopectone.

**10.b. S.pombe;** A 200ml culture of S.pombe was grown to an OD<sub>600</sub> of 0.3-0.7 in minimal medium supplemented with 0.5% glucose. The cells were pelleted as above and washed in 10mls CPE buffer and then digested for 15-60 minutes at 30 C in 5mls CPS buffer and 5mg/ml Novozyme 234. Spheroplast formation was monitored under

the microscope and as soon as the cells started to round up the cells were washed in 40mls of TS buffer and split into two 50 ml falcon tubes. Washing was repeated 3 times (spheroplasts were centrifuged at 2K for 10 mins each time) after which cells were resuspended in TSC buffer at  $2-5 \times 10^8$  spheroplasts/ml. The two tubes of cells were combined and a 100 $\mu$ l aliquot of cells was taken and added to 3-5 $\mu$ g of plasmid DNA. This was incubated at room temperature for 15 minutes after which 1ml of PEG buffer was added and the spheroplasts incubated for a further 15 minutes. The cells were then centrifuged at 2 K for 10 mins, and resuspended in 0.25mls of YE buffer. This was incubated for 1hr and plated onto minimal with 1.2M sorbitol plates with appropriate supplements. The plates were incubated for 4-5 days after which colonies could be picked.

Buffers;

CPE buffer; 20mM Citrate-Phosphate buffer pH 5.6, 40mM EDTA, 10 $\mu$ l  $\beta$ -mercaptoethanol

CPS buffer; 20mM Citrate-Phosphate buffer pH 5.6, 1.2M sorbitol, 10 $\mu$ l  $\beta$ -mercaptoethanol.

TS buffer; 1.2M sorbitol/10mM Tris.HCl pH 7.6

TSC buffer; 10mM Tris pH 7.5, 10mM CaCl<sub>2</sub>, 1.2M sorbitol

PEG buffer; 10mM Tris pH 7.5, 10mM CaCl<sub>2</sub>, 20% Poly ethylene glycol 4000

YE buffer; 1.2M sorbitol, 10mM Tris pH 7.5, 10mM CaCl<sub>2</sub>, 0.5mg/ml yeast extract and 5 $\mu$ g/ml supplements

**10.c. Transformation of *E. coli*;** Preparation of Competent Cells; An overnight culture grown in LB was diluted 100 fold into fresh medium and grown shaking at 37 C with good aeration until the OD<sub>550</sub> was 0.3 to 0.5. Cells were harvested at 3K for 10 minutes at 4 C, and gently resuspended in one half of the original culture volume of ice-cold 50mM CaCl<sub>2</sub>/ 10mM Tris.HCl pH 7.4. After spinning as before, cells were resuspended in 1/15 of the original culture volume of CaCl<sub>2</sub>/ Tris.HCL pH 7.4 and either used immediately or glycerol was added to 15% (v/v) and 400 $\mu$ l aliquots divided into microfuge tubes, frozen on dry ice, and stored at -70 C. 100 $\mu$ l of these cells were used per transformation and gave around  $10^5$  transformants per  $\mu$ g of DNA.

Plasmid Transformation - Frozen Competant cells were thawed on ice and used immediately as if fresh cells. 150 $\mu$ l of cells were added to 1-25 $\mu$ l of DNA in a tube pre-chilled on ice, mixed and left for 10 to 50 minutes, and transferred to a 42 C waterbath for 2 minutes. 200 $\mu$ l of LB was added and incubation was continued for a further 30 minutes. The entire contents of the tube was spread on an appropriate selective plate, and after spreading and drying, incubated inverted overnight at 37 C.

## 11. FISSION YEAST EXTRACTS.

A 1 litre culture of the S.pombe strain Ade 704 h+ was grown to a density of  $8 \times 10^6$  cells/ml ( $OD_{600}$  0.6) in minimal media containing 200mg/l adenine, histidine, leucine and uracil and 0.5% glucose. Yeast spheroplast were then prepared as already described except that Novozym 234 was removed by repeated washing in ice cold 1.2M sorbitol, 10mM Tris pH 7.6 and 1mM PMSF. Nuclei were then prepared from spheroplasts essentially as described by Wiederecht et al. (1987). The washed spheroplast pellet was resuspended in 10mls of buffer A plus 1mM PMSF and left on ice for 20 minutes. The suspension was then homogenised with 10 strokes in a dounce homogeniser. The homogenate was then spun down at 18000g for 10 minutes and the pellet resuspended in 5mls of buffer A + 1mM PMSF with ammonium sulphate at 33mg/ml. The extract was stirred on ice for 15 minutes and then centrifuged at 42K for 2 hrs in an SS55 rotor. Ammonium sulphate was then added at a final concentration of 0.31g/ml to the supernatant from these tubes and the salt mixed in for 20 minutes on ice. This was then centrifuged at 18000g for 45 minutes and the pellet resuspended in 2mls of E buffer. The extract was dialysed against 400mls of the same buffer overnight and then stored at -70 C.

Buffers;

TSP buffer; 1.2M sorbitol, 10mM Tris pH 7.6 and 1mM PMSF

buffer A; 10mM HEPES pH 7.6, 15 mM KCL, MgCl<sub>2</sub>, 0.1 mM EDTA

E buffer; 25mM HEPES pH 7.6, 10% v/v glycerol, 50mM KCL, 0.1mM EDTA, 0.2% Triton x-100 and 1mM PMSF

## 12. PURIFICATION OF BUDDING YEAST AP1 AND ATF.

**12.a. General:** Preparation of large scale yeast extracts and purification of those extracts over heparin agarose and DNA sepharose columns was carried out in collaboration with Dr. Peter Sorger and Dr. David Shore at the Medical Research Council Laboratories in Cambridge. Yeast cultures were grown by Mr.C.Young at the National Institute for Medical Research, Mill Hill, London.

**12.b. Cell culture and breakage;** Cultures of multiply protease deficient Saccharomyces cerevisiae strain BJ2168 (leu2 try1 ura3-52 prb1-1122 pep4-3 prc1-407 gal 2) were grown to late log phase ( $1 \times 10^8$  cells/ml) in YEPD medium in a 400 litre fermenter. 400 litre cultures were harvested in a Sharples centrifuge at the NIMR. The pelleted paste was packed into 50ml syringes and forced out with a mastic gun directly into liquid nitrogen creating a kind of spaghetti. This was then

stored at  $-70\text{ C}$ . Approximately 0.5kg of this frozen yeast and 500ml of liquid nitrogen were loaded into a 5.5 litre blender vessel previously cooled with liquid nitrogen and blended for five minutes to create a yeast powder. Approximately 90% of yeast cells broke open during this process (estimated from microscopic analysis). 1kg batches of this yeast powder was thawed by addition of 1.5 litres of breakage buffer (at room temperature). This was then extracted for 30 minutes in 0.4M ammonium sulphate and centrifuged for 1hr at 45000 rpm in a beckman type 45Ti rotor. The supernatant was precipitated with 0.35 mg/ml ammonium sulphate.

**12.c. Protein purification;** The protein pellet obtained from the 400 litre pellet was resuspended in 3.5 litres of A50 buffer and dialysed overnight against the same buffer. Two litres of dialysed crude extract was loaded overnight onto a 1.5 litre heparin agarose column at a flow rate of 60 mls/hr, washed with 3 litres of A50 and then eluted with a 4 litre linear gradient of A50 to A600 at a flow rate of 180 mls/hr. DNA binding activity was monitored by employing the gel shift assay using specific AP1 or ATF oligonucleotide binding site probes. Specific activity recognising the ATF binding site was obtained in fractions eluted from the first column passage but activity recognising the AP1 oligonucleotide binding site was predominantly observed in the flow through fractions from the first column passage. Therefore 1.5 litres of flow through extract was applied to the column for a second time. Subsequent washing and elution showed that a significant amount of AP1 activity was retained on the column on this passage.

500 mls of pooled ATF peak was precipitated with ammonium sulphate at 0.35 mg/ml, pelleted and redissolved in 100mls of A75. This was then dialysed against A75 and then loaded onto a 50ml DNA sepharose column at a flow rate of 25 ml/hr. The column was washed with 100 mls of A75 buffer and 10 ml fractions eluted with a 400ml linear gradient of A75 to A600 at a flow rate of 30 mls/hr and assayed by gel shift experiments. Peak fractions obtained from this column were then applied to oligonucleotide affinity columns at a rate of 5 mls/hr. These columns were then washed with 20 mls of A100. Step wise elution washes were then carried out using buffers containing KCl concentrations varying from 100mM to 2M. Fractions derived from these columns were then also tested using gel shifts.

**12.d. Preparation of columns:**

**12.d.i. Heparin Agarose:** This was prepared at the MRC Cambridge by Peter Sorger, David Shore and Gustav Ammerer essentially according to the method of Davison et al. (1979). These people had used the column to study proteins unrelated to ATF and AP1. Between each loading of the column two litres of A50N2000 was run through the column in order to regenerate the resin. The column was then equilibrated with A50 prior to actual loading of material.

**12.d.ii. DNA sepharose:** This column was again prepared by the above workers by coupling sheared DNA to cyanogen bromide (CNBr)-activated Sepharose CL-4B using essentially the same method as described for the affinity column.

**12.d.iii. Oligonucleotide affinity columns:** Affinity columns were <sup>structurally</sup> contacted by coupling multimers of either an AP1 site or an ATF site to CNBr-activated Sepharose CL-4B in a manner very similar to that used by Kadadonga and Tjian (1986).

Preparation of single stranded oligonucleotides:

About 1 mg of each synthetic oligo was purified on preparative 12% denaturing acrylamide gels. The DNA was visualised and the band was excised. DNA was then eluted, precipitated and redissolved in 100 $\mu$ l of TE.

Annealing of single stranded oligonucleotides:

50 $\mu$ l of each complementary oligonucleotide were put into a single tube with 20 $\mu$ l of 10x PNK buffer and 30 $\mu$ l dDW. This mixture was then heated to 88 C for 2 minutes, 65 C for 10 minutes, 37 C for 10 minutes and room temperature for five minutes.

Phosphorylation of the 5' ends:

Approximately 10 ng of ds oligo (in 1 $\mu$ l) was added to 10 $\mu$ Ci  $\gamma$ ATP, 1 $\mu$ l 10x PNK buffer (500mM Tris.HCl pH7.5, 100mM MgCl<sub>2</sub>, 50mM DTT, 10mM Spermidine, and 10mM EDTA), 7.5  $\mu$ l dDW and 5 U PNK. This was incubated 37 C for 30 minutes and the free nucleotides separated from those incorporated on a G50 column. The labelled oligonucleotides (in 20 $\mu$ l of 1xPNK buffer) were then added back to the unlabelled pool which were all phosphorylated by adding 15 $\mu$ l 20mM ATP pH 7.5 and 100U of PNK. This reaction was left for 2 hrs at 37 C after which 37 $\mu$ l of 3M sodium acetate, 2.5 $\mu$ l 1M magnesium chloride and 100 $\mu$ l of dDW were added, the tube heated to 65 C for 15 minutes, and 0.75 mls ethanol added and the DNA precipitated. The pellet was then washed twice in 70% ethanol.

**Ligation:**

To each pellet was added 62 $\mu$ l dDW, 10 $\mu$ l 10x linker kinase buffer (700mM Tris.HCl pH 7.5, 100mM MgCl<sub>2</sub>, and 50mM DTT), 20 $\mu$ l 20 mM ATP and 2000 U T4 DNA ligase (BRL). This was incubated overnight at 15 C. 1 $\mu$ l of the ligation reaction was then run on an agarose gel to check the efficiency of the ligation. Typically I achieved products of 200-500 bp in length.

**Coupling DNA to CNBr-activated Sepharose CL-4B;**

Each ligation reaction was phenol/chloroform/isamyl alcohol extracted; 33 $\mu$ l 10M ammonium acetate and 133 $\mu$ l of isopropanol was then added to precipitate the DNA. The pellet was washed twice in 70% ethanol and resuspended in dDW. 10 mls of resin was washed in 300mls ice cold dDW and 100mls 10mM K-phosphate (pH 8.0) in a sintered glass funnel and transferred to a 15 ml screw capped plastic tube. To this was added 4 mls of 10mM K-phosphate (pH 8.0) and the ligated oligonucleotides. This mixture was rotated overnight at room temperature. This was then washed with 200 mls dDW and 100 mls 1M ethanolamine (Aldrich, pH 8.9) and transferred back to a 15 ml tube with 4 mls of 1M Ethanolamine. The mixture was incubated for a further 4 hrs and then washed successively with 100mls 10mM K-phosphate (pH 8.0), 100mls 1M K-phosphate (pH 8.0), 100mls 1M KCL, 100 mls dDW and 100mls of column storage buffer.

**12.e. Buffers:**

**Breakage buffer;** 200mM Tris- HCL pH 8.0 (sigma), 10mM MgCl<sub>2</sub> (BDH), 10% glycerol (FSA), 1mM PMSF (sigma), 500 $\mu$ M TPCK (sigma), 25 $\mu$ M TLCK (sigma) and 2 $\mu$ g/ml Pepstatin A.

**A50;** 50mM Tris-HCl, 1mM EDTA, 50mM (NH<sub>4</sub>)<sub>2</sub> S<sub>0</sub>4, 10% (v/v) glycerol, 1mM dithiotreitol (DTT), pH 8.0, protease inhibitor stock solutions each added at a dilution of 1:1000.

**A75 and A600;** Same as above except (NH<sub>4</sub>)<sub>2</sub> S<sub>0</sub>4 was added at 75mM and 600mM respectively.

Buffers used to elute proteins from the affinity column were A50 based but supplemented with NaCl to bring the final salt concentration upto 100mM, 150mM,

200mM, 300mM, 400mM, 500mM, 600mM, 800mM, 1M, or 2M according to the exact buffer being made.

**A50N2000;** Same as A50 except it is supplemented with NaCl to a final concentration of 2M.

**Protease inhibitor stock solutions;** 100mM phenylmethanesulphonyl fluoride (PMSF) in methanol; 50mM N-tosyl-L-phenylalanine chloromethyl ketone (TPCK), 5mM N $_{\alpha}$ -p-tosyl-L-lysine chloromethyl ketone (TLCK) in methanol; 2mg/ml pepstatin A in 1;1 (v/v) methanol; water. All obtained from Sigma.

**12.f. Small scale budding yeast preparations;** The pellet from a 50ml culture was washed in 25mls of breakage buffer recentrifuged and resuspended in 0.4mls of breakage buffer and this paste was introduced as droplets into a mortar containing liquid nitrogen. Once the nitrogen evaporated the frozen drops were ground typically around 80 strokes with the pestle which led to approximately 80% cell breakage (as observed under the microscope).

The paste containing the broken cells was thawed and ammonium sulphate added to a final concentration of 0.35mg/ml. This was then left on ice for 30 minutes and spun in a microfuge at 4 C for 20 minutes. The supernatant was removed and then either dialysed against 100mM binding buffer or used immediately for gel shift assays.

**Hela cell extracts** were a gift from Helen Hurst and were prepared as described in Hurst and Jones (1987)

### **13. SCREENING A $\lambda$ GT11 EXPRESSION LIBRARY WITH A BINDING SITE PROBE.**

The methods used in this protocol are largely based on previously published methods (Singh et al 1988, Landschulz et al 1988).

**13.a. Plating;** The  $\lambda$ gt11 expression library used in this study was obtained from Clontech; it contained S.pombe (strain Sp 223) genomic inserts, which had an average size of 3.0 kb, with EcoR1 linkers at their ends.

For infection 10 mls of LB supplemented with 0.2% maltose and 100 $\mu$ g/ml ampicillin was inoculated with E.coli strain Y1090 and grown overnight at 37 C. To each of 4 individual 100 $\mu$ l aliquots of the Y1090 cells was added 3-5x 10<sup>4</sup> pfu of the lamda library. The phage was allowed to infect the bacteria for 15 minutes at 37 C. 9mls of top agarose (prepared as phage broth but with 7g of agarose added per litre) was then added (at 47C) to each infection and the mixtures were quickly plated on to 15cm LB/ampicillin plates (prewarmed to 37 C). The 4 plates were then incubated at 42 C for approximately 4 hrs until plaques became visible. Nitrocellulose filters, soaked in 10mM IPTG (to induce expression of the fusion gene) and then dried briefly on Whatmann 3MM paper, were then laid on to each plate and the plates incubated for a further 5 hrs at 37 C. The plates were then cooled for 10 minutes at 4 C, the filters carefully peeled off and immersed in blocking buffer (face up) . A new filter was then applied to each plate and the plates were then incubated for a further 2 hrs. After this period these filters were also removed and placed in blocking buffer. Once in blocking buffer the filters were agitated for 1 hr at room temperature and then transferred into N50 buffer, washed briefly and then placed again into fresh N50 buffer. Filters were stored overnight in this form.

**13.b. Probing;** The four filters were then transferred (face up) into a 15 cm petri dish containing 50 mls of N50 buffer supplemented with 10 $\mu$ g/ml of poly (dI; dC) and 5 x 10<sup>7</sup> cpm of both ATF and AP1 binding site probe. Filters were gently agitated for 1 hr at room temperature. They were then washed four times over a total period of 30 minutes in N50 buffer, dried on Whatman paper and exposed overnight at -70 C with an intensifying screen.

**13.c. Selection and amplification of positive clones;** Only 4 positives appeared which could be observed on both original and replica filters. These were picked using the wide end of a pasteur pipette after aligning the autoradiographs with the LB plates. Plugs were put into SM buffer containing a few drops of chloroform and soaked for atleast 4 hrs. It was estimated that plugs from the primary plates contained around 10-20 plaques (10<sup>7</sup>-10<sup>8</sup> pfu). Secondary infections were carried out on 9 cm plates. Three dilutions of the primary positive were made to ensure a suitable number of plaques could be screened on the second round. The procedure of infection and



screening was the same as the first except that no replica filters were screened. Positives were selected and replated for a tertiary infection at a density of around 100-200 pfu per plate to allow isolation of individual plaques with the fine end of a pasteur pipette. On this screen, filters were tested with probes consisting of either ATF or AP-1 sites and not a mixture. Single plaques were picked and lysate stocks prepared.

**13.d. Preparation of high titre virus stocks;** High titre stocks were prepared by plating out purified virus on 9 cm plates to cause complete lysis. Once lysis was observed 5 mls of SM buffer was placed on the plates and left overnight. The SM buffer was then harvested and titred.

**13.e. Isolation of recombinant phage lysogens;** To obtain recombinant protein for gel shift analysis I decided to prepare extracts from recombinant phage lysogens. This was carried out essentially as Singh et al (1989). 5 mls of E.coli strain Y1089 was grown overnight at 37 C in LB supplemented with 0.2% maltose and 100µg/ml ampicilin. 50 µl of the saturated culture was diluted into 5 mls of LB with 10mM MgCl<sub>2</sub> and 100 µl of this diluted culture was mixed with approximately 10<sup>7</sup>-10<sup>8</sup> pfu (5µl) of purified phage and incubated 20 minutes at 32 C. This was then added to 10 mls of LB and 10-100 µl aliquots were plated out on 9 cm LB, 100mM ampicilin plates which were incubated overnight at 32 C. Colonies arising were checked for the lysogenic state by replica plating onto 100µg/ml ampicilin LB plates grown at 32 C and 42C. The true lysogens grow at 32 C but not 42 C.

Preparation of crude cell extracts from recombinant phage lysogens. 5 ml overnight cultures of recombinant phage lysogens were grown overnight in 100µg/ml ampicilin LB at 32 C. 20 µl of this culture was mixed with 2 mls of LB/ ampicilin medium and grown at 32 C until the OD 600 reached 0.5. 22 µl of 1M IPTG was then added to the culture and the grown at 37 C for 1hr. 1ml of this culture was spun down in a microfuge for 1 minute and the pellet resuspended in 100ml of extract buffer . Cell suspensions were freeze thawed on dry ice and incubated with lysozyme for 15 minutes on ice. Tubes were then spun at 4 C for 30 minutes, the supernatants dialysed against extract buffer and extracts stored at -70 C.

Preparation of phage DNA; Phage DNA derived from A, B, and GC isolates was prepared by N.C.Jones essentially as described in Maniatis et al (1982).

### **13.f. Buffers;**

Extract buffer; 50mM Tris, pH 7.5, 1mM EDTA, 1mM DTT and 1mM PMSF

SM buffer; 100mM NaCl, 8mM MgSO<sub>4</sub>, 1M Tris-HCl (pH 7.5) and 2% gelatin.

N50 buffer; 50mM Tris-HCL Ph 7.5, 50mM NaCl, 1mM EDTA, 5mM MgCl<sub>2</sub>, 1mM DTT.

Blocking buffer; As above but with 5% Marvel (dried milk)

#### 14. $\beta$ -Galactosidase assays;

These were carried out essentially as Harshman et al (1988). Yeast cultures derived from a single colony were grown to an OD<sub>600</sub> of approximately 1.0 and then harvested by centrifugation at 2K for 5 min in a bench top centrifuge. The pellet was washed in 5ml of water, recentrifuged and resuspended in 0.1 M Tris.HCL (pH 7.5), and 0.05% Triton-X-100 approximately 50 OD<sub>600</sub>/ml. Cells were permeabilised by freeze thawing on dry ice at -70 C. The reaction start point initiated when 100 $\mu$ l of this cell extract was added to 400 $\mu$ l of Z buffer and 100  $\mu$ l of 4mg/ml o-nitro-phenyl- $\beta$ -D-galactosidase (sigma). Once a noticeable yellow colour had developed, the reaction was stopped by addition of 0.25ml of 1M Na<sub>2</sub>CO<sub>3</sub> and the time of the reaction noted. The samples were placed on ice until all the reactions were completed, at which point the cells were removed by a 5 min centrifugation at 2.5K and the A<sub>420</sub> determined.  $\beta$ -galactosidase units were calculated as Hope and Struhl (1986) ie 1000x OD<sub>420</sub> / assay time X volume of cells used X OD<sub>600</sub> of cells used.

Z buffer (500mls); 8.05g Na<sub>2</sub>HPO<sub>4</sub>, 2.75g NaH<sub>2</sub>PO<sub>4</sub>, 0.375g KCl, 0.123g MgSO<sub>4</sub>, 1.35 ml  $\beta$ -mercaptoethanol.

#### 15. Primer extensions;

All work involving quantification of mRNA was carried out with great care and gloved hands and was largely based on the methods of Jones et al (1985). An oligonucleotide (5' GTGACCTATATGGTGGCA-3') complementary to the CAT coding sequence (3'-CACTGGATATACACCGT-5') was 5' end labeled with  $\gamma$ -<sup>32</sup>ATP using T4 polynucleotide kinase (Boehringer). 4 ng of labelled oligonucleotide was annealed with 6-25  $\mu$ g of total RNA in 10ml of 10mM Tris.HCL (pH 8.0), 250mM KCl and 1mM EDTA for 1 hr at 50 C. Extension was initiated by addition of 25m 20mM Tris.HCL (pH 8.7), 10mM MgCl<sub>2</sub>, 5mM DTT, 0.25mM of each dNTP and 10U of super RT (Anglian). After 1hr at 37 C the reactions were ethanol precipitated, resuspended in loading dye, and analysed on a 6% acrylamide sequencing gel.

#### 16. CAT assays;

CAT assays were carried out essentially as Gorman et al (1982). A 20 ml fission yeast culture was typically grown overnight to an OD<sub>595</sub> of 0.3-0.5. The cells were then spun down for 5 minutes at 2.5k in a bench top centrifuge and resuspended in 50ml 250mM Tris.HCl (pH 7.5) , 1mM PMSF. An equal volume of acid washed glass beads was then added and the tube vortexed at top speed for 1 minute. The supernatant from this was removed and the beads washed two further times with 250  $\mu$ l of 250mM Tris.HCl (pH 7.5) , 1mM PMSF, and the washes were pooled with the original lysates. 10  $\mu$ l (3-200 $\mu$ g of protein) of lysate was then added to 85  $\mu$ l of ddW, 3  $\mu$ l of 40mM Acetyl Coenzyme A (Sigma) and 2 $\mu$ l <sup>14</sup>C Chloramphenicol (0.25 $\mu$ Ci). The reaction was

incubated at 37 C for 10-30 minutes and then extracted with ethyl acetate. The organic upper phase, containing the chloramphenicol was transferred to a new <sup>Epp</sup>endorf, dried down, resuspended in 20 µl ethyl acetate and spotted on to silica gel thin layer chromatography (TLC) plates. The plates were run in 95%; 5% (v/v) chloroform; methanol and autoradiographed overnight.

## **17. DNase 1 footprinting;**

**17.a. Binding reaction;** The exact conditions for DNase footprinting were determined empirically for each probe and each extract examined. Binding reactions were carried set up in a final volume of 45µl as follows. Using crude fission yeast or Hela cell extracts approximately 200µg (20µl) of protein was added to 0.8ng of DNA probe, 2µg (2µl) of poly (dI: dC) nonspecific competitor and 22 µl DNase buffer (20mM HEPES-KOH (pH 7.9), 50mM KCL, 20% glycerol, 2mM DTT and 2mM MgCl<sub>2</sub>). With DNA sepharose purified budding yeast ATF activity 15µg of protein was used in such reactions. Incubations containing no protein were also prepared. After 20 minutes at room temperature, 5µl of DNase 1 ( Amersham), diluted to 5µg/ml in DNase buffer, was added for 30-60 seconds. For 'DNA only' incubations, a similar reaction was carried out but the DNase was diluted a further 100 fold. The reactions were stopped by the addition of 50 µl proteinase K stop buffer ( 100mM Tris.HCL (pH 7.5), 2% SDS, 20 mM EDTA, 400µg/ml proteinase K) and incubation at 37 C for 10 mins. Samples were extracted 2 times with phenol mix and precipitated in ethanol with tRNA 10mM MgCl<sub>2</sub>. Pellets were washed once in 70% ethanol and dried in a speed vacuum. The dried pellets were cherenkov counted resuspended in sequencing loading buffer and equal counts (as determined by cherenkov counting) loaded on a standard 50% urea, 10% acrylamide sequencing gel together with the appropriate A+G marker lane.

**17.b. A+G marker lane;** approximately 4ng of probe (5µl), 4µl dDW, 1µg salmon sperm DNA (Sigma) and 3µl 10% formic acid were incubated between 13 and 18 minutes. The reaction was stopped by adding 30µl 3M NaOAc, 0.6µl 0.5M EDTA, 1µl of 10mg/ml tRNA and 255µl dDW mixing and then ethanol precipitating. The pellet was washed in 70% ethanol and resuspended in a fresh dilution of 10 % piperidine (BDH). This was then incubated at 90 C for 25 minutes after which was added 60 µl 3M NaOAc, 440 µl dDW. This was mixed and then further added was 6µg tRNA and this was precipitated with 600µl of isopropanol. The pellet was washed in 70 % ethanol, dried, cherenkov counted and resuspended in loading dye at 5000 cpm/µl.

## 18. Gel Shift Assays.

**18.a. General;** Methods are largely based on those described by Fried and Crothers (1981). Typically binding reactions between extracts and endlabelled probes were carried out in a final volume of 25 $\mu$ l. A typical reaction contained 1 $\mu$ l (1-30 $\mu$ g) of extract, 0.1ml (0.2ng) of oligonucleotide probe, 2 $\mu$ l of poly dI:dC (supplied by Pharmacia and made up at 1mg/ml) and 21.9  $\mu$ l gel retardation buffer (20 mM Tris.HCL pH 7.5, 100mM NaCl, 2mM EDTA, 5mM MgCl<sub>2</sub>, 7mM  $\beta$ -mercaptoethanol, 1mM PMSF, and 10% glycerol). Variations in the amount of extract used depended predominantly on its purity. The probe was always added last and the reactions were immediately mixed. Incubations were carried out at room temperature for 20 minutes and then loaded onto a 8% polyacrylamide gel and run at 200V for 90 minutes in 0.5x TBE.

**18.b. Specific competition;** in certain gel shift assays the specificity of the complex was tested by adding unlabelled oligonucleotides to the binding reactions at various molar excesses (see results section for details)

**18.c. Phosphatase reactions;** The following amounts are relevant for reactions which eventually became loaded into a single lane of a gel. Phosphatase reactions were carried out prior to the addition of DNA. During the analysis of AP1 complexes either 10 $\mu$ g of fission yeast or 5 $\mu$ g of HeLa crude nuclear extract was treated with 0, 20, or 40 units of calf intestinal phosphatase (Boehringer) in 5 $\mu$ l gel retardation buffer supplemented with 2mM zinc chloride for 20 mins at room temperature. In the analysis of ATF complexes 1.5 $\mu$ g of budding yeast extract (purified over heparin agarose and DNA sepharose columns) was incubated at 37 C for 15 mins in the same buffer as above supplemented with 10mg Bovine serum albumen (BRL). When phosphatase inhibitors were used they were added at the following concentrations: pyrophosphate, 16mM; vanadate, 16mM; molybdate, 70mM (all supplied by sigma).

**18.d. Kinase Reactions;** Typically 1 unit of the catalytic subunit of protein kinase A (Sigma; catalogue no. p2645) was added to 4ml of affinity purified yATF (first round, peak activity) or 1.5 $\mu$ g DNA sepharose purified yATF and incubated for 15 minutes at 37 C in gel retardation buffer supplemented with 10mg BSA, 25mM spermidine (for affinity purified extracts), and 2mM ATP. Control reactions were carried out in the same buffer without added protein kinase and with or without added ATP. Once again these reactions were carried out prior to the addition of DNA and the amounts outlined above are relevant to what would subsequently constitute a single lane in a gel shift assay.

## 19. OLIGONUCLEOTIDES;

**19.a. Preparation of single stranded oligonucleotides:** In normal circumstances about 0.2 mg of each synthetic oligo was purified on preparative 12% acrylamide (38% acrylamide; 2% bisacrylamide) 50% urea and 1x TBE based gel. These were 1.5 mm thick, 40 cm long, and run for around 90 mins in 1X TBE in a vertical gel tank. The DNA was visualised by illuminating the gel on TLC plates precoated with silica gel 60 F<sup>254</sup> fluorescence indicator. The DNA band was excised, crushed and left to soak in 1ml TE plus 500mM ammonium acetate overnight at 37 C. The DNA, now in solution, was separated from the gel by syringe filtration and placed into 13 x 51mm polyallomer Beckman centrifuge tubes. 0.5 mls of TE and 3.5 mls of ethanol were added and mixed to this and the tube placed on dry ice for 30 minutes. Tubes were then spun for 30 minutes at 49K in a Beckman SW 50.1. Precipitated DNA was redissolved in 100µl of TE.

Annealing of single stranded oligonucleotides: Each complementary oligonucleotide <sup>as</sup> were put into a single tube with 40µl of 10x PNK buffer and 60µl ddW. This mixture was then heated to 88 C for 2 minutes, 65 C for 10 minutes, 37 C for 10 minutes and room temperature for five minutes.

### 19.b. Oligonucleotides used;

<b>HIS-168</b>	5'-3'	TCGACGGATCCAGCGGATGACGCTTTTTTTTGCATG
	3'-5'	GCCTAGGTCGCCTACTGCGAAAAAAC
<b>HIS-wt</b>	5'-3'	TCGACGGATCCAGCGGATGACTCTTTTTTTTGCATG
	3'-5'	GCCTAGGTCGCCTACTGAGAAAAAAC
<b>Col-mut1</b>	5'-3'	TCGACGGATCCTTGAGGAGTCAGCCGCATG
	3'-5'	GCCTAGGAACCTCAGTCGGC
<b>Col-mut2</b>	5'-3'	TCGACGGATCCTTGATGAGTCCTCCGCATG
	3'-5'	GCCTAGGAACTACTCAGGAGGC
<b>Col-AP1</b>	5'-3'	TCGACATCGATTTGATGAGTCAGCCGCATG
	3'-5'	GTAGCTAAACTACTCAGTCGGC
<b>SV40-AP1</b>	5'-3'	GATCCATCTCAATTAGTCAGCAAG
	3'-5'	GTAGAGTTAATCAGTCGTTCTTAG

**E355-TATA**

5'-3' TCGACATCGATGGGCGGCTTTCGTACAGGGTGCGGTGCGGCATGCAG  
 3'-5' GTAGCTACCCGCCGAAAGCAGTGTCCCACGCCAGCGCGTACGTC

ATCTGGGCAGGGTATAACTCACCTGAAAATCAGAGGGCGAGGTATTCA  
 TAGACCCGTCCCATATTGAGTGGACTTTTAGTCTCCCGCTCCATAAGT

GCTG  
 CGACCTAG

**3X E4 ATF**

5'-3' TCGACTTAACCGTTACGTCATTTTTATTAAACCGTTACGTCATTTTTAT  
 3'-5' GAATTGGCAATGCAGTAAAAATAATTGGCAATGCAGTAAAAATA

TAACCGTTACGTCATTTGCATG  
 ATTGGCAATGCAGTAAAC

**3X E4m2**

5'-3' TCGACTTAACCGTTACGCCATTTTTATTAAACCGTTACGCCATTTTTAT  
 3'-5' GAATTGGCAATGCGGTAAAAATAATTGGCAATGCGGTAAAAATA

TAACCGTTACGCCATTTGCATG  
 ATTGGCAATGCGGTAAAC

**3X E4m1**

5'-3' TCGACTTAACCGTTACTCCATTTTTATTAAACCGTTACTCCATTTTTAT  
 3'-5' GAATTGGCAATGAGGTAAAAATAATTGGCAATGAGGTAAAAATA

TAACCGTTACTCCATTTGCATG  
 ATTGGCAATGAGGTAAAC

08

**1X E4 ATF**            5'-3'    GATCCACCGTTACGTCATTTTTG  
                               3'-5'            GTGGCAATGCAGTAAAAACCTAG

**1X fibronectin**    5'-3'    GATCCCCCCTGACGTCACCCGGG  
                               3'-5'            GGGGGACTGCAGTGGGCCCTAG

**1X somatostatin** 5'-3'    GATCCTGGCTGACGTCAGAGAGG  
                               3'-5'            GACCGACTGCAGTCTCTCCCTAG

**1X E4m1**            5'-3'    GATCCACCGTTACTCCATTTTTG  
                               3'-5'            GTGGCAATGAGGTAAAAACCTAG

**1X E4m2**            5'-3'    GATCCACCGTTACGCCATTTTTG  
                               3'-5'            GTGGCAATGCGGTAAAAACCTAG

**Nonsense**            5'-3'    GATCCATATTGGCTTCAATCCAAAATAAG  
                               3'-5'            GTATAACCGAAGTTAGGTTTTATTCTAG

The sequences in the E3-AP1 and nonsense oligomers were as previously described (Hurst and Jones, 1987)

## **CHAPTER 3; RESULTS.**

### **IDENTIFICATION AND CHARACTERISATION OF AN AP1 LIKE ACTIVITY IN YEAST.**



## 1. Summary.

The Simian Virus 40 (SV40) enhancer was the first described and remains the classic model of investigation. Early analysis described its properties but more recent work has attempted to address how it actually works. This has led to the idea that it is comprised of a number of functional elements which on their own have very little or no activity, but together can act synergistically to drive transcription. Reduction in activity due to the loss of one element can often be recovered by the duplication of another (Herr and Clarke 1986, Clarke and Herr 1987, Ondek and Herr 1988). We decided to investigate whether further insights into enhancer function may be gained using genetic analysis to dissect its activity in yeast.

Previous analysis has shown that the SV40 promoter/enhancer can function in fission yeast. However this experimentation was very preliminary and did not address the detailed requirements for efficient expression. Based on this preliminary data we decided to characterise this promoter's activity in more detail using deletion analysis and factor binding assays. We measured the promoter activity by fusing it to the coding sequences of the chloramphenicol acetyl transferase gene and performing CAT assays. From this work we confirmed that the promoter is expressed well in fission yeast and demonstrated that it initiates transcription at the same site as in mammalian cells. The majority of the enhancer sequences, however, could be removed without affecting this activity. DNaseI footprint analysis of the promoter revealed the presence of an AP1-like factor in S.pombe cells that protects a region of the promoter almost identical to that protected by human AP1. The specificity of binding of the yeast and mammalian AP1 proteins was found to be similar. Our experiments also show that an AP1-like factor is present in budding yeast cells. This factor has a very similar binding specificity to the mammalian AP1 and fission yeast AP1-like factors but is clearly distinct from the binding activity of the product of the budding yeast GCN4 gene. In fission yeast the AP1 binding site can act as an upstream activating sequence with an inducing activity similar to that found in mammalian cells. Both the mammalian AP1 and fission yeast AP1-like factors are sensitive to phosphatase treatment indicating that they are phosphorylated. It should be noted that this work was carried out in collaboration with Sergio Moreno and Paul Nurse. Sergio constructed the promoter deletion series and tested them (fig 3.1.1) and also carried out the copy number experiments (fig.3.1.5.B). Apart from this, all data shown was produced by me.

## 2. The SV40 promoter functions efficiently in S.pombe.

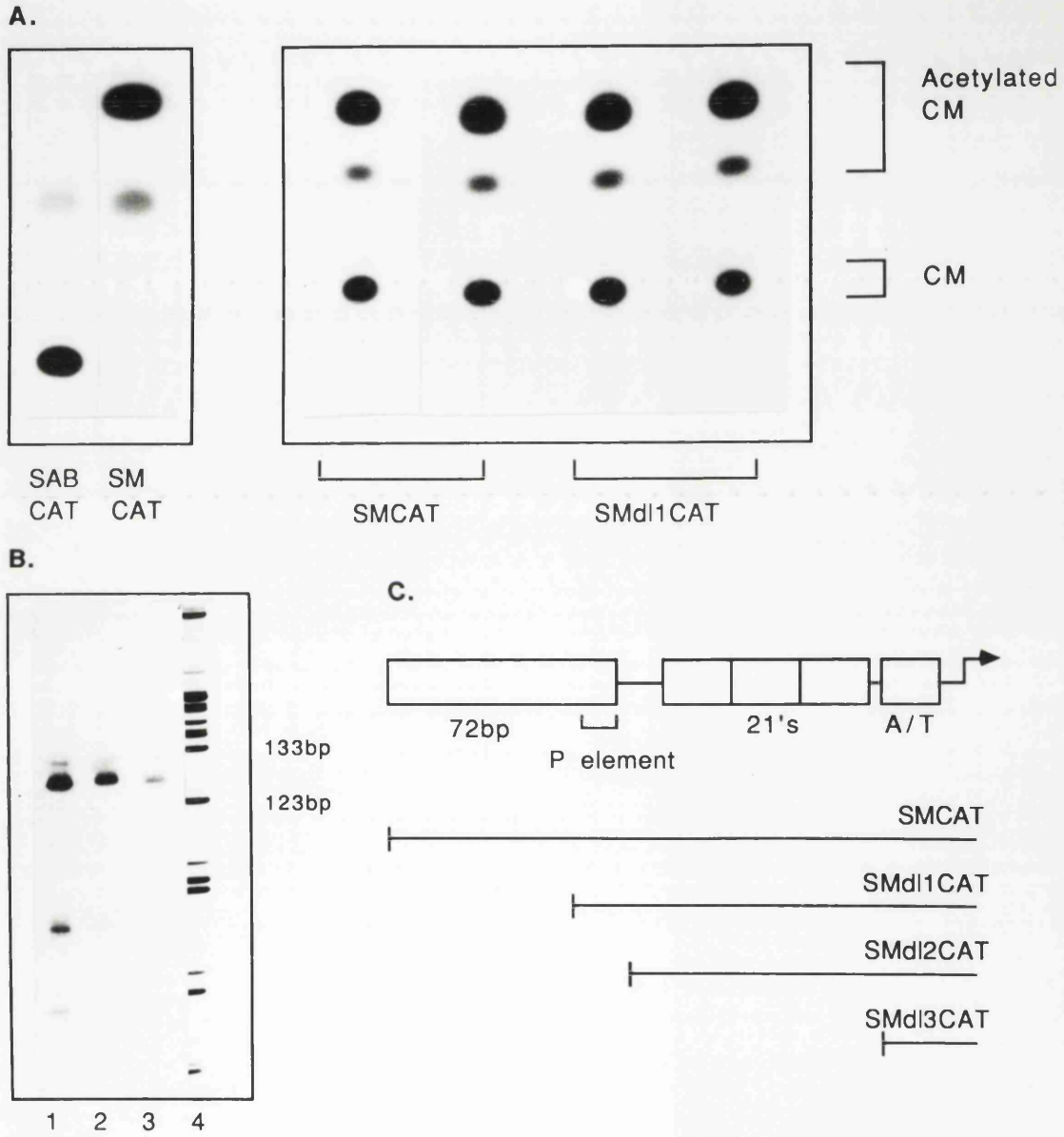
**Fig. 3.1.1. Function of the SV40 early gene promoter in S.pombe**

**A.** The expression of CAT coding sequences in S.pombe cells transformed with the plasmids SABCAT, SMCAT or SMdl1CAT (see 3.1.1.C for details of constructs) was determined by a standard CAT TLC assay. The CAT assays were performed with crude extracts using 10 µg (left panel) and 100 ng (right panel) of the extract supernatant (see experimental procedures) and incubating for 30 mins at 37 C. The right hand panel shows two representative results for each of the SM and SMdl1 constructs.

**B.** 6 µg of total RNA from cells transformed with the plasmids SMdl1CAT (lane 1), SMdl2CAT (lane 2) and SMdl3CAT (lane 3) were analysed by primer extension using a primer oligonucleotide complementary to a sequence within the CAT coding region. Lane 4 contains marker DNAs and the size of the relevant markers that flank the major extension product is indicated.

**C.** Schematic organisation of the SV40 early promoter region which comprises of the 72 bp enhancer element, the GC rich 21 bp repeat region and the AT-rich (TATA) element (Fromm and Berg, 1982, 1983). The location of the 'P' site which is specifically bound by the AP1 transcriptional activator protein is indicated (Lee et al., 1987a). Below the schematic are indicated the extent of these promoter sequences that are present in the various plasmids used in this study. SMCAT contains the complete promoter/enhancer region, SMdl1CAT lacks all of the enhancer except the 'P' site, SMdl2CAT additionally lacks the 'P' site and SMdl3CAT contains only the AT-rich element

**Fig. 3.1.1. Expression of the SV40 Promoter/Enhancer in Fission yeast.**



A chimeric gene containing the SV40 promoter/enhancer region fused to the reporter chloramphenicol acetyl transferase (CAT) sequences was inserted into a yeast LEU2 containing plasmid. The plasmid (SMCAT) was transformed into *S.pombe* cells and expression of the reporter CAT sequences in a number of individually isolated transformants measured. As shown in fig.3.1.1A, expression of this chimeric gene was very efficient whereas the level of CAT expression from a plasmid containing the same CAT sequences but lacking the SV40 elements (SABCAT) was hardly detectable under the same assay conditions. Surprisingly, the removal of most of the SV40 enhancer sequences had no effect on the level of CAT expression. SMdl1CAT has lost all enhancer sequences except the 'P' motif, which in mammalian cells acts as a binding sequence for the transcription factor AP1 (Lee et al., 1987a). In many different SMdl1CAT transformants, the level of CAT expression was equivalent to that found when the complete enhancer was present (Fig.3.1.1A).

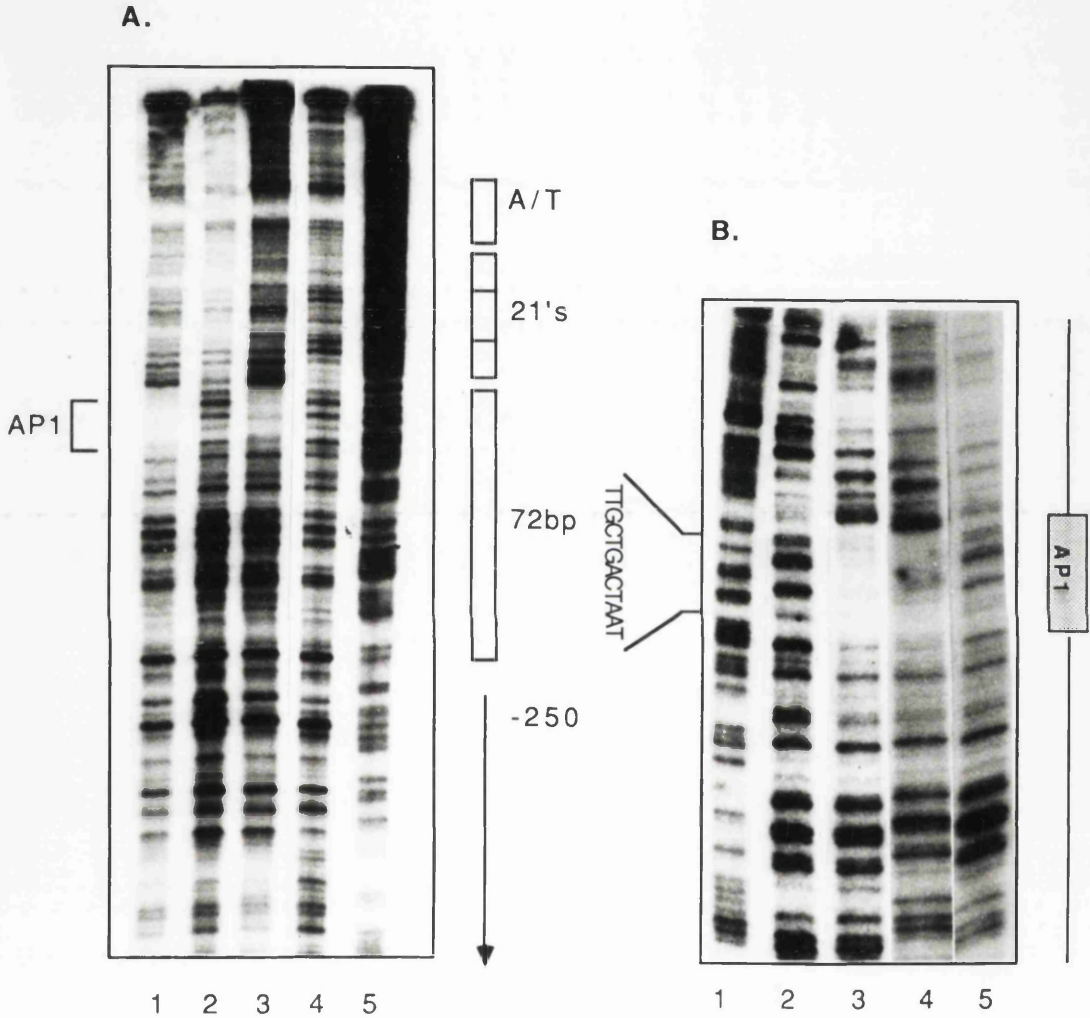
The nature and location of the transcriptional start site was analyzed by primer extension of RNA from SMdl1CAT transformed cells. In mammalian cells, SV40 early gene transcription is predominantly initiated from a set of closely adjacent sites 24-27 bp downstream of the TATA sequence (Gosh et al., 1981; Benoist and Chambon, 1981) which are 125-130 nucleotides upstream of the 5' end of the CAT primer used in this analysis. The major extended product obtained was 128 bp in length (Fig.3.1.1B, lane 1) showing that the same start site is utilized in both fission yeast and mammalian cells. Fig.3.1.1.B also shows the primer extension analysis of RNAs from two other fission yeast transformants. One transformant contains the plasmid SMdl2CAT which is identical to SMdl1CAT except that the 'P' motif has been removed (Fig.3.1.1.C). The other transformant contains SMdl3CAT which has additionally lost the 21bp repeat region of the promoter; the A/T rich element is thus the only promoter element that remains. In both these cases transcription initiated from the same start site (Fig.3.1.1B, lanes 2,3), indicating that neither the 'P' element nor the 21bp repeat region determined the site of initiation. The level of expression in these transformants however was lower than that obtained with SMdl1CAT; removal of the P site reduced expression, and this was reduced further to a very low level when the 21bp repeat region was deleted. These data indicate that both the 'P' element and 21bp repeat region contribute to SV40 promoter activity in fission yeast. The efficient utilization of the SV40 promoter in *S.pombe* is likely to involve the interaction of yeast transcription factors with specific promoter sequences.

### **3. Fission yeast contains an AP1 like binding activity.**

**Fig. 3.1.2. DNase 1 footprint analysis of the SV40 early promoter/enhancer in S.pombe.**

DNase 1 footprint analysis was carried out using a plus strand probe prepared by 3' end-labelling of the EcoRI site of the plasmid pSVGC-0 (a gift from T.Williams.) with [ $\alpha$ -<sup>32</sup>P] ATP and CTP and reverse transcriptase. 0.8  $\mu$ g of this labelled probe was used in each reaction. Panel A, lanes 1-3 and panel B, lanes 3-5 were incubated with 200  $\mu$ g of crude S.pombe nuclear extract; and panel A, lane 4 and panel B, lane 2 were incubated without extract. Panel A, lane 2 and panel B, lane 5 were also incubated with 300 fold molar excess of oligonucleotide competitors containing the SV40 AP1 binding site; panel A, lane 3 and panel B, lane 4 with similar excess of an unrelated nonsense oligonucleotide. Panel A, lane 5 and panel B, lane 1 represent the A + G ladder. The footprint reactions were carried out as described in materials and methods. The locations of the cis-acting promoter/enhancer elements are depicted at the right of each figure, and the location and sequence of the specific footprint obtained shown at the left.

Fig. 3.1.2. Footprinting analysis of the SV40 Promoter/Enhancer.



I investigated these interactions by DNAaseI footprint analysis. Incubation of the SV40 promoter/enhancer region with a nuclear extract from S.pombe cells resulted in a single region of protection (Fig.3.1.2.A, lane 1). This region corresponds to the 'P' site of the enhancer (Fig.3.1.2.B, lane 1) which is also protected by purified mammalian AP1 protein (Lee et al., 1987b; Angel et al., 1987). The footprint is lost in the presence of an oligonucleotide containing the AP1 binding site (Fig.3.1.2A, lane 2; Fig.3.1.2B, lane 5) but not by an oligonucleotide consisting of completely unrelated sequences (Fig.3.1.2A, lane 3; Fig.3.1.2B, lane 4). The binding of protein to this site is therefore specific. No other protected regions were evident although it is quite possible that other regions of protection would be found if fractionated rather than crude extracts were used. I used the gel shift assay to investigate the specificity of the yeast factor for the AP1 binding site. This assay relies on the fact that specific protein-DNA complexes migrate more slowly in the native polyacrylamide gel than the free unbound DNA. A <sup>32</sup>P-labelled oligonucleotide containing the SV40 AP1 binding site (SV-AP1) was added to crude nuclear extract from S.pombe and HeLa cells in the presence of an excess of unlabelled competitor oligomers. The relevant sequence of these oligomers is shown in Fig.3B. ColAP1 contains the core AP1 binding site of the human collagenase promoter (Angel et al., 1987); it differs by a single nucleotide from the SV40 site, resulting in a more perfect palindromic sequence. Human AP1 protein binds efficiently to both the SV40 and collagenase sites (Angel et al., 1987). Col-mut1 and Col-mut2 differ from the Col-AP-1 site by one and two nucleotides respectively, which significantly decrease binding of human AP1 protein (Angel et al., 1987). Incubation of the labelled SV-AP1 oligomer with the S.pombe nuclear extract resulted in a specific retarded complex that was efficiently competed by Col-AP-1 (Cwt, Fig.3.1.3), but not by either of the two mutants (Cm1, Cm2) nor by an unrelated oligonucleotide of similar size (N). A similarly behaving complex was obtained with the HeLa extract (Fig.3.1.3), although some competition was observed with Col-mut2. Thus mutations that resulted in decreased binding of human AP1, also lowered binding of the yeast factor. The two complexes differed significantly in their mobility which may reflect different sizes of the cognate binding proteins or differences in post-translation modification (see below). These data indicate that S.pombe contains a factor with the same binding specificity as AP1.

#### **4. S.cerevisiae contains two AP1 like factors.**

The SV40 AP1 binding site resembles the UAS sequence located upstream of a number of co-ordinately regulated amino acid biosynthetic genes of S.cerevisiae that is specifically bound by the activator protein GCN4 (Hill et al., 1986). I therefore tested the ability of this site to interact with budding yeast factors. Incubation of labelled SV-

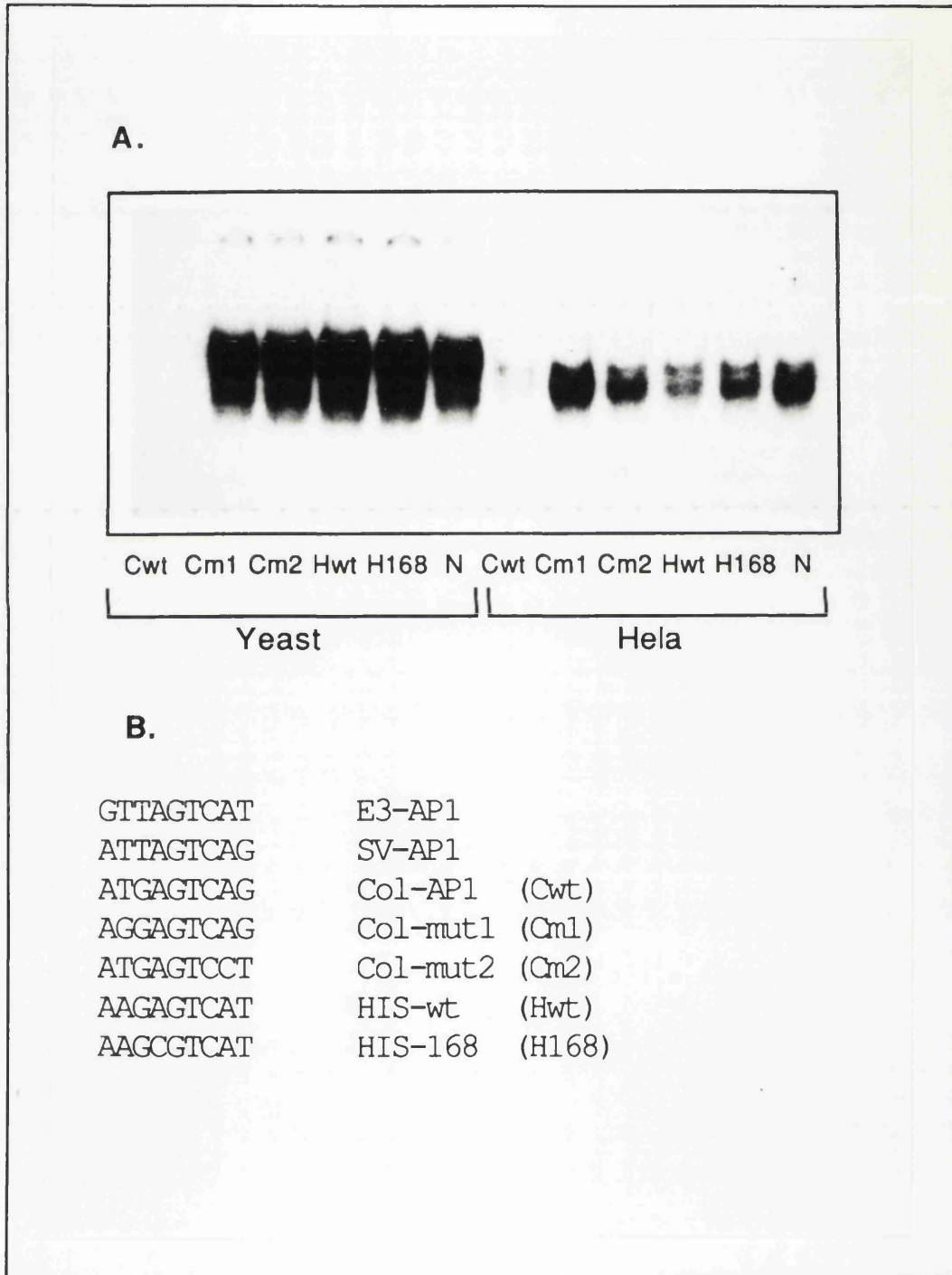
**Figure 3.1.3. The AP1-like factor of S.pombe and HeLa AP1 have similar binding specificities.**

**A.** Competition analysis of binding to the SV40 AP1 binding sites by HeLa and S.pombe nuclear factors. A radiolabelled oligonucleotide probe containing the SV40 AP1 binding site (SV-AP1) was incubated in a standard binding reaction with either 15  $\mu$ g S.pombe nuclear extract (lanes 1-6) or 7  $\mu$ g HeLa nuclear extract (lanes 7-12), 100 ng of nonsense oligonucleotides and a 100 fold molar excess of an unlabelled specific oligonucleotide competitor. The specific competitor DNAs used were Col-AP1 (lanes 1,7), col-mut1 (lanes 2,8), col-mut2 (lanes 3,9) his-wt (lanes 4,10) and his-168 (lanes 5,11) and a nonsense oligonucleotide of unrelated sequence (lanes 6,12). The binding reactions were loaded onto low-ionic strength 6% polyacrylamide gels, electrophoresed to separate DNA-protein complex from free probe and autoradiographed. Details on preparation of extract and probe, and the binding conditions are given in experimental procedures.

**B.** The AP1 binding site or related sequence present in each of the oligonucleotides used in this study are shown. SV-AP1 contains the AP1 binding site of the SV40 promoter and Col-AP1 the binding site of the human collagenase genes. Col-mut1 and Col-mut2 are two mutant forms of the collagenase AP1 site that decrease binding of pure human AP1 protein (Angel et al., 1987). HIS-wt contains the GCN4 binding site of the HIS 3 promoter of S.cerevisiae. HIS-168 is a mutant form of this site that has significantly decreased GCN4 binding (Hill et al., 1986). The complete sequence of these oligonucleotides is shown in materials and methods.



Fig. 3.1.3. Competition analysis of Fission yeast yAP1.



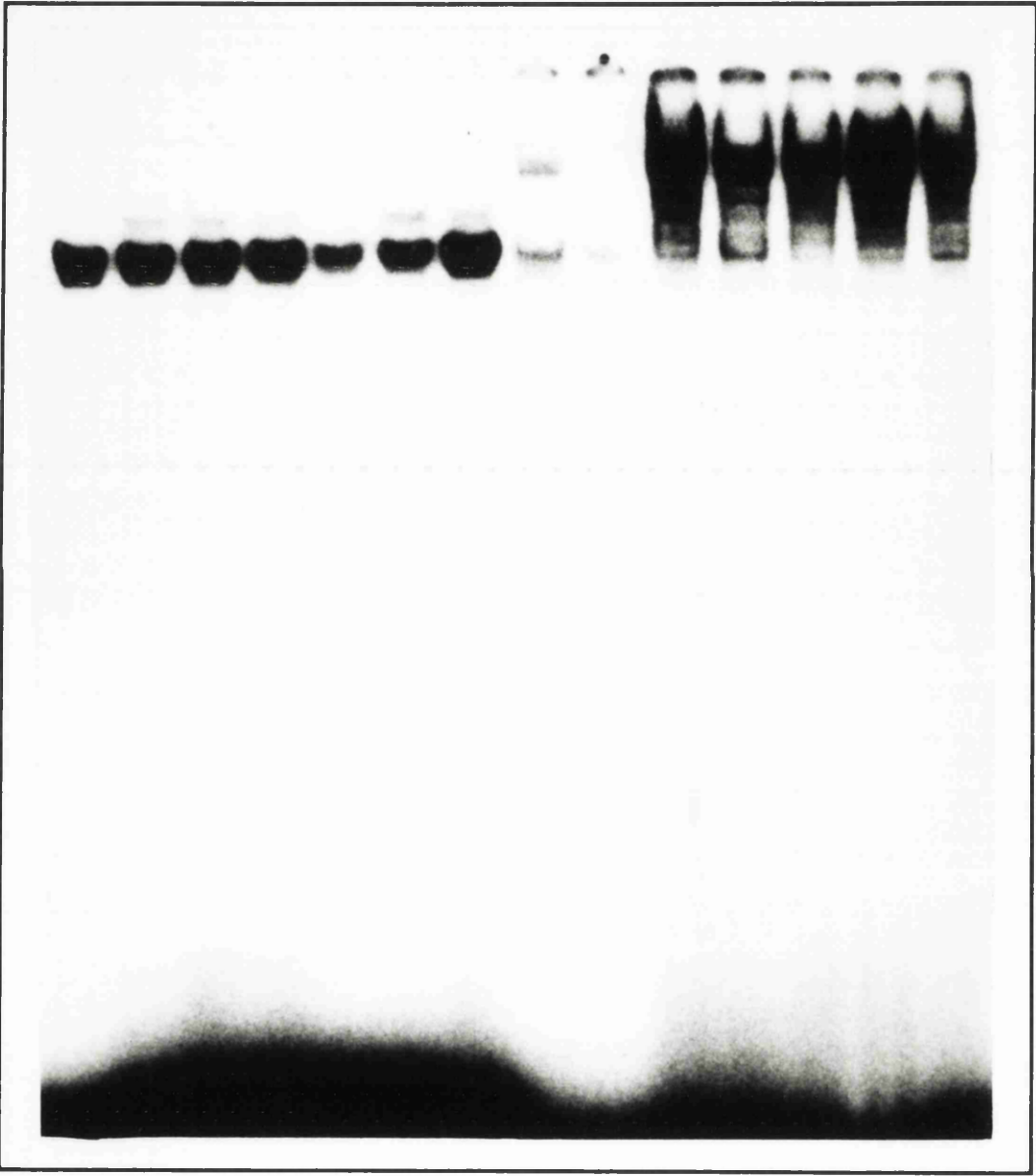
AP1 oligomer with a whole cell extract from S.cerevisiae resulted in a specific complex that was competable by an excess of unlabelled SV-AP1 or Col-AP1 oligomer (Fig.3.1.4, lanes 9,8), but not by an unrelated oligomer of similar size (lane 14). No competition was obtained with the two mutant oligomers, Col-mut1 and Col-mut2 (lanes 10 and 11). Therefore the binding specificity of the S.cerevisiae factor was similar to the factor identified in S.pombe and to human AP1. Surprisingly, an oligonucleotide containing the GCN4 binding site of the S.cerevisiae HIS 3 promoter (HIS-wt) did not compete for this factor even at a high molar excess (Fig.3.1.4, lane 12); this same oligomer also failed to compete for the AP1-like factor of S.pombe (Fig.3.1.3, lane 4) whereas slight competition for the human AP1 protein was observed (Fig.3.1.3, lane 10). An interpretation of these results was that the binding seen in the S.cerevisiae extract was not due to GCN4. In order to investigate this possibility further, retention assays were carried out using the HIS 3-derived oligonucleotide as the probe. Incubation of this probe with S.cerevisiae extract gave two distinct complexes. The faster migrating complex was not competable in a specific fashion and probably represents non-specific binding to the probe. The slower migrating, weaker complex however was competed in a pattern expected for the GCN4 protein. It was competed efficiently by the AP1 site of collagenase (Fig.3.1.4, lane 1) but not by the SV40 AP1 site (lane 2). The two mutant oligomers col-mut1 and col-mut2 failed to compete (lanes 3,4). Unlike the SV-AP1 derived complex however, this complex was efficiently competed by an excess of unlabelled HIS-wt (lane 5); it was not competed by an oligomer containing a point mutation of the HIS-3 GCN4 binding site (HIS-168) (lane 6) that has previously been shown to severely decrease binding of pure GCN4 protein (Hill et al., 1986). The specificity of this complex therefore exactly mirrors the specificity of GCN4 binding, strongly suggesting that this weak complex results from the GCN4 protein. The weakness of the band might also be expected since the levels of this protein in unstarved cells is very low (Arndt et al., 1987). Clearly therefore, at least two proteins exist in S.cerevisiae that have very similar binding specificities; one binds well to the SV40 AP1 site but not to the HIS-3 GCN4 site and the other binds well to the HIS-3 site but not the SV40 AP1 site. However both proteins can bind to the core sequence that constitutes the collagenase AP1 site. The ability of GCN4 to recognise this site is predictable from the mutagenesis studies of Hill et al (1986). These workers demonstrated that only one out of a large number of point mutations within the HIS3 site actually generated a sequence possessing a higher affinity for GCN4 than the wild type site itself. This mutation (TGA~~CT~~CA) produces a sequence that is identical to the core of the collagenase site thereby explaining the above result. An interesting observation at the time noted that although this site had a higher affinity for the GCN4 protein it could not be found in any of the promoters under general amino acid control. However the results

**Figure 3.1.4. S.cerevisiae contains specific proteins that bind to the AP1 and GCN4 binding sites.**

Radiolabelled HIS-wt oligonucleotide (lanes 1-7) or SV-AP1 oligonucleotide (lanes 8-14) was incubated in a standard binding reaction with 150 µg of S.cerevisiae whole cell extract and a 200 fold molar excess of oligonucleotide competitor. The competitors used were Col-Ap1 (lanes 1,8), SV-AP1 (lanes 2,9), Col-mut1 (lanes 3,10), Col-mut2 (lanes 4,11), HIS-wt (lanes 5,12), HIS-168 (lanes 6,13) and nonsense (lanes 7,14). These oligonucleotides are described in Figure 3. The nonsense oligomer contains sequences from -107 to -128 of the E3 promoter. The binding reactions were analysed by standard gel shift assay with 4 µg of poly dI:dC used in each reaction.

**Fig. 3.1.4. Budding yeast contains yAP1 and GCN4.**

1 2 3 4 5 6 7 8 9 10 11 12 13 14



I have described offer a possible explanation for this apparent paradox ie that such a sequence would also interact with yAP-1 and therefore no longer be specific for GCN4 binding and regulation.

Together these results indicate that at least two proteins in budding yeast can recognise the AP1 consensus sequence. Such observations are indicative of a possible transcription factor family. However at present it is impossible to determine whether the activity we see associating with the SV40 AP1 site is attributable to more than one protein.

## **5. The AP1 binding site can serve as a UAS in both yeast systems.**

Deletion of the AP1 binding site of the SV40 promoter resulted in lower levels of expression in S.pombe (Fig.3.1.1B). This suggested to us that the AP1-like factor may act as a trans-acting transcriptional regulator. In order to study this in more detail, we tested the ability of the AP1-binding site to serve as a UAS element. A single copy of the oligonucleotide containing the collagenase AP1 binding site was inserted immediately upstream of a TATA-box sequence that was fused to the CAT coding region. Expression of CAT in cells transformed with this plasmid was compared to those transformed with the plasmid that lacked the insert. The presence of the Col-AP1 binding site resulted in increased levels of CAT expression. A number of different transformants were analyzed in this way and the average increase due to the Col-AP1 insert was 5-fold. A typical example is shown in Fig.3.1.5.A. This difference was not due to differences in plasmid copy number in the transformants as determined by Southern analysis of transformed cell DNA (Fig.3.1.5.B). We also tested the ability of the AP1-binding site of the Adenovirus E3 promoter (Hurst and Jones, 1987) to stimulate expression in a similar assay. Increased expression was observed but in this case however, the increase was only about 2-fold. This result prompted us to compare the efficiency of binding of the S.pombe AP1-like factor to the Col-AP1 and the E3 AP1 binding sites. Competitor gel shift assays showed that binding of the factor to SV-AP1 oligomer was competed significantly less well by an excess of oligonucleotide containing the E3 binding site than one containing the collagenase site (Fig.3.1.5.C). This difference in affinity of the two sites for the AP1-like protein probably explains why the E3 binding site acts as a weaker UAS element.

## **6. The AP1-like factor of S.pombe is sensitive to phosphatase treatment.**

**Fig. 3.1.5. The AP1 binding site can act as a UAS element in S.pombe**

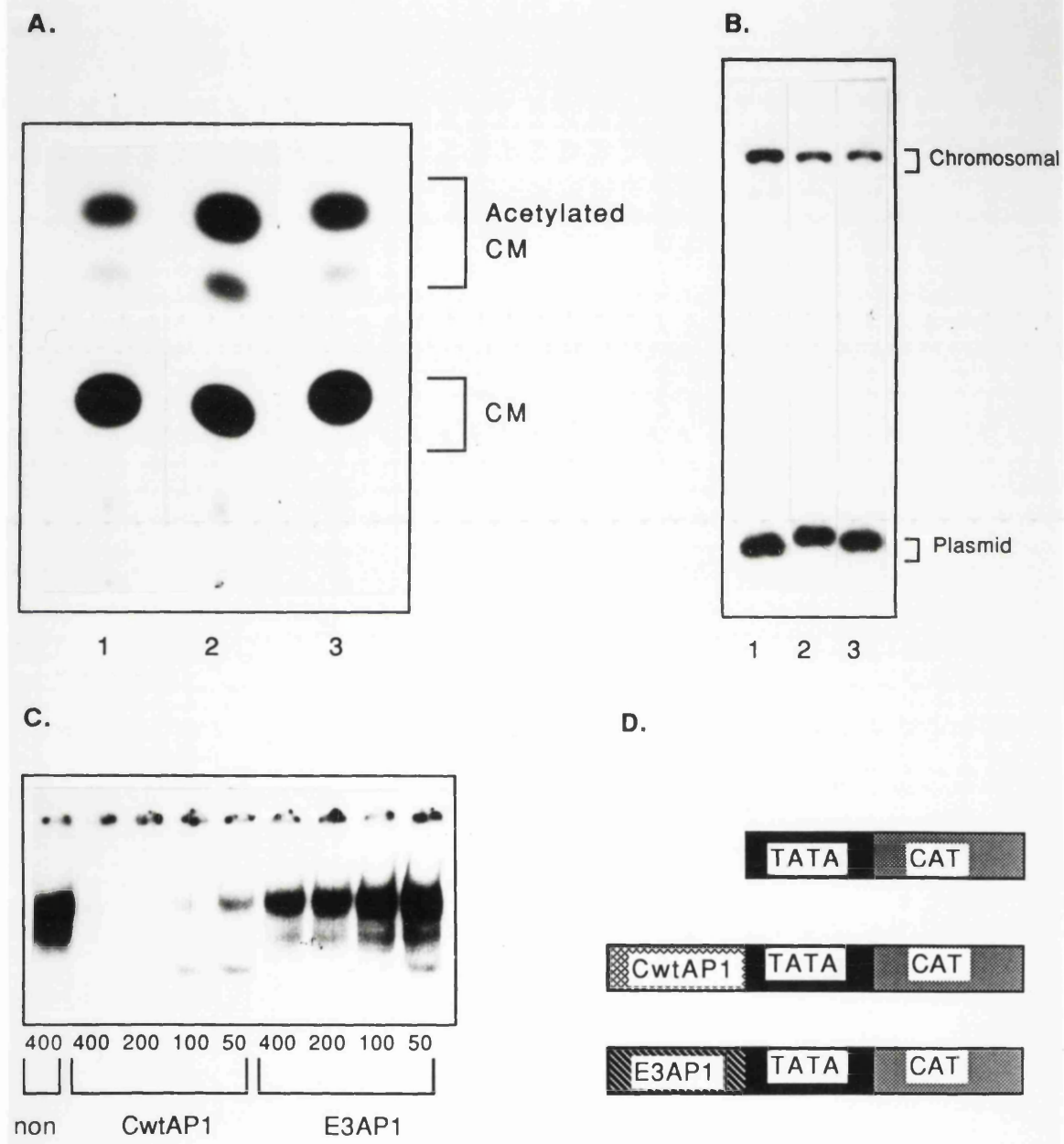
**A.** The expression of CAT coding sequences in S.pombe cells transformed with the plasmids pYT3 (lane 1), pYT3CP (lane 2) or pYT3E3 (lane 3). pYT3 contains a TATA box region fused to the CAT coding sequence in the yeast vector pURA4. pYT3CP is identical to pYT3 except it contains a single copy of the Col-AP1 oligomer upstream of the TATA box and pYT3E3 has an oligonucleotide containing the adenovirus E3 promoter AP1 binding site, inserted upstream of the TATA box. The CAT assays were as described in Fig.3.1.1, using 60 µg of extract protein per reaction.

**B.** Copy number of the transforming plasmid in each of the transformed cells described in A. 10 µg of S.pombe DNA was digested with SalI and BamHI, Southern blotted and probed with <sup>32</sup>P labelled CAT DNA (plasmid). As an internal control the blot was also probed with the S.pombe single copy gene CDC2 (chromosomal)

**C.** Affinity of the S.pombe AP1-like factor for the collagenase and E3 promoter AP1 binding sites. Radiolabelled SV-AP1 oligomer was incubated with 10 µg of S.pombe nuclear extract and an excess of unlabelled nonsense oligomer, Col-AP1 oligomer or on oligomer containing the adenovirus E3 AP1 binding site. The molar excess of competitor in each of the reactions is indicated. The binding reactions were analysed as described in Figure 3.1.3.

**D.** Schematic representation of the chimeric CAT genes contained in the plasmids used in this study.

Fig. 3.1.5. The AP1 site can act as a UAS in Fission yeast.



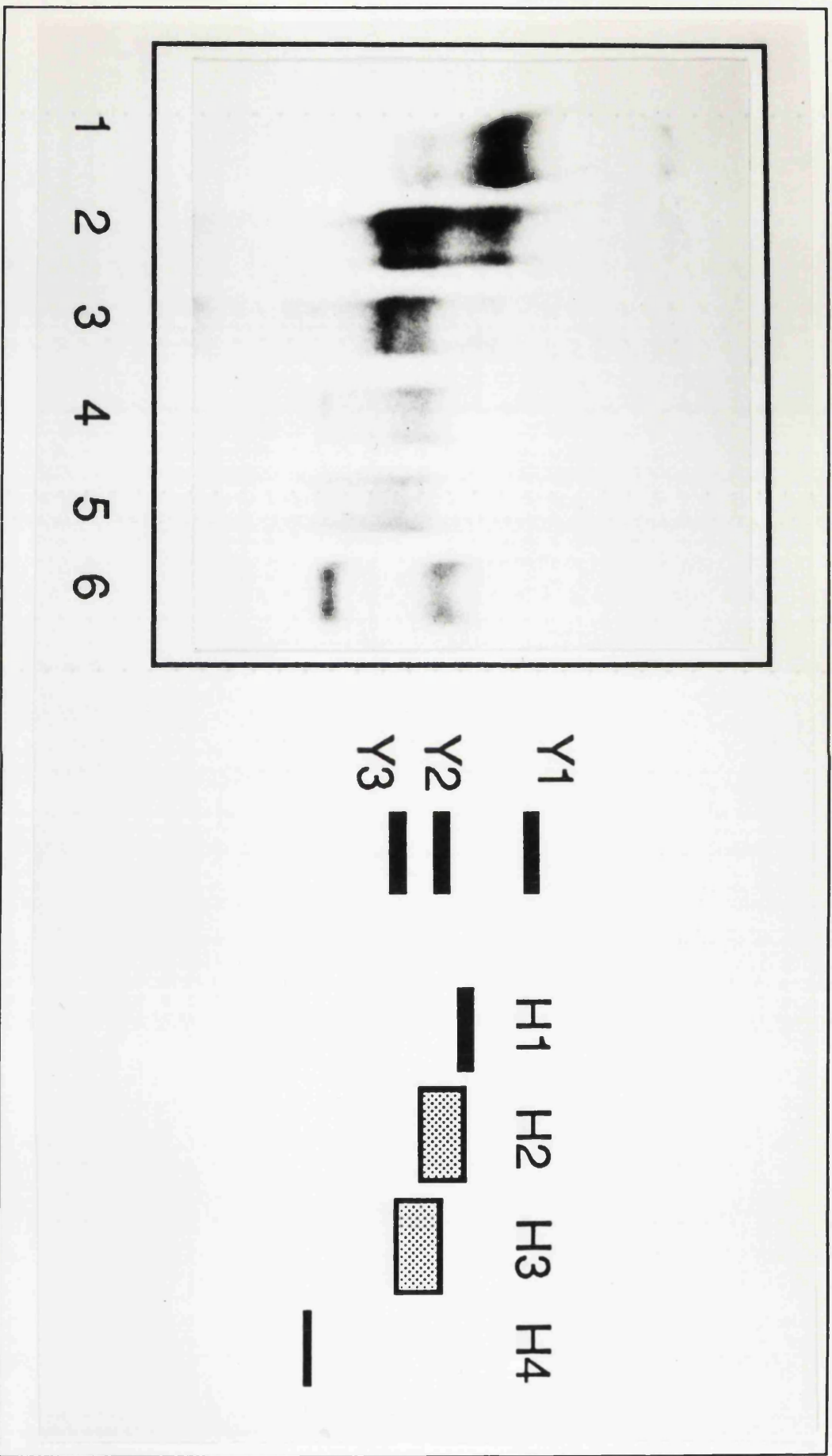
The distinct mobility differences between the S.pombe and HeLa cell derived complexes could be due to different sizes of the AP1 and AP1-like factors. Alternatively however, the differences could reflect the consequences of post-translational modification of the factors. If modification is involved a possible candidate would be phosphorylation. In S.cerevisiae the complex between the HSTF and its cognate binding site has a slower mobility following heat shock; this change in mobility appears to be due to increased phosphorylation (Sorger et al., 1987). In order to test this possibility we pretreated the S.pombe and HeLa extracts with calf intestinal phosphatase prior to complex formation. Treatment of the yeast extract with 20 units of phosphatase resulted in a decrease of the normal complex Y1 and the appearance of two different complexes Y2 and Y3 with increased mobility (Fig.3.1.6, lane 2). With 40 units of phosphatase Y1 virtually disappears and there is an increase in Y3 in relation to Y2 (lane 3). The most likely explanation for these results is that the S.pombe AP1-like factor is phosphorylated, possibly at multiple sites, and that this phosphorylation induces some conformational change that alters mobility of the protein. The human AP1 factor also appears to be phosphorylated. Treatment of the HeLa extract with 20 units of phosphatase results in the disappearance of the normal complex H1 (the other complex H4 obtained with the untreated extract is due to non-specific binding, data not shown) and the appearance of a diffuse complex H2 that has an increased mobility (lanes 4 and 5). An even faster migrating, diffuse complex (H3) is obtained with 40 units of phosphatase. Although the H2 and H3 complexes do not migrate as a sharp band, it is interesting to note that the mobilities of Y2 and H2 and of Y3 and H3 are similar. Further ideas considering what may be occurring during phosphatase treatment are addressed in the discussion section.



**Fig. 3.1.6. The S.pombe and HeLa AP1 factors are sensitive to phosphatase treatment.**

Radiolabelled SV-AP1 oligomer was incubated with 15 µg of S.pombe nuclear extract (lanes 1-3) or 7µg of HeLa nuclear extract (lanes 4-6) that had been preincubated with 0 (lanes 1,6), 20 (lanes 2,4) or 40 (lanes 3,5) units of calf intestinal phosphatase for 10 mins at room temperature. The binding reactions were analysed as described in Figure 3. Also shown is a schematic of the yeast and HeLa cell derived protein complexes. Y1-Y3 are yeast derived and H1-H4 are HeLa cell derived; all complexes except H4 are competed in a specific fashion with an excess of unlabelled SV-AP1 oligomer. H4 is due to non-specific binding to the probe; it's intensity is variable from reaction to reaction and is not competed in a specific fashion.

Fig. 3.1.6. Phosphatase treatment of yAP-1.



**CHAPTER 4; RESULTS.**

**IDENTIFICATION AND CHARACTERISATION OF A  
YEAST ATF-LIKE BINDING ACTIVITY.**

## **1. Summary.**

Although I have presented work carried out on the SV40 promoter/enhancer as the first results chapter in my thesis, my initial investigations actually focussed on the expression of the adenovirus E2A promoter. This promoter is inducible by E1A in mammalian cells and so I tested whether this feature of regulation could also be observed in yeast. The work I present in this chapter demonstrates that the E2A promoter is functional in fission yeast but that its activity does not appear to change if E1A protein is expressed in the same cells. However the promoter uses a transcriptional start site also utilised in mammalian cells and, in addition, a number of fission yeast DNA binding proteins recognise E2A sequences previously shown to interact with mammalian transcription factors. One of these mammalian factors is called ATF and recognises an element whose consensus sequence is only 1 bp different to that of the consensus AP1 site. I show that yATF and yAP1 are also capable of distinguishing this single nucleotide difference thereby demonstrating a remarkable conservation of DNA binding specificity. In addition the budding yeast ATF activity is likely to be a transcriptional activator.

## **2. Expression of the Adenovirus E2A promoter in fission yeast.**

The original aim of my thesis was to determine whether the E1A early protein of adenovirus could activate transcription in yeast; this would ascertain whether a genetic investigation of the activation mechanism, using the fission yeast system, was feasible. To address this experimentally a fusion gene, consisting of the E1A inducible E2A early promoter linked to the chloramphenicol acetyl transferase coding sequences, was cloned into the yeast episomal vector pURA4 to create pU4E2CAT. In addition the adenovirus 13s cDNA, capable of mediating the transactivation effect in mammalian cells (Berk 1986), was cloned downstream of the SV40 promoter in another episomal plasmid pSM1 creating pPB13s.

To test the E2A promoter activity and inducibility pU4E2CAT was transformed into fission yeast (as described in materials and methods). Two of the resulting transformants, E2C1 and E2C2, were grown up and then transformed with the pPB13s plasmid. Extracts from a number of these transformants were tested for CAT activity alongside the parental strains; fig. 4.1.1.A shows representative data. From this it can be seen that a strain harbouring the 13s E1A expression plasmid (strain E2C113a; lane 3) did not have a notably higher CAT activity than the parental strain

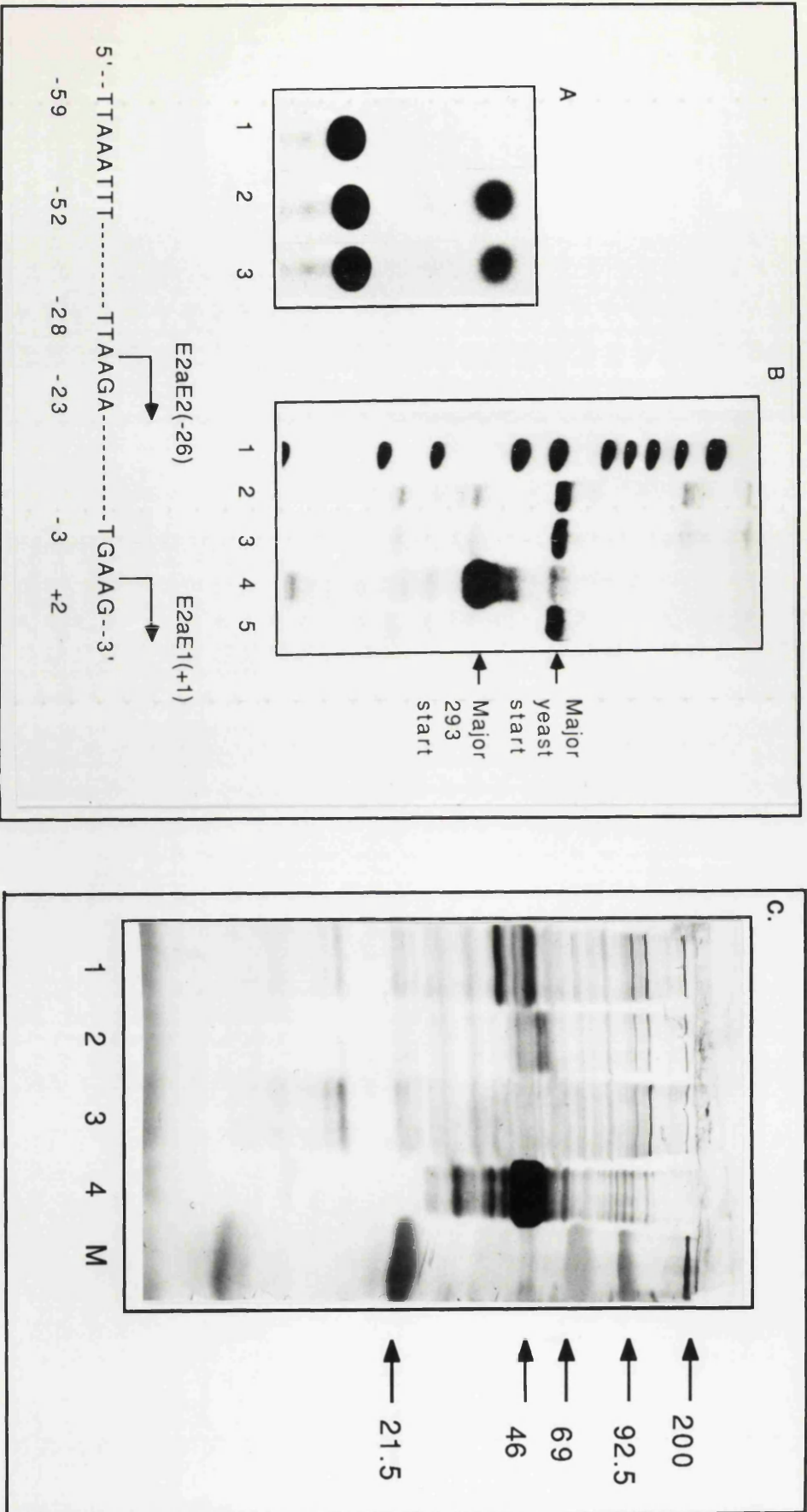
**Fig. 4.1.1. Expression of E2ACAT in S.pombe.**

**A.** S.pombe cells were stably transformed with plasmids containing the E2A promoter fused to the chloramphenicol acetyl transferase (CAT) coding sequences together with or without a second plasmid containing E1A. Expression of the E2ACAT gene was determined by a standard CAT TLC assay using 3µg of protein and incubating for 10 minutes at 37 C in each reaction. Lane 1, extract from cells containing the stably replicating yeast vector pURA4; Lane 2, extract from cells stably transformed with pU4E2CAT, which contains the pE2ACAT gene inserted into pURA4; Lane 3, extract from cells stably transformed with pU4E2CAT and pPB13S, a plasmid containing the E1A 13s DNA sequence linked to the SV40 early promoter and enhancer.

**B.** Primer extension analysis of RNA prepared from S.pombe and 293 cells expressing the E2ACAT gene. Lane 1, markers (pBR322 cut with Msp 1; New England Biolabs) ; Lanes 2, 3 and 5, analysis of 13 µg of total RNA prepared from S.pombe stably transformed with pU4E2CAT alone (lane 5) or together with pPB12S (expressing the 12s E1A cDNA) (lane 2) or pPB13S (expressing the 13s E1A cDNA) (lane 3). Lane 4, 25 µg of cytoplasmic RNA prepared from 293 cells transiently transfected with pE2ACAT. At the bottom of the figure is shown the positions on the E2A promoter of the major (+1) and minor (-26) transcription start sites together with sequences between positions -23 and -28 and between -52 and -59 that represent 'TATA'-like elements.

**C.** Western blot analysis of E1A producing S.pombe strains. 40µl of crude S.pombe cell extracts (at approximately 15µg/µl) from strains E2C112a (lane 1) E2C113a (lane 2) or E2C1 (lane 3) were loaded with an equal volume of laemli loading buffer onto a denaturing polyacrylamide gel. In addition 10µl of insect cell lysate (lane 4; from cells overproducing E1A) and protein size markers were also loaded (Amersham).

Fig. 4.1.1. Analysis of E1A and an E1A inducible Promoter in Fission Yeast.



(strain E2C<sub>1</sub>; lane 2) which contained only the E2CAT test construct.. This indicated that E1A was having no activating affect in this system but to ensure that the E1A expression plasmid was actually producing E1A protein I carried out western blot analysis on the extracts. As a control for this experiment I used E1A protein that had been overexpressed in insect cells using the baculovirus expression system which is known to migrate at almost exactly the same mobility as the E1A protein synthesised in HeLa cells (donated by G.Patel and described in Patel and Jones 1990). Fig.4.1.1.C shows that protein species roughly comigrating with the insect cell produced E1A protein (lane 4) appear specifically in yeast strains which have been stably transformed with either pPB13s (strain E2C<sub>1</sub>13a; lane 2) or pPB12s (strain E2C<sub>1</sub>12a; lane 1) and not with a strain not harbouring an E1A expression plasmid (lane 3). The bands which do appear in lane 3 are thought to be due to background interactions.

The reason for the slight discrepancy in mobility between the insect cell produced protein and the fission yeast derived protein is not clear but may result from differential modifications such as phosphorylation. Whether this differential mobility is important for the lack of transactivating activity is also not known and it was difficult to address this experimentally.

However to characterise the E2A promoter activity further I decided to carry out primer extension analysis to map the start site of transcription in yeast. As a control RNA was also prepared from 293 cells transformed with the PU4E2CAT plasmid. 293 cells express E1A constitutively and therefore provide an active environment to test E1A inducible promoters (Graham et al 1977). Fig.4.1.1.B. shows the results of the primer extension analysis. As described in the introduction the E2A early promoter has two transcriptional start sites, one minor and one major. These two starts are marked in the figure and it can be clearly seen that 293 cells predominantly use the major site. In contrast the RNA derived from fission yeast appears to predominantly utilise the 'minor' site. To test whether E1A could qualitatively alter the start site in yeast RNA was also prepared from E2C<sub>1</sub>13a and E2C<sub>1</sub>12a strains. Lanes 2 and 3 demonstrate that E1A still has no noticeable effect. The reason why the minor start site is used in fission yeast is not clear and possible explanations are addressed in the Discussion. However, by demonstrating that a 'real' start site was being utilised the analysis was encouraging and it prompted an investigation into the yeast factors interacting with the promoter.

### **3. Fission yeast factors interacting with the E2A promoter.**

DNAseI footprinting of the promoter with nuclear extracts prepared from S.pombe and HeLa cells is shown in Figure 4.1.2. Previous studies demonstrated that a HeLa cell factor termed activating transcription factor or ATF (see introduction) binds to

sequences between positions -68 to -80 relative to the E2 major start site (see fig.4.1.2.A, lane 5) and is functionally very important for optimal expression of this promoter. A footprint in this same region was also obtained with yeast nuclear extract (Fig.4.1.2.A, lane 4), which was efficiently competed with an excess of oligonucleotide containing an ATF binding site (Fig.4.1.2.A, lanes 1,2) but not with an oligonucleotide comprised of unrelated sequences (Fig.4.1.2.A, lane 3). It appears, therefore, that S.pombe cells contain a factor that specifically interacts with the ATF binding site.

Two other footprints in common between the yeast and HeLa extracts were also observed, namely binding site 2 located between nucleotides -128 and -142 and a weak binding site 3 between nucleotides -60 and -68. A factor which binds to site 2 has been previously described (Boeuf et al 1987) and it may be related to a factor termed E4F2 that binds to an identical sequence in the E4 promoter (K. Lee, pers.comm). However the functional significance of this factor for promoter activity in mammalian cells is not clear. The nature of the yeast cellular factor that binds to site 3 is not known but may be related to the E1A inducible factor E2F, also essential for full activity of this promoter in mammalian cells (Kovesdi et al 1986, Loeken and Brady 1989).

#### **4. Yeast ATF can interact with the Adenovirus E4 promoter.**

The presence of a factor in yeast that binds to the ATF sequence was confirmed by footprinting the E4 promoter. An ATF binding site within the E4 promoter is located between sequences -40 and -58 (Lee and Green 1987) and fig.4.1.2.B, lane 2, confirms that this region is protected by extracts from HeLa cells. This region is similarly protected by fission yeast extracts (lane 3) and this protection is competed by an ATF site oligonucleotide (lane 4) but not by a nonsense sequence (lane 5). The extent of the protected region is exactly the same for both fission yeast and HeLa cell extracts; however, there are subtle differences in hypersensitive cutting sites that flank the protected region, possibly reflecting different sizes of the yeast and HeLa factors, leading to different interactions with DNase 1.

#### **5. S.cerevisiae also contain an ATF like binding activity.**

I also tested the ability of budding yeast extracts to protect the ATF site in the E4 promoter. A similar region of protection was indeed observed in budding yeast extracts which could be competed away with an excess of ATF site oligonucleotide (lane 8) but not by the same excess of an oligonucleotide bearing unrelated sequences (lane 9).



## Fig. 4.1.2.

### A. DNase 1 protection of the E2AE promoter by yeast and HeLa nuclear extracts.

The boxes in the central diagram represent regions protected by both yeast and HeLa cell extracts. The upper and lower peripheral sequences refer to the actual nucleotides protected in these shared regions on the coding and noncoding strands. Lanes 1-7 demonstrate protections on the noncoding strand and lanes 8-10 on the coding strand. Lanes 1, 2 and 3 are each protected with 200 µg of yeast extract but with the addition of a 150 fold molar excess over labelled probe of E4 ATF, E2 ATF and nonsense site oligonucleotides. Lanes 4 and 8 are protected with 200 µg of yeast extract and lanes 5 and 9 by 150 µg of HeLa extract without oligonucleotide competitors. Lanes 6 and 10 represent unprotected DNA and lane 7 is the A+G ladder.

The E2 promoter probes used in Fig.4.1.2A were generated by labelling at the BssH2 site (-17) of pE2A-E-CAT (Murthy et al 1985) on either the 3' or 5' strand.

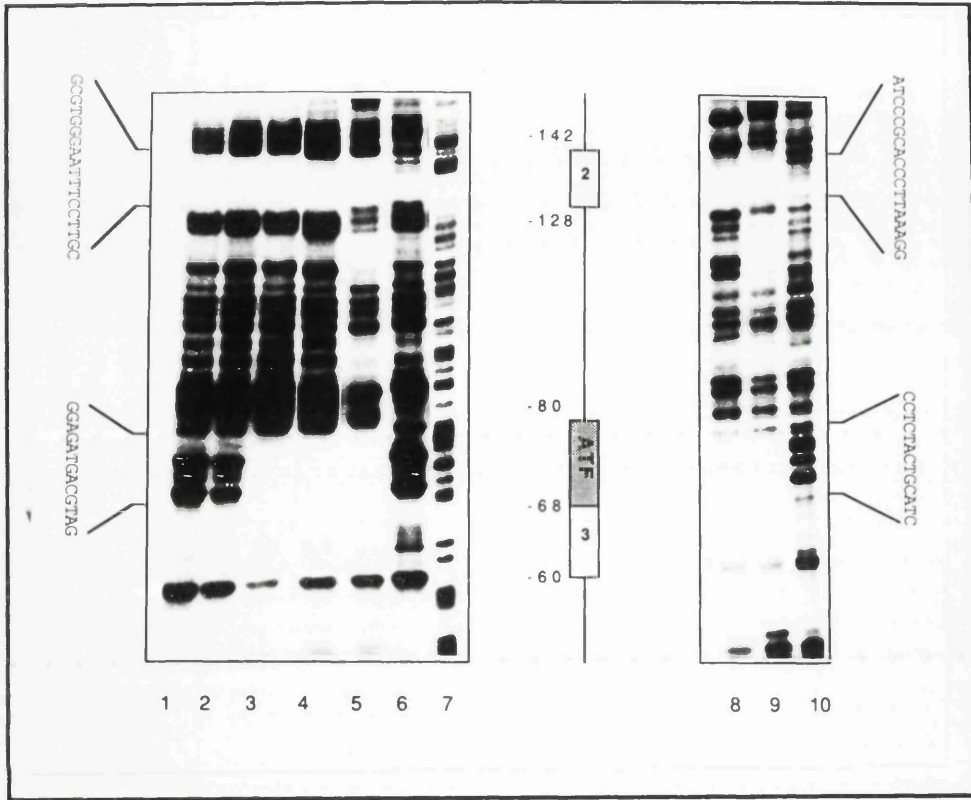
Subsequent cleavage by EcoRV (at -235) produced probes for the non-coding and coding strand respectively. Footprinting reactions using poly dI:dC as the nonspecific competitor were performed as described in the materials and methods .

### B. DNase 1 footprint analysis of the Adenovirus E4 early promoter with yeast and human extracts.

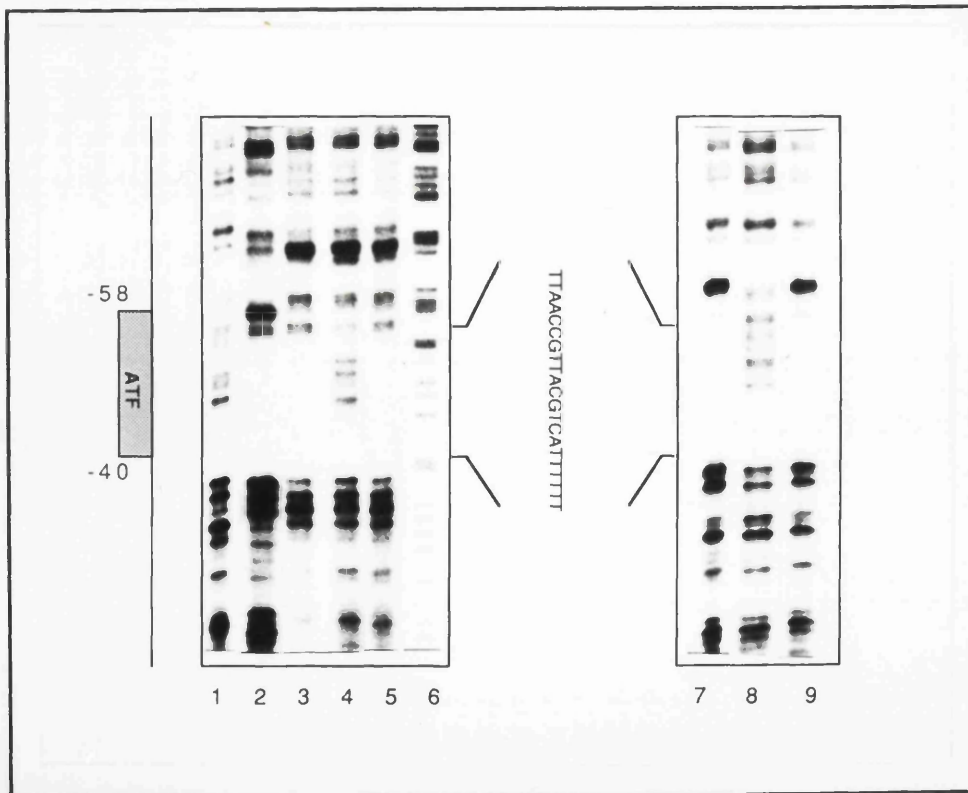
DNase 1 footprint analysis was carried out on a 3' end labelled E4 promoter probe extending from +15 to -325 Footprint reactions were carried out as already described for S.pombe and Hela cell extracts. 1.5 µg of DNA sepharose purified S.cerevisiae extracts were used in budding yeast protections. Lane 1 is unprotected DNA, lane 2 is protected with Hela cell extracts, lanes 3 to 5 are protected with S.pombe cell extracts, lanes 7 to 9 are protected with S.cerevisiae extracts and lane 6 is the A+G ladder. A 300 fold molar excess of oligonucleotide containing from the E4 ATF binding site was added to reactions 4 and 8 and the same molar excess of a nonsense oligonucleotide was added to reactions 5 and 9. The schematic ATF box on the left hand side shows the extent of the previously characterised human ATF protection of the E4 promoter. The sequence of the protected region is shown.

Fig. 4.1.2. Footprinting analysis of the Adenovirus E2A and E4 promoters

A.



B.



Therefore both yeast systems contain factors very similar in DNA binding specificity to ATF.

## **6. Yeast ATF and Yeast AP-1 are distinguishable.**

The consensus binding sites for ATF (T/G T/A CGTCA) and human transcription factor AP1 (TT/G AGTCA) are very similar, although in mammalian cells the factors that preferentially bind to these two sites *in vitro* are clearly distinct (see N.Jones et al 1988 for review). Since I had shown the existence of yeast factors that bind to the AP1 site, I investigated whether the ATF binding activity detected in yeast was due to yAP1 or, as in human cells, to a distinct factor.

Budding yeast extracts used in this analysis had been purified through heparin agarose and DNA-sepharose columns (see chapter 5). Gel retardation assays gave rise to three major complexes (a,b and c) following the incubation of budding yeast extracts with an oligonucleotide probe containing the ATF binding site (see Fig 4.1.3.A, lane 6); the relative ratios of the different complexes varied from one extract preparation to the next, probably due to differences in phosphorylation of the complex components (see chapter 6). All three complexes were efficiently competed by an excess of unlabelled oligonucleotide containing the E4 ATF site or by an oligonucleotide containing the CRE from the somatostatin promoter (Montminy et al 1986); they were not efficiently competed by oligonucleotides containing the AP1 sites from the SV40 or collagenase promoters (compare lane 7 and 8 with lanes 9 and 10, Fig 4.1.3.A). Identical results were obtained with fission yeast extracts (lanes 7, 8, 9 and 10, Fig 4.1.3.B). In contrast, complexes obtained by incubating yeast extracts with a labelled probe containing the SV40 AP1 binding site were efficiently competed by the two AP1 - containing oligonucleotides but not by the oligonucleotides containing the E4 ATF site or somatostatin CRE (compare lanes 4 and 5 with lanes 2 and 3, Figs. 4.1.3.A and 4.1.3.B). This data argues that two different binding activities, yATF and yAP1, exist in both budding and fission yeast. In addition the activity binding to the ATF site is unlikely to be GCN4 as it is not efficiently competed with the collagenase AP-1 site, a sequence which represents the optimal binding site for the GCN4 factor (Hill et al 1986)

A close comparison of the consensus ATF and AP1 sites reveals a consistent difference at the third residue, which exists as a C for ATF and an A for AP1. The importance of this change in switching specificity has been shown previously in human systems (Hurst and Jones 1987) where a single basepair change converting the CGTCA of the adenovirus E3 ATF binding site to AGTCA altered the preference of this site from binding ATF to AP1. I therefore tested the importance of this change on binding in

**Fig. 4.1.3 Yeast ATF- and AP1-like binding activities are distinct.**

**A.** S.cerevisiae extracts were incubated in a standard binding reaction with a radiolabelled oligonucleotide probe containing the SV40 AP1 binding site (lanes 1-5) or the E4 ATF binding site (lanes 6-10), and a 50 fold molar excess of unlabelled oligonucleotide competitor. Competitor DNAs used were E4 ATF (lanes 2 and 7), Som-CRE (lanes 3 and 8), SV40-AP1 (lanes 4 and 9), and Col-AP1 (lanes 5 and 10). The binding reactions were loaded onto low ionic strength 8% polyacrylamide gel, electrophoresed to separate DNA-protein complex from free probe and autoradiographed.

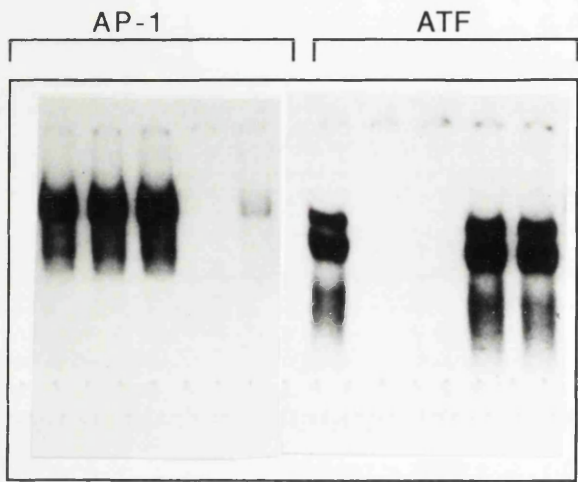
**B.** S.pombe extracts were assayed by gel retardation analysis; the addition of probe and specific oligonucleotide competitors were as in A.

**C.** Radiolabelled E4-ATF oligonucleotide (lanes 1-3) or SV40-AP1 oligonucleotide (lanes 4-6) were incubated with S.cerevisiae extracts as in A. Oligonucleotides bearing the E3-ATF site (lanes 2 and 4) or a mutated E3-ATF site (lanes 3 and 5) were added at a 250 fold molar excess.

**D.** The core sequences of oligonucleotide competitors used in the competition analysis.

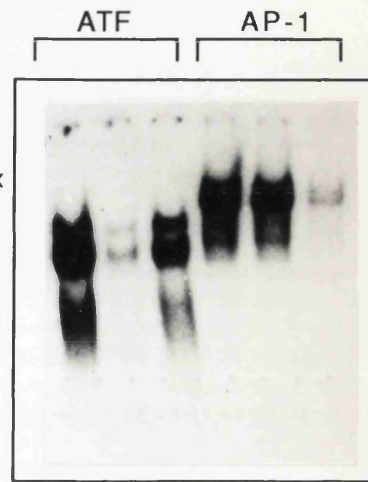
**Fig. 4.1.3. Detailed analysis of  $\gamma$ ATF and  $\gamma$ AP1 binding specificity**

**A.**



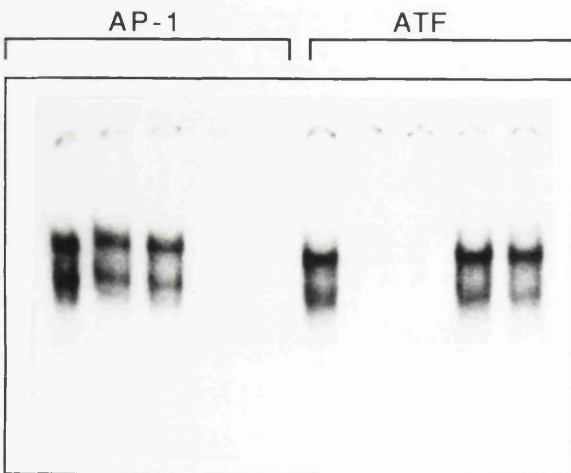
- E4 Som SV Col - E4 Som SV Col  
1 2 3 4 5 6 7 8 9 10

**C.**



- E3 E3m - E3 E3m  
1 2 3 4 5 6

**B.**



- E4 Som SV Col - E4 Som SV Col  
1 2 3 4 5 6 7 8 9 10

**D.**

E4 -50 to -42 TACGTCAT  
Som -42 to -49 GACGTCAG  
E3 -60 to -53 TTCGTCAC

Consensus TTCGTCA  
GA

SV40 120 to 113 TTAGTCAG  
Col -72 to -65 TGAGTCAG  
E3mt TTAGTCAC

Consensus TTAGTCA  
G

budding yeast extracts. An oligonucleotide containing the C to A change in the E3 ATF site (E3m) competed very poorly for yATF binding when compared to the wild-type site (lanes 2,3, Fig.4.1.3.C). In contrast E3m was an efficient competitor for yAP1 binding compared to the wt E3ATF site. The single basepair change therefore resulted in a switch in binding specificity from yATF to yAP1, demonstrating that the nucleotide requirements for both these factors has been conserved from yeast to humans.

## 7. The yATF binding site can act as an efficient UAS

I investigated the ability of the ATF binding site to act as a transcriptional activating sequence *in vivo*. In budding yeast a sensitive assay for UAS activity has been developed and involves inserting the sequence of interest into a basal promoter and monitoring expression levels as compared to the basal promoter (Guarente and Mason 1983). Oligonucleotides containing three tandem copies of the wild-type E4 ATF binding site or mutant sites were cloned upstream of the *S.cerevisiae* cytochrome c (CYC1) TATA element fused to the  $\beta$ -galactosidase marker gene (Guarente and Mason 1983).  $\beta$ -galactosidase enzyme levels were measured in exponentially growing cultures of cells transformed with the various plasmids. Insertion of wild-type ATF sites (pB3E4) resulted in an approximate 70-fold stimulation of  $\beta$ -galactosidase expression as compared to the TATA- $\beta$ -gal gene containing no insert (pBT) (see Fig. 4.1.4.A). To test whether this stimulation was due to yATF binding, two different mutant templates were used, M1 and M2. Competition binding experiments showed that both templates were defective for yATF binding, M1 being more defective than M2 (lanes 3, 4, Fig. 4.1.4.B). M1 was also more defective than M2 in binding human ATF factor (lanes 7 and 8, Fig.3B), once again emphasising the conserved binding specificity between the human and yeast factors. Both mutant binding sites were defective in stimulating  $\beta$ -galactosidase expression compared to the wild-type site. Three copies of the M2 site, partially defective for yATF binding, stimulated expression approximately 10-fold (pBE4M2) above basal levels whereas three copies of the M1 site, a highly defective binding site, failed to stimulate expression (pBE4M1) (see Fig. 4.1.4.A). Therefore, the relative ability of the three oligonucleotides to stimulate  $\beta$ -galactosidase expression correlated with their ability to bind yATF, arguing strongly that yATF can mediate transcriptional activation.

Further experiments addressing the possible role of cAMP in the regulation of this yeast factor are addressed in the next chapter and the Discussion.

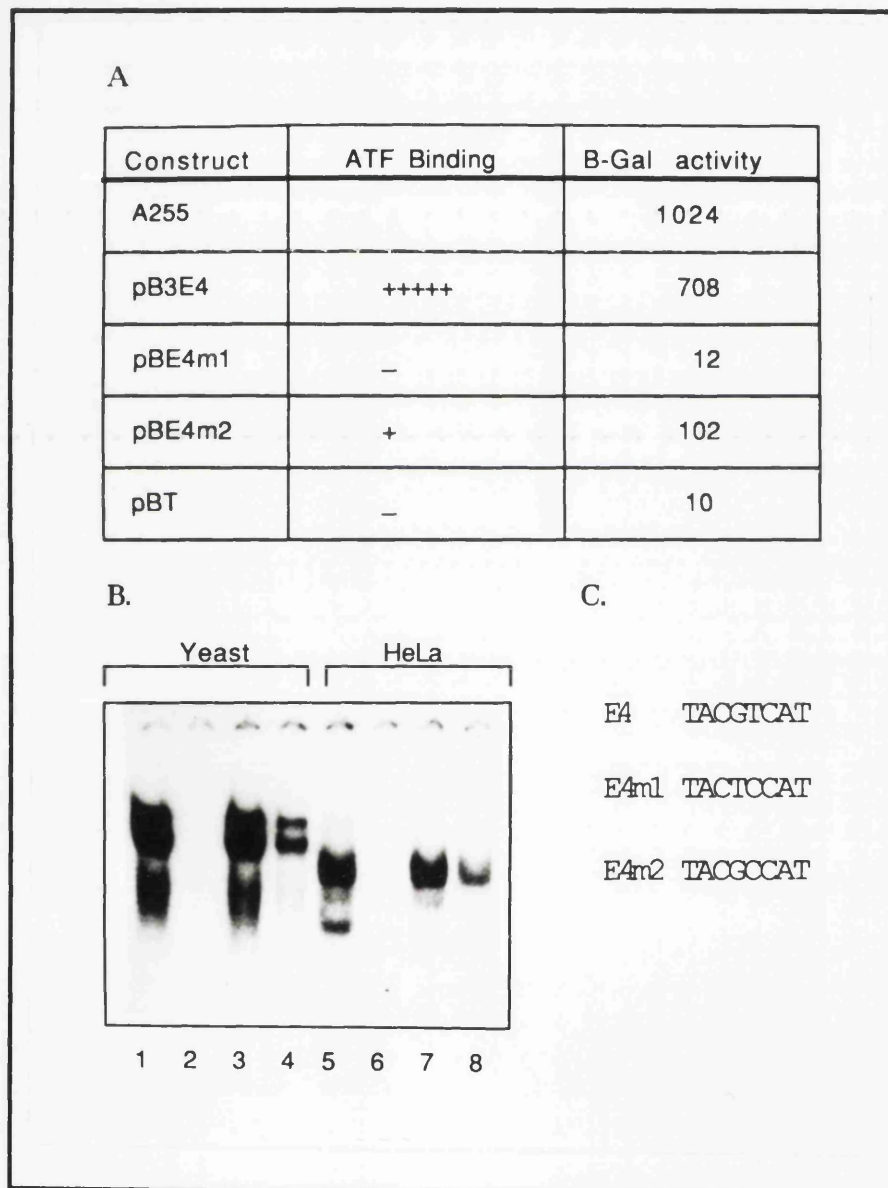
**Figure. 4.1.4. The yATF binding site is an efficient UAS sequence.**

**A.** Oligonucleotides containing ATF wild type or mutant binding sites were inserted upstream of a CYC1:lacZ fusion gene as described in materials and methods. Plasmid pBT contains no UAS. Plasmid pB3E4 contains three tandem copies of the E4 ATF binding site and plasmids pBE4M1 and pBE4M2 contain three tandem copies of single point mutated ATF sites. Yeast strains were transformed with these constructs as well as a construct (A255) containing the wild-type CYC1 UAS elements.  $\beta$ -galactosidase activity was quantitated for cultures grown from single colonies as described in materials and methods.

**B.** The ability of the mutant E4 ATF binding site to bind yATF factor was tested by the gel retardation assay. *S.cerevisiae* or HeLa cells extracts were incubated in a standard binding reaction with radiolabelled oligonucleotide probe containing the E4 ATF binding site without (lanes 1,5) or with a 50-fold molar excess of unlabelled oligonucleotide containing the wild-type E4 ATF site (lanes 2,6), M1 ATF site (lanes 3, 7) or M2 ATF site (lanes 4, 8).

**C.** The core sequence of the wild-type and M1 and M2 mutant ATF binding sites are shown.

**Fig. 4.1.4. ATF binding sites can act as a UAS in budding yeast**





## **CHAPTER 5; RESULTS**

### **PURIFICATION OF BUDDING YEAST AP1 AND ATF.**

## 1. Introduction and summary.

The previous chapters have described the characterisation of AP1 and ATF like activities in both fission and budding yeast. Due to the importance of these proteins as transcription factors in mammalian cells it was of interest to gain further information on the yeast factors. We therefore decided to purify both yAP1 and yATF from S.cerevisiae in an attempt to gain enough pure protein to obtain peptide sequence information. Peptide sequence information can be used to generate degenerate oligonucleotides which are suitable tools for obtaining DNA clones (see Gonzalez et al 1989a for example). For this protocol we could use S.cerevisiae or S.pombe as a starting material. At this point the only criteria by which the yeast proteins had been shown to be homologous to their human counterparts was in their DNA binding specificity. In this respect there was little to choose between S.pombe and S.cerevisiae (see fig.4.1.3 chapter 4) and therefore there was no reason to favour either system. However a number of practical considerations made S.cerevisiae a more favourable candidate.

1. To purify sufficient quantities of transcription factor to obtain sequence information requires a large amount of starting material. At the time this project was considered there was an immediate supply of S.cerevisiae available from Dr. Colin Young at the N.I.M.R, Mill Hill, London.
2. During the early stages of extract preparation and purification handling large quantities of yeast is very difficult but becomes more manageable with the help of other people. In using the budding yeast system I was able to collaborate with other workers interested in unrelated transcription factor proteins (see materials and methods). Thus we were able to cooperate during these early stages.
3. As a result of this collaboration I was able to use chromatographic equipment set up at the MRC laboratories in Cambridge. This included both Heparin agarose and DNA sepharose columns which had already been successfully employed to isolate yeast transcription factors (eg Sorger and Pelham 1987).
4. A obvious problem in protein purification protocols is proteolysis of the factor of interest. The availability of a S.cerevisiae strain that was deficient of multiple proteases (BJ2168; Sorger and Pelham 1987) was therefore an obvious advantage.

Starting material for the purification of ATF and AP1 came in two forms.

1. Heparin agarose column fractions derived from 2kg of yeast which had not been processed further from a previous extract preparation.
2. Frozen yeast cells which had not been processed at all.

The purification procedure was essentially as Sorger (1987b) involving a three step process.

Whole cell extracts were initially subjected to salt gradient elution Ion exchange chromatography by fractionation over a heparin agarose column. This step proved to separate most of the  $\gamma$ ATF from most of the  $\gamma$ AP1 enabling further purification to deal with each factor separately. Peak activities from this separation were loaded onto a non specific DNA column before specific oligonucleotide affinity chromatography was carried out.

The purification involved processing this starting material in three batches. As the procedures employed and results obtained were essentially the same each time, I present data from only the first of these as an example analysis.

## **2. Purification of $\gamma$ AP1.**

### **2.a. Heparin Agarose Chromatography;**

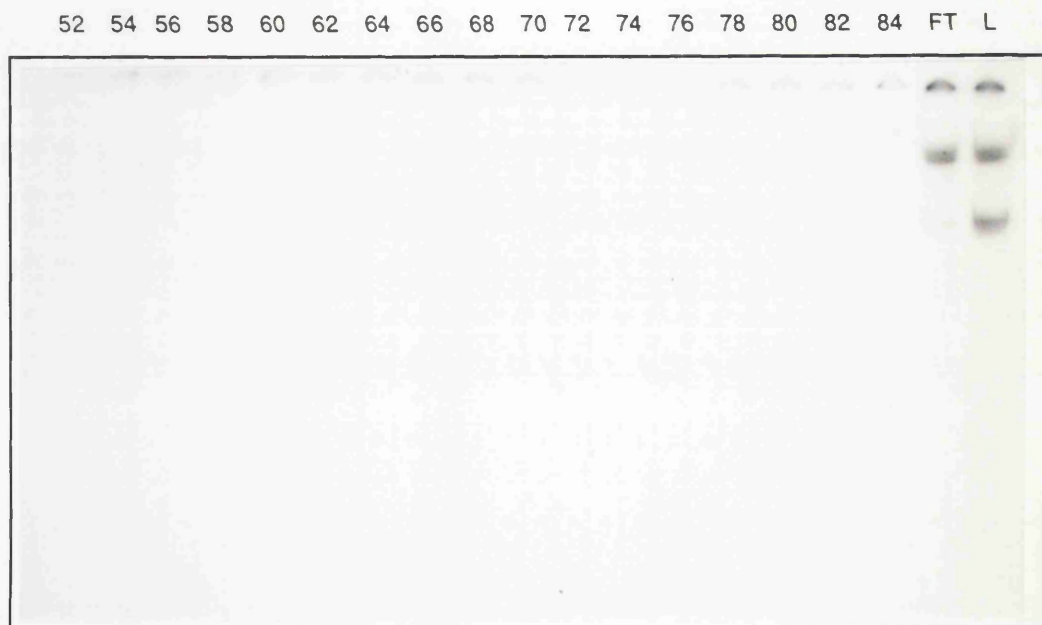
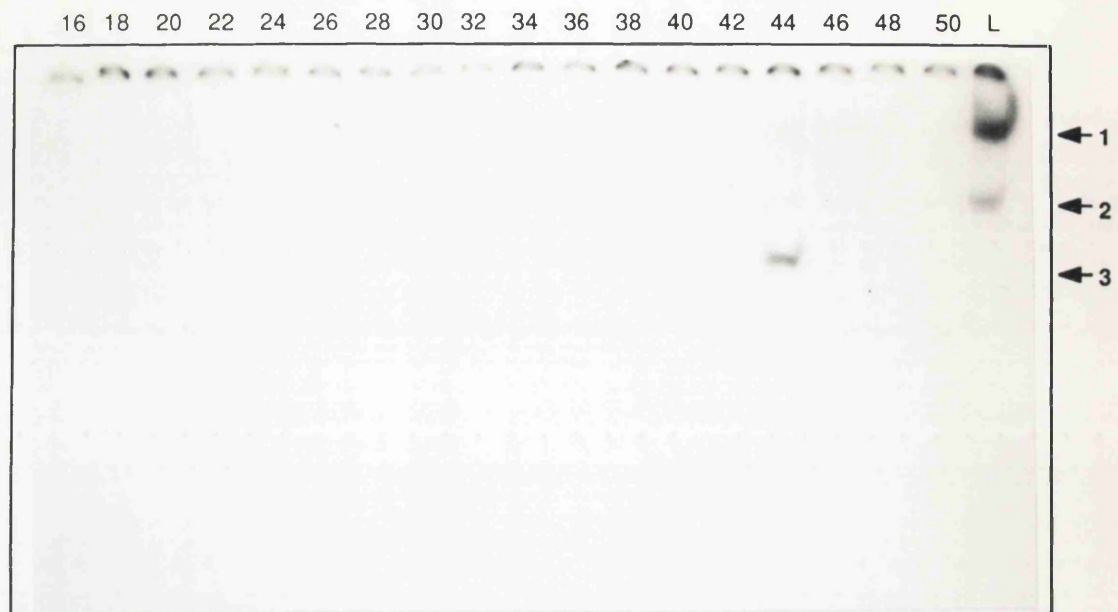
Ion exchange separates biomolecules based on differences in their charge characteristics and is a useful first separation step as large columns can be easily and cheaply prepared to allow a high sample capacity. Heparin itself is an anionic linear polysaccharide which when coupled to agarose it provides a cationic exchange matrix. (Silva and Dietrich, 1975).

An ammonium sulphate salt gradient of 50mM to 600mM was used to elute protein fractions from the heparin agarose column. Having originally characterised  $\gamma$ AP1 by its ability to bind the SV40 AP1 site, a labelled oligonucleotide probe bearing these sequences was used to detect binding activity of the fractions in a gel shift assay. Fig. 5.1.1. shows that very little  $\gamma$ AP1 activity appears to have been eluted in any of the fractions. To simplify the issue I have labelled the activities 1, 2, and 3. A faint peak of activity which comigrates with the major form of  $\gamma$ AP-1 activity, as seen in crude extracts (activity 1), appears in fractions 40 to 46 (240mM-290mM ammonium sulphate); however most of this activity appears in the flow through (FT). A rapidly migrating band (activity 3) which was not observed in crude cell extracts can be seen to peak in lane 44. In order to test the specificity of all these bands oligonucleotide competition analysis was carried out. Fig. 5.1.2, lanes 4-6, confirms the activity 1 observed in both the flow through and fraction 44 is specific, being competed efficiently by an AP1 site (lanes 2 and 5) but not by a nonsense oligonucleotide (lanes

**Fig. 5.1.1. First round heparin agarose purification of  $\gamma$ AP1.**

A selection of column fractions representing the elution profile of the 50-600mM ammonium sulphate gradient were assayed by standard gel shift analysis. 5 $\mu$ l of each fraction was incubated with the SV40 AP1 site probe in the presence of 2 $\mu$ g of poly (dI;dC). Numbers above each lane represents the number of the column fraction assayed. FT is the flow through fraction from this column and L is what was loaded on. According to the salt concentration of the fraction tested the salt in the binding buffer was varied. Between 50-200 mM (fractions 16-32) a 100mM NaCl binding buffer was used, between 200-400 mM (fractions 33-62) a 50mM buffer, and between 400-600mM (fractions 62-84) a 0mM buffer. Both FT and load contained a 100mM based buffer. The figure represents an eight hour exposure of this experiment. Major  $\gamma$ AP1 mobility forms of activity are numbered as 1, 2, and 3 on the right of the figure.

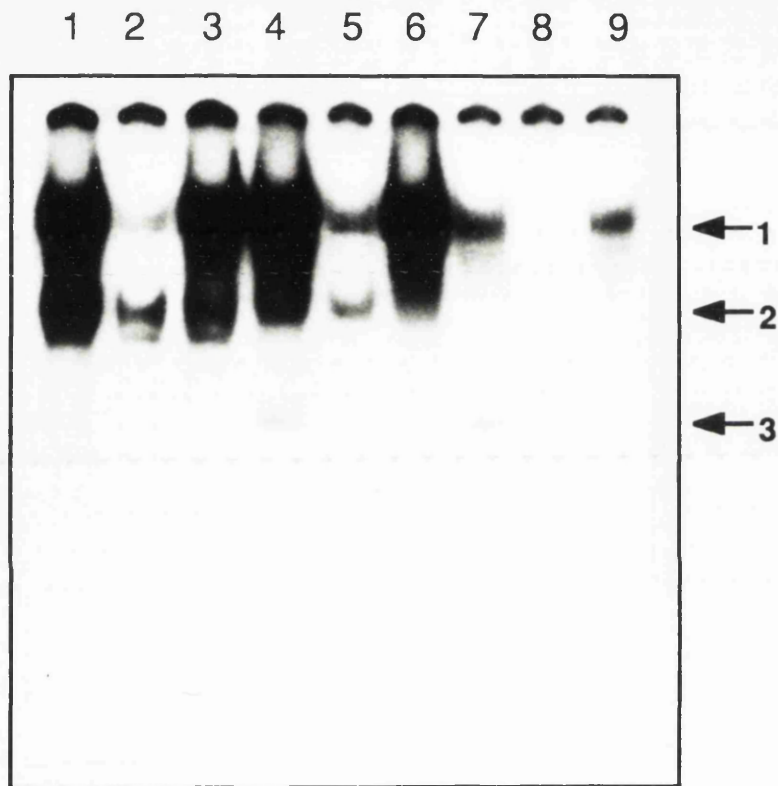
Fig. 5.1.1. First round Heparin Agarose Purification of  $\gamma$ AP1.



**Fig. 5.1.2. Competition analysis of  $\gamma$ AP1 heparin agarose activity.**

Apart from crude extract (lanes 1-3) and the flow through fraction (lanes 4-6), only fraction 44 (lanes 7-9) was analysed. To the standard reaction a 100 molar excess of either cold SV40 oligonucleotide site (lanes 2, 5, and 8) or a nonsense oligonucleotide (lanes 3, 6, and 9) was added. The gel was exposed for 24 hours. Once again the different mobility forms of  $\gamma$ AP1 are labelled on the right.

**Fig. 5.1.2. Competition analysis of  $\gamma$ AP1 Heparin Agarose fractions**

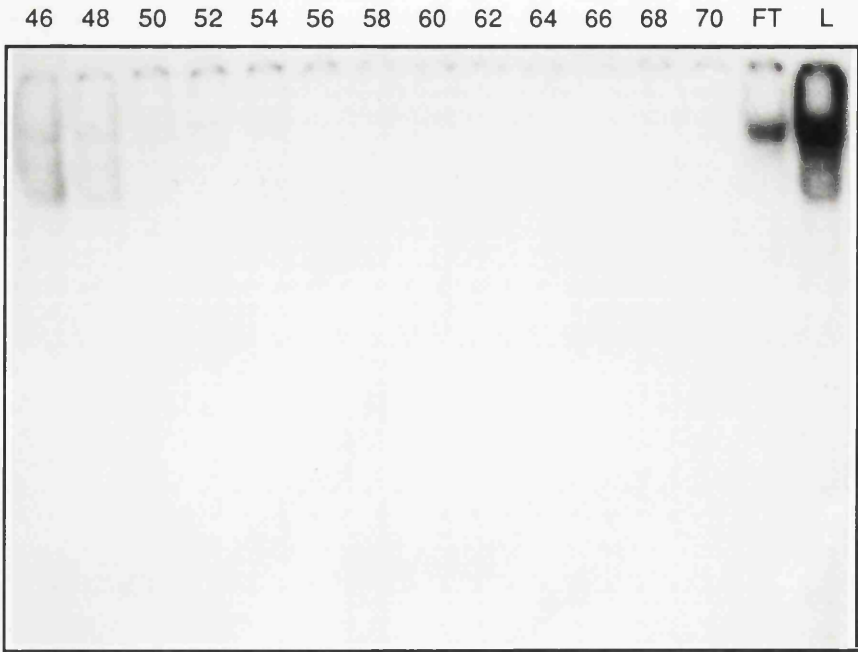
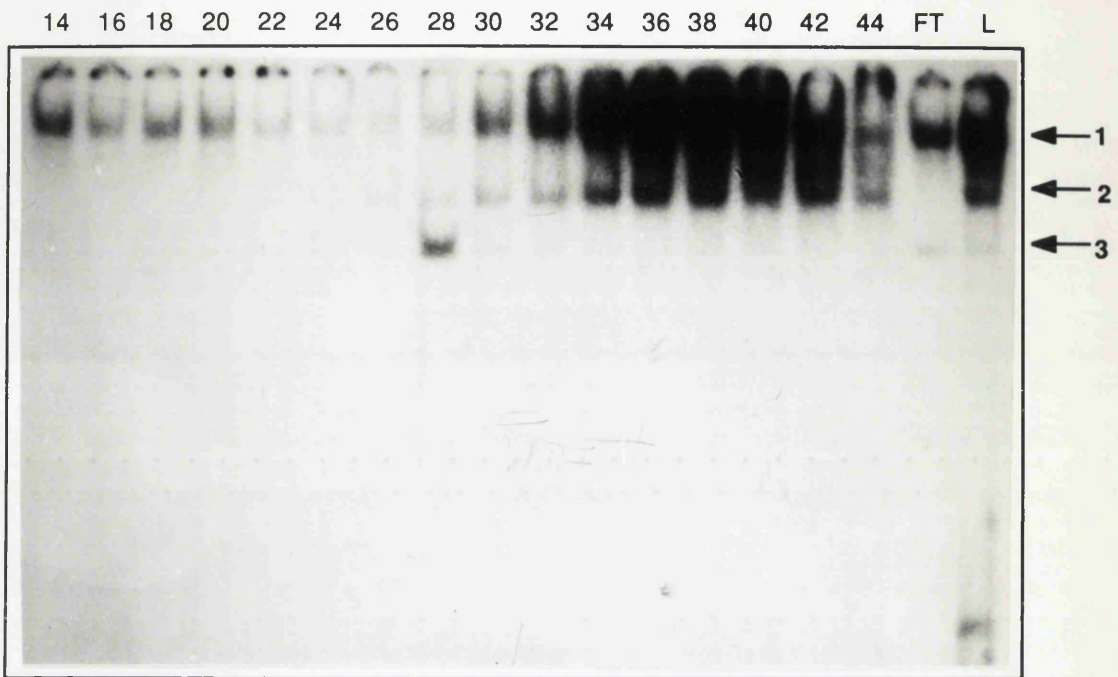


**Fig. 5.1.3. Second round heparin agarose purification of yAP1.**

5 $\mu$ l samples of fractions derived from a second loading and elution of the heparin agarose column were again tested for yAP1 activity. The numbers above each lane indicates the fraction number used in that reaction, FT and L representing flowthrough and load respectively; the reaction conditions themselves varied just as in fig. 5.1.1. In this case it meant a 100mM NaCl based buffer was used for fractions 14-28, a 50mM buffer for fractions 30-44, and a 0mM buffer for fractions 46-70. Dried gels were exposed for approximately 16 hrs.



Fig. 5.1.3. Second Round Heparin Agarose Purification of yAP1.



3 and 6) . By comparison of the same lanes it is also possible to see that activities 2 and 3 are nonspecific. This figure also emphasizes how little eluted specific activity was apparent even in the peak fraction (compare levels of activity 1 in lanes 4 and 7). Thus it appears that the majority of yAP1 was not retained on the initial passage. We postulated that the column may have been overloaded on this initial passage and so the flow through fractions were pooled and reloaded onto the same column. Retained protein was eluted using a similar salt gradient as employed for the first column and the fractions again analysed by gel shift assay. Fig.5.1.3 shows that a clear peak of activity was obtained this time with the majority of yAP1 collecting in fractions 32-42 (250mM-350mM), with a reduced activity in the flow through. A small amount of activity was also seen in fractions 14 to 20 (<100mM). However fractions 30 to 43 inclusive were judged to contain the main peak of yAP1 and were therefore pooled ready for further purification. This pool was ammonium sulphate precipitated, resuspended in A75 and dialysed against the same buffer ready for application to the DNA-sepharose column.

## **2.b. DNA-sepharose;**

This step in the purification was employed to try and separate proteins which bind tightly but relatively non specifically to DNA from specific transcription factors which bind tightly only to their cognate site.

The pooled samples derived from the heparin agarose column were loaded on to the column, washed and eluted with a linear ammonium sulphate gradient of 75mM to 600mM and collected as 10ml fractions. The samples were again tested with a gel shift assay to detect peak activities. Fig.5.1.4.A. shows the the main peak of eluted activity appears in fractions 12-19 (125mM-190mM) but that the mobility of the eluted activity appears to increase with salt concentration within these fractions. The specificity of the different mobility forms was therefore tested. Fig. 5.1.4.B. lanes 1-3 and 4-6 confirms that both upper and lower forms are specific, being efficiently competed by an AP1 site (lanes 2 and 5) but not by a nonsense site oligonucleotide (lanes 3 and 6). A similar comparison<sup>n</sup> also demonstrates that the activity in lanes 25-31 (fig.5.1.4.A) is nonspecific (see lanes 7-9, fig. 5.1.4.B). The reason why yAP1 activity manifested itself in multiple mobility forms was not clear, but in case this was important I did not pool these peaks fractions at this stage. Instead each was considered separately for oligonucleotide affinity purification.

## **2.c. Oligonucleotide affinity chromatography;**

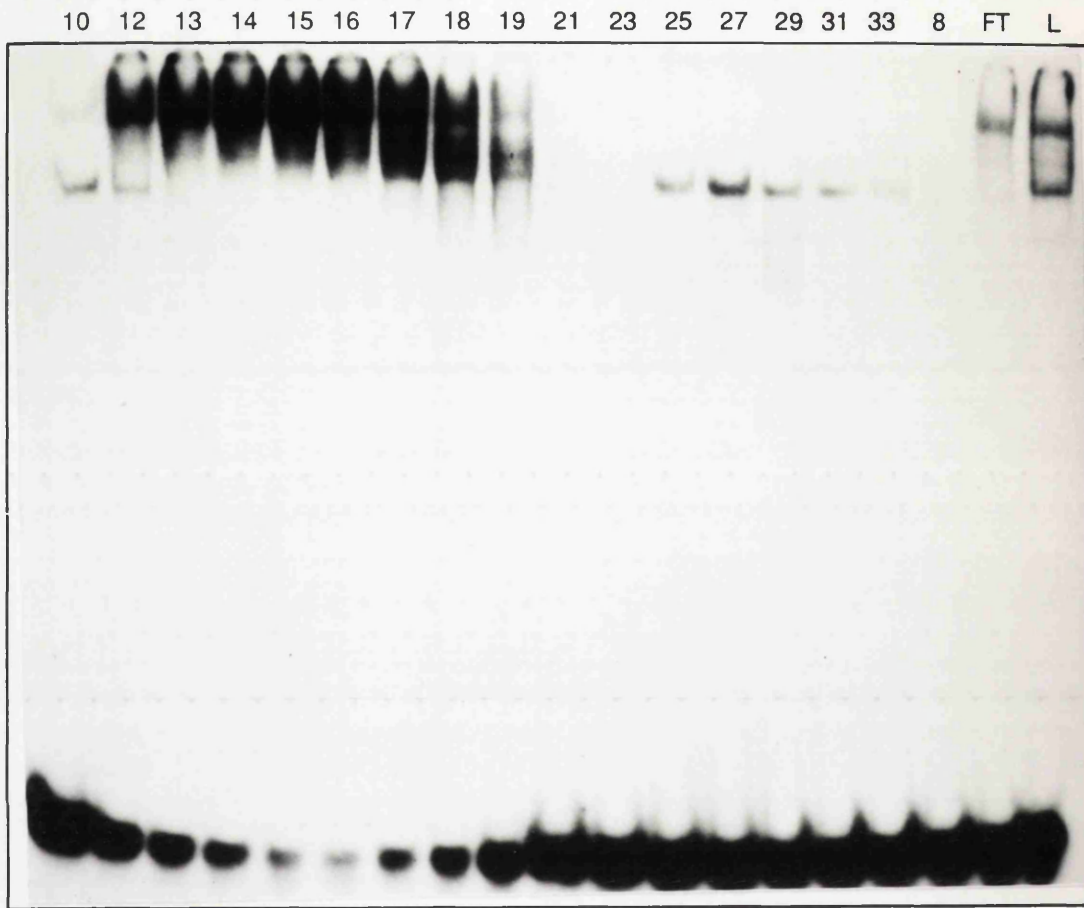
**Fig. 5.1.4. DNA sepharose purification of yAP1.**

**A.** Peak activities from the second heparin agarose column were loaded on to and eluted from a DNA-sepharose column as described in the text. 2  $\mu$ l samples of fractions representing the elution profile were then tested for yAP1 activity in the presence of 1  $\mu$ g of poly (dI:dC) competitor. The same criteria as in fig 5.1.1 was used to determine the salt concentration of the binding buffer. This meant fractions 8-19 were based in 100mM NaCl and fractions 21-33 in 50mM. Once again the numbers above each lane represent the number of the fraction, with FT and L being flowthrough and load material respectively. The figure represents a 4 hour exposure.

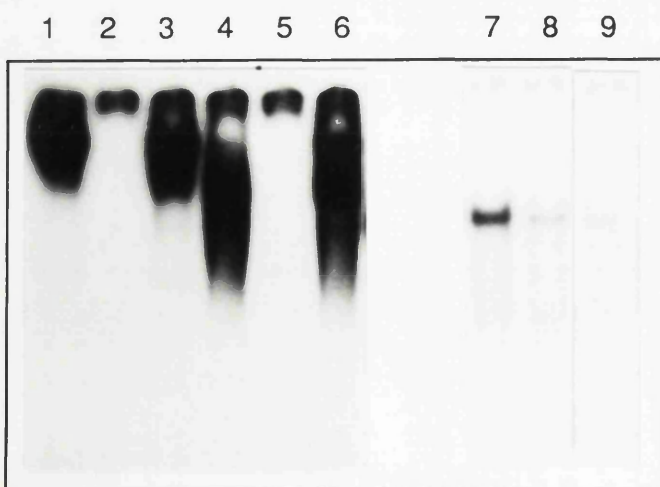
**B.** Competition analysis was carried out on various peak activities derived from the DNA sepharose column. Lanes 1-3 contain extracts from fraction 14, lanes 4-6 extracts from fraction 19, and lanes 7-9 extracts from fraction 27. Assays were carried out as fig.5.1.4.A. except that each reaction contained 20  $\mu$ g of BSA. In addition a 100 molar excess of competitor oligonucleotides bearing SV40 site (lanes 2, 5, and 8) or nonsense sequences (lanes 3, 6, and 9) were also used.

Fig. 5.1.4. DNA Sepharose purification of yAP1.

A.



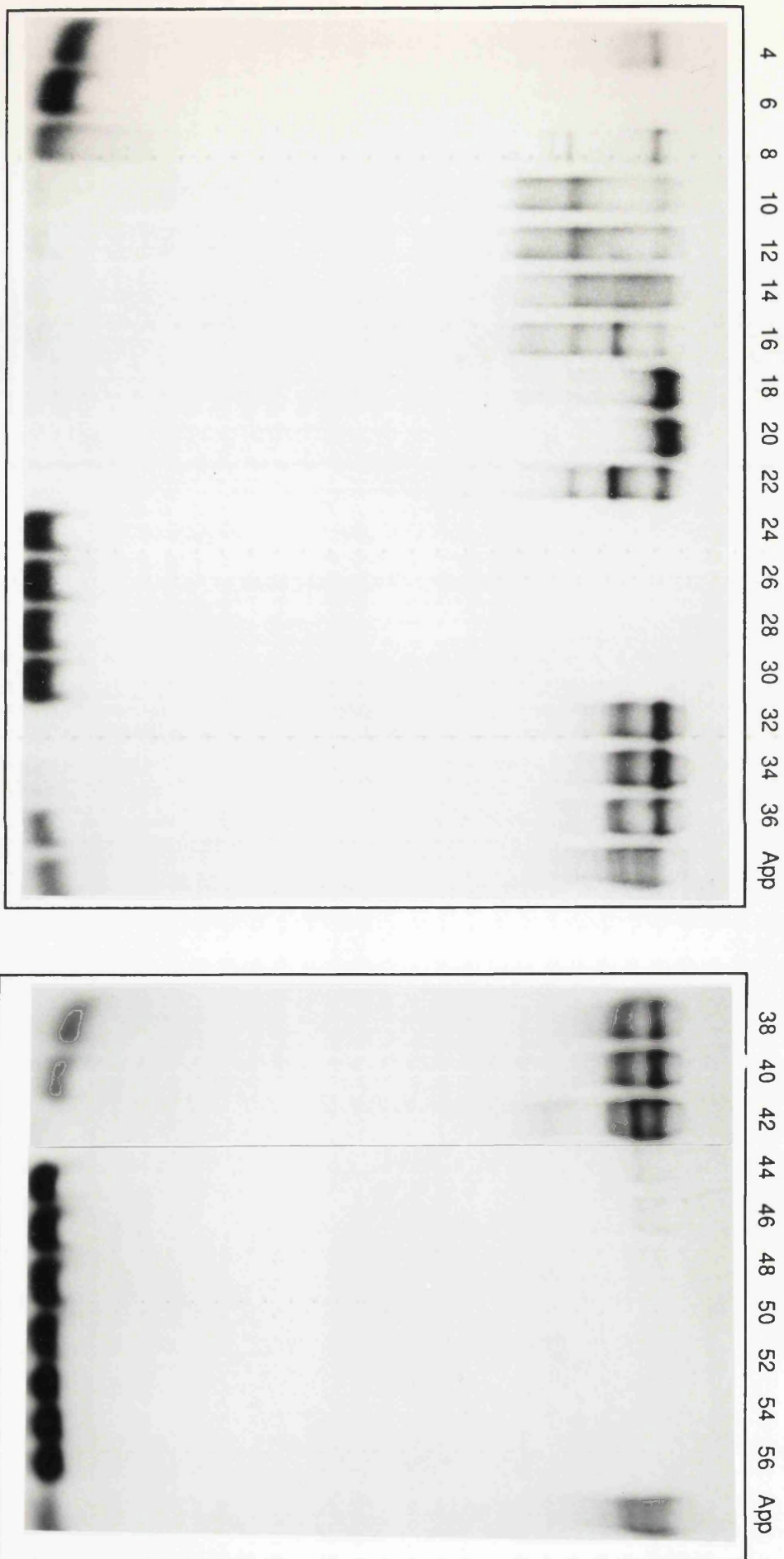
B.



**Fig. 5.1.5. Affinity purification of  $\gamma$ AP1.**

1 $\mu$ l samples of fractions obtained from a stepwise elution of an AP1 site affinity column were tested for specific activity. Fraction numbers tested are indicated above each lane in the figure and 'app' represents 2 $\mu$ l of what was applied to the column. Because a suitable conductivity meter was unavailable to test fractions eluted from the column, salt concentrations in this assay are estimates. However the estimates are: 0.1M wash, fractions 4-16; 0.2M wash, fractions 18-30; 0.4M wash, fractions 31-40; 0.8M wash, fractions 41-50; 2M wash 51-58. In the binding assays both the salt concentration and the amount of nonspecific competitor was varied. Lanes 4-30 contained 100mM NaCl, lanes 31-50 contained 50mM NaCl, and lanes 52-56, 0mM NaCl. In the other variable, 0.1 $\mu$ g of non specific DNA was added to lanes 4-30 and 50ng to each of the rest. In addition 20 $\mu$ g of BSA was added to each binding assay as protein concentration were very low in a number of cases. The photograph represents a 2 hour exposure.

Fig. 5.1.5 Affinity purification of YAP1

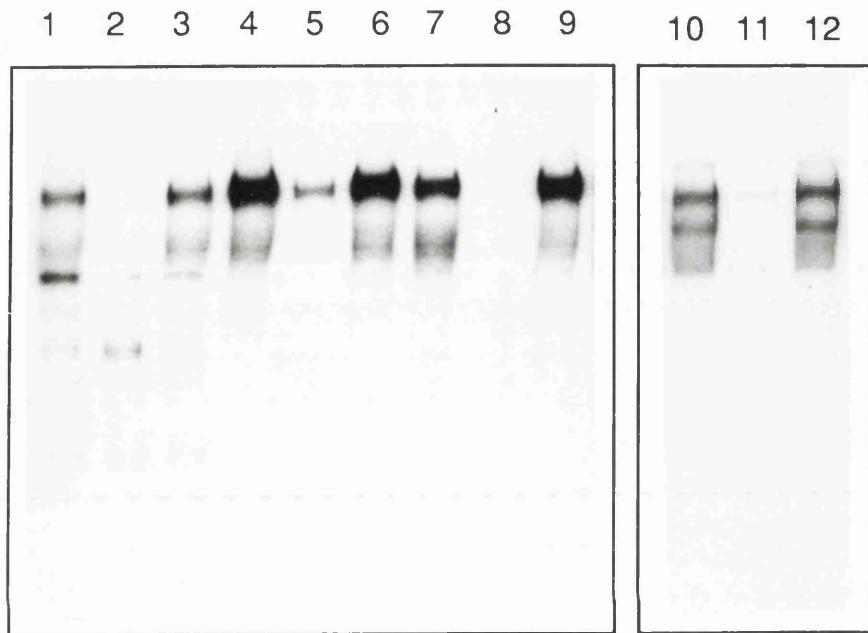


**Fig.5.1.6. Competition analysis of affinity purified yAP1.**

Binding activity observed in a number of the affinity column fractions were tested for absolute specificity. Lanes 1-3 contain 1 $\mu$ l fraction 22, lanes 4-6 contain 1 $\mu$ l fraction 34, lanes 7-9 contain 1 $\mu$ l fraction 38 and lanes 10-12 contain 0.2 $\mu$ l fraction 42.

Reaction conditions employed were the same as those outlined for the same fractions in fig.5.1.5. except a 100 molar excess of either cold SV40 AP1 oligo (lanes 2, 5, 8, and 11) or the E4 ATF oligo (lanes 3, 6, 9, and 12) were added. The photo represents a 2 hour exposure.

Fig. 5.1.6. Competition analysis of affinity purified yAP1.





Having largely removed proteins which bind DNA with a high affinity but nonspecifically this step is employed to separate proteins which recognise a specific oligonucleotide with a far greater affinity than they recognise random DNA. Theoretically it should act as a very powerful purification step and indeed it is normally extremely successful (See Treisman 1988 and Sorger 1987b for examples). In an attempt to apply this to yAP1 6mls of fraction 16 (see fig.5.1.4.A) was loaded onto an affinity column containing multimerised copies of the SV40 AP1 site covalently linked to sepharose (as described in Materials and Methods). The column was washed extensively in 100mM NaCl column wash buffer and fractions then eluted with a NaCl step gradient. All fractions from this column including the flow through were collected as 0.5-0.6 ml aliquots in 1.5 ml eppendorf tubes. A suitable conductivity meter was not available to test salt concentration of fractions derived from this column. Therefore estimations were made as to where the step elutions emerged from the column. Subsequent to fraction collection, samples were tested by gel shift assay. Fig.5.1.5. shows that DNA binding activity was observed in a number of fractions. To test the specificity of activity in the higher salt elutions, fractions 22, 34, 38, and 42 were challenged with an excess of cold oligonucleotide containing the same sequences as the probe or sequences derived from the ATF binding site of the adenovirus E4 promoter. From previous analysis yAP1 specific activity should only be competed efficiently by the AP1 site (see chapter 4). Fig.5.1.6. shows the results of this assay. By this criteria activity in fraction 22, representing a 0.2M NaCl salt elution, appears to be a mixture of specific and nonspecific binding (Compare lanes 2 and 3). The specific activity comigrates with both upper and lower forms of yAP1 (as characterised from previous analysis). However fractions 34 (compare lanes 5 and 6), 38 (compare lanes 8 and 9), and 42 (compare lanes 11 and 12), representing 0.4M and 0.8M elutions, appear to contain only specific yAP-1 activity. Samples from these fractions were then tested for their purity by running a denaturing protein gel and staining it with a colloidal gold protocol (as described in Materials and Methods).

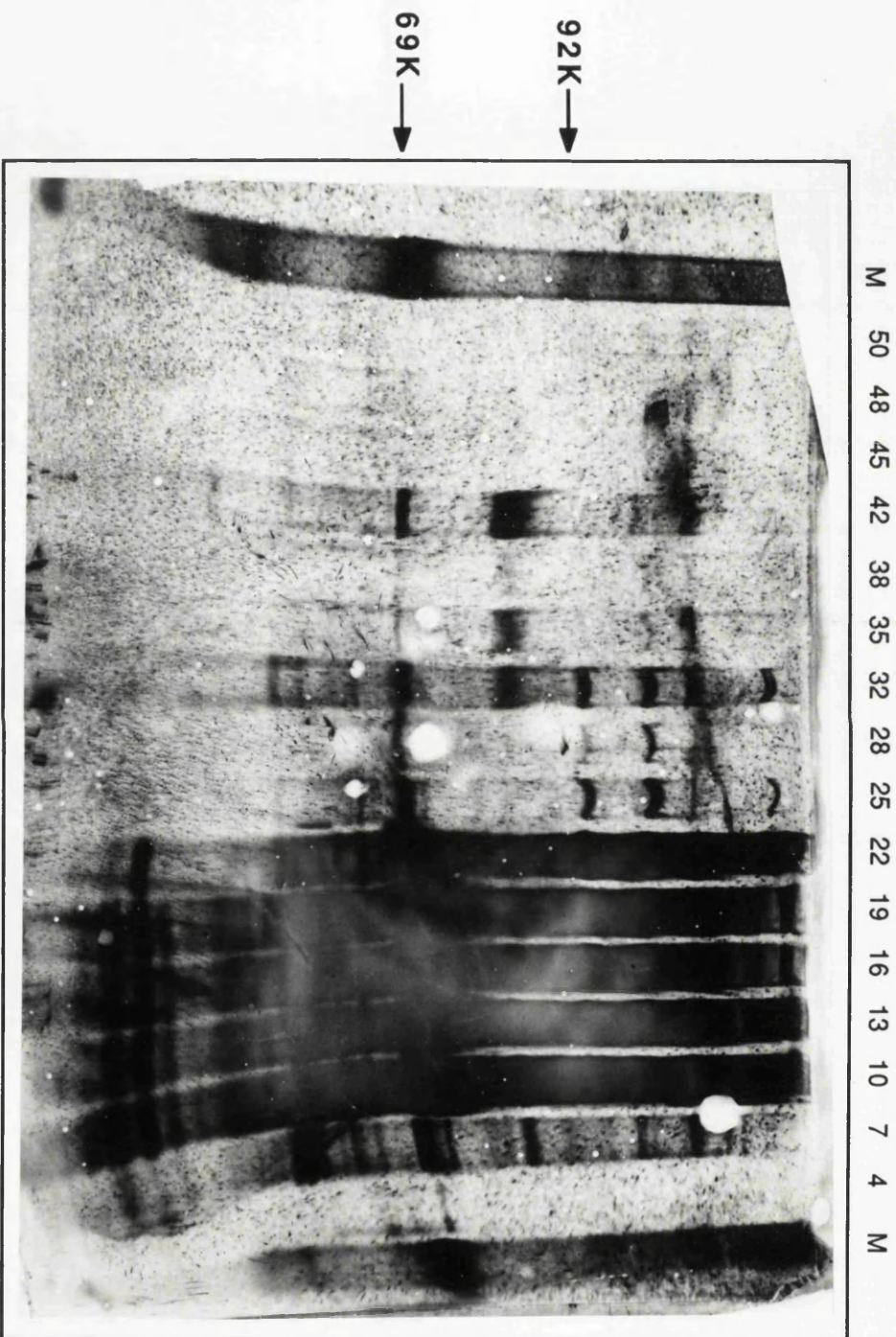
#### **2.d. Colloidal gel staining;**

To try and visualise yAP1 as an actual protein species I carried out colloidal gold staining of selected affinity purified fractions that had been run on a denaturing acrylamide gel. Colloidal gold staining initially involves blotting the protein gel onto nitrocellulose and then treating the nitrocellulose filter. Fig.5.1.7. shows the results of this analysis. As expected the low salt elutions (lanes 7-22) contain multiple proteins which cannot be easily identified as separate species by the staining protocol.

**Fig. 5.1.7. Colloidal gold staining of affinity purified yAP1.**

To identify yAP1 as a protein species 40µl samples of affinity purified fractions were run on a protein gel and visualised by colloidal gold staining as described in materials and methods. In this figure the numbers above the lanes again correspond to fraction numbers and are directly comparable to the gel shift assay data. Markers are loaded on both flanks of the gel and the important marker sizes are indicated on the left side of the figure.

Fig. 5.1.7. Colloidal Gold Staining of Affinity purified yAP1 fractions.



However in the 0.4M elution (eg lane 32 and 35) and the 0.8M elution (lane 42) individual species could be visualised, the abundance of which could be compared to the binding activity in the gel shift assays to identify potential yAP1 candidates. The most striking band observed by colloidal gold staining, whose abundance correlates with gel shift activity, runs as a 85kd species (see lanes 42 and 35 fig.3.1.8.). This is therefore a likely yAP1 candidate but there is another band of approximately 150kd which should also be considered.

### **2.e. Southwestern analysis;**

In order to help with this analysis one protein gel (see fig.5.1.8) was stained with colloidal gold and also subjected to southwestern analysis. Southwestern analysis involves blotting a protein gel onto nitrocellulose and probing the nitrocellulose with a labelled binding site (see chapter 6 and materials and methods for details). Southwestern analysis with DNA sepharose purified yAP1 (fig.5.1.8. lane 2) gives rise to only a single band of ~85kd when probed with the SV40 AP-1 site, with no signal corresponding to a 150kd protein. This species exactly comigrated with the 85kd band observed by colloidal gold staining (lane 1) and this data therefore supports the notion that yAP1 itself is 85kd. Budding yeast yAP1 has also been identified and purified as a 85kd species by Harshman et al (1988). This is essentially in agreement with the observations published by Harshman et al (1988). However these workers also see lower molecular weight species of 40-45kd. Whether these lower molecular weight species are really different proteins or simply degradation products of the 85Kd species is presently not clear.

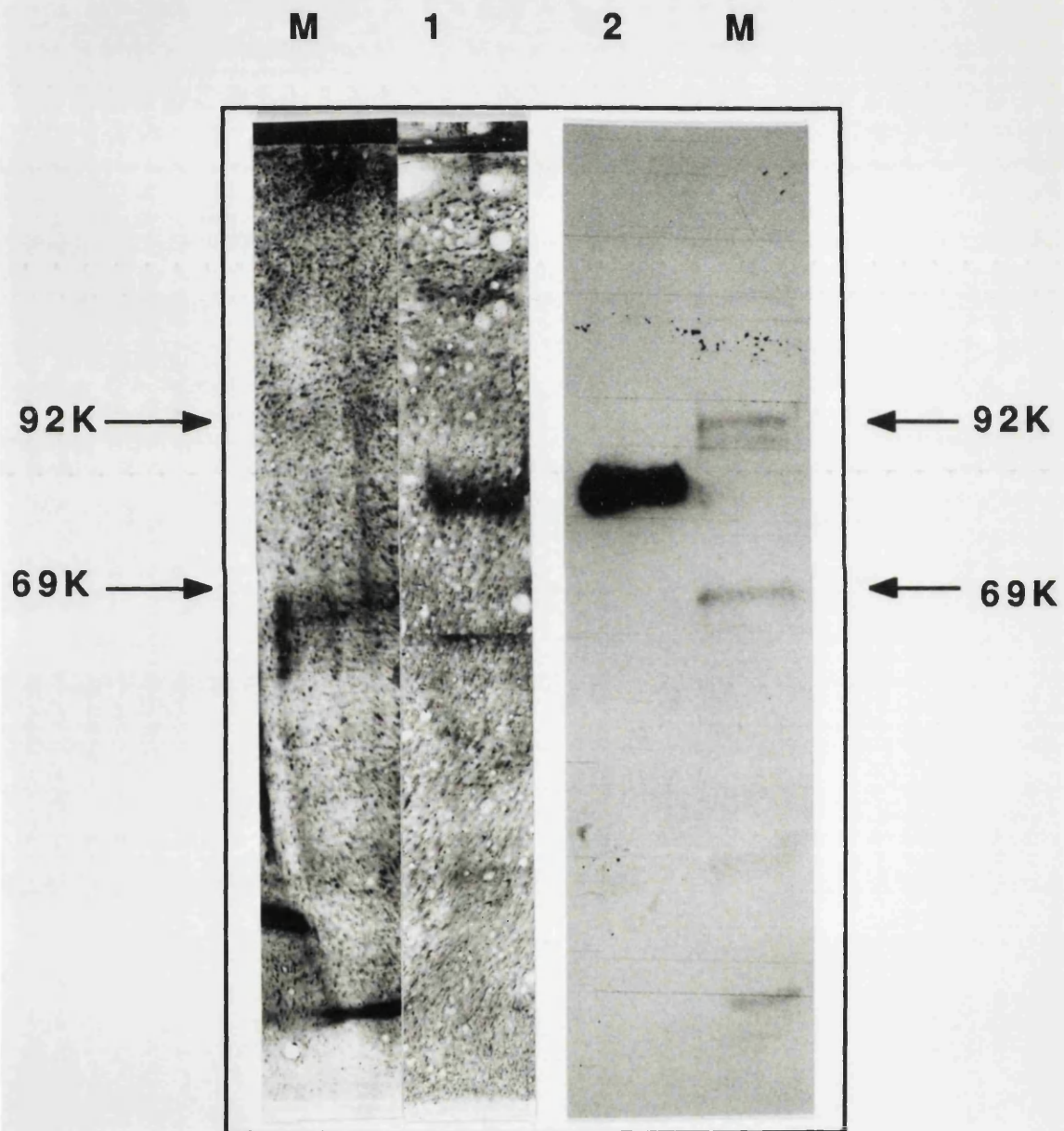
Another point to note from the southwestern analysis is that the fraction tested is fraction 17 (fig. 5.1.4.A) which contains both high and low mobility forms of the yAP1 activity. By southwestern analysis only a single sized binding species can be visualised. This argues that the DNA binding component in all these forms is the same. It is possible that different binding forms are differentially amenable to visualisation by southwestern analysis but a more likely explanation is that they represent different kinds of protein-protein interactions. Whether these interactions are multimers of the 85Kd species alone or involve other proteins, possibly the 150kd species, is presently unclear.

### **2.f. Peptide sequencing;**

**Fig 5.1.8. Comparison of southwestern and purified yAP1.**

Southwestern and colloidal gold staining. A single protein gel was blotted and divided into two equal sections. The 2 left hand lanes have been stained with colloidal gold and the two right hand lanes have been subjected to southwestern analysis as described in materials and methods. Lane 1 contains 30  $\mu$ l of extract from fraction 42 (fig 5.1.5.) and lane 2 contains 20 $\mu$ l of DNA sepharose purified fraction 17 (fig.5.1.4.). M lanes contains markers.

**Fig. 5.1.8 Comparisome of Southwestern  
and purified yAP1**



I proceeded to prepare 10 $\mu$ g of the 85kd species to a purity of that shown in lane 42, fig.5.1.7. This was then given to the protein sequencing facilities at the ICRF operated by Ron Brown and Kate Perks. Although the 85kd species was not the only protein in this preparation it was considered to be sufficiently pure to gain instructive peptide sequence information. Sequences from three peptides was then used to generate degenerate oligonucleotides for Polymerase Chain Reaction (PCR; Saiki et al 1985) analysis. As these experiments were initiated a clone for yAP1 was published (Moye-Rowley et al 1989). Comparison of the peptide sequences from my pure preparation of yAP1 showed one of them had an 8/8 match with amino acids 178-185 of the published sequence and another had a 7/11 match. It was believed that the 4 mismatches may have been due to protein sequencing errors. The last peptide however had no recognisable match with any of the published sequence (see fig5.1.9.). At this point I postulated whether the matches meant that the protein we had isolated was the same as that cloned by Moye-Rowley et al (1989) or a related family member. Homologies that exist between yAP-1 and GCN4 occur in the respective DNA binding domains of these proteins and not elsewhere. However the actual peptide matches we observe do not occur in regions of the yAP-1 clone important for DNA binding. Considering this and the size similarity I concluded that the protein I had purified was almost certainly that purified and cloned by Moye-Rowley et al (1989). We concluded that the peptide sequence which had no match to the predicted sequence of the yAP1 clone was most likely derived from another of the protein species in the purified preparation.

### Fig. 5.1.9. Peptide sequences from purified yAP1.

3 peptide sequences obtained from my purified yAP1 preparations (top) were compared to the predicted amino acid of a *S.cerevisiae* yAP1 clone (bottom) isolated by Moye-Rowley et al (1989)

Peptide 1	<b>G P K V L E Y L A F V</b>
Predicted yAP1 sequence (aminoacids 124-134)	<b>D S K V L E Y L A R R</b>
Peptide 2	<b>L G F D M S A N</b>
Predicted yAP1 sequence (aminoacids 378-385)	<b>L G F D M S A N</b>
Peptide 3	<b>Y P Y F I P Q N P V Q</b>

### **3. Purification of Budding Yeast ATF.**

#### **3.a. Summary;**

The purification of yATF and yAP1 was carried out side by side. Therefore the techniques employed, the salt gradients applied to the columns and the assay method used to detect activity were essentially the same as already described for yAP1. The only difference was that the probe used to detect binding activity in the gel shift assay contained sequences derived from the Adenovirus E4 promoter. This oligonucleotide had already been used to characterise the specificity of yATF binding activity demonstrated in chapter 4. The yATF activity was successfully purified over heparin agarose and DNA sepharose columns, but problems arose when specific oligonucleotide affinity chromatography was attempted.

#### **3.b. Crude extracts and Heparin Agarose purification;**

Incubation of the ATF site in crude whole cell extracts gave rise to four distinguishable complexes (see lane 1, fig.5.2.1.A) and addition of either ATF site (lane 2) or nonsense sequence (lane 3) oligonucleotides demonstrated that the upper most of these was convincingly competed in a specific fashion.

Crude extracts were separated by ion exchange chromatography with a heparin agarose column. Fig.5.2.1.B shows gel shift assay analysis searching for yATF activity in a 50-600mM ammonium sulphate elution profile. In contrast to yAP1, very little specific activity was observed in the flow through fraction (lane ft) and instead there is a strong peak of activity in fractions 28-48 (200mM -300mM ammonium sulphate). This activity again manifests itself as multiple complexes which I have designated ha1, ha2, ha3 and ha4. Competition analysis using fraction 34 demonstrated that only two of these, ha1 and ha2, were specific (see lanes 5-7, fig.5.2.1A). Ha1 comigrated with the specific band observed in crude extracts (compare lanes 4 and 5 fig.5.2.1A) but ha2 appeared to be a novel new form not previously observed. The reason for the appearance of this new complex was not clear. It could represent a proteolytic product, a different protein becoming unmasked, or the dissociation of a multimeric complex into a simpler form. At this stage it was decided to pool the peak fractions and proceed with the purification. Unfortunately, because this work was being carried out as a collaboration it was not possible to take all the fractions desired as there was a certain amount of overlap between peak activities of different transcription factors. For this reason only fractions 34-44 inclusive were pooled.



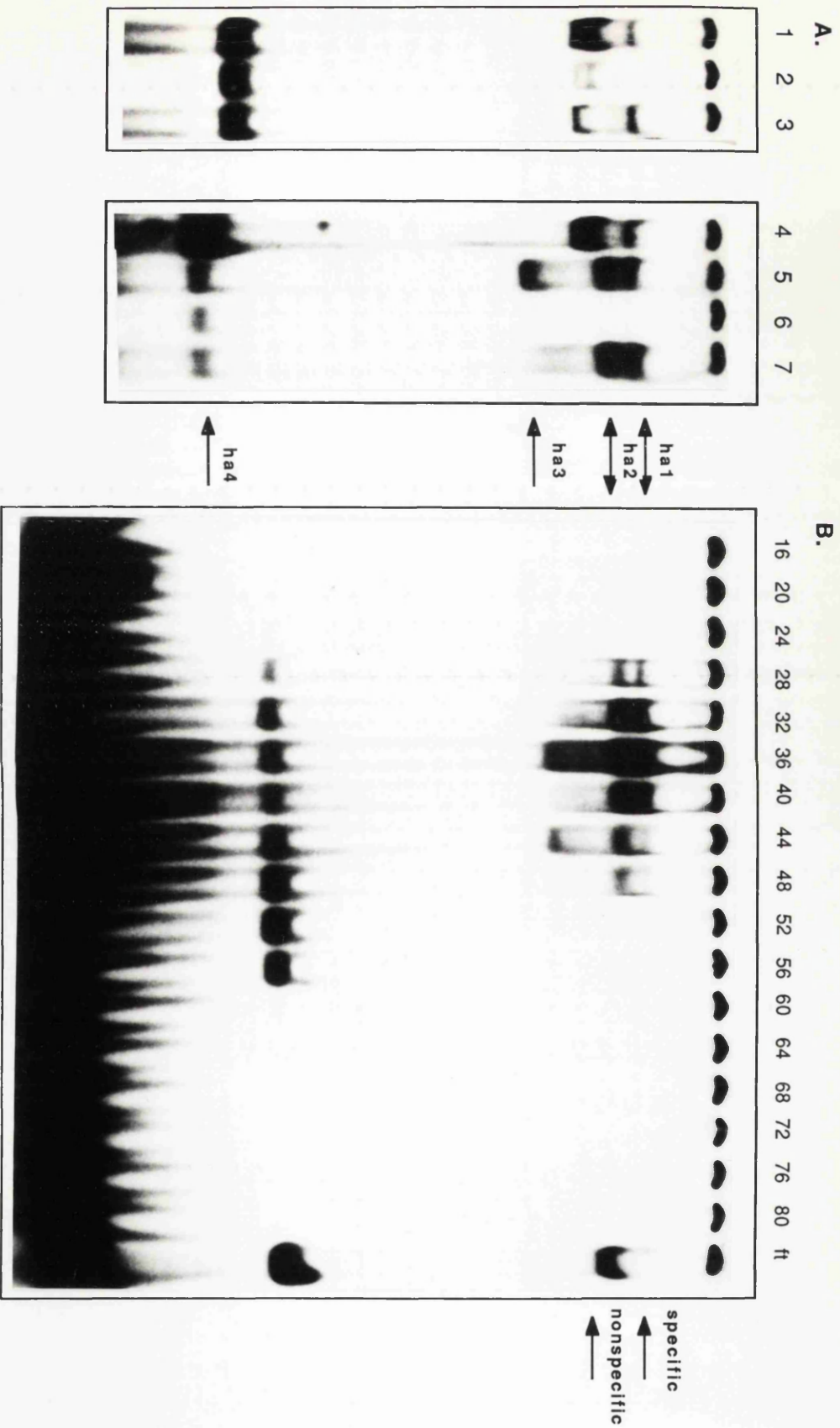
**Fig. 5.2.1. Heparin agarose purification of  $\gamma$ ATF.**

**B.** A selection of column fractions representing the elution profile of the 50-600mM ammonium sulphate gradient were assayed by standard gel shift analysis. 5 $\mu$ l of each fraction was incubated with the E4 ATF site probe in the presence of 2 $\mu$ g of poly (dI;dC). Numbers above each lane represents the number of the column fraction assayed. The fractions are the same as those assayed in fig.5.1.1. and similar variations in gel shift conditions were employed. The arrows labelled specific and nonspecific indicate respective activities in the flow through fraction.

**A.** Testing for specific  $\gamma$ ATF activity. Lanes 1-4 contain 5 $\mu$ l (150 $\mu$ g) of crude extract and lanes 5-7 contain 5 $\mu$ l of heparin agarose fraction 36. Standard binding conditions were used with the addition of a 100 molar excess of cold competitor oligonucleotide containing sequences from the adenovirus E4 site (lanes 2 and 6) or a nonsense site (lanes 3 and 7). Arrows labelled ha -1, -2, and -3 indicate the positions of complexes described in the text.

Both figures are derived from 16 hr autoradiograph exposures.

Fig. 5.2.1. Heparin Agarose purification of YATF



**Fig 5.2.2. DNA sepharose purification of yATF activity.**

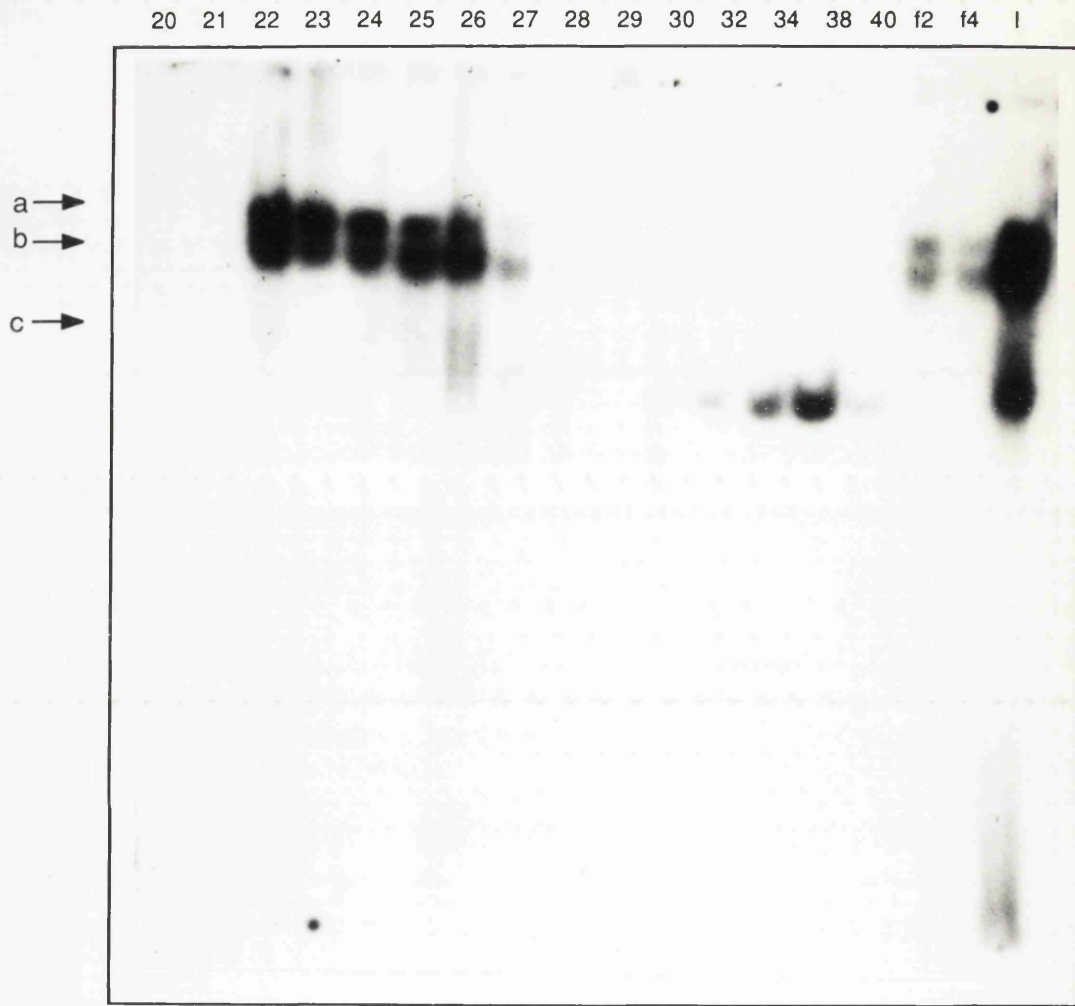
**A.** Available peak activities from the heparin agarose column were loaded on to and eluted from a DNA-sepharose column as described in the text. 2 $\mu$ l samples of fractions representing the elution profile were then tested for yATF activity in the presence of 1 $\mu$ g of poly (dI:dC) competitor. The same criteria as in fig 5.1.1 was used to determine the salt concentration of the binding buffer. This meant fractions 20-26 were based in 100mM NaCl and fractions 27-40 in 50mM. The figure shown is derived from a 16 hr autoradiograph exposure.

**B.** Fractions 26 lanes (1-3) and 34 (lanes 4-6) were tested for their specificity. Reaction conditions were set up as in fig. 5.2.2.A except that a 100 molar excess of an E4 ATF site oligonucleotide (lanes 2 and 5) or a nonsense oligonucleotide (lanes 3 and 6) were added to certain reactions. The photograph in the figure is derived from a 6 hr autoradiograph exposure.

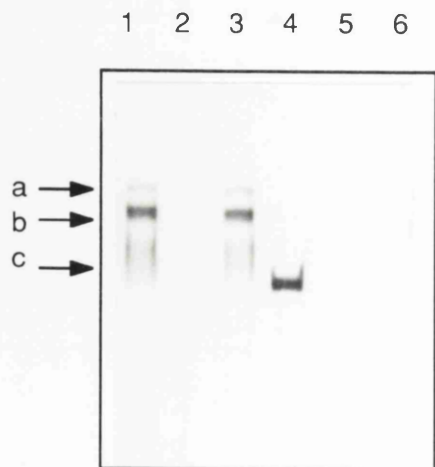
The labelled arrows indicate the relative positions of yATF complexes designated a, b, and c in the text.

**Fig. 5.2.2. DNA sepharose purification of yATF activity.**

**A.**



**B.**



### 3.c. DNA sepharose purification;

The pooled fractions were precipitated, resuspended, dialysed and loaded onto the DNA sepharose column as already described. A 400 ml ammonium sulphate gradient of 75mM to 600mM was applied and the fractions assayed. Fig. 5.2.2.A. shows the activity being separated across the profile. Complexes of the same mobility as ha1 and ha2 can be seen to elute between fractions 22 and 30 (120mM-200mM) with the strongest activity occurring between 22 and 26; very little activity is seen in the flow through (see f2 and f4). In addition to these complexes a smear can be seen below the ha2 like complex (see particularly lane 26). Oligonucleotide competition analysis confirmed that this activity was specific being competed by an ATF site but not a nonsense sequence (compare lanes 2-3, fig. 5.2.2.B) and the bands were called yATF -a, -b, and -c. A faster mobility band, which comigrated with the nonspecific ha4, was seen to elute in fractions 32-36. Subsequent competition analysis confirmed that this was indeed nonspecific being competed by both AP1 and nonspecific oligonucleotides (see lanes 4-6, fig. 5.2.2.B). Fractions 22-27 were pooled<sup>dmcl</sup> dialysed against 100mM binding buffer before the final stage in the purification protocol.

### 3.d. Affinity chromatography;

The theory behind affinity chromatography is outlined in the section describing yAP1 purification. Due to the success of this protocol in purifying yAP1 and other proteins it seemed logical to apply it in the purification of yATF. The results presented in this section involve the passage of protein over an oligonucleotide affinity column bearing sequences derived from the E4 adenovirus early promoter. As will shortly become apparent, the success of this protocol was very limited. However it should be noted that attempts to purify yATF also employed similar columns containing an ATF/CRE site derived from the cAMP inducible fibronectin promoter. Whichever ATF site was used the degree of success gained in purification was essentially the same.

Fig. 5.2.3.A. and B. shows the results of this analysis when 8 mls of pooled yATF activity derived from the DNA sepharose column was loaded onto and eluted from an affinity column containing oligomerised E4 ATF sites (as described in Materials and Methods). Most binding activity is being eluted in the flow through (fractions 5-10) and the low salt washes (fractions 10-15). The activity of fraction 18 was a mystery as a number of attempts to repeat this observation failed. However, as with the AP1 site column, the major interest lies in the activity eluted in the higher salt elutions. Unfortunately fraction 43 (0.5M salt, Fig. 5.2.3.B) represents the most stringent

### **Fig 5.2.3. Affinity purification of yATF.**

**A and B.** 5 $\mu$ l samples of fractions obtained from a stepwise elution of an E4 ATF site affinity column were tested for specific activity. Fraction numbers tested are indicated above each lane in the figure and the applied material is designated App. Because a suitable conductivity meter was unavailable to test fractions eluted from the column, salt concentrations in this assay are estimates. However the estimates are: 0.15M wash, fractions 11-20; 0.3M wash, fractions 21-30; 0.4M wash, fractions 31-40; 0.5M wash, fractions 41-50; 0.7M wash 51-60. In the binding assays both the salt concentration and the amount of nonspecific competitor was varied. Lanes 6-20 contained 100mM NaCl, lanes 21-30 contained 50mM NaCl, and lanes 31-50, 25mM NaCl. In the other variable, 0.1 $\mu$ g of non specific DNA was added to lanes 4-30 and 50ng to each of the rest. In addition 20 $\mu$ g of BSA was added to each binding assay as protein concentration was very low in a number of cases. The photograph represents a 16 hour exposure. Labelled arrows refer to different yATF mobility forms.

**C.** Shows competition analysis of the activity eluted at the 0.4M salt step.

Extracts taken from fraction 36 (fig.5.2.3.A.) were assayed as above but competed with a 30 molar excess of either an E4 ATF site (lane 3) or SV40 AP-1 site (lane 2) oligonucleotide.

**Fig. 5.2.3. Affinity Purification of yATF**

**A.**

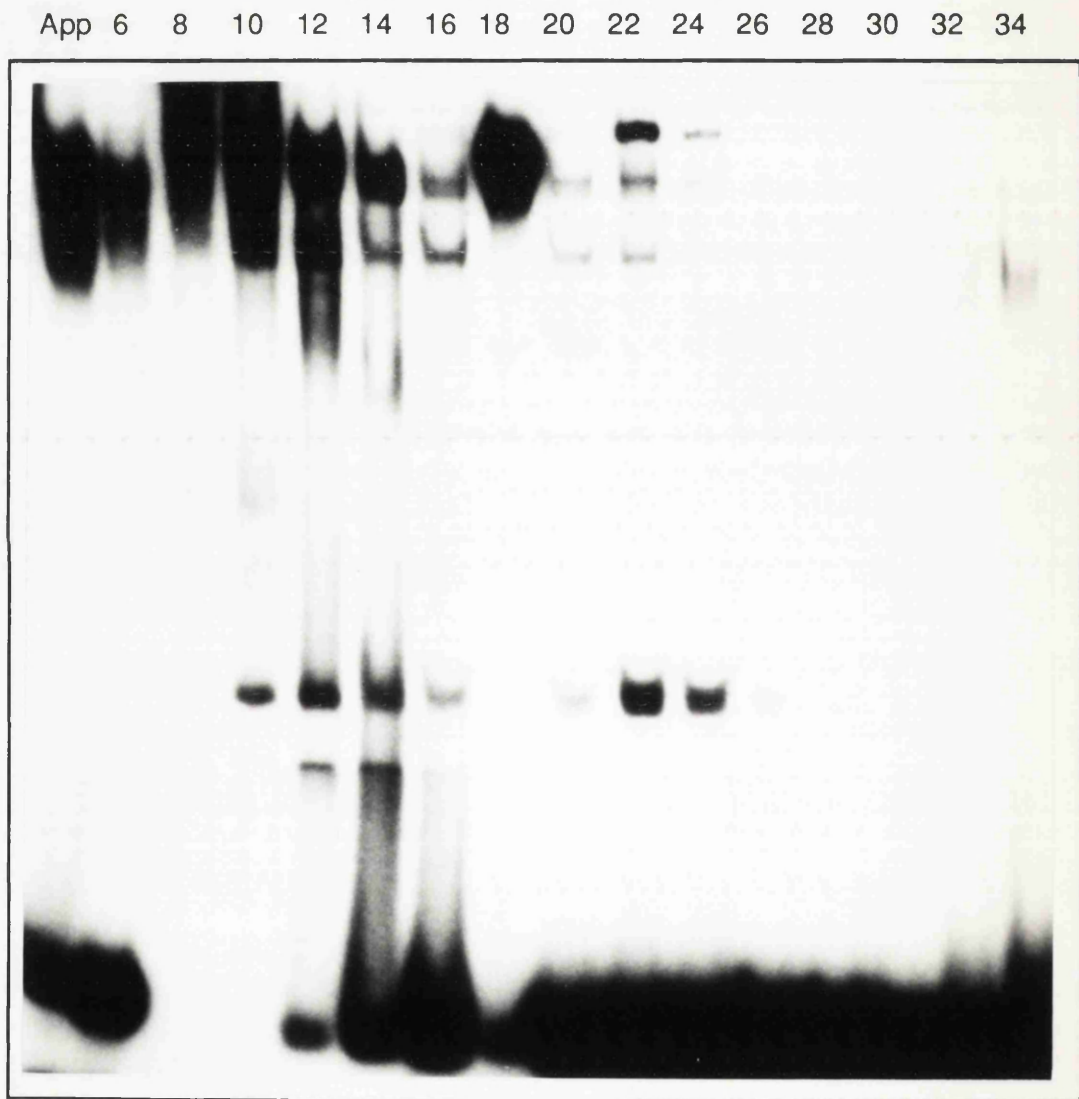
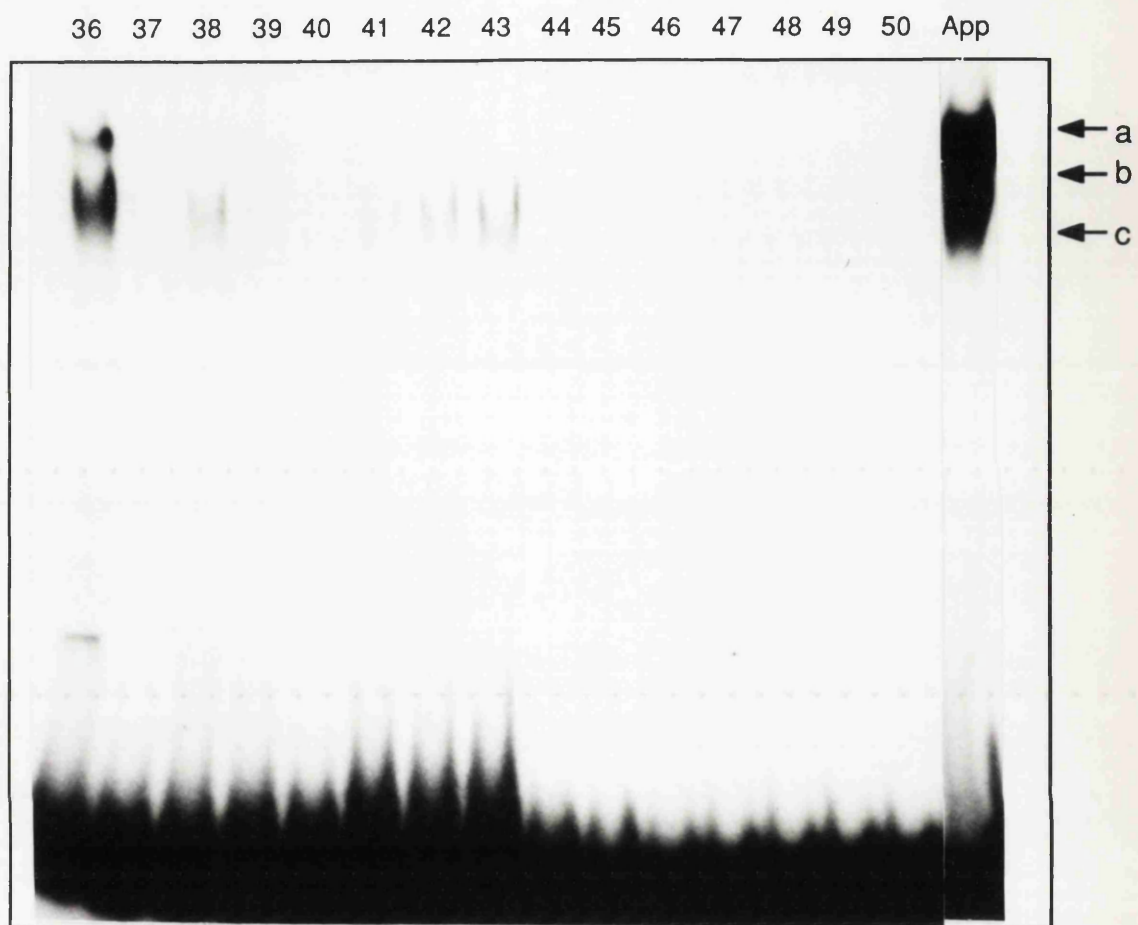
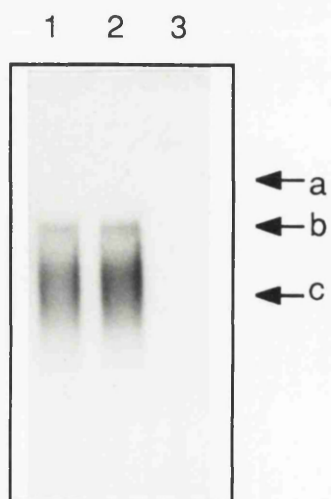


Fig. 5.2.3. Affinity purification of yATF.

B.



C.





elution that still contains binding activity. There is also a 'peak' at fraction 36 (0.4M NaCl). Apart from the relatively low salt tolerance manifested in this elution profile it is also rather disturbing to consider the amount of apparent activity in the higher salt elutions. At best the activity derived could be described as modest but the success of this purification step can be put into true perspective by comparing it to the equivalent stage in the purification of yAP1. For example compare fig. 5.1.5 and 5.1.6 with fig. 5.2.3. The activity in lane 10 of fig 5.1.6 is the peak of the yAP-1 purification and represents a two hour exposure of a gel shift containing 0.2µl of extract. In contrast the activity in lane 36 of fig.5.2.3 is an overnight exposure (approximately 16 hours) of a gel shift containing 5µl of extract. In addition a comparison of the relative activities of extract applied to the affinity column to that observed in the peak elutions reinforces how much more successful the yAP1 purification was, especially considering that the probe has been totally depleted in some of the yAP1 fractions (eg lanes 42 and 32). Silver stained and colloidal gold stained gels containing fractions from a number of different affinity chromatography attempts were carried out but there was never a convincing and repeatable yATF candidate that was identified. The reasons why yATF did not successfully purify remain largely unknown. It is not a manifestation of this particular affinity column as affinity columns bearing different CREs gave similar results. In addition this same affinity column was successfully used by a postdoctoral fellow in the laboratory to purify human ATF. However some clues lay in the increased mobility that was observed in the 'peak' yATF affinity fractions. Further analysis indicated that protein phosphorylation may be important for the manner in which yATF interacts with DNA.

### **3.e. Phosphorylation of yATF;**

Certain reports have indicated that human CREB/ATF is a phosphoprotein (Montminy and Bilezjian 1987, Yamamoto et al 1988, Hai et al 1988) and can be phosphorylated by the catalytic subunit of cAMP inducible protein kinase *in vitro* (Yamamoto et al 1988). This prompted me to investigate the phosphorylation state of the yATF factor. DNA sepharose purified yATF extracts were treated with alkaline phosphatase, and the ability of yATF to bind to DNA analysed (Figure 5.2.4.A). Incubation of untreated extracts with a labelled ATF binding site probe resulted in 3 main complexes which I have already designated a, b and c as shown in Figure 4.1.3. Increasing phosphatase treatment led to the reduction and disappearance of complexes a and b with a concurrent increase in complex c (see Fig. 5.2.4.A, lanes 1, 2 and 3). I also saw a complex with a slightly accelerated mobility to that of complex c (lane 3). These alterations were largely inhibited by the addition of phosphatase inhibitors to the reactions (see lanes 4-

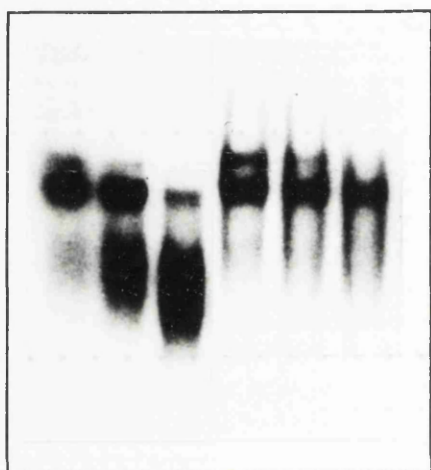
**Fig. 5.2.4 Phosphatase treatment of the yATF complex.**

**A.** S.cerevisiae DNA sepharose purified extracts were incubated at 37 C for 15 min in the absence (lanes 1 and 4) or in the presence of 5 units (lanes 2 and 5) or 10 units (3 and 6) of calf alkaline phosphatase. Reactions 4-6 were incubated in the presence of phosphatase inhibitors. Extracts were then tested in standard gel retardation assays with the E4 ATF oligonucleotide probe.

**B.** yATF activity derived from either affinity purified fraction 36 of fig.5.2.3. (lanes 1-4) or nonspecific DNA sepharose chromatography (lanes 5-7) were incubated at 37 C for 15 min under varying conditions and then assayed by the gel retardation assay using the E4 ATF oligonucleotide probe. Lanes 1 and 5 show control incubations in standard gel retardation buffer, lane 2 shows the addition of 1mM ATP, and lanes 3 and 6 have 1mM and 1-2 units of protein kinase A (catalytic subunit). Lanes 4 and 7 are treated with kinase A and ATP, but subsequently dephosphorylated with 20 unit of alkaline phosphatase for 15 min at 37C. The three distinct complexes of different mobilities are referred to as a,b and c.

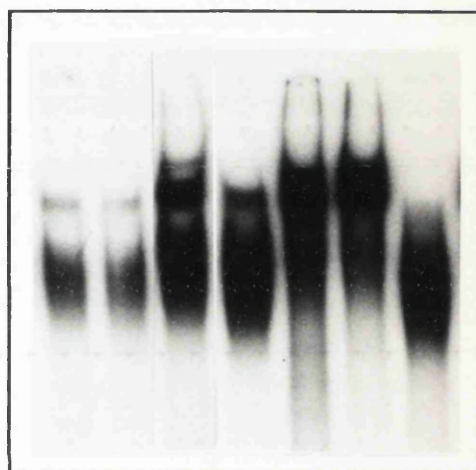
Fig. 5.2.4. Phosphatase and Kinase treatment of  $\gamma$ ATF.

A.



1 2 3 4 5 6

B.



1 2 3 4 5 6 7

— a —  
— b —  
— c —

6), and were therefore very likely to be due to dephosphorylation of the yATF/DNA complex (see Discussion).

Such an increase in mobility was reminiscent of the complexes formed in some of the affinity purified yATF fractions (see fig 5.2.3.B, lane 38 for example), suggesting that the yATF may have been dephosphorylated during the purification. Since the catalytic subunit of the cAMP inducible kinase has been shown to phosphorylate the CREB-1 factor *in vitro* (Yamamoto et al 1988) I determined whether this kinase could also phosphorylate affinity enriched yATF fraction 36. Addition of the kinase and ATP resulted in a marked increase in yATF binding activity and the appearance of the slower migrating complexes (compare lanes 1 and 3, Fig. 5.2.4.B) reminiscent of unphosphorylated DNA sepharose extracts. Addition of ATP alone did not result in these changes (see lane 2, Fig. 5.2.4.B). Phosphatase treatment of the re-phosphorylated extracts resulted in reversal of these mobility changes, (see lane 4, Fig. 5.2.4.B). This data therefore suggests that the yATF complex is phosphorylated and that this phosphorylation can influence the way it binds to the ATF site. It should be noted that phosphatase treatment *in vitro* did not result in a loss of DNA binding activity. It is not entirely clear why this is the case but it may simply be a consequence of subtle differences in specificity between the yeast phosphatase and the commercial activity obtained from Boehringer. Irrespective of the detailed explanations this data could be regarded as encouraging for the affinity purification of yATF as the apparent activity of the 'peak' fractions has now been dramatically increased. However the activity observed is still substantially lower than that observed in the yAP1 peaks. In addition, the inclusion of either phosphatase inhibitors or ATP and kinase with the extracts which were then applied to the affinity columns did not result in a notable increase in retained activity. Thus I could not simply extrapolate from the band shift data to improve conditions on the column indicating additional unknown factors were important. Thus it was decided to attempt an alternative strategy for the isolation of further yeast transcription factor clones.

**CHAPTER 6; RESULTS.**

**USING  $\lambda$  GT11 SCREENS TO CLONE YEAST  
DNA BINDING PROTEINS.**

## **1. Introduction and Summary.**

The previous chapter outlines the attempts to purify S.cerevisiae yATF and yAP-1. By purifying sufficient protein it is possible to obtain peptide sequence information which in turn can facilitate the isolation of a genomic or cDNA clone. This was a suitable protocol to follow for the isolation of the budding yeast clones for the reasons outlined at the beginning of chapter 5. However for the attempted isolation of fission yeast transcription factor clones I decided to follow an alternative methodology. This protocol was first published by Singh et al (1988) and Vinson et al (1988) and variations of it have since been used by a number of laboratories in the successful isolation of DNA binding protein clones (See Singh et al 1989 for review). The technique itself is described in full detail in the Materials and Methods section .

The main advantage of the technique is that circumvents the necessity to purify protein from large amounts of starting material and instead employs a direct screen of an expression library. The selection in the screen utilises the property of transcription factors to recognise their cognate site in a specific fashion. The probe is therefore a labelled binding site which, just as in other DNA binding protocols, should be able to associate specifically and selectively with its cognate factor. In this way positive plaques may be picked and enriched just as in other cloning procedures.

My interest lay in both the yAP1 and yATF proteins of S.pombe. Therefore for my initial analysis I used a mixture of both binding site probes with the intention of distinguishing the preferential specificity of the positive signals at a later stage. Once purified a plaque can be used to prepare a high titre viral stock which can in turn be used to prepare viral lysogens and viral DNA.

This chapter describes the identification and purification of two positive clones which were both shown to preferentially recognise an ATF site over an AP1 site when the protein was immobilised on a nitrocellulose filter. Partial sequence analysis of one of these clones has revealed a homology with the basic DNA binding domain of the mammalian CREB/ATF proteins. The characterisation of these clones is incomplete and therefore further analysis will be required to firmly establish their identity and function.

## **2. Picking and enriching positive signals.**

The screening protocol was carried out as described in Materials and Methods. Fig. 6.1.1 shows exposures from the two plates which gave positive candidate signals. The criteria I employed to pick positives was that a spot observed on the original filter must be accompanied by a lighter but visible spot on the replica filter. From previous

**Fig. 6.1.1. Filter hybridisation analysis of  $\lambda$ gt11 expression library.**

Nitrocellulose filters were prepared from four 15cm plates and probed as described in materials and methods. The upper panels are radiograph exposures from original (3, 4) and replica (3', 4') filters which were deemed to contain potential positive candidates by the criteria outlined in the text. No positive candidates could be detected on filters derived from plates 1 and 2. The four positive signals themselves are highlighted with > signs and have been numbered 1-4. The figure is derived from a 16 hour radiograph exposure.

**Fig.6.1.2. Tertiary screens of expression library.**

Positive agar plugs derived from the secondary enrichment screen were replated at approximately 100 pfu/plate. Shown in this figure is an example result of this analysis. Plate 1i is a derivative of positive 1 (fig 6.1.1), and 4ii a derivative of 4 (fig 6.1.1). At this stage it was very easy to align single plaques over autoradiograph spots to allow absolute plaque purification of any positive. Examples of positive signals aligning over plaques are indicated with >.

Fig.6.1.1. Primary expression Screen

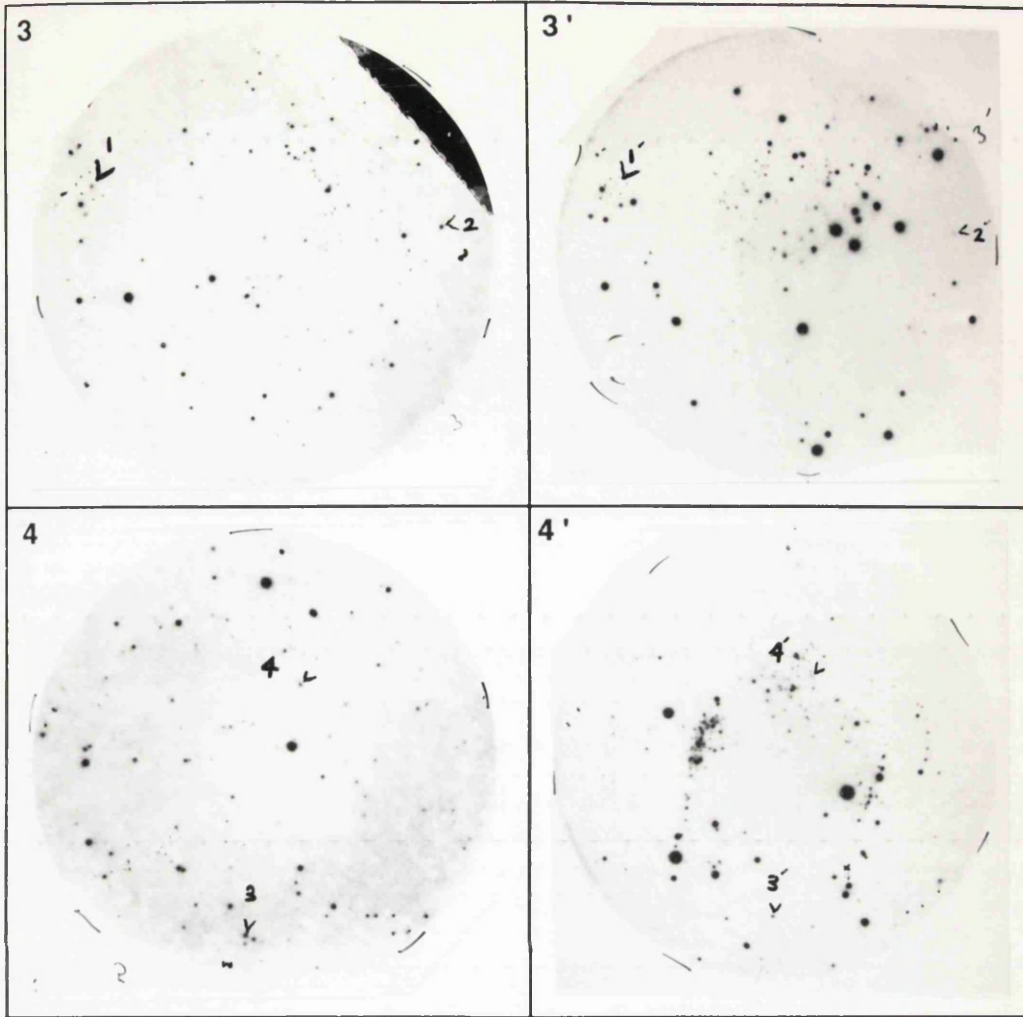
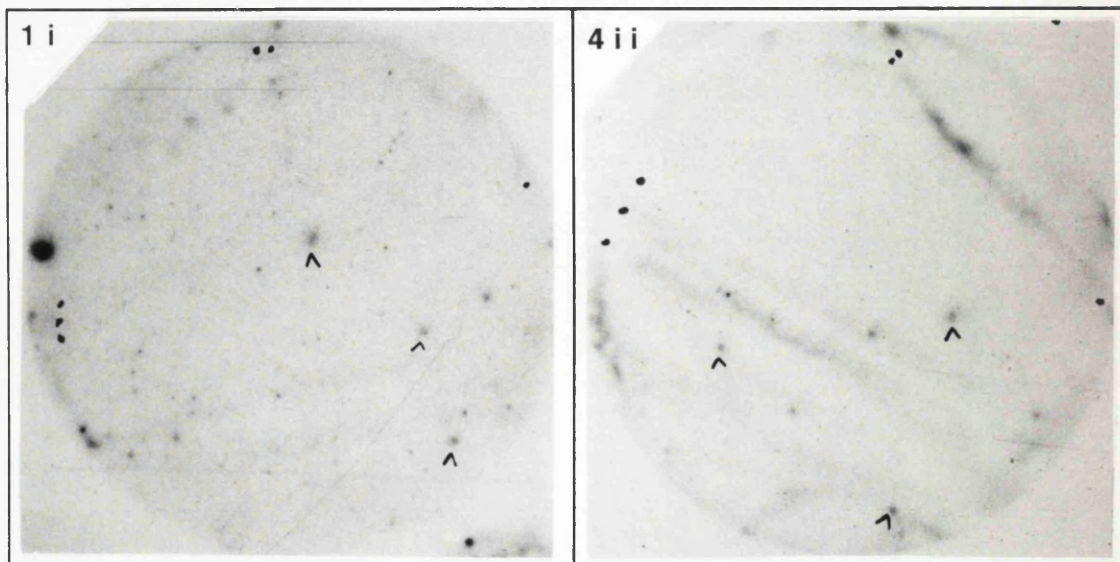


Fig 6.1.2. Tertiary Screens of expression library.





experience with other screening attempts this approach appeared the optimum in minimising artefacts without missing real positives. The amount of apparent background varies from one screen to another and the exact cause of this variation is unclear. However by carefully comparing primary and replica lifts it was possible to isolate four positive signals, two from plate 3 (labelled 1 and 2) and two from plate 4 (labelled 3 and 4 in fig 6.1.1).

Attempts were then made to enrich these four positives through secondary and tertiary screens, both of which involved probing with a mixed preparation of AP1 and ATF sites. At the end of the tertiary screen only two of the original four positives (positives 1 and 4 from fig. 6.1.1) had enriched successfully and these are shown in fig.6.1.2. At this stage single plaques could easily be distinguished and so isolates were picked from each plate to prepare high titre viral stocks which were used for further experimentation. From this point the isolate derived from 1i will be called A and that from 4ii will be called B.

One experiment of immediate interest was to determine whether these two clones bound preferentially to an ATF or AP1 site.

### **3. Clones A and B preferentially bind an ATF site.**

To analyse the exact binding specificity of clones A and B, I probed nitrocellulose filters separately with either an ATF or AP1 site. As a control for this experiment I used a phage isolate expressing the GCN4 protein. I had previously isolated a clone for GCN4 using a southwestern screening protocol on a S.cerevisiae cDNA library donated by Jane Mellor. It was unfortunate that GCN4 was the only clone I obtained in this screen but during its characterisation I demonstrated its preference for an AP1 site over an ATF site. It thus served as a control in this analysis.

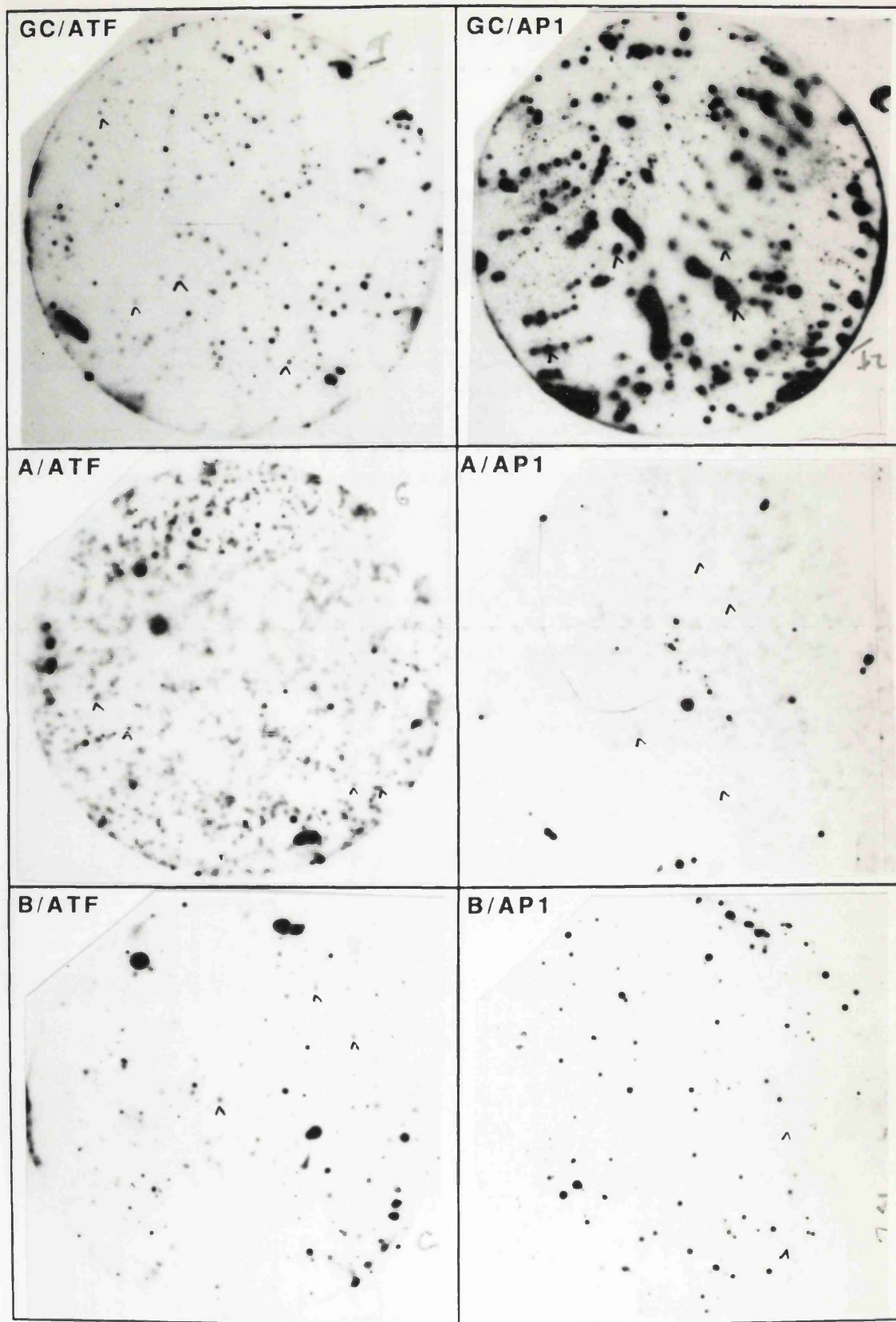
Before I discuss these results I should point out that not all the spots observed in fig. 6.1.3. represent true positive signals. The very dark punctate spots are generally artefacts which do not line up over plaques. For this reason I have highlighted (with >) a number of representative spots from each plate which do line up over plaques and are thus suitable for discussion purposes.

The first point to note is that the control GCN4 assay (GC) has given the predicted pattern of recognising the AP-1 site better than the ATF site, in agreement with previous assays. When a similar comparison<sup>A</sup> is made between the plaques derived from A or B it can be seen that in both these cases the tendency is to bind the ATF site better than the AP1 site, quite the reverse of GCN4. On first glance this is more obvious for A rather than B. However once it is realised that the dark punctate spots seen with the AP1 site

**Fig. 6.1.3. Individual ATF and AP-1 site screen of expression library.**

Positive plaques derived from fig. 6.1.2. were plated and probed with either a multimerised fibronectin ATF site (left) or SV40 AP1 site (right) probes (see materials and methods). GC is a phage expressing a GCN4 clone which we had obtained from a previous screen; A is a phage picked from plate 1i (fig.6.1.2.) and B from plate 4ii. All filters were treated in the same way apart from the probing step. Highlighted with > are representative plaques from each plate. Figs. are derived from 12 hr radiograph exposures.

Fig. 6.1.3. Individual ATF and AP1 site screen of expression library.



probing of B do not line up over plaques, and one examines only those signals that do, it becomes more apparent that the B clone shows a very similar probe preference as A. Thus in this type of assay both S.pombe clones bind the ATF site better than the AP1 site.

#### **4. Gel shift analysis on lysogen derived extracts.**

As gel shift analysis had originally demonstrated specific yATF activity in yeast extracts it seemed pertinent to test the specific binding properties of the phage clones this same assay.

For this purpose I prepared protein extracts derived from bacterial lysogens of both A and B viruses. Lysogens are a useful way of preparing large amounts of recombinant protein for such assays as every bacterial cell in the culture harbours an inducible phage. Fig.6.1.4.A. shows results obtained when A, B, and GC lysogen extracts were incubated with either an ATF site probe derived from the E4 promoter or an AP1 site derived from the SV40 promoter. A number of bands were observed in all the lysogen preparations and were assumed to be due to bacterial proteins interacting with the probe. However some activities were specific to particular lysogen extracts and therefore represented activities most likely to be derived from the recombinant phage. These bands have been indicated with arrows and labelled A, B, and GC according to the particular lysogen extract in which they are observed.

By concentrating on these bands it is clear that the GCN4 protein prefers binding to an AP1 site than to an ATF site (compare lanes 6 and 5). The results for A and B are less convincing but there does seem to be a slight preference for the ATF site probe over the AP1 site (compare lane 1 with 2 and lane 3 with 4). The competition analysis in fig.6.1.4.B. is similarly suggestive but unconvincing. In both cases competition by an ATF site is only marginally more efficient than that observed for an AP1 site (compare lanes 2 and 3 for clone B, and lanes 5 and 6 for clone A). This is in contrast to the situation seen in yeast extracts where the ATF site is a significantly more efficient competitor for yATF activity than an AP1 site (compare lanes 7 and 9 in fig. 4.1.3.A)

There are a number of experimental reasons which could explain why the bacterially produced protein does not share the stringent specific binding properties of the yeast activity. For example to form a complex which is stable during electrophoresis may require certain modifications, such as phosphorylation, which it does not receive in the bacterial cell. The phosphorylation data concerning budding yeast yATF indicates that such a consideration should be taken into account. In addition the specific activity observed in yeast extracts may involve a complex between more than one protein

**Fig.6.1.4. Gel shift assays testing lysogen extracts derived from phages GC, A and B.**

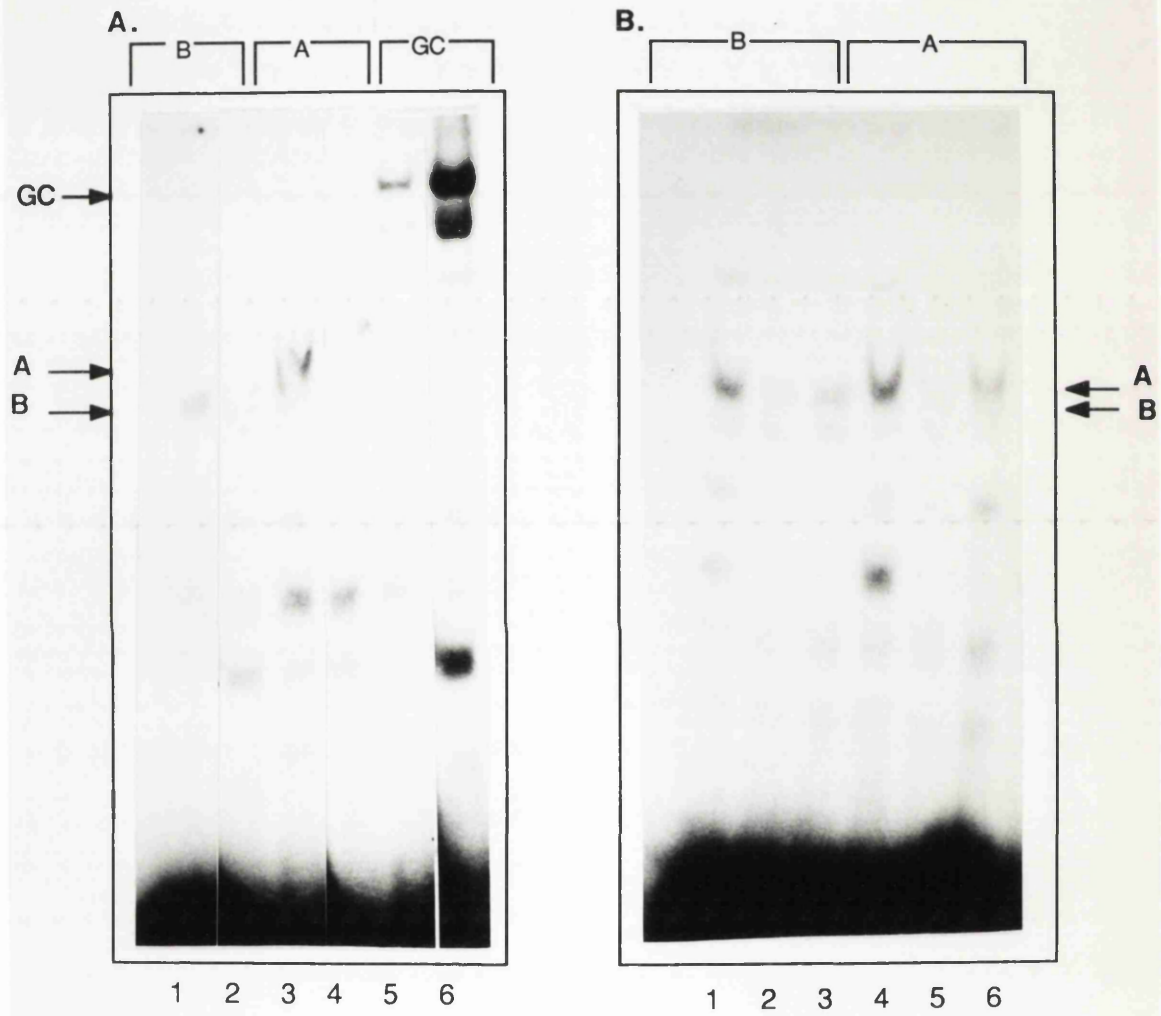
Assays were carried out under standard conditions using a 100mM NaCl based binding buffer, 0.3 µg Poly dI;dC and 5µl of bacterial extract in each reaction.

A. shows incubation of the extracts prepared from lysogens harbouring phage; B (lanes 1 and 2), A (lanes 3 and 4), and GC (lanes 5 and 6) incubated with either the SV40 AP1 site (lanes 2, 4, and 6) or the E4 ATF (lanes 1, 3, and 5) site probes. N.B. All probes in gel shift assays contain only a single factor binding site.

B. shows competition analysis of lysogens A and B when incubated with the E4 ATF site probe. Lanes 1-3 contain extracts from lysogen B and lanes 4-6 contain extracts derived from lysogen A. The E4 ATF site cold oligonucleotide was added as a competitor to lanes 2 and 5 and the SV40 AP1 site was similarly added to lanes 3 and 6 at a 50 fold molar excess.

Arrows indicate the position of complexes which are specific to particular lysogen extracts (either A, B or GC).

Fig. 6.1.4. Gel shift assays using lysogen derived extracts.



species, which will obviously not form in the bacterial system. However another reason why clone B in particular was only binding very inefficiently became apparent from preliminary sequence analysis.

## 5. Sequence analysis of clone B.

Viral DNA, prepared as described in the Materials and Methods, was restricted with EcoR1 to release a 1.5 kb fragment of yeast DNA. This fragment was subcloned into Bluescript (Stratagene) and sequenced entirely on both strands. Overlapping flanking sequences were isolated by hybridisation of this fragment to a partial SAU3A *S.pombe* genomic library obtained from Paul Nurse's laboratory. Sequence data thus far obtained is shown in fig.6.1.5. The region between the Eco R1 sites is the original fragment and the only region, so far, to have been sequenced on both strands. An analysis of potential open reading frames which are themselves in frame as a fusion with  $\beta$ -galactosidase of  $\lambda$ gt11 showed a limitation of either 23 or 3 amino acids according to the end. I therefore concluded that the protein responsible for the binding activity I observed was most likely translated from a start codon somewhere within the isolated fragment. To identify possible start positions I looked for methionine residues with a Shine-Dalgarno (S-D) like sequence shortly upstream of them. The closest match to the S-D sequence of AGGAGG in a suitable region was AGGAGA located 11 base pairs upstream of a Met codon at residue 1495 in the sequence. Although this has been considered as the most likely translation product we have no proof that it is responsible for the DNA binding activity observed. None the less it is itself part of a 510 amino acid ORF which extends over the whole length of this fragment, indicating that it may be biologically relevant. Additional information which I shall outline in the discussion also indicates that it is translation of this particular ORF which is probably responsible for the activity I observe.

If translated this region would produce a polypeptide of 92 amino acids before entering vector sequence. This length of sequence, although rather small, is large enough to include a potential DNA binding domain. For example GCN4 can still bind DNA as a deletion product of 60 amino acids (Hope and Struhl 1986). However it would have been extremely fortunate to randomly encompass the whole functional unit intact within these 92 amino acids and it is possible that the poor specific binding seen in gel shift assays is due to the limited size of the product. Observations relating to this and other aspects of the sequence are addressed in the Discussion section.

**Fig. 6.1.5. DNA sequence of the B clone and its predicted amino acid sequence in a reading frame that extends the whole length of the clone.**

Included in this figure are regions which have not been sequenced on both strands. However the 1532 bp between the two EcoR1 sites indicated, represent the original B fragment and have been sequenced on both strands. The potential translational start is indicated as an outlined and underlined methionine residue, and the putative Shine-Dalgarno sequence is also highlighted.



**Fig. 6.1.5. DNA Sequence of the B clone and it's predicted aminoacid sequence in one reading frame (numbering increases in 5' to 3' direction)**

27 54  
 GAC TAT GCT ACA AAC AAG GAT CCG TGC TTT CTT CAA AGA AAA ATT TTT CAA AAG  
 D Y A T N K D P C F L Q R K I F Q K

81 108  
 CGA TTG AAG GAC ATA CAA GGC ATC AAA TTA CTA CAA GCA AAT TTG CAA GTG TAC  
 R L K D I Q A I K L L Q A N L Q V Y

135 162  
 AAT GAA TTT AGA ACT TTT CCT TGG GCT AAG TTA TTT TTC AAT CTA AGG CCC TTA  
 N E F R T F P W A K L F F N L R P L

189 216  
 TTA AGT AGT ACT CAA AAT GAT AAG CAG NIT AAA AAG CGA GAT GCT GAA ATC ATA  
 L S S T Q N D K Q K K R D A E I I

243 **Eco RI** 270  
 GAG CTG AAG TAC GAG CTG AAA AAA CAA CAG AAT TCA AAA TCT GAG GTA GAG CGT  
 E L K Y E L K K Q Q N S K S E V E R

297 324  
 GAT CTC GTC GAG ACC AAC AAT TCT TTA ACT GCT GTT GAA AAT TTG CTT ACT ACT  
 D L V E T N N S L T A V E N L L T T

351 378  
 GAA CGT GCG ATA GGC CTT GAT AAG GAG GAA ATA TTA CGT CGA ACA CAG GAA AGG  
 E R A I A L D K E E I L R R T Q E R

405 432  
 CTT GCA AAC ATA GAA GAT TCG TTT AGC GAA ACA AAA CAG CAA AAT GAA AAT TTG  
 L A N I E D S F S E T K Q Q N E N L

459 486  
CAA CGA GAG TCA GCG TCA TTG AAA CAG TTA ATT AAC GAA TTA GAA AGT GAG TTA  
Q R E S A S L K Q L I N E L E S E L

513 540  
CTA GAA AAA ACT TCC AAG GTT GAG ACT TTG TTA TCA GAA CAG AAC GAA CTG AAA  
L E K T S K V E T L L S E Q N E L K

567 594  
GAA AAA CTG AGT TTG GAA GAG AAG GAT TTG CTA GAT ACA AAG GGA GAA CTT GAG  
E K L S L E E K D L L D T K G E L E

621 648  
TCT TTA CGT GAG AAC AAT GCA ACT GTT TTA TCA GAA AAA GCA GAG TTC AAT GAA  
S L R E N N A T V L S E K A E F N E

675 702  
CAA TGT AAA AGT TTA CAG GAA ACT ATC GTT ACC AAG GAT GCT GAA CTT GAT AAA  
Q C K S L Q E T I V T K D A E L D K

729 756  
CTC ACT AAA TAT ATT TCT GAC TAT AAA ACT GAA ATT CAA GAA ATG AGA CTG ACA  
L T K Y I S D Y K T E I Q E M R L T

783 810  
AAT CAG AAG ATG AAT GAA AAG AGT ATT CAG CAG GAA GGT TCT TTA TCT GAG AGT  
N Q K M N E K S I Q Q E G S L S E S

837 864  
CTC AAA CGA GTC AAA AAA TTA GAA CGT GAA AAT TCT ACT CTT ATA AGC GAC GTC  
L K R V K K L E R E N S T L I S D V

891 918  
TCT ATT TTA AAA CAG CAG AAA GAA GAG TTA AGC GTC TTG AAA GGA GTG CAA GAG  
S I L K Q Q K E E L S V L K G V Q E

945 972  
TTA ACA ATA AAT AAC TTA GAA GAG AAG GTT AAT TAT TTG GAA GCC GAT GTC AAA  
L T I N N L E E K V N Y L E A D V K

999 1026  
CAG CTT CCG AAA TTG AAG AAA GAA CTA GAA TCC TTG AAT GAC AAA GAT CAA CTT  
Q L P K L K K E L E S L N D K D Q L

1053 1080  
TAT CAG CTC CAG GCT ACC AAA AAT AAA GAA CTG GAA GCC AAA GTT AAA GAA TGT  
Y Q L Q A T K N K E L E A K V K E C

1107 1134  
TTA AAT AAC ATC AAA TCG TTA ACG AAA GAG CTT GAA AAT AAG GAA GAG AAA TGC  
L N N I K S L T K E L E N K E E K C

1161 1188  
CAA AAT TTG TCA GAT GCT TCT TTA AAG TAC ATT GAA TTG CAA GAA ATC CAT GAG  
Q N L S D A S L K Y I E L Q E I H E

1215 1242  
AAT TTA CTT CTT AAA GTT AGC GAC TTG GAG AAT TAC AAG AAA AAA TAT GAA GGA  
N L L L K V S D L E N Y K K K Y E G

1269 1296  
CTT CAG CTG GAC TTC CAG GGG TTA AAA GAC GTC GAT ACA AAC TTT CAG GAA CTT  
L Q L D F Q G L K D V D T N F Q E L

1323 1350  
TCG AAA AAA CAT CGT GAT TTG ACC TTT AAT CAT GAA AGT CTT CTT CGG CAA TCG  
S K K H R D L T F N H E S L L R Q S

1377 1404  
GCG TCC TAC AAG GAA AAG CTT TCG TTG GCG AGC TCC GAG AAT AAA GAT CTC TCT  
A S Y K E K L S L A S S E N K D L S

1431 1458  
AAT AAG GTA TCG AGT TTG ACG AAG CAA GTA AAT GAA CTT TCT CCA AAA GCA TCT  
N K V S S L T K Q V N E L S P K A S

### Shine-Dalgarno

1485 1512  
AAG GTT CCC GAA CTT GAG AGG AAG ATT ACT AAC TTA ATG CAC GAG TAC AGT CAA  
K V P E L E R K I T N L M H E Y S Q

1539 1566  
CTC GGT AAG ACC TTT GAA GAT GAA AAG CGT AAA GCA TTG ATA GCT AGT CGA GAT  
L G K T F E D E K R K A L I A S R D

1593 1620  
AAC GAA GAA TTA CGC TCT TTG AAA TCA GAG TTG GAG AGT AAA CGA AAG CTT GAA  
N E E L R S L K S E L E S K R K L E

1647 1674  
GTT GAG TAT CAA AAG GTA CTA GAA GAA GTA AAA ACA ACT AGA AGC CTC CGT TCG  
V E Y Q K V L E E V K T T R S L R S

1701 1728  
GAA GTC ACC CTT TTA AGG AAC AAA GTT GCT GAT CAT GAA AGT ATA CGT AGT AAG  
E V T L L R N K V A D H E S I R S K

1755 **Eco RI** 1782  
CTT TCA GAA GTA GAA ATG AAG CTG GTT GAT ACT CGA AAG GAA CTG AAT TCT GCT  
L S E V E M K L V D T R K E L N S A

1809 1836  
CTT GAT TCC TGC AAG AAG CGA GAG GCA GAG ATA CAT CGC CTT AAA GAC CAT AGA  
L D S C K K R E A E I H R L K D H R

1863 1890  
CCG TCC GGA AAG GAG AAT AAT ATA CCG GCT GTC AAA ACT ACG GAA CCC GTT TTG  
P S G K E N N I P A V K T T E P V L

1917 1944  
AAA AAT ATT CCA CAA CGA AAG AGA ATT TTT GAT TTA CAG CAA AGA AAT GCA AAT  
K N I P Q R K R I F D L Q Q R N A N

1971  
CAA GCG TTA TAC GAG AAC TTG AAA CGA GAC TAC GAC CGT CTT ATC TTG AG  
Q A L Y E N L K R D Y D R L I L

## **CHAPTER 7; DISCUSSION**

The work in this thesis has demonstrated a degree of homology between yeast and mammalian transcriptional machinery. The original experiments which led to these observations being made involved the expression of viral promoters in fission yeast cells. I will therefore discuss this work first. Subsequent experiments involved a more detailed analysis of specific phenomena such as the specificity of DNA binding proteins recognising these promoters, the ability of particular promoter elements to mediate activating potentials, and the properties of the yeast transcription factors which may be important for their regulation. As these experiments were carried out other laboratories were generating additional information which influenced the perspective with which I could view my results. In this section I shall discuss the data I have obtained and attempt to place it in context with developments which have taken place since I started my work. Finally I shall attempt to give an interpretation on the biological significance of the developments that have taken place and outline areas where fruitful information may be gained in the future.

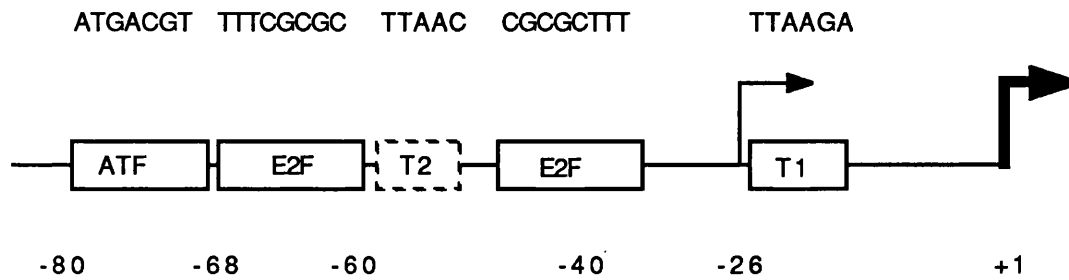
## **1. Expression of viral promoters in fission yeast.**

In Chapters 3 and 4 I have described the expression of the Adenovirus E2A promoter and the SV40 promoter/enhancer in fission yeast cells. The SV40 promoter/enhancer had previously been shown to be functional in expressing cDNA products, but detailed analysis had not been carried out; the E2A promoter had never been expressed at all. The initial characterisation of both promoters involved determining the site of transcriptional initiation.

### **1.1. The Adenovirus E2A promoter;**

In the case of the E2A promoter there is a difference in the preferred major initiation site between fission yeast and mammalian cells. Fission yeast prefers to use a start site that has been characterised as the minor start site in mammalian cells, utilised only 20% of the time. Both these start sites have TATA boxes located at appropriate distances upstream but in both cases they bear only a very weak resemblance to the consensus (see fig. 8.1.1. below).

Fig. 8.1.1. Schematic representation of factor binding sites on the E2 promoter



Shown in the boxes are the names of factors known to interact with regions of the E2 promoter. The hatched box indicates the position of the distal TATA element with which no factor has been definitively shown to associate. The important sequences within each element are shown above and their position relative to the major transcriptional start site in mammalian cells is shown below. The arrows represent major (thick) and minor (thin) sites of mammalian initiation.

A sequence specific DNA binding protein, presumably a TATA box factor, has been shown to associate with the proximal (T1) but not distal (T2) 'TATA' region. One might therefore infer from these results that *S.pombe* does not possess a factor which can recognise this site and therefore constitutively uses the 'minor' site. It is not clear which factors are involved in directing transcription from the minor site in mammalian cells but it could possibly involve E2F, ATF or both. The ability of supposed 'regulatory' factors to direct transcriptional start has recently been demonstrated (Chen and Struhl 1989, and Pei and Berk 1989). As fission yeast factors have been shown to bind both the ATF and E2F sites it is possible that it is these factors which are dominant in determining the start site used from this promoter in these cells.

## 1.2. The SV40 promoter/enhancer;

Analysis of the SV40 promoter/enhancer gave a more straight forward result. In this promoter the TATA element is 28-33 bp upstream of the major initiation site. In mammalian cells, it determines the site of initiation and its removal results in heterogenous starts (Benoist and Chambon, 1981). The demonstration that transcription initiates from the same site in *S.pombe* suggests that this TATA element can be recognised by a fission yeast TATA factor and thus functions in determining the position of initiation. Examination of a number of other *S.pombe* genes reveals that



TATA elements are consistently positioned 25-30 bp upstream of the start site very reminiscent of most mammalian promoters. (Russell, 1983). In contrast, TATA box elements in S.cerevisiae genes are located at variable distances from the start site, usually 60-120 bp upstream, and they appear not to direct transcription initiation at a fixed distance downstream (see introduction and Struhl, 1987a). With respect to this aspect of transcription therefore, it appears that there is greater similarity between the promoters of higher eukaryotes and S.pombe than with those of S.cerevisiae.

The SV40 promoter/enhancer was also subjected to deletion analysis. This demonstrated that the majority of the enhancer sequences can be removed with little or no effect on expression levels. However expression is affected to some extent by removal of both the 'P' element and the 21bp repeat region. In mammalian cells the P element binds an activity called AP-1 (Lee et al 1987, Angel et al 1987) whereas the 21bp repeat region has been shown to bind two transcription factors, Sp1 (Gidoni et al., 1984) and AP2 (Mitchell et al., 1987). The enhancer region of SV40 as a whole contains binding sites for a number of transcription factors including AP-2.

The lack of enhancer activity implies that either fission yeast factors are not interacting efficiently with the enhancer or factors that are interacting are not capable of activating in the context of this assay. The inability to obtain convincing footprints over most of the enhancer is suggestive of the former explanation but it should be noted that protein/DNA interactions could not be detected over the 21 bp repeat region which was shown to contribute to transcriptional activity. Thus it is possible that certain regions of the enhancer would be able to activate transcription if they were placed in a different context.

Collectively these studies laid the basis for all further experimentation described in my thesis. They provided a clear demonstration that promoters normally expressed in mammalian cells can be successfully utilised in fission yeast. The most impressive aspect of this illustrated that both systems chose similar start sites of transcription and as such represented the first work of its type to be reported. Subsequent studies which have been carried out in collaboration with the laboratory of Dr. B. Thimmapaya have served to establish further functional parallels. These workers have previously functionally dissected the E2A promoter in mammalian cells (Murthy et al 1985) and have now done the same thing in fission yeast. Intriguingly the same regulatory elements, ie the ATF and E2F sites shown in fig. 8.1.1, are similarly important for expression in both systems. Another study using the Adenovirus E3 promoter has reached similar conclusions. Thus despite the fact that viral promoters have evolved to be expressed exclusively in mammalian cells, the transcriptional machinery has been sufficiently conserved to allow productive interactions with yeast factors.

## **2. Conservation in DNA binding specificity between yeast and mammalian proteins.**

Previous chapters have described the identification of DNA binding activities in both fission yeast and budding yeast that can specifically recognise a consensus TRE or CRE. Stringent competition analysis has reinforced the concept that the yeast proteins possess a DNA binding specificity that has been remarkably conserved throughout evolution into human cells. However the concept of AP1 and ATF/CREB binding activity in mammalian cells has changed dramatically in the last couple of years. In the following sections I shall discuss these developments and try to gauge the implications which they hold for the apparently homologous proteins that exist in yeast.

### **2.1. Multiple proteins binding to the TRE;**

In the Introduction I described the cloning of c-jun, a gene which encodes a transcription factor capable of recognising the TRE consensus when bound as a homodimer. Since that time the number of cloned factors shown to recognise this site has increased dramatically. This has come about, in part, since the realisation that C-JUN binds far more tightly to an AP-1 site when it forms a heterodimer with the C-FOS protein. The indications for a role of C-FOS stemmed from the observations of Distel et al (1987) who showed that C-FOS was associated with a complex that bound to a control element in the adipocyte P2 promoter, the element's sequence closely resembling an AP1 site. In subsequent analysis C-FOS was shown to be present in preparations of affinity purified AP1 (Rauscher et al 1988, Franza et al 1988). Indeed it became apparent that a number of other proteins were also present in these preparations, some such as JUN-B and JUN-D were related to C-JUN, and others such as FOS B and FRA-1 were related to C-FOS (Nakabeppu et al 1988, Franza et al 1988). Sequence comparisons have revealed that, not surprisingly, all these proteins have strong sequence homologies within their DNA binding domains, each with a basic region adjacent to a leucine zipper. Within each of the JUN and FOS families there is also homology outside this region but the similarities are not so significant.

In vitro translational analysis has shown that the JUN proteins can all heterodimerise with each other to bind an AP-1 site but in each case do so far more efficiently when heterodimerised with one of the FOS protein family (see Kouzarides and Ziff 1989 for

review). The FOS proteins themselves will only associate with a TRE in the presence of a JUN protein. The selective synergism apparent between ~~that~~ the FOS and JUN families in vitro can also be reproduced as an activation potential in vivo. For example in F9 cells, which do not possess high levels of endogenous AP1 activity, a synergistic activation of transcription can be observed from a multimerised TRE element if c-jun and c-fos are coexpressed (Chiu et al 1989b). Such results generated a keen interest in the molecular basis for such synergism and subsequent studies have shown the FOS and JUN proteins directly interact, even in the absence of DNA. This interaction occurs through their respective leucine zippers and the synergistic nature of C-FOS-C-JUN DNA binding can be largely ascribed to the heterodimeric preference of these structures. For example O'Shea et al (1989) have shown that when synthetic peptides corresponding to the zipper regions of C-JUN and C-FOS are mixed, the heterodimer form is 1000 times as abundant as either of the homodimers. Indeed if the basic domain of C-FOS is fused to the zipper of C-JUN then the fusion product can heterodimerise with a wild type C-FOS protein to bind an AP1 site almost as efficiently as a normal C-FOS/C-JUN heterodimer (Neuberg et al 1989).

The upshot of all these discoveries is that the AP1 activity observed either in binding assays or in functional assays could be due to any one or more of a number of different proteins. The concept of C-FOS and C-JUN as singularly responsible for AP1 binding activity is therefore outdated and incorrect. The implications for these developments at a functional and regulatory level are discussed in a later section.

However how does this relate to what is known about yeast AP1 activity? It appears that a familial scenario, albeit less complex also exists in this system. At least two clones, yAP-1 and GCN4 have been shown to recognise the consensus TRE and there could in fact be more. Moye-Rowley et al (1989) have indicated as such in their southwestern analysis where specific AP1 binding activity can be seen even when the yAP1 gene has apparently been disrupted. However it should be noted that the disruption performed by this group simply inserted a marker into the middle of the yAP1 locus. In so doing it is still feasible that a truncated product is produced. As the DNA binding domain is located at the aminotermminus, it is quite feasible that what they are seeing is a truncated version of yAP1.

With regard to heterodimerisation there has been no work reporting the ability of yAP-1 and GCN4 to heterodimerise with each other. However it should be noted that GCN4 will not heterodimerise with either C-JUN or C-FOS in vitro. Molecular mutagenesis studies have addressed which residues within the zipper region are important in determining whether association will occur, but as yet there is no clear reason for selective heterodimerisation.

## 2.2. Many factors can bind CRE/ATF sites;

The CREB/ATF field has been similarly explosive. The apparent convergence of two scientific fields to the CREB/ATF factor resulted in the rapid isolation of a clone by two groups. One approach utilised a direct expression screen of a human library (Hoeffler et al 1988), as described in Materials and Methods, and the other generated degenerate oligonucleotides from peptide sequencing of purified rat protein (Gonzalez et al 1989a). The two clones isolated by these greatly variant protocols were highly homologous and the differences put down to species variation.

Thus a key player acting at the apparent converging point of E1A and cAMP stimulatory pathways became available for genetic manipulation. However at this stage it was still not entirely clear that CREB-1 was a single player in the game. Indeed I outlined a number of studies in the Introduction which indicated that multiple DNA binding proteins may be involved. This theory gained added support with the isolation of a second CREB factor by Maekawa et al (1989) that was quite clearly divergent from the those previously described. Subsequently Hai et al (1989) added another 5 proteins (which I shall call ATF- 1, 3, 4, 7, and 8) to this family all of which were isolated by a direct expression screen. In total, therefore, six clones have been isolated which preferentially bind a CRE to a TRE. All are highly homologous within their DNA binding domains and all contain leucine zippers. Functional analysis has shown that CREB-1, CREB-2, ATF-1, and ATF-3 can all form homodimers and CREB-2 and ATF-3 can form heterodimers with each other (Dwarki et al 1990, Hai et al 1989). The whole scenario is somewhat reminiscent to the AP1 story.

Not surprisingly the CREB/ATF proteins are also highly homologous, at least within their DNA binding domains, to the JUN/FOS protein family. In addition there is at least one example of a JUN/CREB heterodimer. Benbrook and Jones (1990) have shown that CREB-1 can selectively heterodimerise with c-jun to form a complex that binds a CRE with a higher affinity to a TRE, thereby bringing C-JUN to a site it would not preferentially recognise on its own or in the presence of C-FOS. Unfortunately the physiological significance of such selective interactions are not yet clear and it is unlikely that a full understanding will be gained for quite some time.

It is also unclear exactly how to interpret these findings with regard to the yeast system. This is largely due to the fact that, apart from the candidate clone I have reported in my thesis, no yATF genes have been isolated. With regard to binding activity, budding yeast extracts clearly contain several forms of yATF activity as assayed by band shift analysis. There are several interpretations to this data but it could in theory reflect the association of different sized proteins with the ATF site. Beyond this it is somewhat

difficult to speculate. Indeed one would not have predicted the apparent plethora of CREB/ATF and for that matter JUN/FOS proteins, which have now been isolated, from the fairly unpolymorphic band shift activity observed in crude HeLa cell extracts (see Hurst and Jones 1987 for example). However even if there is more than a single yATF or yAP-1 protein in either yeast system it is unlikely that the numbers will compare with the mammalian system. Such evolutionary divergence probably reflects the development of a complicated regulatory network which can be variant in different cell types according to the proteins expressed. Such a sophisticated system is probably beyond the requirements of a single cell organism.

### **2.3. Is Clone B a member of the CREB/ATF transcription factor family?**

The B clone derived from the S.pombe expression library was isolated due to its apparent ability to specifically recognise a CRE/ATF site. As I have already discussed a number of other proteins from mammalian systems have been isolated by similar criteria. Sequence analysis of such clones has not surprisingly demonstrated a block of homology in their respective DNA binding domains which is comprised of a leucine zipper preceded by a basic domain. Beyond this region there is generally no strong sequence homology (see Ziff 1989 for review). In analysing the B clone sequence it therefore seemed pertinent to try and identify a region homologous to the DNA binding domains of these other family members. Such a domain was identified at the carboxy terminus of the originally isolated fragment and it is aligned with some other zipper proteins in Table. 8.1.1. The significance of this finding with regard to the possible implications for function is discussed below.

#### **2.3.1. Homology in the DNA binding domain;**

The alignment shown in table. 8.1.1 is quite indicative of the B clone being a member of the CREB/ATF protein family. For example in the 26 amino acids preceding the first leucine of the zipper 12 residues are conserved (regarding lysine and arginine as conserved) between CREB-1 and B. By the same criteria GCN4 and C-JUN are conserved at 13 residues, and yAP1 and C-JUN at 9. In addition there is a block of homology apparent for almost all family members between residues 8 and 12. However in spite of this apparently striking homology there are also some notable deviations in the B clone which may be cause for concern. For example residue 17 is arginine in all family members apart from B where it is isoleucine, and residue 19 is basic in all cases apart from B where it is serine. The functional importance of the basic residue at position 19 has been definitively demonstrated for CREB-1 where its substitution to

**Table. 8.1.1. Sequence alignment of a region of the B-Clone with the DNA binding domains of a number of CREB/ATF and AP-1 proteins.**

Visual comparison of the carboxyterminal region of fragment B revealed a homology with the DNA binding domains of a number of other zipper proteins. This region is subdivided into the 'Basic domain' and 'Leucine Zipper' as indicated. Within the basic domain the numbering arbitrarily assigns positions of amino acids so they can be referred to in the text; a region of particular homology is set in isolation from its flanking sequences. In the zipper region the numbers refer to the positions of the leucines.

**Table 8.1.1. Comparison of a region of B clone sequence with the DNA binding domains of CREB/ATF and AP1 proteins**

	Basic Domain		Leucine Zipper				
	1	2 6	1	2	3	4	
CREB-1	KREVRIM	KNREA	ARECRKKKKEYVKC	L ENRVAV	L ENQNK	L IEEI	KA L KDLYCHK
CREB-2	KRRKYLE	RNRPA	ASRCRQRRKVVVS	L EKKAED	L SSINGQ	L QSEVTL	L RNEVAQL
ATF-3	RKKRRRE	RNKIA	AAKCRNKKKKEKTEC	L QKESEK	L ESVNAE	L KAQIEE	L KNEKQHL
B-CLONE	RSEVTLI	RNKVA	DHESIRSKLSEVEMKL	VDTRKE	L NSALDS	C KKREAE	I HRLIKDH
YAP 1	TKOKRTA	QNRPA	QPAFRERKEREKME	L EKKVVS	L ESIOQQ	N EVEATE	L RDQLITL
GCN4	PALAKRA	RNTEA	ARRSRARKLIQRMKQ	L EDKVEE	L LSKNYH	L ENEVAR	L KKLVGEE
C-JUN	KAERKRM	RNRPA	ASKCRKKKLERIAR	L EEKVKT	L KAQNSE	L ASTANM	L REQVAQL
JUN-B	KVERKRL	RNRPA	ATKCRKKKLERIAR	L EDKVKV	L KAENAG	L SSAAGL	L REQVAQL
JUN-D	KAERKRL	RNRPA	ASKCRKKKLERISR	L EEKVKT	L KSQNT	L ASTASL	L REQVAQL
C-FOS	KRRIRRE	RNKMA	AAKCRNRRRELTD	L QAETDQ	L EDEKSA	L QTEIAN	L LKEKEKL

glutamic acid (negatively charged) totally abolishes the proteins ability to bind to DNA (Dwarki et al 1990). However because this experimental substitution is so radical (ie switching a negatively charged residue for a positively charged one) it is impossible to assess the full implications of this result for the B clone.

The overall basicity of the region is also a little worrying. B has only got an overall charge of +1 whereas yAP-1 has a charge of +8, CREB-1 of +8, and FOS of +9. However the overall charge may not be very important as long as the 'right' charge is in the 'right place'. For example Neuberg et al (1989) and Gentz et al (1989) have shown that at least three basic residues of C-FOS can be simultaneously mutated to uncharged residues without interfering with its ability to bind DNA with C-JUN. Indeed Gentz et al (1989) further demonstrate that replacement of three alanines with three lysines totally abolishes the ability of C-FOS to form a DNA binding complex with C-JUN.

Therefore with the field where it stands it is impossible to predict whether the deviations in the binding domain of B which I have discussed are important enough to eliminate the sequence homology that exists as simply fortuitous and without any functional significance.

### **2.3.2. Leucine Zipper;**

The putative leucine zipper of B is also somewhat problematic. Notably the zipper itself is 1 amino acid residue staggered compared to the other family members; additionally "leucine" positions 3 and 4 are not actually leucine residues in the B clone (see table. 8.1.1). The significance of the first observation has not really been experimentally tested but the disruption of leucines within the zipper has been a thriving topic of investigation. The consensus which has emerged from these analyses indicates that the disruption caused by mutating leucine residues in the zipper depends on; how many leucines are disrupted, what they are changed to, the position of the leucines changed, the particular protein which is being mutagenized, and the protein the mutagenized product dimerises to. For example Smeal et al (1989) show that a double mutation in the C-JUN zipper totally abolishes its ability to dimerise with a jun partner whereas the mutated product can dimerise with c-fos with almost wild type activity. Other studies have indicated that single substitutions with other hydrophobic residues can generally be tolerated but that this toleration is lost when either the number of changes is increased or the single change is more radical. Therefore the substitution of isoleucine for leucine at position 4 is probably not that detrimental as isoleucine itself is a hydrophobic residue whereas the presence of cysteine at position 3 is somewhat more concerning. With regard to this latter change it should be noted that single leucine changes in C-JUN seem to be best tolerated at position 3 (Smeal et al 1989) and in



another yeast leucine zipper factor, yAP1, this same position is occupied by an asparagine residue in the wild type protein (Moye-Rowley et al 1989).

Through these and other experiments (see Introduction) the original model of the leucine zipper has been somewhat modified and the structure is now thought to resemble a leucine based coiled coil. A consequence of this updated zipper model is that the leucine position is not the only important consideration. Instead if leucine occupies residue 4 on an idealised helix then position 1 is also important and should be hydrophobic. A notable exception is C-FOS protein where in two cases this position is occupied by lysine, a positively charged residue. Likewise there are two instances of charged residues in equivalent positions of B (Arginine and Glutamic acid). C-FOS cannot homodimerise and it is believed that the alignment of repulsive charges which would be required for such a structure to exist may be preventative (Kouzarides and Ziff 1989). A similar repulsion would occur in a coiled coil dimer of B where helices are aligned in a parallel fashion which could mean B requires another protein to form a stable DNA binding complex and this could explain its poor performance in gel shift assays. However it should be stressed that a clear understanding of the molecular constraints which determine the preferential choice of partner in dimer formation has not yet been attained. Therefore to predict how B could make its own selection in the cell, based only on its primary sequence, is somewhat premature.

In discussing these results it should be noted that the B clone which has been tested in DNA binding assays is restricted half way through its putative zipper. If the putative DNA region discussed above is actually used by the protein, then the truncation could easily explain the poor performance of B in gel shift assays.

### **2.3.3. Perspectives;**

The CREB/ATF family of proteins have been defined by their DNA binding specificities and this in turn was the criteria used to clone the S.pombe factor. By definition therefore the B clone should be regarded as a member of this family. However before this can be confidently stated further demonstrations of both its ability to specifically recognise an ATF site *in vitro* and *in vivo* should be carried out. By producing the product in an *in vitro* transcription system one may overcome some of the problems which may be associated with a bacterial product. I did not do this experiment with the originally identified fragment as the sequence analysis indicated that the putative DNA binding domain was incomplete. Now that I have cloned flanking sequences this experiment should become more instructive.

However a more pertinent experiment is to actually disrupt the endogenous allele of the gene by standard gene replacement technology and observe the phenotypic effects. This

can be immediately assayed, assuming the knock out is not lethal, by testing the binding activity in extracts prepared from the null mutation. If the null mutation is lethal it will be of interest to test whether it can be rescued by expressing one of the mammalian CREB proteins in the mutant cell. I have constructed a plasmid capable of generating the null phenotype and this will be tested in due course.

## **2.4. Yeast homologues to another family of transcription factors;**

When I started my project there were no reports of yeast transcription factors possessing DNA binding specificities homologous to mammalian proteins. In this way my work was amongst the pioneering studies in this field. However during my thesis a number of other interesting homologies have become apparent and I shall describe some of these in this section.

### **2.4.1. CCAAT box proteins;**

Interestingly, another family of mammalian transcription factors, the CCAAT box binding proteins, also have a homologue in yeast (Chodosh et al 1988b). In mammalian cells there are at a number of different CCAAT box binding activities which can be distinguished chromatographically and by subtle variations in DNA binding specificity (see Johnson and McKnight 1989 for review). One of these, CP1, is itself a heterogeneous complex. It has two components, called CP1a and CP1b, which are unable to efficiently bind DNA on their own but when mixed can associate with a CCAAT box site as a heterodimer (Chodosh et al 1988a). This is similar to the situation found in yeast where a CCAAT box binding complex also contains two components, HAP2 and HAP3. Remarkably by mixing yeast and mammalian components it is possible to generate a heterodimer consisting of proteins from both species i.e. HAP2/CP1a and HAP3/CP1b. This provides an example where both protein-DNA and protein-protein interactions have been conserved.

However despite the fact that there are a number of mammalian CCAAT box binding proteins which have subtle differences in specificity, the HAP2/3 gene products represent the only yeast activity to have been reported. In addition rather than distinguishing between the different types of CCAAT box elements, as the mammalian proteins do, it appears that the yeast HAP2/3 activity can recognise them all (Chodosh et al 1988b). This is somewhat different to what we have found where the small difference in specificity between ATF and AP1 is as apparent in yeast as it is in humans. Mammalian clones for some of the CCAAT box binding proteins, including CTF/NF-1 (Santoro et al 1988) and C-EBP (Landschultz et al 1988a), have recently

become available for analysis. It was anticipated that proteins which recognised such similar DNA sites might well possess homologies in their DNA binding domains. Surprisingly the DNA binding motifs do not appear conserved between these family members (for review see Johnson and McKnight 1989). In addition there is no notable homology between HAP2/3 and these clones although it should be noted that the clones for CP1a and CP-1b have not yet been published. This is in contrast to the family of TRE/CRE binding proteins which all use a basic/leucine zipper motif and also the octamer binding proteins OCT-1 and OCT-2 which use a helix turn helix motif (for review see Johnson and McKnight 1989).

Therefore one might speculate that whereas the AP-1/CREB family of proteins could have arisen from divergent evolutionary processes, where a single progenitor motif has given rise to a number of proteins with the same or slightly varied DNA binding specificity, the CCAAT box binding protein family are more likely a consequence of convergent evolution so that variant proteins from divergent origins have come under selective pressure to recognise the same or similar element. The physiological significance of generating such transcription factor families is discussed in a later section.

#### **2.4.2. SRF and MCM1;**

A number of other yeast and mammalian transcription factors also appear to have related specificity. Human serum response factor, SRF and yeast MCM1 provide an interesting example. Both these proteins recognise similar DNA sites and sequence analysis has demonstrated that they are homologous in their DNA binding domain (Norman et al 1989, Jarvis et al 1989, Hayes et al 1988). However in addition to recognising similar DNA elements both these proteins appear to recruit other factors to the DNA to form a regulatory complex. The example of MCM1 and  $\alpha 1$  has been discussed in the introduction. Similarly it appears SRF can recruit a factor called p62<sup>TCF</sup> to associate with SRE and this association may be critical to the SRE's activating potential (see Treisman 1990 for review). Interestingly a recent experiment has shown that the product of MCM1 gene can also recruit the p62<sup>TCF</sup> factor to DNA indicating that just as in the zipper proteins and the CAAT box proteins, protein-protein as well as protein DNA interactions have been conserved (R. Treisman personal communication).

#### **2.4.3. Heat Shock Factor;**

The first and perhaps best example of a conserved factor is the Heat Shock Factor, HSF. Although this was not formally demonstrated until 1987 (Wiederrecht and Parker 1987), such a result was predicted as the elements mediating the response were known to be conserved among many eukaryotes (Pelham 1985). The interest in this factor is

therefore two fold; firstly with respect to its conserved DNA binding specificity but secondly due to its apparent conservation in regulatory function. Both yeast and mammalian factors are phosphorylated and the phosphorylation state, at least in yeast, probably plays an important role in its regulation (Sorger et al 1987, Sorger and Pelham 1988). The recent cloning of the yeast factor has facilitated preliminary genetic analysis (Sorger and Pelham 1988, Wiederrecht et al 1988); this has shown that it associates with its cognate site as a trimer, the protein subunits interacting via a putative coiled coil structure (Sorger and Nelson 1989). Despite these discoveries it is still unclear exactly how the protein is regulated. Unlike its mammalian counterpart it appears that it can bind DNA in either unstimulated or stimulated cells although the reasons for this apparent difference are also unclear (Sorger et al 1987). In this respect it will be of interest to determine whether the human protein can functionally substitute in a yeast null mutation.

#### **2.4.4. TATA factor;**

The factors discussed in this section have all been implicated with activation/regulation processes. However in the introduction I outlined that the largest subunit of RNA polymerase II also showed a degree of conservation throughout eukaryotes, indicating other initiation factors may be similarly homologous. This was formally demonstrated by Cavalini et al (1988) and Buratowski et al (1988) when they showed that human nuclear extracts deficient of endogenous TFIID could be functionally complemented by a *S.cerevisiae* activity which was shown to recognise the TATA element in a sequence specific manner. In yeast transcriptional start occurs 40-120 bp downstream of the TATA box. However when the yeast activity was placed in a mammalian *in vitro* transcription system it mediated transcriptional start 25-30 bp downstream of the TATA box even if a yeast promoter was used. This indicates that in mammalian cells components other than TFIID are capable of ensuring the correct start site is used. The *S.cerevisiae* TFIID factor has since been purified and cloned by a number of laboratories (Schmidt et al 1989, Horikoshi et al 1989, Eisenman et al 1989, and Hahn et al 1989). Strangely the cloned product encodes only a 27kd polypeptide which is much smaller than the 120-140 kd predicted size of the homologous mammalian protein.

This section has predominantly discussed a number of examples of yeast and mammalian proteins which manifest similarities in DNA binding specificity. In most cases, this also entails the ability of proteins from heterologous systems to interact with each other. Beyond this, however, only the heat shock factor has been shown to have any conservation in regulatory function. The next section discusses whether there is

any potential for  $\gamma$ ATF and  $\gamma$ AP1 to possess regulatory activities which are homologous to their mammalian counterparts. This is a very important consideration as in many ways it represents the critical question which my studies were aimed at addressing.

### **3. Conservation in activation and regulatory properties.**

#### **3.1. AP1 and ATF binding sites can act as UASs in yeast;**

In the results I have shown that the AP1 site can act as a UAS in fission yeast and the ATF site can act as a UAS in budding yeast. Three ATF sites can stimulate expression 70 fold compared to basal levels. Similar tests using three copies of an AP1 site shows a 150 fold stimulation over basal levels in the same circumstances (S.Kuge personal communication). This is considerably more impressive than the 5-fold stimulation seen with an AP1 site in S.pombe, even accepting the fact only a single site was used in this case.

At first glance this could indicate that the budding yeast factor is a stronger constitutive activator; either because it has received an activating modification such as phosphorylation (which the S.pombe factor hasn't) or because it has a stronger constitutive activation motif. Although these explanations should not be discounted it should be noted that the context in which the AP1 site finds itself is different in each test system. A number of experiments have shown that the activating potential of a particular transcription factor can depend on the exact TATA sequence present in the promoter (Harbury and Struhl 1989, Simon et al 1988). In budding yeast the AP1 sites are upstream of a TATA element known to be highly inducible by a number of different UASs. However studies of transcriptional processes in S.pombe are very limited and therefore it is not clear whether the E3 promoter TATA element in the fission yeast constructs is similarly responsive. Therefore it may be the contextual differences which are responsible for the differential activity observed.

#### **3.2. Regulatory Considerations;**

Information about how the CREB/ATF and AP1 families of proteins mediate a regulated transcriptional response is somewhat scarce at present. This is a consequence of a number of factors.

In mammalian cells clones encoding many of these proteins have only recently become available and studies addressing their regulatory functions are, for the most part, at a very preliminary stage. The multitude of clones that have been isolated, and the demonstration that at least some can form heterodimers, indicates that there will be interplay between different family members in response to a particular regulatory stimulus. Although intrinsically interesting this also serves to complicate the issue and reinforces the notion that far more than a preliminary analysis is required to obtain an informed understanding. However in a few cases pertinent progress has been achieved and I will highlight these findings in due course.

In yeast the problem is not so much an overwhelming complexity but rather a lack of it. More specifically the only clones that are presently available for analysis encode GCN4 and yAP1, the latter one only recently having been isolated. No yATF clone, the potential candidate in the form of clone B apart, has yet been reported.

For these reasons a comparison of the regulatory roles between yeast and mammalian systems is somewhat premature. However due to the nature of my thesis I feel obliged to at least try discuss possibilities even if the ideas put forward are of a somewhat simplistic and speculative nature.

### 3.2.1. AP1;

In mammalian cells both c-fos and c-jun have been cloned by virtue of the fact that transforming retroviruses have transduced a portion of their coding regions and can consequently act to transform cells. In addition the mRNA levels of c-jun, jun-b, c-fos, fos-b and fra1 are all rapidly induced by different mitogen<sup>ic</sup> stimulation of quiescent cells (Bravo et al 1990). This indicates an important role for these genes in the control of cell growth presumably by regulating the transcription of other genes. Beyond this however a comprehension of how the effect is mediated is largely unknown. In addition, apart from changes in absolute protein levels it is also unclear how the activity of the products are themselves regulated. In this sense therefore it is very difficult to extrapolate instructively to the yeast system. The GCN4 factor in budding yeast mediates transcriptional activation in response to amino acid starvation and therefore probably bears little regulatory relevance to the mammalian AP1 proteins. In addition GCN4 can not form dimers with either C-FOS or C-JUN indicating that even within the conserved zipper region there are functional dissimilarities.

However with regard to yAP-1 very little is known. It was not cloned by genetic means, so its physiological role is not immediately apparent. Even so it is worth noting that a yAP-1 null mutation is viable indicating that if it is involved in stimulating cell growth, as its mammalian counterparts seem to be, it is redundant in this role. Another

question that immediately comes to mind is the possible existence of yeast proteins which heterodimerise with yAP1 just as FOS does with JUN. The purification of yAP-1 as described in this thesis and by Harshman et al (1988) indicates a predominant protein of 85kd as responsible for yAP1 band shift activity. Therefore although it is conceivable that such proteins do exist in budding yeast we don't have any evidence at present to indicate this is so.

In the case of fission yeast even less is known. However I have demonstrated that both human and fission yeast AP1 activity is sensitive to phosphatase treatment. Due to the multiplicity of JUN/FOS proteins it is difficult to make instructive conclusions on the human data. In the case of fission yeast however the mobility changes are quite large, which is indicative of the dissociation of a multimeric complex. It may be of interest to establish whether such a complex is heterogenous in its content.

### 3.2.2. ATF/CREB;

The recent isolation of multiple ATF/CREB clones would indicate that mammalian ATF activity is capable of pleiotropic effects. Two of these, activation by either cAMP or E1A, have already been identified and characterised but it is likely there will be more. If and when future regulatory roles for the CRE become apparent they can also be analysed in yeast, but for the present one is limited to determine whether either cAMP or E1A can mediate their effects through a CRE in yeast. With respect to the latter of these two mechanisms I have shown that E1A does not appear to activate the expression of the E1A inducible E2A promoter in fission yeast. In retrospect, expecting a viral protein to function in a yeast cell may have been asking a little too much. However a more reasonable conception is that yATF may be regulated by fluctuating cAMP levels. In mammalian cells cAMP mediated induction is mediated through a cAMP inducible kinase (see introduction). CREB-1 can be phosphorylated by the catalytic subunit of this enzyme *in vitro* and a phosphorylation site has been found in the predicted amino acid sequence of the CREB-1 clone (Gonzalez et al 1989a). When this site is mutated, so it can no longer be phosphorylated, activity can no longer be induced from basal levels. Interestingly transcription remains at a basal level even if this serine is mutated to an acidic residue (Gonzalez et al 1989b). This argues against the idea of some workers (see Ptashne 1988) that phosphorylation may potentiate the activation of transcription factors by simply increasing a local concentration of negative charge. The cAMP response in budding yeast is also mediated by cAMP inducible protein kinases indicating that transcriptional responses may also be mediated in a similar way (reviewed in Broach and Deschenes 1990). I have presented data in chapter 6 indicating that the budding yeast yATF complex is phosphorylated and that the

phosphorylation state of the factor can be increased by the catalytic subunit of the cAMP inducible kinase in vitro. The phosphorylation state itself seems to determine the gel shift mobility observed when this factor binds to DNA. A number of reasons may explain these differences. One possibility is that  $\gamma$ ATF binds DNA as a monomer and that the different complexes represent different phosphorylation states of the monomer. It is also possible that more than one protein molecule associates directly or indirectly with the DNA template in a phosphorylation dependent manner. In this case phosphatase treatment and subsequent dephosphorylation of the components would lead to a dissociation of the complex and consequently an increase in electrophoretic mobility. The apparent dependence of protein/DNA interactions on the phosphorylation state of the extract can be explained by a similar scenario. In this case however the modification of either a monomeric DNA binding protein or a phosphorylation dependent protein-protein complex would have a direct effect on DNA binding activity. Another possibility is that each complex is due to totally different proteins each one having a different phosphorylation dependence for binding DNA.

The biological significance of this data remains to be tested. Indeed there is no evidence that the ATF activity I see in vitro is at all involved with transcriptional regulation in response to cAMP. Unfortunately cAMP inducible promoters have not yet been characterised in yeast and so it is difficult to gain the kind of physiological insight possible in mammalian cells. However the genetic approach of investigating cAMP pathways in budding yeast has resulted in the isolation of numerous mutant strains and regulatory products. Thus one can ask a direct question of whether a CRE is responsive to fluctuating cAMP levels by assaying the activity of the ATF site constructs described in chapter 4 in a number of mutant strains which possess different constitutive kinase activities. Such experiments are presently being carried out by a postdoctoral fellow in the laboratory.

### **3.3. Why have multiple proteins recognising the same site ?**

Throughout the discussion I have described the identification of multiple proteins which can recognise the same DNA element either but have not really outlined why such a phenomenon<sup>oA</sup> may have evolved. In this section I shall discuss work that gives an insight into this question and speculate about the biological processes where it may be important.

#### **3.3.1. Molecular regulatory interplay;**



The reason for having multiple proteins recognising the same site either as homo- or hetero-dimers is not immediately clear. However a likely explanation is that it allows regulatory interplay between multiple factors to facilitate pleiotropic regulation of transcription by both positive and negative effectors. The facility of forming heterodimers would simply serve to increase the variety. With regard to heterodimer formation one can imagine, in a simple scenario, three possible consequences of interactions; an increase in activating potential, a decrease in activating potential, and a changed specificity of activation. In the case of C-JUN it appears it's choice of partner can influence all three criteria.

I have already discussed examples of two of the above; the heterodimerisation of C-FOS with C-JUN increases both the ability of C-JUN to interact with a TRE and also increases it's ability to activate transcription (Kouzarides and Ziff 1989); on the other hand heterodimerisation with CREB-1 changes the apparent specificity of C-JUN so that rather than binding a TRE, as it would on its own or with increased affinity with C-FOS, the CREB/JUN heterodimer preferentially associates with a CRE (Benbrook and Jones 1990).

It has also recently become apparent that negative controls can also be mediated by heterodimerisation. For example expression of C-JUN in F9 cells can activate transcription through a single AP-1 site but this activation is repressed by co-expression of JUN B; expression of JUN B alone has no effect on a single AP-1 site (Chiu et al 1989b). Strangely a different result is obtained if a number of AP-1 sites are used. In this case both C-JUN and JUN B can activate transcription when expressed alone and this effect is enhanced when either is coexpressed with C-FOS. Importantly however if C-JUN is expressed in the presence of both C-FOS and JUN B then essentially no activation of transcription is observed (Schutte et al 1989). All these proteins can form heterodimers with each other *in vitro* and so a logical conclusion of this data is that the C-JUN/JUN B heterodimer forms in preference to that of C-FOS/JUN or each respective homodimer and acts as a dominant repressing influence. This is somewhat strange considering Nakabeppu et al (1989) have shown that when these proteins are produced *in vitro* JUN/C-FOS heterodimers can bind an AP-1 site at least 10-fold better than C-JUN/JUN B heterodimers. To accommodate all the data it must be concluded that C-JUN/JUN B once formed has a low affinity for DNA.

### 3.3.2. Developmental considerations;

This type of regulation could theoretically occur for any group of transcription factors which can interact with the same site or with each other. Indeed recent experiments involving the helix-loop-helix (HLH) protein family have indicated both positive and

negative effects consequential to heterodimerisation. In the Introduction I discussed how heterodimers between different HLH family members can recognise the same DNA element. However a recently discovered member of this family, the ID protein, is somewhat different to the others as it lacks a basic domain adjacent to its HLH motif. This means it cannot bind DNA as a homodimer and importantly can act as a negative regulator of transcription by heterodimerising with other HLH proteins, such as MyoD, thereby preventing them from binding DNA. Interestingly during myogenesis ID levels decrease which presumably allows MyoD, whose levels do not change, to dissociate from non-binding heterodimers and form functional activating complexes with specific promoters to direct differentiation.

It will be no surprise to find zipper proteins in similar kinds of differentiation and developmental processes. It is certainly true that c-jun and Jun-B are differentially expressed during embryogenesis (Wilkinson et al 1989) and molecular demonstrations of the kind described above will in all probability be quick to follow. Once again this indicates that the complexity involved is unlikely to be as apparent in a single celled organism such as yeast. However it is equally pertinent to point out that regulatory parallels between human and yeast systems can not be discounted until a role for all the family members have been described.

### **3.4. Conservation of activation in general;**

A number of studies have addressed whether the mechanism by which transcriptional activation occurs is conserved throughout eukaryotes. For the most part this has simply involved expressing mammalian activators in yeast and yeast activators in mammalian cells. The first successful demonstration of such heterologous activation came from studies using the GAL4 protein. Expression of this factor in mammalian cells led to the activation of a promoter containing its cognate binding site. Mutants of GAL4 defined as defective activators in yeast were also defective at activating transcription in mammalian cells (Kakadini and Ptashne 1988, Webster et al 1988). In vitro analysis has served to reinforce this concept. In a footprinting assay similar to that which demonstrated ATF-TFIID interactions (Horikoshi et al 1988a), Horikoshi et al (1988b) have shown that GAL4 can also alter the way in which mammalian TFIID associates with DNA. Interestingly mutants of GAL4 which can bind DNA but cannot activate transcription do not qualitatively change the TATA footprint, suggesting the factor interactions detected in binding assays are directly relevant to the activation process (Horikoshi et al 1988b).

Reciprocal type experiments demonstrated that both the oestrogen receptor and the glucocorticoid receptors can activate transcription in yeast (Metzger et al 1988, Schena

and Yamamoto 1988). One of the activation domains of the glucocorticoid receptor is an acidic helical domain, reemphasizing the apparent importance of this motif for transcriptional activation (Hollenberg and Evans 1988). However there are no recognised activation motifs in the oestrogen receptor.

The last point is of some interest as the best characterised yeast activating domains all appear to be acidic (Hope and Struhl 1986, Ma and Ptashne 1987a). Indeed the only motif that was found when random sequences were tested for activating potential was an acidic helix (Ma and Ptashne 1988). Although this result may have been biased by the limitations of the assay it is suggestive of a very restricted type of activation mechanism in yeast which may exclude the proline rich and glutamine rich sequences which have been characterised in mammalian factors. It will be of interest to see whether mammalian activators such as SP1 and AP-2, which do not contain acidic activating domains, can function efficiently in yeast.

An interesting development which has been consequential of this type of experimental approach has been the identification that activator proteins can be of differing strengths. The strongest activator so far identified is the VP16 product of the Herpes Simplex Virus which, with other cellular factors, forms a specific DNA binding complex which activates early viral gene expression (see Goding and O'Hare for review). When this product is artificially brought to DNA by fusing it with the DNA binding region of the GAL4 product it can activate transcription in HeLa cells 100-fold more efficiently than the wild-type GAL4 product (Sadowski et al 1988). A similar result has been demonstrated for the Adenovirus E1A product, although in this case there is no definitive evidence that the product normally transactivates by forming a complex with DNA (Lillie and Green 1989, Nevins 1990). The molecular basis for these differences in activation potential is not clear but it probably involves differential abilities of activators to form associations with components of the initiation machinery.

#### **4. Summary and perspectives.**

The studies I have described in my thesis have convincingly demonstrated that the binding specificities of a number of transcription factors have been remarkably conserved between yeast and mammalian cells. As well as yATF and yAP-1 yeast factors have been identified which have a similar specificity to mammalian CCAAT box binding proteins, heat shock factor, serum response factor, TFIID, and there are also a number of others. In some of these cases, such as TFIID and the CCAAT box binding proteins, proteins from heterologous systems can productively interact with each other

as well as the same DNA sequence; and in one case, the heat shock factor, the yeast and mammalian proteins are also involved in the same regulatory response.

The question now becomes: what experiments are instructive to expand this analysis? In addressing this problem I will speculate on topics directly relevant to my field of study, but for the most part this will be generally relevant.

A major goal is to decipher the regulatory roles of yATF and yAP1. In mammalian cells the CRE/ATF site is responsive to E1A and fluctuating levels of cAMP. The cAMP pathway in yeast provides a particularly instructive starting point as it is clearly biologically relevant and its effects appear to be mediated by cAMP inducible protein kinases. Such kinases are also present in mammalian cells where they have been shown to phosphorylate and activate the transcription factor CREB-1. The type of experimental approach suitable for testing the ability of a CRE to respond to fluctuating cAMP levels in yeast has been outlined in an earlier section.

However cAMP regulated transcription in yeast as a whole has not yet been well defined. Genetic analysis which has been so useful in isolating so many of the regulatory products involved in this pathway has not yet identified cAMP inducible transcription factors or promoters. In this respect it may be instructive to test the feasibility of such a transcriptional response by introducing the mammalian CREB-1 protein. To eradicate the background activity of endogenous yATF proteins CREB-1 could be linked to a heterologous DNA binding such as lex-A. In this way one can test for cAMP inducibility of the fusion protein in the same way as we are presently testing the effect of cAMP on a CRE, but instead using lexA binding sites in the test gene (Brent and Ptashne 1985). In addition we could mutate the serine residue, which is critical and becomes phosphorylated during cAMP induction of transcription in mammalian cells (Gonzalez et al 1989b) and see whether it is similarly disruptive in yeast. In this way we can establish whether there is at least the potential for a cAMP stimulated transcriptional response even if we cannot yet identify the endogenous transcription factors responsible.

However at this point it should be noted that, apart from CREB-1 which is clearly activated by cAMP, the regulatory roles for most of the mammalian CREB/ATF clones that have been isolated have not been determined. It is thought that this family may have evolved through duplication of a progenitor DNA binding domain with its subsequent insertion adjacent to heterologous regulatory domains, thus producing a series of proteins with the same or a similar DNA binding specificity but with pleiotropic regulatory effects (Ziff 1989). Obviously some of these domains are responsive to cAMP, and/or E1A but it is likely that other forms of regulation will also become apparent.

Yeast represents a more primitive form of this family and so to test specifically for cAMP mediated stimulation through a CRE in this system presupposes that the CREB DNA binding motif had already attained a region regulatable by cAMP at this early stage. Although this is entirely possible it is not inevitable and so it may be more instructive to ask what yATF does with a more open ended approach. This would involve integrating the CRE- $\beta$ -galactosidase construct, described in chapter 4, into the yeast genome and isolating mutants with altered  $\beta$ -galactosidase activity.

Characterisation of the products obtained from such an analysis would be directly relevant to the regulation of yATF. In addition the findings may be informative to the regulation of one or more of the mammalian CREB/ATF family.

Most of this type of experimental analysis is more suited to the budding yeast system at present simply because transcriptional response mechanisms and the cAMP pathway have been more comprehensively characterised in *S.cerevisiae*. However our understanding of such regulatory processes is continually expanding in *S.pombe*; the recent characterisation of a highly inducible *S.pombe* promoter (Hoffman and Winston 1989) provides a good starting point from which to develop the kind of assay system used to test UAS sequences that is available in *S.cerevisiae*.

Once available this type of assay will greatly facilitate instructive transcriptional studies and this will be very useful in testing, for example, the functional consequences of disrupting the B clone locus. If this results in a marked decrease in ATF site driven promoter activity and the disruption is not a lethal event, then a B-clone null may provide a convenient environment in which to isolate other CREB/ATF protein activators from heterologous systems including *S.cerevisiae*.

Once isolated and fully characterised it may be interesting, despite the arguments I have levelled against such experimentation, to test whether yeast clones can mediate cAMP stimulation of transcription in mammalian cells. I had at one stage hoped to carry out all these types of experiments myself during the period of my thesis. Unfortunately I happened upon a major stumbling block in obtaining a yeast clone in the first place.

Once there is a clear indication that this problem has indeed been overcome then there is a very exciting series of investigations in prospect.

**CHAPTER 8; REFERENCES**

Aebi, M., Furtur, R., Prantl, F., Niederberger, P. and R. Hutter (1984) *Curr. Genet.* 8 165-172.

Anderson, J.E., Ptashne, M., and S.C.Harrison. (1987). *Nature.* 326. 846-852.

Angel,P., Imagawa,M., Chiu,R., Stein,B., Imbra,R.J., Rhamsdorf,H.J., Jonat,C., Herrlich,P. and Karin,M. (1987) *Cell.* 49. 729-739

Allison, L.A., Wong J.K.-C, Fitzparick, V.D., Moyle, M. and C.J. Ingles.1988. *Mol.Cel.Biol.* 8. 321-329

Allison, L.A., and C.J. Ingles (1989) *Proc. Natl. Acad. Sci. USA.* 86. 2794-2798.

Angel,P., Allegretto,E., Okino,S., Hattori,K., Boyle,W.J., Hunter,T. and Karin,M. (1988) *Nature* 332. 166-171

Antequera, F., Macleod, D., and A.P.Bird (1989) *Cell.* 58. 509-517

Arndt,K.T., Styles,C. and Fink,G.R. (1987) *Science.* 237. 874-880

Beach,D. and Nurse,P. (1981) *Nature.* 200. 140-142

Bembrook, D.M., and N.C.Jones (1990) *Oncogene.* 5. 295-302.

Benoist,C. and Chambon,P. (1981) *Nature.* 290. 304-310

Berk, A.J. (1986) *Annu. Rev. Genet.* 20 45-79.

Biggin, M.D., and R.Tjian (1989) *T.I.G.* 5. 377-383

Bird, A.P. (1987) *TIG.* 3. 342-347.

Bohmann, D., Bos, T.J., Admon, A., Tetsuji, N., Vogt, P.K. and R.Tjian. (1987) *Science.* 238. 1386-1392

- Bohmann, D., R.Tjian (1989) *Cell*. 59. 709-717.
- Boeuf, H., Zajchoowski, P.A., Tamura, T., Hauss, C., and C.Kedinger, (1987) *Nucleic Acid Res.* 15. 509-527
- Bolivar, F., and K. Backman. 1979. *Methods Enzymology*. 68; 245
- Bram, R.J. and R.D.Kornberg (1985) *Proc. Natl. Acad. Sci. USA*. 82. 43-47
- Bravo, R., (1990) *Seminars in Cancer Biology*. in press.
- Brandl, C.J., and Struhl, K (1989) *Proc.Natl. Acad.Sci.USA*. 86. 2652-2656.
- Brent, R., (1985) *Cell*. 42. 3-4
- Brent, R., and M.Ptashne. (1985) *Cell* 43. 729-736
- Breathnach, R.,and P.Chambon.(1981) *Annu. Rev. Biochem.* 50. 349-383.
- Broach, J.R., and R.J.Deschenes (1990) *Adv. Can. Res.* 54. 79-139.
- Buratowski, S., Hahn, S., Guarente, L and P.Sharp (1989) *Cell*.56. 2509-2516.
- Buratowski S, Hahn S, Sharp PA, and L.Guarente (1988) *Nature*. 334. 37-42.
- Busslinger, M., Hurst, J., and R.A.Flavell. (1983) *Cell*. 34. 197-206.
- Cavallini B, Huet J, Plassat J, Sentenac A, Elgy J, Chambon P (1988) *Nature* 334:77-80.
- Cameron, S., Levin, L., Zoller., and M.Wigler (1988) *Cell*. 53. 555-566
- Cannon, J., and K.Tatchel (1987) *Mol. Cell. Biol.* 7. 2653-2663.
- Cedar, H. (1988) *Cell* 53. 3-4
- Chambon, P., *Ann.Rev.Biochem.* 44. 613-638.

Chen,W. and Struhl,K. (1985) EMBO J. 4. 3273-3280.

Chen, W., and K.Struhl (1989) EMBO J. 8. 261-268.

Chiu, R., Boyle, W.J., Meek, J., Smeal, T., Hunter, T., and M.Karin (1988) Cell. 54. 541-552.

Chodosh LA, Baldwin AS, Carthew RW, Sharp PA (1988a) Cell. 53. 11-24.

Chodosh LA, Olesen J, Hahn S, Baldwin AS, Guarente L, Sharp PA (1988b) Cell. 53. 25-35.

Chiu R, Boyle WJ, Meek J, Smeal T, Hunter T, Karin M (1988) Cell 54. 541-552.

Clarke, J., and W.Herr (1987) J.Virol. 61. 3536-3557.

Cohen DR, Ferreira PCP, Gentz R, Franza BR, Curren T (1989) Genes. Dev. 3. 173-184.

Cohen, C. and D.A.D.Parry. (1986) TIBS 11. 245-248.

Cortes, P., Buckbinder, L., Leza, M., Rak, N., Hearing, P., Merino, A., and D. Reinberg (1988) Genes. Dev. 2. 975-990.

Courey, A.J., and R.Tjian (1988) Cell. 55. 887-898.

Crick, F.H.C (1952) Nature. 170. 882-883.

Davis, R.L., Lassar, A.B., and H.Weintraub (1987) Cell. 51. 987-1000.

Davis, R.L., Cheng, P.F., Lassar, A.B., and H.Weintraub (1990) Cell 60. 733-746.

Deutsch, P.J., Jameson, J.L., and J.F.Habener (1987) J.Biol. Chem. 262. 12169-12174.

Dingham, J.D., Lebowitz, R.M., and R.G.Roeder. (1983) Nucl. Acids Res. 11, 1475-1489.



- Distel, R.J., Ro, H.S., Rosen, B.S., Groves, D.L., and B.M.Spiegelmann (1987) Cell .49. 835-844.
- Donahue. T.F., Daves, R.S., Lucchini, G and G.R.Fink (1983) Cell 32 . 89-98.
- Durkacz, B., Beach,D., Hayles,J. and P.Nurse (1985) Mol.Gen.Genet. 201. 543-545
- Dwarki, V.J., Montminy, M.R., and I.M.Verma (1990) EMBO J. 9. 225-232.
- Eisenmann DM, Dollard C, Winston F (1989) Cell. 58. 1183-1191.
- Elkaim, R., Goding, C., and C.Kedinger, (1984) Nucleic Acids Res. 11, 7105-7117
- Engel, D.A., Hardy., S., and T. Shenk (1988) Genes. Dev. 2. 1517-1528
- Fire, A ., Samuels, M., and P.A.Sharp (1984) J.Biol. Chem. 259. 2509-2516.
- Franza B.R, Rauscher F.J, Josephs S.F, and T.Curran (1988) Science. 239. 1150-1153.
- Freedman, L.P., Luisi,B.F., Koorszun, Z.R., Basavappa, R., Sigler, P.Ba and K.R. Yamamoto (1988) Nature 334, 543-546.
- Fried M, Crothers DM (1981) Nucl.Acids Res. 9. 6505-6525.
- Fromm,M. and Berg,P. (1982) J.Mol.Appl.Genet. 1, 457-481
- Fromm,M. and Berg,P. (1983) Mol.Cell.Biol. 3. 991-999.
- Galas DJ, Schmitz A (1978) Nucl.Acids Res. 5:3157-3170.
- Gentz, R., Rauscher, F.J, Abate, C., and T.Curran (1989) Science 243. 1695-1699.
- Gibson T J (1984) PhD Thesis, Cambridge
- Gidoni,D., Dynan,W.S. and R.Tjian (1984) Nature. 312. 409-413

- Gill, G., and M.Ptashne (1987) *Cell*. 121-126.
- Giniger, E., Varnum, S.M., and M.Ptashne. (1985) *Cell*. 40. 767-774.
- Giniger, E., and M.Ptashne (1987) *Nature* 330. 670-672.
- Goding, C.R., and P.O'Hare (1989) *Virology*. 173. 363-367.
- Gonzalez, G.A., Yamamoto, K.K., Fischer, W.H., Karr, D., Menzel, P., Biggs, W., Vale, W.W, and M.R.Montminy (1989) *Nature*. 337. 749-752.
- Gonzalez, G.A., and M.R.Montminy (1989) *Cell*. 59. 675-680.
- Gosh,P.K., Lebowitz,P., Frisque, R.J. and Y.Gluzman. (1981)  
*Proc.Natl.Acad.Sci.USA*. 78, 100-104
- Goodbourne, S. (1990) *Seminars in cancer biology*. in press.
- Gorman, C.M., Moffat, L., and B.H.Howard (1982) *Mol.Cell.Biol*. 2. 1044-1051
- Gorman, C.M., Rigby, P.W.J. and Lane, D.P. (1985) *Cell* 42. 519-526
- Graham, F.L., Smiley, J., Russell, W.C., and R.Nairn (1977) *J. Gen. Virol*. 36, 59-68.
- Greenberg, M.E., and E.B.Ziff (1984) *Nature* 311. 433-438.
- Gross, D.S., and W.T.Garrard (1988) *Ann. Rev. Biochem*. 57. 159-197.
- Guarente, L., Nye, J.S., Hochschild, A., and M.Ptashne. (1982) *Proc. Natl. Acad. Sci. USA*. 2236-2239.
- Guarente, L., and T. Mason (1983) *Cell*. 32. 1279-1286.
- Hahn, S., E.T.Hoar and L.Guarente.(1985). *Proc. Natl. Acad. Sci. USA*. 82.8562-8566.
- Hahn S, Buratowski S, Sharp PA, Guarente L (1989) *Cell* 58:1173-1181.

- Hai, T., Horikoshi, M., Roeder, R.G., and M.R.Green (1988a) *Cell*. 54. 1043-1051.
- Hai, T., Liu, F., Allegreto, E.A., Karin, M., and M.R.Green (1988b) *Genes. Dev.* 2. 1216-1226.
- Hai, T., Liu, F., Coukos, W.J., and M.R.Green (1989) *Genes. Dev.* 3. 2083-2090.
- Halazonetis TD, Georgopoulos K, Greenberg ME, Leder P (1988) *Cell*. 55. 917-924.
- Han, M., and M.Grunstein. (1988) *Cell*. 55. 1137-1145.
- Harbury, P.A.B., and K.Struhl. (1989) *Mol. Cell. Biol.* 9. 5298-5304.
- Hardy, T., and T.Shenk (1988) *Proc. Natl. Acad. Sci. USA*. 85. 4171-4175.
- Harshman, K.D., Scott Moye-Rowley and Parker,C.S. (1988) *Cell*. 53. 321-330.
- Hawley, D.K., and R.G.Roeder (1987) *J.Biol.Chem.* 262. 3452-3461.
- Hayes, T.F., Sengupta, P., and B.H.Cochran (1988) *Genes. Dev.* 2. 1713-1722.
- Herr, W., and J. Clarke (1986) *Cell*. 45. 461-470.
- Herskowitz, I (1989) *Nature*. 342. 749-757.
- Hill, D.E., Hope, I.A., Mackie, J.P., and K.Struhl. (1986) *Science* 234, 451-457
- Hinnebush, A.G. and G.R.Fink (1983a). *Proc. Natl. Acad. Sci. USA*. 80. 5374-5378
- Hinnebush, A.G. and G.R.Fink (1983b). *J.Biol.Chem.* 258. 5238-5247
- Hinnebush, A.G. (1984) *Proc. Natl. Acad. Sci. USA*. 81. 6422-6426.
- Hinnebush,A.G., Lucchini,G. and G.R.Fink (1985) *Proc.Natl.Acad.Sci.USA*. 82. 498-502

- Hirai SI, Ryseck RP, Mechte F, Bravo R, and M.Yaniv (1989) EMBO J. 8:1433-1439
- Hoeffler JP, Meyer TE, Yun Y, Jameson JL, and J.F. Habener (1988) Science 242:1430-1433.
- Hoffman, C.S., and F.Winston (1989) Gene. 84. 473-479.
- Hollenberg, S.M., and R.M.Evans (1988) Cell. 55. 899-906
- Holler, M., Westin, G., Jiricny, J., and W.Schaffner.(1988) Genes Dev. 2. 1127-1135.
- Hope,I.A. and K.Struhl. (1985) Cell. 43. 177-188
- Hope, I.A. and K.Struhl (1986) Cell. 46. 885-894.
- Hope, I.A. and K.Struhl (1987) EMBO J . 6. 2781-2784.
- Hope, I.A., Mahadevan, S., and K.Struhl. (1988). Nature. 333. 635-640.
- Horikoshi, M., Hai, T., Lin, Y.S., Green, M.R., and R.G.Roeder (1988a) Cell. 54. 1033-1042.
- Horikoshi, M., Carey, M.F., Kakidoni, H., and R.G.Roeder (1988b) Cell 54:665-669.
- Horikoshi, M., Wang, C.K., Fujii, H., Cromlish, J.A., Weil, P.A., and R.G.Roeder (1989) Nature. 341. 299-302.
- Hurst, H.C., and N.C.Jones. (1987) Genes. Dev. 1, 1132-1146
- Imbra, R.I and Karin, M. (1986) Nature. 323. 555-558
- Imigawa, M.M Chiu, R., and M.Karin (1987) Nature 329, 648-651.
- Imperiale, M.J., and J.R.Nevins (1984) Mol. Cell. Biol. 4. 875-882.

- Jarvis, F.F., Clark, K.L., and G.F.Sprague (1989) *Genes. Dev.* 3. 936-945.
- Johnson, A.D and I.Herskowitz (1985) *Cell* 42. 237-247.
- Johnson, M. (1987) *Nature* 328. 353-355.
- Johnson, M. *Microbiological Reviews* (1987) 51. 458-476.
- Johnson, P.F., Landschultz, W.H., Graves.B.J., and S.L.McKnight (1987) *Genes Dev* 1. 133-146.
- Johnson, P.F., and S.L.McKnight (1989) *Ann. Rev. Biochem.* 58. 799-839.
- Jones, K.A., Yamamoto, K.R., and R.Tjian (1985) *Cell.* 42. 559-572.
- Jones, K.A., Lucciw, P.A., and R.Tjian (1988) *Genes. Dev.* 2. 1101-1114.
- Jones, N.C., Rigby, P.W.J., and E.Ziff (1988) *Genes. Dev.* 2. 267-281
- Jordan, S.R., and C.O.Pabo (1988) *Science* 242. 893-899.
- Kakidani,H. and M.Ptashne (1988) *Cell* 52. 161-167
- Kadonaga J.T, and R.Tjian (1986) *Proc.Natl.Acad.Sci.USA.*83. 5889-5893.
- Kaufner,N.F., Simanis,V., and P.Nurse (1985) *Nature* 318, 78-80.
- Keegan, L., Gill, G., and M.Ptashne. (1986) *Science* 231. 699-704.
- Kouzarides, T., and E.Ziff (1988). *Nature* 336. 645-651.
- Kouzarides, T., and E.Ziff. (1989). *Nature.* 340. 568-571.
- Kouzarides, T., and E.Ziff. (1989). *Cancer Cells* 1, 71-76
- Kovesdi, I., Reichel, R., and J.R.Nevins (1987) *Proc. Natl. Acad. Sci. USA.* 84. 2180-2184.

- Landschulz, W.H., P.F.Johnson, E.Y.Adashi, B.J.Graves and S.L.Mcknight.  
(1988) Genes. Dev. 2: 786-800.
- Landschulz, W.H., P.F.Johnson, and S.L.Mcknight. (1988a) Science 240. 1759-1764.
- Landschulz, W.H., P.F.Johnson, and S.L.Mcknight (1989b) Science 243. 1681-1688.
- Laughton, A. and M.P.Scott (1984) Nature. 310. 25-31.
- Lee, K.A.W., and M.R.Green (1987) EMBO J. 6. 1345-1353.
- Lee, K.A.W., Hai,T.Y., Siva Raman, L., Thimmappaya, B., Hurst,H.C., Jones, N.C., and M.R.Green (1987) Proc.Natl.Acad.Sci.USA. 84, 8355-8359
- Lee, K.A.W., Fink, J.S., Goddman, R,H., and M.R.Green (1989) Mol. Cell. Biol. 9. 4390-4397.
- Lee, M.G., and P.Nurse (1987) Nature. 327. 31-35
- Lee,W., Haslinger, A., Karin, M., and R.Tjian (1987a) Nature. 325. 368-372
- Lee,W., Mitchell, P. and R.Tjian (1987b) Cell. 49. 741-752
- Leza, M.A., and P.Hearing (1988) J.Virol. 62. 3003-3013.
- Lillie, J.W., and M.R.Green (1989) Nature. 338. 39-44.
- Loeken, M.R. and J.Brady (1989) J.Biol.Chem. 264, 6572-6579.
- Lucchini,G., Hinnenbusch, A.G., Chen.C. and G.R.Fink (1984) Mol.Cell.Biol..4.1326-1333.
- Ma, J. and M.Ptashne (1987a) Cell. 48. 847-853
- Ma, J. and M.Ptashne (1987b) Cell. 50. 137-142

Ma, J. and M.Ptashne (1987c) Cell. 51. 113-119

Maekawa, T., Sakura, H., Kanei-Ishii, C., Sudo, T., Yoshimura, T., Fujisawa, J., Yoshida, M., and S.Ishii (1989) EMBO J. 8:2023-2028.

Maki, Y., Bos, T.J., Davis, C., Starbuck, M., and P.K.Vogt (1987) Proc. Natl. Acad. Sci. 84. 2848-2852.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: a laboratory manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory)

Maniatis, T., Goodbourn, S. and J.A.Fischer (1987) Science. 236. 1237-1244

Manley, J.L., Fire, A., Cano, A., Sharp, P.A., and M.L.Gefter. (1980) Proc. Natl. Acad. Sci. USA. 77. 3855-3859.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and A.P.Bird. (1989) Cell. 58. 499-507.

Mermod, N., O'Neill, E.A, Kelly, T.J., and R.Tjian (1989) Cell. 58. 741-753.

Metzger, D., White, H.H., P.Chambon (1988) Nature. 334. 31-36.

Miller, J., McLachlan, A.D., and A.Klug. (1985) EMBO J. 4. 1609-1614

Mitchell, P.J., Wang, C. and R.Tjian. (1987) Cell. 50. 847-861

Mitchell, P.J., and R.Tjian (1989) Science. 245. 371-378.

Montminy, M.R., and L.M.Bilezikjian (1987) Nature. 328. 175-178.

Montminy, M.R., Sevarino, K.A., Wagner, J.A., Mandel, G., and R.H.Goodman, (1986) Proc. Natl. Acad. Sci. USA. 83. 6682-6686.

Moye-Rowley, W.S., Harshman, K.D., and C.S.Parker (1989) Genes. Dev. 283-292.

- Murray, E.J., and F.Grosveld (1987) EMBO. 6. 2329-2335.
- Murthy,S.C.S., Bhat,G.P. and B.Thimmappaya. (1985) Proc.Natl.Acad.Sci.USA. 82, 2230-2234
- Murre, C., McCaw, P.S., and D.Baltimore (1989a) Cell 56. 771-783.
- Murre, C., McCaw, P.S., Vassein, H., Caudy, M., Jan, Y.N., Cabrera, C.V., Buskin, J.N., Hauschka, S.D., Lassar,A.B., Weintraub, H., and D.Baltimore (1989b) Cell 58. 537-544.
- Nagawa,F. and Fink,G.R. (1985) Proc. Natl. Acad. Sci. USA. 82. 8557-8561.
- Nahon, J.L., Venetianer, A., and J.M.Sala-Trepat. (1987) Proc. Natl. Acad. Sci. USA. 84. 2135-2139.
- Nakabeppu Y, Ryder K, and D.Nathans (1988) Cell 55. 907-915.
- Nevins, J.R., (1990) Seminars in cancer biology (in press)
- Neuberg, M., Adamkiewicz, J., Hunter, J.B., and R.Muller (1989) Nature. 341. 243-245. n
- Nishizuka, Y., (1984) Nature 308. 693-698.
- Norman, C., Runswick, M., Pollock, R., and R.Treisman (1988) Cell 55:989-1003.
- Ondek, B.A., Shepherd, A., and W.Herr (1987) EMBO J. 6. 461-470.
- O'Shea E.K., Rutkowski, R., and P.S.Kim (1989) Science 243. 538-542.
- Otwinski, Z., Schevitz, R.W., Zhang, R.G., Lawson, C.L., Joachomimic, A., Marmorstein, R.Q., Luisi, B.F., and P.B.Sigler (1988) Nature. 335. 321-329.
- Pelham, H.R.B. (1985) Trends in Genet. 1. 31-35



- Penn, M.D, Galgoci, B. and H. Greer. (1983) Proc. Natl. Acad. Sci. USA. 80. 2704-2708.
- Pfeifer, K., Arcangioli, B., and L.Guarente. (1987) Cell. 49. 9-18.
- Ptashne, M (1988) Nature 335. 683-689.
- Rauscher, F.J., Cohen, D.R., Curren, T., Bos, T.J., Vogt, P.K., Bohmann, D., Tjian, R.,and B.R. Franza (1988) Science. 240. 1010-1016.
- Rauscher FJ, Voulalas PJ, Franza BR, and T.Curren (1988) Genes Devel. 2. 1687-1699.
- Raychaudhuri, P., Rooney, R., and J.R.Nevins (1987) EMBO J. 6. 4073-4081.
- Reinberg, D., Horiskoshi, M. and R.G.Roeder (1987) J.Biol.Chem. 262. 3322-3330
- Roeder, R.G. and W.J. Rutter (1969) Nature 224; 234-7
- Rosenfeld, P.J., and T.J. Kelly (1986) J.Biol.Chem. 261:1398-1408.
- Roussou, I., Thiereos.G., and B.Hauge. (1988) Mol.Cell Biol. 8. 2132-2139.
- Russell, P.R. (1983) Nature. 301. 167-169
- Russell,P. and P.Nurse (1986) Cell. 45. 145-153
- Ryder, K., Lau, L.F., and D.Nathans (1988) Proc.Natl.Acad.Sci.USA. 85:1487-1491.
- Ryder K, Lanahan A, Perez-Albuerne E, Nathans D (1989) Proc.Natl.Acad.Sci.USA. 86. 1500-1503.
- Sadowski, I., Ma, J., Triezenberg, S., and M.Ptashne. (1988) Nature. 335. 563-564.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and N.Arnheim (1985) *Science*. 230. 1350-1354.

Sawadago, M., and R.G.Roeder. (1985) *Proc. Natl. Acad. Sci. USA*. 4394-4398.

Saltzman, A.G and R. Weinman. *FASEB*. 3. 1723-1733

Santoro, C., Mermod, N., Andrews P.C., and R.Tjian (1988) *Nature*. 334. 218-224.

Sassone-Corsi, P. (1988) *Proc. Natl. Acad. Sci. USA*. 85. 7192-7196.

Schena M, and K.Yamamoto (1988) *Science*. 241. 965-967.

Schmidt MC, Kao CC, Pei R, and A.J.Berk (1989) *Proc.Natl.Acad.Sci.USA*. 86. 7785-7789.

Schuermann M, Neuberger M, Hunter JB, Jenuwein T, Ryseck RP, Bravo R, and R.Muller (1988) *Cell*. 56. 507-516.

Schutte, J., Viallet, J., Nau, M., Segal, S., Fedorko, J., and J.Minna (1989) *Cell*. 59. 987-997.

Sentenac,A. (1985) In: *CRC Cit.Rev.Biochem. Eukaryotic RNA Polymerases*. 18, 31-91

Serfling,E., Jasin,M. and W.Schaffner (1985) *Trends in Genet*. 1. 224-230

Shepherd, J.C.W., McGinnis, W., Carrasco , A.E., De Robertis, E.M., and W.J.Gehring. (1984) *Nature* 310. 70-71

Silva, M.E., and C.P.Dietrich (1975) *J.Biol.Chem*. 250, 6841-6846

Simon, M.C., Fisch, T.M., Benecke, B.J., Nevins, J.R., and N.Heintz (1988) *Cell* 52. 723-729.

Singh, H., J.H.Lebowitz, A.S.Baldwin and P.A.Sharp (1988) *Cell*. 52. 415-423.

SivaRaman, L., Subramanian, S., and B.Thimmapaya. (1986) Proc. Natl. Acad. Sci. USA. 83. 5194-5198.

Smale, T., and D.Baltimore (1989) Cell. 57. 103-113.

Smeal, T., Angel, P., Meek, J., and M.Karin (1989) Genes. Dev. 3. 2091-2100.

Sorger, P.K., Lewis, M.J. and H.R.B.Pelham (1987) Nature. 329. 81-84

Sorger, P.K., and H.R.B.Pelham (1988) Cell. 54. 855-864.

Sorger, P.K., and H.C.M.Nelson (1989) Cell 59. 807-813.

Struhl, K., and R.W.Davies (1981) J. molec. Biol. 152. 553-568.

Struhl, K (1982) Nature. 300. 284-287.

Struhl, K (1983) Nature. 305. 391-397.

Struhl, K. (1987a) Cell 49, 295-297

Struhl, K. (1987b) Cell 50, 841-856

Struhl, K. (1989) Annu. Rev. Biochem. 58. 1051-1077.

Sturm, R., Das, G., and W.Herr (1988) Genes Dev. 2. 1582-1599.

Svaren, J., and R.Chalkley (1989) TIG. 6. 52-56.

Thireos, G., Driscoll-Penn, M., and H.Greer. (1984) Proc. Natl. Acad. Sci. USA.81. 5096-5100.

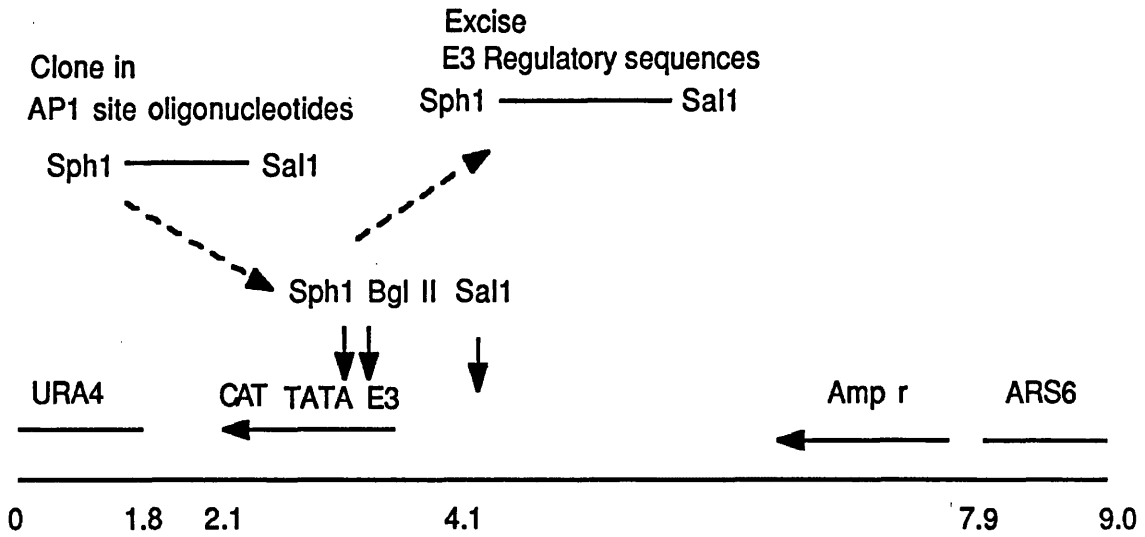
Turner R, and R.Tjian (1989) Science. 243. 1689-1694.

Tzamaris, D., Roussou, I., and G.Thieros (1989) Cell. 57. 947-954.

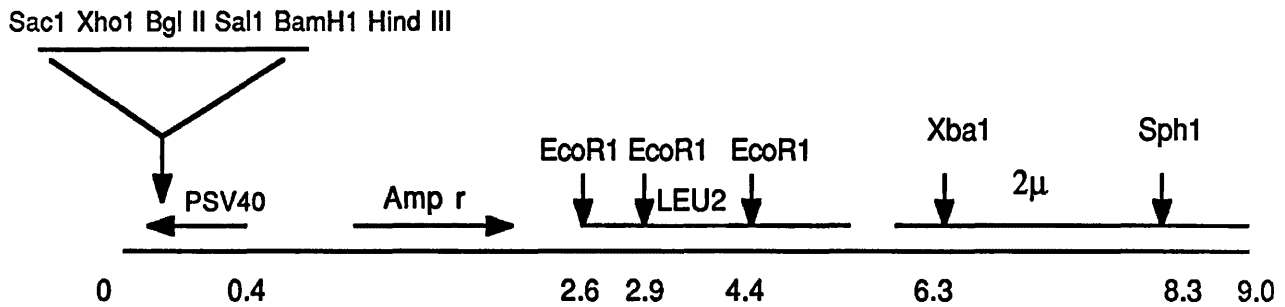
Van Dyke, M.W, Roeder, R.G., and M.Sawadago (1988) Science 241. 1335-1338

- Vogt, P.K., Bos, T.J. and R.F.Doolittle. (1987) Proc.Natl.Acad.Sci.USA. 84. 3316-3319
- Watt, F., and P.L.Molley (1988) Genes Dev. 2. 1136-1143.
- Webster, N., Jin,J.R., Green,S., Hollis,M. and P.Chambon. (1988) Cell. 52. 169-178
- Weil, P.A., Luse, D.S., Segall, J., and R.G.Roeder (1979) Cell. 19. 469-484
- Weintraub, H., Beug, H., Groudine, M., and T.Graf (1982) Cell. 28. 931-940.
- Wiederrecht, G., Shuvey,D.J., Kibbe,W.A. and C.S.Parker (1987) Cell. 48. 507-515
- Wiederrecht , G., Seto, D., and C.S.Parker (1988) Cell. 54 841-853.
- Wilkinson, D.G., Bhatt, S., Ryseck, R.P., and R.Bravo (1989) Development. 106. 465-471.
- Williams, T.J., Admon, A., Luscher, B., and R.Tjian. (1988) Genes Dev. 2. 1557-1569.
- Wolfner, M., Yep, D., Messenguy, F. and G.R.Fink. J.Mol.Biol (1975) .96. 273-290
- Yamamoto, K.K., Gonzalez, G.A., Biggs, W.H., and M.R.Montminy. (1988) Nature 334. 494-498.
- Yanisch-Perron, C., Vieira, J., and J.Messing (1985) Gene.33(1): 103-119.
- Yisraeli, J., Adelstein, R.S., Melloul, D., Nudel, U., Yaffe, D., and H.Cedar (1986) Cell 46. 409-416.
- Zalkin, H., and C. Yanofsky (1982) J.Biol.Chem. 257, 1491-1500

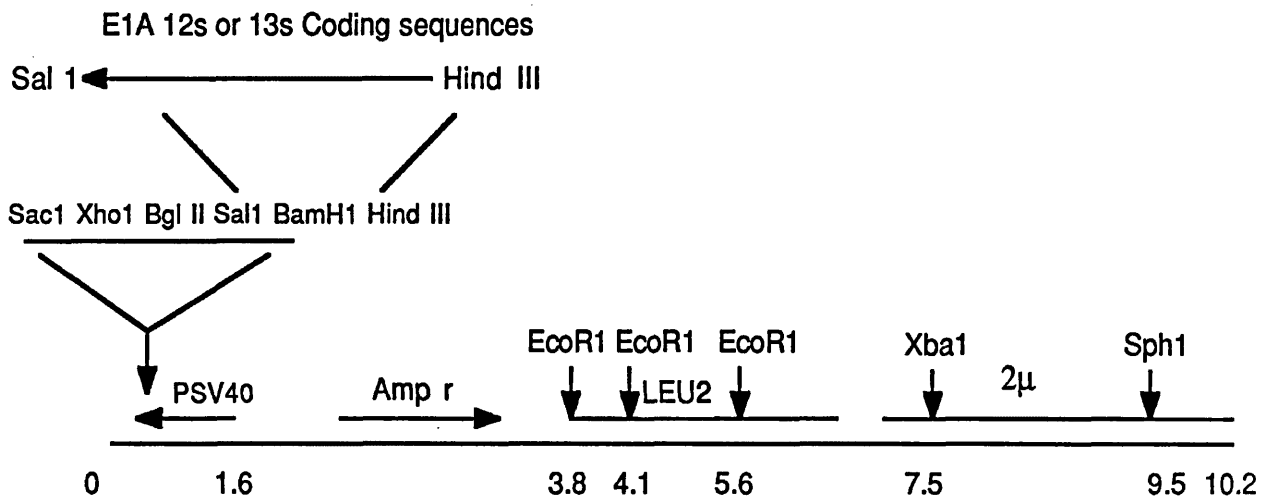
B) Excision of E3 regulatory sequences and insertion of AP1 site oligonucleotides



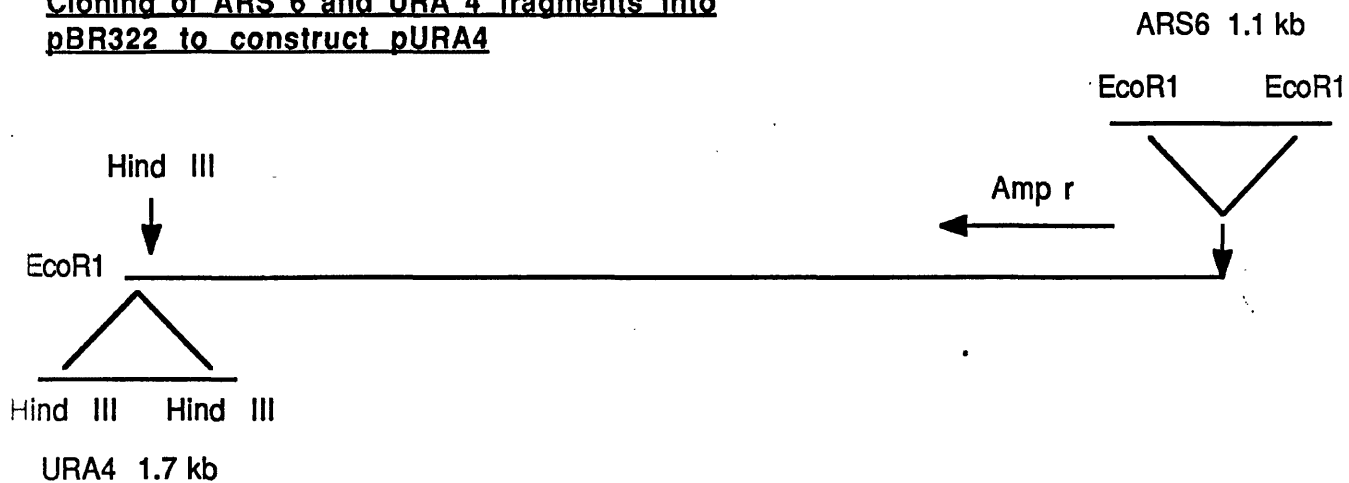
Map of pSM1



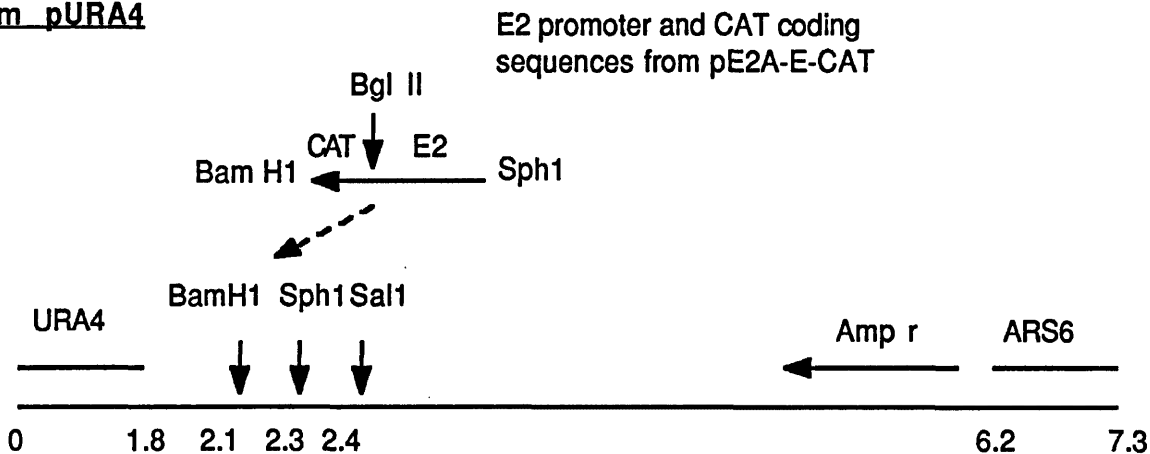
Construction of pPB12s and pPB13s from pSM1.



Cloning of ARS 6 and URA 4 fragments into pBR322 to construct pURA4

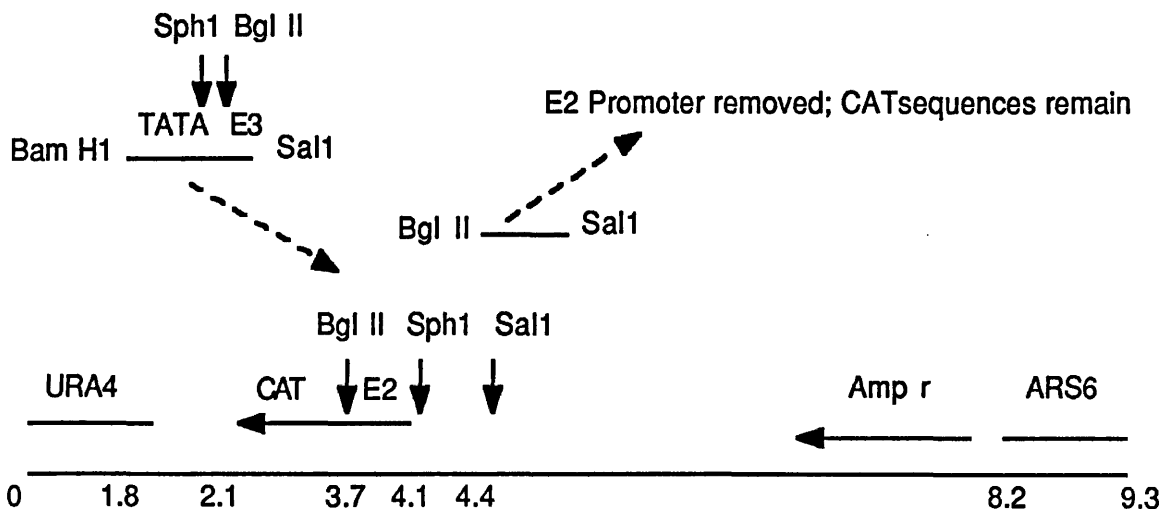


Construction of pU4E2CAT from pURA4



Construction of PYT3, PYT3CP and PYT3E3

A) Insertion of E3 regulatory and TATA elements.



# Mammalian cAMP-responsive element can activate transcription in yeast and binds a yeast factor(s) that resembles the mammalian transcription factor ATF

(*Saccharomyces cerevisiae*/*Schizosaccharomyces pombe*/DNA-binding protein/protein phosphorylation/protein kinase A)

ROBERT H. JONES AND NICHOLAS C. JONES

Gene Regulation Laboratory, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, England

Communicated by George R. Stark, December 30, 1988

**ABSTRACT** The human ATF and AP1 transcription factors bind to highly related DNA sequences. Their consensus binding sites differ by a single nucleotide, but this single change is crucial in determining factor binding specificity. We have previously identified an AP1 (yAP1) binding activity in yeast. In this report we identify a yeast ATF (yATF) binding activity whose specificity can be distinguished from that of yAP1 by the same crucial nucleotide that distinguishes binding of human ATF and AP1. The ATF binding site can act as an efficient upstream activating sequence *in vivo*, suggesting that yATF is a transcriptional activator. The yATF DNA-binding complex is phosphorylated and the binding activity of partially purified yATF can be enhanced *in vitro* by the addition of protein kinase A, indicating that the phosphorylation state of yATF may be important in determining its ability to bind DNA.

Certain components regulating basal and inducible transcription appear to be conserved from yeast to humans (1). It has been known for some time that the large subunit of RNA polymerase II has been conserved during evolution (2), and more recently it has been shown that yeast contains a "TATA box"-binding factor that can substitute for the mammalian TF11D factor in *in vitro* transcription systems (3, 4). In addition, the yeast activator protein GAL4 stimulates transcription from a promoter containing the GAL4 binding site in human HeLa cells (5, 6), and the human estrogen receptor can activate a promoter containing the appropriate binding site in yeast (7). Certain promoters from viruses that infect mammalian cells can function efficiently in yeast, using the same transcriptional start site in both yeast and mammalian cells (8, 9). As a consequence of these studies, some examples of yeast and mammalian activating factors recognizing identical or very similar DNA binding sequences have been described (8-14). One example of a cis-acting promoter element recognized by both human and yeast proteins is the AP1 consensus binding site, which has been shown to be a phorbol 12-tetradecanoate 13-acetate (TPA)-responsive element (TRE) in mammalian cells (15, 16). This site is very similar to the consensus binding site for the yeast transactivating protein GCN4; both GCN4 and another yeast protein, yAP1, can specifically recognize this site *in vitro* (8, 10, 13).

The consensus sequence of the TRE differs by only a single nucleotide from the consensus sequence of another promoter element called the cAMP-responsive element (CRE) (for review, see ref. 17). The activity of promoters containing the CRE can be stimulated by elevated levels of cAMP. Two proteins that bind to the CRE have been described, CREB (CRE-binding protein; ref. 18) and ATF (activating transcription factor; ref. 19); the two proteins have similar molecular weights and identical binding specificities *in vitro*, and it is

therefore likely that CREB and ATF are identical. The single nucleotide difference in the TRE and CRE consensus sequence implies that ATF/CREB and AP1 have evolved in human cells to mediate quite distinct regulatory responses by being able to recognize very subtle differences in promoter elements.

This report demonstrates that the capacity of different factors to distinguish between TRE and CRE elements has been conserved from yeast to humans. We have identified an activity in both budding and fission yeast that has a DNA binding specificity identical to that of mammalian ATF but quite distinct from that of yAP1 and GCN4. In addition we show that an ATF/CREB binding site can act as an efficient upstream activating sequence (UAS) in budding yeast, that the yATF binding complex can be phosphorylated by the cAMP-inducible protein kinase, and that phosphorylation is important for efficient binding.

## MATERIALS AND METHODS

**Yeast Strains and Plasmids.** *Saccharomyces cerevisiae* extracts were prepared from a protease-deficient strain, BJ2168 (20).  $\beta$ -Galactosidase activity was assayed in strain W303 (*ade2-1, trp1, can1-100, leu2-3,-112, his3-11,-14, ura3*; ref. 21).

The plasmid pLG559 $\Delta$ -312 (A255; ref. 22) was the parent of plasmids used in this study. The *Sma* I-*Xho* I fragment that contains the *CYC1* wild-type UAS elements was replaced with an oligonucleotide linker containing sites for the restriction enzymes *Sph* I and *Sal* I. Oligonucleotides containing wild-type or mutant ATF binding sites were subsequently inserted between these two sites of the linker. The sequence of the ATF oligonucleotide containing three copies of the wild-type ATF binding site was 5'-TCGACTTAACCGT-TACGTCATTTTTTATTAACCGTTACGTCATTTTTAT-TAACCGTTACGTCATTTGCATG-3'. The two mutant oligonucleotides were identical except for single-base-pair changes in the three ATF core binding sequences (see Fig. 3).

**Extract Preparation.** Uninfected HeLa cell nuclear extract was a gift of H. Hurst (23). *Schizosaccharomyces pombe* extract was prepared as described (8). *S. cerevisiae* extract was prepared and partially purified by chromatography on heparin-agarose and nonspecific (calf thymus) DNA-Sepharose columns (20). The extract was subsequently passed over an oligonucleotide affinity column containing ligated ATF binding sites from the adenovirus E4 early promoter.

**DNase I Protection (Footprint) and Gel Retardation Assays.** For footprint reactions, the *Bam*HI (+15) to *Eco*RI (-325) E4 promoter fragment contained in the plasmid pE4RA (a gift

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ATF, activating transcription factor; yATF, yeast ATF; CRE, cAMP-responsive element; CREB, CRE-binding protein; TRE, phorbol 12-tetradecanoate 13-acetate (TPA)-responsive element; SV40, simian virus 40; UAS, upstream activating sequence.

from K. Lee, Imperial Cancer Research Fund, London), was 3'-end-labeled at the *Bam*HI site with [ $\alpha$ - $^{32}$ P]dATP and [ $\alpha$ - $^{32}$ P]dCTP by reverse transcriptase. Each reaction mixture contained 0.8 ng of probe, 2  $\mu$ g of poly(dI)·poly(dC) nonspecific competitor, and crude nuclear extract from HeLa or *Sch. pombe* cells or partially purified (heparin-agarose and DNA-Sepharose chromatography) *S. cerevisiae* extract. Reaction conditions were as described (8, 23).

Gel retardation assays were carried out as described (8) with crude nuclear extracts from HeLa or *Sch. pombe* cells or DNA-Sepharose-purified extracts from *S. cerevisiae* cells.

**$\beta$ -Galactosidase Activity Assays.**  $\beta$ -Galactosidase assays were performed on yeast cultures grown from single colonies in minimal medium essentially as described by Harshman *et al.* (13).  $\beta$ -Galactosidase units were calculated as described by Hope and Struhl (24), and values represent the average of four independent determinations.

**Phosphatase and Kinase Reactions.** Partially purified *S. cerevisiae* extract (1.5  $\mu$ g of protein from heparin-agarose and DNA-Sepharose chromatography) was incubated at 37°C for 15 min with 0, 5, or 10 units of alkaline phosphatase (Boehringer) in gel retardation buffer (8) containing 10  $\mu$ g of bovine serum albumin. When phosphatase inhibitors were used they were added at the following concentrations: pyrophosphate, 16 mM; vanadate, 16 mM; and molybdate, 70 mM. After phosphatase treatment, binding reactions and gel electrophoresis were carried out as normal.

For kinase reactions, typically 1 unit of the catalytic subunit of protein kinase A (Sigma; catalogue no. P 2645) was added to a yATF fraction purified by oligonucleotide affinity chromatography (first passage, peak activity) or by DNA-Sepharose chromatography (1.5  $\mu$ g of protein). Incubation was for 15 min at 37°C in gel retardation buffer supplemented with 25  $\mu$ M spermidine (for affinity-purified fractions) and 2 mM ATP. Control incubations were carried out in the same buffer without added protein kinase and with or without ATP.

## RESULTS

**Yeast Contains a Factor(s) That Binds Specifically to the ATF Site.** The E4 early promoter of adenovirus contains multiple ATF binding sites (25). A segment of this promoter containing one such site was analyzed by DNase I footprinting using extract from either fission (*Sch. pombe*) or budding (*S. cerevisiae*) yeast. In both cases, protection was observed over the ATF site (Fig. 1, lanes 3 and 7) and the region of protection was identical to that obtained with HeLa cell extract (lane 2), although small differences in the pattern of hypersensitive cutting around the protected region did occur. Footprints generated by the budding (lane 8) and fission (lane 4) yeast extracts were eliminated by the inclusion in the reaction of an excess of unlabeled oligonucleotide containing the E4 ATF binding site but not by a similarly sized oligonucleotide of unrelated sequence (lanes 5 and 9). Therefore, yeast contain a factor(s) whose binding specificity is similar to that of mammalian ATF. The consensus binding sites for ATF (KWCCTCA, where K is T or G and W is T or A) and human transcription factor AP1 (TKAGTCA) are very similar, although in mammalian cells the factors that preferentially bind to these two sites *in vitro* are clearly distinct (for review, see ref. 17). Since we (8) and others (13) had previously shown the existence of yeast factors that bind to the AP1 site, we investigated whether the ATF binding activity we detected in yeast was due to yAP1 or, as in human cells, to a distinct factor.

By gel retardation analysis, three major complexes (a–c) were observed after incubation of budding yeast extracts with an oligonucleotide probe containing the ATF binding site (Fig. 2A, lane 6); the relative amounts of the different complexes varied from one extract preparation to the next,

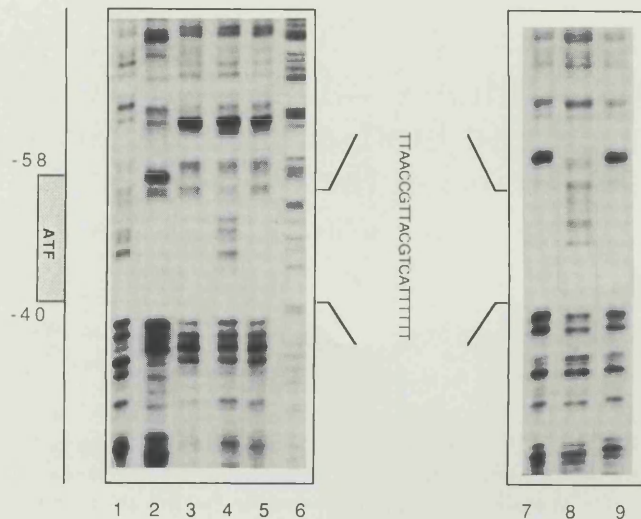


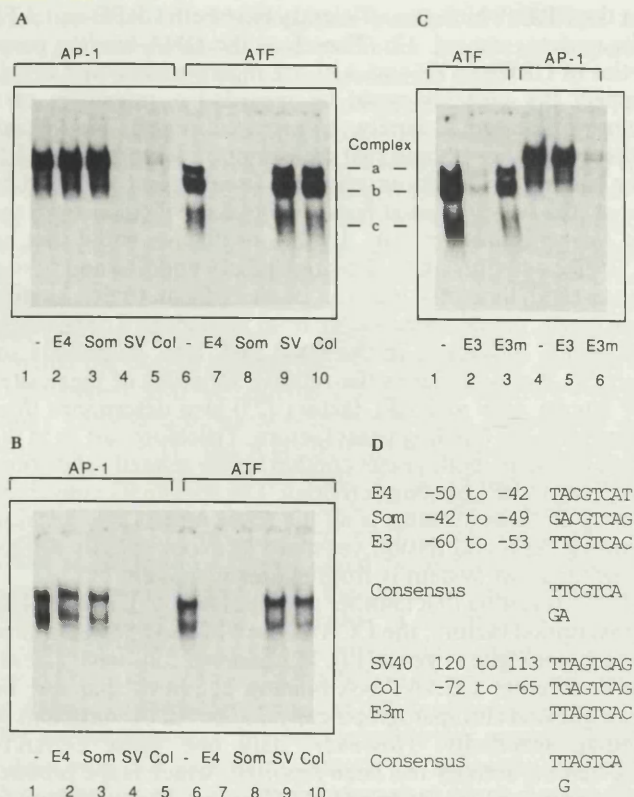
FIG. 1. DNase I protection (footprint) analysis of the adenovirus E4 early promoter with yeast and human extracts. Analysis was carried out on a 3'-end-labeled E4 promoter probe extending from +15 to -325. Footprint reactions were carried out as previously described for *Sch. pombe* (8) and HeLa (23) cell extracts. For protection with *S. cerevisiae* proteins, 1.5  $\mu$ g of DNA-Sepharose-purified extract was used. Lane 1, unprotected probe; lane 2, probe protected with HeLa cell extract; lanes 3–5, probe protected with *Sch. pombe* extract; lanes 7–9, probe protected with *S. cerevisiae* extract; lane 6, products of A+G sequencing reactions. A 300-fold molar excess of oligonucleotide containing from the E4 ATF binding site was added to reaction mixtures 4 and 8, and the same molar excess of a nonsense oligonucleotide (23) was added to reaction mixtures 5 and 9. The schematic ATF box on the left shows the extent (-40 to -58) of the previously characterized human ATF protection of the E4 promoter. The sequence of the protected region is shown.

probably due to differences in phosphorylation of the complex components (see below). Formation of all three labeled complexes was efficiently inhibited by an excess of unlabeled competitor oligonucleotide containing the E4 ATF site or containing the somatostatin CRE but not by oligonucleotides containing the AP1 sites from the SV40 or collagenase promoters (Fig. 2A, compare lanes 7 and 8 with lanes 9 and 10). Identical results were obtained with fission yeast extract (Fig. 2B, lanes 7–10). In contrast, when yeast extracts were incubated with a labeled probe containing the SV40 AP1 binding site, the two AP1-containing oligonucleotides competed efficiently with the probe but the oligonucleotides containing the E4 ATF site or somatostatin CRE did not (Fig. 2A and B, compare lanes 4 and 5 with lanes 2 and 3). These data argue that two different binding activities, yATF and yAP1, exist in both budding and fission yeast.

A close comparison of the consensus ATF and AP1 sites reveals a consistent difference at the third residue, which is cytosine for ATF and adenine for AP1. The importance of this change in switching specificity has been shown in human systems (23), where a single-base-pair change converting the CGTCA of the adenovirus E3 ATF binding site to AGTCA altered the binding preference of this site from ATF to AP1. We therefore tested the importance of this change on binding in budding yeast extracts. An oligonucleotide containing the C  $\rightarrow$  A change in the E3 ATF site (E3m) competed very poorly for yATF binding when compared to the wild-type site (Fig. 2C, lanes 2 and 3). In contrast, E3m was an efficient competitor for yAP1 binding compared to the wild-type E3 ATF site. The single-base-pair change therefore resulted in a switch in binding specificity from yATF to yAP1, demonstrating that the nucleotide requirements for both these factors has been conserved from yeast to humans.

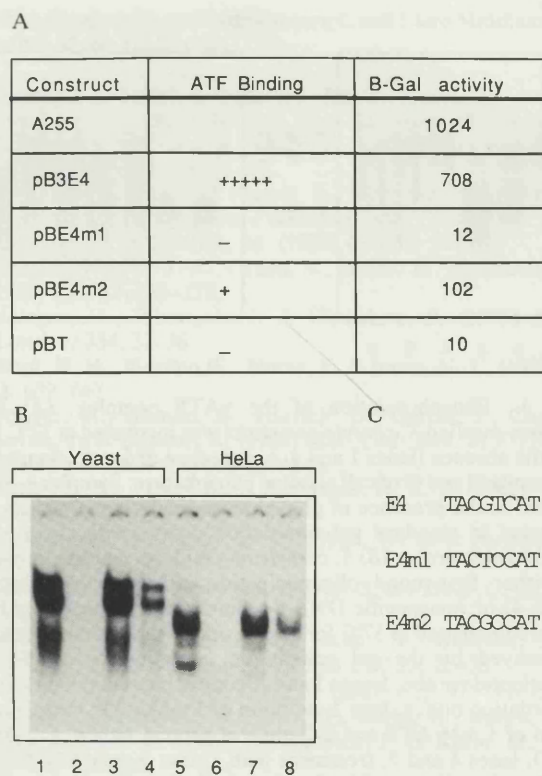
**The yATF Binding Site Can Act as an Efficient UAS.** We investigated the ability of the ATF binding site to act as a





**FIG. 2.** Gel retardation analysis showing that yeast ATF- and AP1-like binding activities are distinct. (A) *S. cerevisiae* extract was incubated in a standard binding reaction mixture with a radiolabeled oligonucleotide probe containing the simian virus 40 (SV40) AP1 binding site (lanes 1–5) or the E4 ATF binding site (lanes 6–10) and, where indicated, a 50-fold molar excess of unlabeled oligonucleotide competitor containing the E4 ATF site (lanes 2 and 7), the somatostatin CRE (lanes 3 and 8), the SV40 AP1 site (lanes 4 and 9), or the collagenase AP1 site (lanes 5 and 10). The reaction mixtures were loaded onto low-ionic-strength 8% polyacrylamide gel, electrophoresed to separate DNA–protein complex from free probe, and autoradiographed. (B) *Sch. pombe* extract was assayed as described (8). The probe and specific oligonucleotide competitors were as in A. (C) Radiolabeled E4 ATF oligonucleotide (lanes 1–3) or SV40 AP1 oligonucleotide (lanes 4–6) was incubated with *S. cerevisiae* extract as in A. Oligonucleotides bearing the E3 ATF site (lanes 2 and 4) or a mutated E3 ATF site (lanes 3 and 5) were added at 250-fold molar excess. (D) Core sequences of oligonucleotide competitors used in the gel retardation assays. SV40 120 to 113 contains the AP1 binding site of the SV40 promoter, and Col -72 to -65 contains the AP1 binding site of the human collagenase gene. The E4 -50 to -42 and E3 -60 to -53 oligonucleotides are derived from the ATF sites of the adenovirus E4 and E3 early promoters, respectively, and E3m is a point-mutated ATF site from the E3 promoter (BS2/3; ref. 23). Som -42 to -49 the CRE from the somatostatin promoter.

transcriptional activating sequence *in vivo*. Oligonucleotides containing three tandem copies of the wild-type E4 ATF binding site or mutant sites were cloned upstream of the *S. cerevisiae* cytochrome *c* (*CYC1*) TATA element fused to the  $\beta$ -galactosidase marker gene.  $\beta$ -Galactosidase enzyme levels were measured in exponentially growing cultures of cells transformed with the various plasmids. Insertion of wild-type ATF sites (pB3E4) resulted in an  $\approx 70$ -fold stimulation of  $\beta$ -galactosidase expression as compared to the TATA- $\beta$ -galactosidase gene containing no insert (pBT; see Fig. 3A). To test whether this stimulation was due to yATF binding, two different mutant templates were used, m1 and m2. Competition binding experiments showed that both templates were defective for yATF binding, m1 being more defective than m2 (Fig. 3B, lanes 3 and 4). Mutant m1 was also more defective than m2 in binding human ATF factor (Fig. 3B, lanes 7 and 8), once again emphasizing the con-



**FIG. 3.** The yATF binding site is an efficient UAS. (A) Oligonucleotides containing wild-type or mutant ATF binding sites were inserted upstream of a *CYC1-lacZ* (cytochrome *c*- $\beta$ -galactosidase) fusion gene as described in *Materials and Methods*. Plasmid pBT contains no insert. Plasmid pB3E4 contains three tandem copies of the E4 ATF binding site, and plasmids pBE4m1 and pBE4m2 contain three tandem copies of single-point-mutated ATF sites. Yeast strains were transformed with these constructs as well as the construct (A255) containing the wild-type *CYC1* UAS elements.  $\beta$ -Galactosidase activity was quantitated for cultures grown from single colonies. (B) Ability of the mutant E4 ATF site to bind yATF factor as tested by gel retardation. *S. cerevisiae* or HeLa cell extract was incubated in a standard binding reaction mixture with radiolabeled oligonucleotide probe containing the E4 ATF binding site without (lanes 1 and 5) or with a 50-fold molar excess of unlabeled oligonucleotide containing the wild-type E4 ATF site (lanes 2 and 6), m1 ATF site (lanes 3 and 7), or m2 ATF site (lanes 4 and 8). (C) Core sequences of the wild-type and m1 and m2 mutant ATF binding sites.

served binding specificity between the human and yeast factors. Both mutant binding sites were defective in stimulating  $\beta$ -galactosidase expression compared to the wild-type site. Three copies of the m2 site, partially defective for yATF binding, stimulated expression  $\approx 10$ -fold (pBE4m2) above basal levels, whereas three copies of the m1 site, a highly defective binding site, failed to stimulate expression (pBE4m1; see Fig. 3A). Therefore, the relative ability of the three oligonucleotides to stimulate  $\beta$ -galactosidase expression correlated with their ability to bind yATF, arguing strongly that yATF can mediate transcriptional activation. Similar results were obtained in a *GCN4*<sup>-</sup> strain, indicating that the activity was not due to the *GCN4* gene product (data not shown).

**Phosphorylation of yATF.** Certain reports have indicated that human CREB/ATF is a phosphoprotein (18, 26, 27) that can be phosphorylated by the catalytic subunit of cAMP-inducible protein kinase A *in vitro* (ref. 26; H. Hurst and N.C.J., unpublished data). This prompted us to investigate the phosphorylation state of the yATF factor. DNA-Sepharose-purified yATF extracts were treated with alkaline phosphatase, and the ability of yATF to bind to DNA was analyzed (Fig. 4A). Incubation of untreated extracts with a labeled ATF-binding-site probe resulted in three main com-

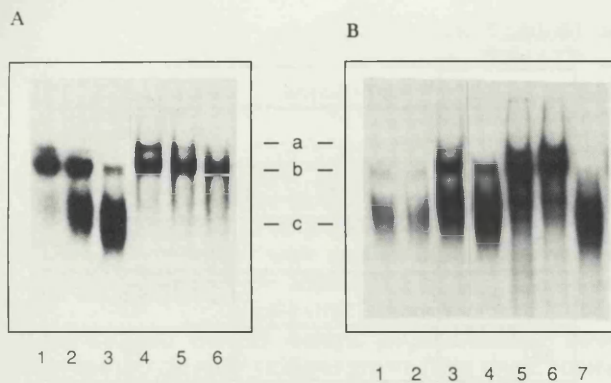


FIG. 4. Phosphorylation of the yATF complex. (A) DNA-Sepharose-purified *S. cerevisiae* extract was incubated at 37°C for 15 min in the absence (lanes 1 and 4) or presence of 5 units (lanes 2 and 5) or 10 units (lanes 3 and 6) of calf alkaline phosphatase. Samples 4–6 were incubated in the presence of phosphatase inhibitors. Extracts were then tested in standard gel retardation assays with the E4 ATF oligonucleotide probe. (B) *S. cerevisiae* yATF peak activity fractions from either first-round oligonucleotide affinity chromatography (lanes 1–4) or nonspecific DNA-Sepharose chromatography (lanes 5–7) were incubated at 37°C for 15 min under various conditions and then assayed by the gel retardation assay using the E4 ATF oligonucleotide probe. Lanes 1 and 5, control incubations in standard gel retardation buffer; lane 2, addition of 1 mM ATP; lanes 3 and 6, addition of 1 mM ATP and 1–2 units of protein kinase A (catalytic subunit); lanes 4 and 7, treatment with kinase A and ATP followed by dephosphorylation with 20 units of alkaline phosphatase for 15 min at 37°C. The three distinct complexes of different mobilities are referred to as a–c.

plexes, a–c (Fig. 2). Increasing phosphatase treatment led to the reduction and disappearance of complexes a and b with a concurrent increase in complex c (Fig. 4A, lanes 1–3). We also saw a complex with a slightly greater mobility than complex c (lane 3). These alterations were largely inhibited by the addition of phosphatase inhibitors to the reactions (lanes 4–6), and were therefore very likely to be due to dephosphorylation of the yATF–DNA complex (see Discussion). Longer phosphatase treatment led to the generation of a slightly faster migrating complex and an eventual reduction in binding activity (data not shown). When yATF-containing extract was passed over oligonucleotide affinity columns containing multimerized ATF binding sites, the activity recovered from the column gave protein–DNA complexes of similar mobility to those obtained with the phosphatase-treated DNA-Sepharose-purified extract (Fig. 4B, compare lanes 1 and 7), suggesting that the yATF may have been dephosphorylated. In addition, substantial loss of binding activity was observed, consistent with our findings on the phosphatase-treated extracts. Since the catalytic subunit of the cAMP-inducible kinase has been shown to phosphorylate CREB *in vitro* (ref. 26; H. Hurst and N.C.J., unpublished data), we determined whether this kinase could also phosphorylate affinity-enriched yATF. Addition of the kinase and ATP resulted in a marked increase in yATF binding activity and the appearance of the slower migrating complexes (Fig. 4B, compare lanes 1 and 3). Addition of ATP alone did not result in these changes (Fig. 4B, lane 2). Phosphatase treatment of the rephosphorylated extract resulted in reversal of these mobility changes (Fig. 4B, lane 4). It therefore appears that the yATF complex is phosphorylated and that this phosphorylation contributes to efficient binding to the ATF site.

## DISCUSSION

The consensus sequence of the TRE, which is specifically recognized by the AP1 transcription factor, differs by only a single nucleotide from the consensus sequence established

for the CRE, which can efficiently bind both CREB and ATF (for review, see ref. 17). Therefore the DNA-binding properties of CREB/ATF and AP1 are highly related and in this respect the proteins could be regarded as members of a transcription factor family. In previous reports we (8) and others (13) have shown that transcription factors with AP1-like binding specificity exist in both fission yeast and budding yeast, the budding yeast factor being quite distinct from the *GCN4* gene product (10). In this report we show that an ATF-like activity can be detected in both budding and fission yeasts and that this activity, with respect to its binding specificity, is indistinguishable from human ATF. The single nucleotide difference in the CRE and TRE consensus sequences that determines the relative affinities of these sites for human ATF and AP1 factors (23) also determines their affinity for the budding yeast factors. Therefore just as in the human system, both yeasts contain highly related yet distinct yATF and yAP1 binding activities. The specificity conserved among all three systems is all the more remarkable because budding yeast and fission yeast are as evolutionarily distant as either yeast system is from mammalian cells.

It is interesting that another potential family of mammalian transcription factors, the CCAAT-box-binding proteins, also has a homologue in yeast (14). Mammalian cells have at least three different CCAAT-box-binding activities that can be distinguished chromatographically and by subtle variations in binding specificity. However, only one yeast CCAAT-box-binding activity has been reported, which is the product of the previously identified *HAP2/3* genes. The specificity of *HAP2/3* is most homologous to the mammalian CP1 CCAAT-box-binding activity; however, rather than distinguishing between the different types of CCAAT-box elements, as the mammalian proteins do, it appears that the yeast *HAP2/3* activity can recognize them all. This is somewhat different from what we have found for ATF and AP1, where the small difference in specificity between the two factors is as apparent in yeast as it is in humans. Therefore, in terms of a speculative evolutionary model, one might regard the *HAP2/3* proteins as representing an ancestral activity that subsequently evolved to produce a more complex array of mammalian CCAAT-box-binding proteins, whereas if the genes encoding ATF/CREB and AP1 arose from a common ancestor, the divergence occurred before the separation of the two yeast and mammalian lineages.

The conserved binding specificity poses the intriguing question of whether yATF and human ATF are also conserved at the functional and regulatory level. Our data strongly suggest that the yATF of budding yeast can act as a transcriptional activator. The ATF binding site can act as an efficient UAS element, and the introduction of point mutations in the binding site that decreased the affinity of the site for yATF also decreased its efficiency as an UAS.

With regard to a regulatory conservation, the situation is unclear. In mammalian cells ATF/CREB has been implicated in mediating the response of promoters to both the transactivator E1A and conditions that result in elevated cAMP levels (18, 19). In either case the mechanism of the response is largely unknown. However, in the case of the cAMP response it is likely that it operates through the cAMP-inducible kinase pathway. Yeast also contains such kinases, the genes for which have been cloned and analyzed (28–30). Consequently, a number of mutant strains have been isolated that have altered kinase (31) or adenylate cyclase (32) activity. It would be of interest to assay the effect of cAMP on yATF activity in these strains. The involvement of a kinase pathway implicates phosphorylation as a potential mechanism by which CREB/ATF may be modulated. Indeed, it has been reported that ATF/CREB is a phosphoprotein (18, 27), and its phosphorylation state is increased with elevated cAMP levels (18). In addition the cAMP-

dependent kinase phosphorylates purified ATF/CREB *in vitro* (26) and causes altered mobilities in gel-shift experiments (H. Hurst and N.C.J., unpublished data).

We present data in this report indicating that the yATF complex is also phosphorylated. Treatment of DNA-Sepharose-purified extracts with alkaline phosphatase led to the disappearance of slow migrating forms with the concurrent shift of faster migrating forms down the gel; an eventual reduction of binding activity was also observed on longer treatments. Affinity-purified extracts manifested many of the characteristics of phosphatase-treated extracts: the predominant protein-DNA complex existed in a rapidly migrating form and the overall binding activity was reduced compared to that loaded onto the column. It is unclear why first-round affinity chromatography should enrich for a dephosphorylated form of yATF, but it is possible that this enrichment is due to the activation of an endogenous phosphatase that copurified with yATF. Consistent with our interpretation that dephosphorylation had occurred, binding activity as well as slower migrating complexes could be recovered, at least partially, by incubation with the catalytic subunit of the cAMP-dependent kinase. A number of reasons may explain the mobility shifts observed on phosphatase treatment. One possibility is that yATF binds DNA as a monomer and that the different complexes represent different phosphorylation states of the monomer. It is also possible that more than one protein molecule associates directly or indirectly with the DNA template in a phosphorylation-dependent manner. In this case phosphatase treatment and subsequent dephosphorylation of the components would lead to a dissociation of the complex and consequently an increase in electrophoretic mobility. The apparent dependence of protein-DNA interactions on the phosphorylation state of the extract can be explained by a similar scenario. In this case, however, the modification of either a monomeric DNA-binding protein or a phosphorylation-dependent protein-protein complex would have a direct effect on DNA-binding activity. A less likely possibility is that each complex is due to totally different proteins, each one having a different phosphorylation dependence for binding DNA.

The biological significance of these data remains to be tested. The immediate implication is that phosphorylation may be important for the transactivation potential of this factor *in vivo*, and as such the ability to bind DNA is an obvious regulatory stage. However, a much more extensive analysis *in vitro* and *in vivo* will be needed before meaningful conclusions can be made.

We are particularly grateful to Peter Sorger for help and advice on the preparation and chromatographic fractionation of the *S. cerevisiae* extracts. We thank Helen Hurst for HeLa cell extracts, Iain Goldsmith for oligonucleotides, Richard Treisman and Steve Good-

born for critical reading of the manuscript, and Clare Middlemiss for preparation of the manuscript.

1. Guarente, L. (1988) *Cell* **52**, 303-305.
2. Sentenac, A. (1985) *CRC Crit. Rev. Biochem.* **18**, 31-91.
3. Buratowski, S., Hahn, S., Sharp, P. A. & Guarente, L. (1988) *Nature (London)* **334**, 37-42.
4. Cavallini, B., Huet, J., Plassat, J., Sentenac, A., Egly, J. & Chambon, P. (1988) *Nature (London)* **334**, 77-80.
5. Kakidani, H. & Ptashne, M. (1988) *Cell* **52**, 161-167.
6. Webster, N., Jin, J. R., Green, S., Hollis, M. & Chambon, P. (1988) *Cell* **52**, 169-178.
7. Metzger, D., White, J. H. & Chambon, P. (1988) *Nature (London)* **334**, 31-36.
8. Jones, R. H., Moreno, S., Nurse, P. & Jones, N. C. (1988) *Cell* **53**, 659-667.
9. Kornuc, M., Altman, R., Harrich, D., Garcia, J., Chao, J., Kayne, P. & Gaynor, R. (1988) *Mol. Cell. Biol.* **8**, 3717-3725.
10. Struhl, K. (1987) *Cell* **50**, 841-846.
11. Sorger, P. K., Lewis, M. J. & Pelham, H. R. B. (1987) *Nature (London)* **329**, 81-84.
12. Wiederrecht, G., Shuey, D. J., Kibbe, W. A. & Parker, C. S. (1987) *Cell* **48**, 507-515.
13. Harshman, K. D., Moye-Rowley, W. S. & Parker, C. S. (1988) *Cell* **53**, 321-330.
14. Chodosh, L. A., Olesen, J., Hahn, S., Baldwin, A. S., Guarente, L. & Sharp, P. A. (1988) *Cell* **53**, 25-35.
15. Lee, W., Mitchell, P. & Tijan, R. (1987) *Cell* **49**, 741-752.
16. Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R. J., Rahmsdorf, H. J., Jonat, C., Herrlich, P. & Karin, M. (1987) *Cell* **49**, 729-739.
17. Jones, N. C., Rigby, P. W. J. & Ziff, E. B. (1988) *Genes Dev.* **2**, 267-281.
18. Montminy, M. R. & Bilezikjian, L. M. (1987) *Nature (London)* **328**, 175-178.
19. Lee, K. A. W., Hai, T. Y., Siva Raman, L., Thimmappaya, B., Hurst, H. C., Jones, N. C. & Green, M. R. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8355-8359.
20. Sorger, P. K. & Pelham, H. R. B. (1987) *EMBO J.* **6**, 3035-3041.
21. Shore, D. & Nasmyth, K. A. (1987) *Cell* **51**, 721-732.
22. Guarente, L. & Mason, T. (1983) *Cell* **32**, 1279-1286.
23. Hurst, H. C. & Jones, N. (1987) *Genes Dev.* **1**, 1132-1146.
24. Hope, I. A. & Struhl, K. (1986) *Cell* **46**, 885-894.
25. Lee, K. A. W. & Green, M. R. (1987) *EMBO J.* **6**, 1345-1353.
26. Yamamoto, K. K., Gonzalez, G. A., Biggs, W. H. & Montminy, M. R. (1988) *Nature (London)* **334**, 494-498.
27. Hai, T., Liu, F., Allegretto, E. A., Karin, M. & Green, M. R. (1988) *Genes Dev.* **2**, 1216-1226.
28. Tody, T., Cameron, S., Sass, P., Zoller, M., Scott, J. D., McMullen, B., Hurwitz, M., Krebs, E. G. & Wigler, M. (1987) *Mol. Cell. Biol.* **7**, 1371-1377.
29. Cannon, J. & Tatchel, K. (1987) *Mol. Cell. Biol.* **7**, 2653-2663.
30. Kunisawa, R., Davies, T. N., Urdea, M. S. & Thorner, J. (1987) *Nucleic Acids Res.* **15**, 368-369.
31. Cameron, S., Levin, L., Zoller, M. & Wigler, M. (1988) *Cell* **53**, 555-556.
32. Matsumoto, K., Uno, I., Oshima, Y. & Ishikawa, T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2355-2359.

# Expression of the SV40 Promoter in Fission Yeast: Identification and Characterization of an AP-1-like Factor

Robert H. Jones,\* Sergio Moreno,† Paul Nurse,† and Nicholas C. Jones\*

\* I. C. R. F. Gene Regulation Laboratory  
Lincoln's Inn Fields

London, WC2A 3PX, England

† I. C. R. F. Cell Cycle Control Laboratory

Microbiology Unit

Department of Biochemistry

University of Oxford

Oxford, OX1 3QU, England

## Summary

The SV40 promoter is expressed well in the fission yeast *S. pombe*, and it initiates transcription at the same site as in mammalian cells. The majority of the enhancer sequences, however, do not contribute to this activity. DNAase I footprint analysis of the promoter revealed the presence of an AP-1-like factor in *S. pombe* cells that protects a region of the promoter almost identical to that protected by human AP-1. The specificity of binding of the yeast and mammalian AP-1 proteins was found to be similar. We have found two AP-1-like binding activities in budding yeast cells, one of which appears quite distinct from the binding activity of the product of the budding yeast *GCN4* gene. We also demonstrate that in fission yeast the AP-1 binding site can act as an upstream activating sequence. The DNA-protein complexes containing the mammalian AP-1 and fission yeast AP-1-like factors are sensitive to phosphatase treatment, indicating that they may be phosphorylated.

## Introduction

Mutational analysis of viral and cellular genes has revealed two basic types of transcriptional regulatory elements, promoters and enhancers. These elements are complex and consist of multiple sequence motifs that are specifically recognized and bound by *trans*-acting regulatory factors, some of which appear to mediate the response of specific genes to various physiological stimuli (for reviews, see Serfling et al., 1985; Maniatis et al., 1987; Jones et al., 1988).

Recent experiments have suggested that the mechanism of transcriptional activation in *Saccharomyces cerevisiae* and higher eukaryotes may be very similar. A yeast upstream activation site (UAS) and its cognate DNA-binding protein have been shown to stimulate mammalian cell promoters in transfected HeLa cells; the characteristics of this stimulation are identical to the stimulation of the same promoters by a mammalian enhancer element (Kakidani and Ptashne, 1988; Webster et al., 1988). This result strongly suggests that the molecular mechanism of transcription stimulation has been conserved and that the yeast activator protein can productively interact with the mammalian cell transcriptional apparatus. Indeed, there

is conservation between RNA polymerase II of yeast and higher eukaryotes (for review, see Sentenac, 1985). In addition, there appears to be some conservation between the binding site specificities of mammalian and yeast *trans*-acting factors. The best example concerns the factor involved in the well-conserved heat shock response (for review, see Pelham, 1985). This response is mediated through the heat shock element (HSE), which is highly conserved between yeast and mammalian cells, suggesting that considerable conservation of the heat shock transcription factor (HSTF) must exist, at least in the DNA-binding domain.

In addition to conservation of the HSE and its cognate binding protein, which are associated with a limited set of coordinately expressed genes, there also appears to be conservation of a more general mammalian transcription factor, AP-1. This factor is a 47 kd protein probably encoded by the human proto-oncogene *c-jun* (Bohmann et al., 1987; Angel et al., 1988). It binds to a number of different promoters, many of which can be transcriptionally stimulated by phorbol esters such as 12-O-tetradecanoylphorbol-13-acetate (TPA), and the AP-1 binding site can confer TPA responsiveness to heterologous promoters (Lee et al., 1987b; Angel et al., 1987). The binding site of AP-1 is very similar to the UAS element that is specifically bound by the GCN4 protein of *S. cerevisiae*, and comparison of the predicted amino acid sequences of *v-jun* and *GCN4* reveals a certain degree of homology between their respective DNA-binding domains (Vogt et al., 1987; Struhl, 1987b).

We report here the identification of a fission yeast *Schizosaccharomyces pombe* factor whose binding specificity more closely resembles mammalian AP-1 than the *GCN4* product. This factor appears to act as a transcriptional activator and efficiently binds to the AP-1 site of the SV40 early gene promoter. This promoter functions very efficiently in *S. pombe*, with transcription initiating from the same start site used in mammalian cells. A factor with similar specificity was also found in budding yeast *Saccharomyces cerevisiae*, which appears to be distinct from *GCN4*.

## Results

### The SV40 Promoter Functions Efficiently in *S. pombe*

A chimeric gene containing the SV40 promoter/enhancer region fused to the reporter chloramphenicol acetyltransferase (CAT) sequences was inserted into a yeast *LEU2*-containing plasmid. The plasmid (SMCAT) was transformed into *S. pombe* cells, and expression of the reporter CAT sequences in a number of individually isolated transformants was measured. As shown in Figure 1A, expression of this chimeric gene was very efficient whereas the level of CAT expression from a plasmid containing the same CAT sequences but lacking the SV40 elements (SABCAT) was hardly detectable under the same assay conditions. The level of expression of the SV40CAT gene

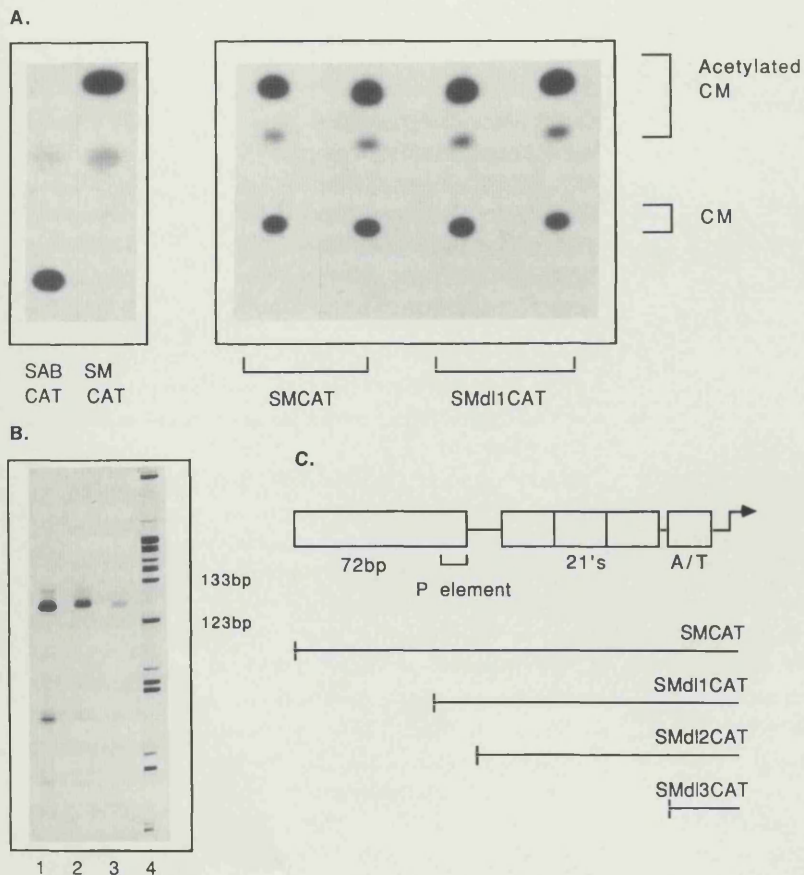


Figure 1. Function of the SV40 Early Gene Promoter in *S. pombe*

(A) The expression of CAT coding sequences in *S. pombe* cells transformed with the plasmids SABCAT, SMCAT, or SMd1CAT (see C for details of constructs) was determined by a standard CAT TLC assay. The CAT assays were performed with crude extracts using 10 µg (left panel) and 100 ng (right panel) of the extract supernatant (see Experimental Procedures) and incubating for 30 min at 37°C.

(B) Total RNA from cells transformed with the plasmids SMd1CAT (lane 1), SMd2CAT (lane 2), and SMd3CAT (lane 3) were analyzed by primer extension using a primer oligonucleotide complementary to a sequence within the CAT coding region. Lane 4 contains marker DNAs, and the size of the relevant markers that flank the major extension product is indicated.

(C) Schematic organization of the SV40 early promoter region which consists of the 72 bp enhancer element, the GC-rich 21 bp repeat region, and the AT-rich (TATA) element (Fromm and Berg, 1982, 1983). The location of the "P" site which is specifically bound by the AP1 transcriptional activator protein is indicated (Lee et al., 1987a). Indicated below the schematic are the extent of these promoter sequences present in the various plasmids used in this study. SMCAT contains the complete promoter/enhancer region, SMd1CAT lacks all of the enhancer except the "P" site, SMd2CAT additionally lacks the "P" site, and SMd3CAT contains only the AT-rich element.

was on the same order as the level obtained when expression of CAT was directed by the fission yeast *Adh* promoter, a very efficient constitutively active *S. pombe* promoter (data not shown; Russell and Nurse, 1986). Surprisingly, the removal of most of the SV40 enhancer sequences had no effect on the level of CAT expression. SMd1CAT has lost all enhancer sequences except the "P" motif, which in mammalian cells acts as a binding sequence for the transcription factor AP-1 (Lee et al., 1987a). In many different SMd1CAT transformants, the level of CAT expression was equivalent to that found when the complete enhancer was present (Figure 1A); the plasmid copy number was the same in both sets of transformants (data not shown).

The nature and location of the transcriptional start site was analyzed by primer extension of RNA from SMd1CAT transformed cells. In mammalian cells, SV40 early gene transcription is predominantly initiated from a set of closely adjacent sites 24–27 bp downstream of the TATA sequence (Gosh et al., 1981; Benoist and Chambon, 1981), which are 125–130 nucleotides upstream of the 5' end of the CAT primer used in this analysis. The major extended product obtained was 128 bp in length (Figure 1B, lane 1), showing that the same start site is utilized in both fission yeast and mammalian cells. Figure 1B also shows the primer extension analysis of RNAs from two other fission yeast transformants. One transformant contains the plasmid SMd2CAT, which is identical to SMd1CAT except that the "P" motif has been removed (Figure 1C). The other transformant contains SMd3CAT, which has additionally

lost the 21 bp repeat region of the promoter; the AT-rich element is thus the only promoter element that remains. In both cases, transcription initiated from the same start site (Figure 1B, lanes 2 and 3), indicating that neither the "P" element nor the 21 bp repeat region determined the site of initiation. The level of expression in these transformants, however, was lower than that obtained with SMd1CAT; removal of the P site reduced expression, and this was reduced further to a very low level when the 21 bp repeat region was deleted. These data indicate that both the "P" element and 21 bp repeat region contribute to SV40 promoter activity in fission yeast.

#### A Factor in *S. pombe* That Has the Same Binding Specificity as the Mammalian Factor AP-1

The efficient utilization of the SV40 promoter in *S. pombe* is likely to involve the interaction of yeast transcription factors with specific promoter sequences. We investigated these interactions by DNAase I footprint analysis. Incubation of the SV40 promoter/enhancer region with a nuclear extract from *S. pombe* cells resulted in a single region of protection (Figure 2A, lane 1). This region corresponds to the "P" site of the enhancer (Figure 2B, lane 1), which is also protected by purified mammalian AP-1 protein (Lee et al., 1987b; Angel et al., 1987). The footprint is lost in the presence of an oligonucleotide containing the AP-1-binding site (Figure 2A, lane 2; Figure 2B, lane 5), but not by an oligonucleotide consisting of completely unrelated sequences (Figure 2A, lane 3; Figure 2B, lane 4). The

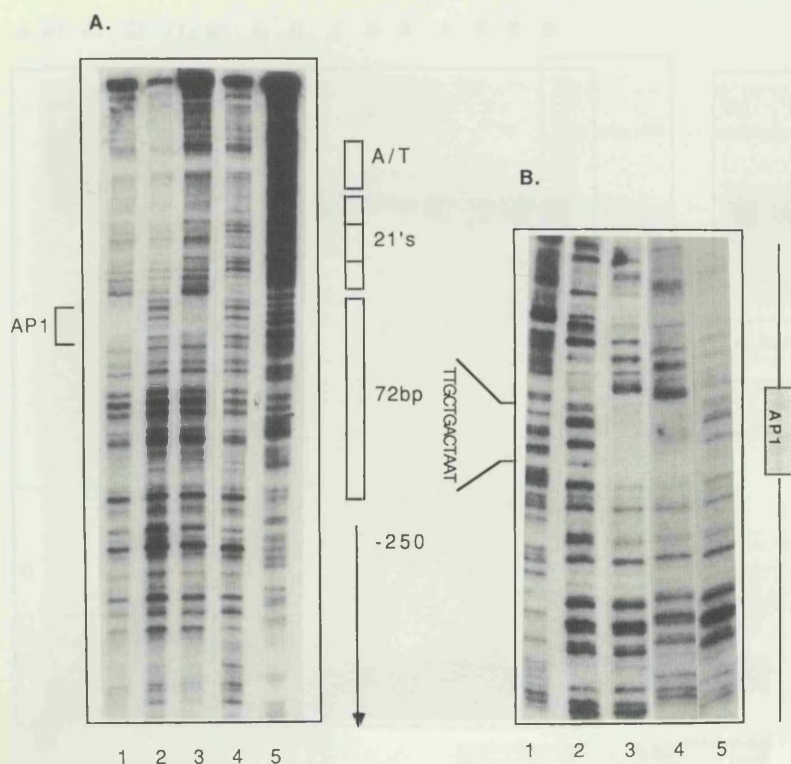


Figure 2. DNAase I Footprint Analysis of the SV40 Early Promoter/Enhancer in *S. pombe*. DNAase I footprint analysis was carried out using a plus strand probe prepared by 3' end-labeling of the EcoRI site of the plasmid pSVGC-0 with [ $\alpha$ - $^{32}$ P]ATP and CTP and reverse transcriptase. Eight-tenths micrograms of this labeled probe was used in each reaction. (A), lanes 1–3, and (B), lanes 3–5, were incubated with 200  $\mu$ g of crude *S. pombe* nuclear extract, and (A), lane 4, and (B), lane 2, were incubated without extract. (A), lane 2, and (B), lane 5, were also incubated with 300-fold molar excess of oligonucleotide competitors containing the SV40 AP-1-binding site; (A), lane 3, and (B), lane 4, were incubated with a similar excess of an unrelated nonsense oligonucleotide. (A), lane 5, and (B), lane 1, represent the AG ladder. The footprint reactions were carried out as described in Experimental Procedures. The locations of the *cis*-acting promoter/enhancer elements are depicted at the right of each figure, and the location and sequence of the specific footprint obtained are shown at the left.

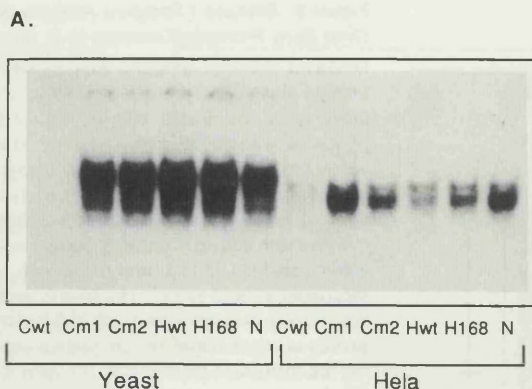
binding of protein to this site is therefore specific. No other protected regions were evident, although it is quite possible that other regions of protection would be found if fractionated rather than crude extracts were used.

We used the gel retention assay to investigate the specificity of the yeast factor for the AP-1-binding site. This assay relies on the fact that specific protein–DNA complexes migrate more slowly in the native polyacrylamide gel than the free unbound DNA. A  $^{32}$ P-labeled oligonucleotide containing the SV40 AP-1-binding site (SV-AP-1) was added to crude nuclear extract from *S. pombe* and HeLa cells in the presence of an excess of unlabeled competitor oligomers. The relevant sequence of these oligomers is shown in Figure 3B. Col-AP-1 contains the core AP-1-binding site of the human collagenase promoter (Angel et al., 1987); it differs by a single nucleotide from the SV40 site, resulting in a more perfect palindromic sequence. Human AP-1 protein binds efficiently to both the SV40 and collagenase sites (Angel et al., 1987). Col-mut1 and Col-mut2 differ from the Col-AP-1 site by one and two nucleotides respectively, which significantly decreases binding of human AP-1 protein (Angel et al., 1987). Incubation of the labeled SV-AP-1 oligomer with the *S. pombe* nuclear extract resulted in a specific retarded complex that was efficiently competed by Col-AP-1 (Cwt, Figure 3), but not by either of the two mutants (Cm1, Cm2), nor by an unrelated oligonucleotide of similar size (N). A similarly behaving complex was obtained with the HeLa extract (Figure 3), although some competition was observed with Col-mut2. Thus, mutations that resulted in decreased binding of human AP-1 also lowered binding of the yeast factor. The two complexes differed significantly in their mobility, which

may reflect different sizes of the cognate binding proteins or differences in posttranslational modification (see below). These data indicate that *S. pombe* contains a factor with the same binding specificity as AP-1.

#### ***S. cerevisiae* Contains Two AP-1-like Factors**

The SV40 AP-1-binding site resembles the UAS sequence located upstream of a number of coordinately regulated amino acid biosynthetic genes of *S. cerevisiae* that are specifically bound by the activator protein *GCN4* (Hill et al., 1986). The *HIS3* gene UAS has been analyzed in considerable detail by saturation mutagenesis. Almost all single base-pair changes within the *GCN4* binding site resulted in reduced binding affinities. One change, however, which increases the palindromic nature of the site, also generated a sequence with increased affinity (Hill et al., 1986). This change produces a site that closely resembles the collagenase AP-1 core site. Notably however, none of the *GCN4* regulated biosynthetic genes contain the high efficiency palindromic site. Incubation of labeled SV-AP-1 oligomer with a whole cell extract from *S. cerevisiae* resulted in a specific complex that was competed by an excess of unlabeled SV-AP-1 or Col-AP-1 oligomer (Figure 4, lanes 8 and 9), but not by an unrelated oligomer of similar size (lane 14). No competition was obtained with the two mutant oligomers, Col-mut1 and Col-mut2 (lanes 10 and 11). Therefore the binding specificity of the *S. cerevisiae* factor was similar to the factor identified in *S. pombe* and to human AP-1. Surprisingly, an oligonucleotide containing the *GCN4* binding site of the *S. cerevisiae* *HIS3* promoter (*HIS*-wt) did not compete for this factor, even at a high molar excess (Figure 4, lane 12); this same oligomer

**B.**

GTTAGTCAT	E3-AP1
ATTAGTCAG	SV-AP1
ATGAGTCAG	Col-AP1 (Cwt)
AGGAGTCAG	Col-mut1 (Cm1)
ATGAGTCCT	Col-mut2 (Cm2)
AAGAGTCAT	HIS-wt (Hwt)
AAGCGTCAT	HIS-168 (H168)

Figure 3. The AP-1-like Factor of *S. pombe* and HeLa AP-1 Have Similar Binding Specificities

(A) Competition analysis of binding to the SV40 AP-1-binding sites by HeLa and *S. pombe* nuclear factors. A radiolabeled oligonucleotide probe containing the SV40 AP-1-binding site (SV-AP-1) was incubated in a standard binding reaction with either 15  $\mu$ g *S. pombe* nuclear extract (lanes 1–6) or 7  $\mu$ g HeLa nuclear extract (lanes 7–12), 100 ng of nonsense oligonucleotides, and a 100-fold molar excess of an unlabeled specific oligonucleotide competitor. The specific competitor DNAs used were Col-AP-1 (lanes 1 and 7), Col-mut1 (lanes 2 and 8), Col-mut2 (lanes 3 and 9), HIS-wt (lanes 4 and 10), and HIS-168 (lanes 5 and 11) and a nonsense oligonucleotide of unrelated sequence (lanes 6 and 12). The binding reactions were loaded onto low ionic strength 4% polyacrylamide gels, electrophoresed to separate DNA-protein complex from free probe and autoradiographed. Details on preparation of extract and probe, and the binding conditions, are given in Experimental Procedures.

(B) The AP-1-binding site or related sequence present in each of the oligonucleotides used in this study are shown. SV-AP-1 contains the AP-1-binding site of the SV40 promoter and Col-AP-1 the binding site of the human collagenase genes. Col-mut1 and Col-mut2 are two mutant forms of the collagenase AP-1 site that decrease binding of pure human AP-1 protein (Angel et al., 1987). HIS-wt contains the GCN4 binding site of the *HIS3* promoter of *S. cerevisiae*. HIS-168 is a mutant form of this site that has significantly decreased GCN4 binding (Hill et al., 1986). The complete sequence of these oligonucleotides is shown in Experimental Procedures.

also failed to compete for the AP-1-like factor of *S. pombe* (Figure 3, lane 4), whereas slight competition for the human AP-1 protein was observed (Figure 3, lane 10). An interpretation of these results was that the binding seen in the *S. cerevisiae* extract was not due to GCN4. In order to investigate this possibility further, retention assays were carried out using the *HIS3*-derived oligonucleotide as the probe. Incubation of this probe with *S. cerevisiae* extract gave two distinct complexes. The faster migrating complex was not competable in a specific fashion and probably represents nonspecific binding to the probe. The same complex was obtained with a probe containing a mutation

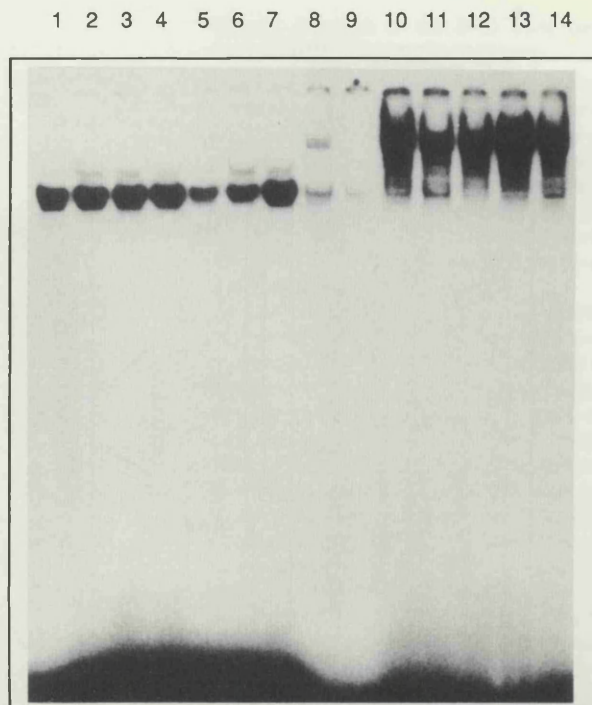


Figure 4. *S. cerevisiae* Contains Specific Proteins That Bind to the AP-1- and GCN4-Binding Sites

Radiolabeled HIS-wt oligonucleotide (lanes 1–7) or SV-AP-1 oligonucleotide (lanes 8–14) was incubated in a standard binding reaction with 150  $\mu$ g of *S. cerevisiae* whole cell extract and a 200-fold molar excess of oligonucleotide competitor. The competitors used were Col-Ap-1 (lanes 1 and 8), SV-AP-1 (lanes 2 and 9), Col-mut1 (lanes 3 and 10), Col-mut2 (lanes 4 and 11), *HIS*-wt (lanes 5 and 12), *HIS*-168 (lanes 6 and 13), and nonsense (lanes 7 and 14). These oligonucleotides are described in Figure 3. The binding reactions were analyzed as described in Experimental Procedures except that 4  $\mu$ g of poly (dI:dC) was used in each reaction.

in the GCN4 binding site (HIS-168) (data not shown). The slower migrating, weaker complex however was competed in a pattern expected for the GCN4 protein. It was competed efficiently by the AP-1 site of collagenase (Figure 4, lane 1) but not by the SV40 AP-1 site (lane 2). The two mutant oligomers col-mut1 and col-mut2 failed to compete (lanes 3 and 4). Unlike the SV-AP-1-derived complex however, this complex was efficiently competed by an excess of unlabeled HIS-wt (lane 5); it was not competed by an oligomer containing a point mutation of the *HIS3* GCN4 binding site (HIS-168) (lane 6) that has previously been shown to severely decrease binding of pure GCN4 protein (Hill et al., 1986). The specificity of this complex therefore exactly mirrors the specificity of GCN4 binding, strongly suggesting that this weak complex results from the GCN4 protein. The weakness of the band might also be expected since the levels of this protein in unstarved cells are very low (Arndt et al., 1987). It appears therefore that at least two proteins exist in *S. cerevisiae* that have very similar binding specificities; one binds well to the SV40 AP-1 site but not to the *HIS3* GCN4 site, and the other binds well to the *HIS3* site but not to the SV40 AP-1 site. Both proteins

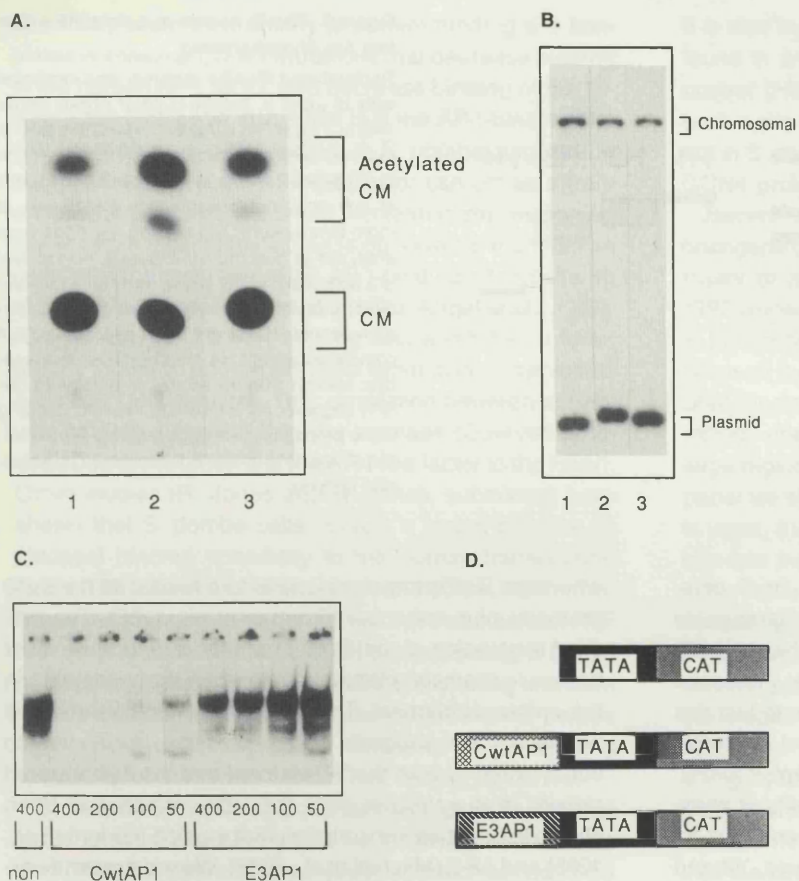


Figure 5. The AP-1-Binding Site Can Act as a UAS Element in *S. pombe*

(A) The expression of CAT coding sequences in *S. pombe* cells transformed with the plasmids pYT3 (lane 1), pYT3CP (lane 2), or pYT3E3 (lane 3). pYT3 contains a TATA box region fused to the CAT coding sequence in the yeast vector pURA4. pYT3CP is identical to pYT3 except it contains a single copy of the Col-AP-1 oligomer upstream of the TATA box, and pYT3E3 has an oligonucleotide containing the adenovirus E3 promoter AP-1-binding site, inserted upstream of the TATA box. The CAT assays were as described in Figure 1, using 600 ng of extract protein per reaction.

(B) Copy number of the transforming plasmid in each of the transformed cells described in (A). Ten micrograms of *S. pombe* DNA was digested with *Sal*I and *Bam*HI, Southern blotted, and probed with  $^{32}$ P-labeled CAT DNA (plasmid). As an internal control the blot was also probed with the *S. pombe* single copy gene *cdc2*<sup>+</sup> (chromosomal).

(C) Affinity of the *S. pombe* AP-1-like factor for the collagenase and E3 promoter AP-1-binding sites. Radiolabeled SV-AP-1 oligomer was incubated with 10  $\mu$ g of *S. pombe* nuclear extract and an excess of unlabeled nonsense oligomer, Col-AP-1 oligomer, or on oligomer containing the adenovirus E3 AP-1-binding site. The molar excess of competitor in each of the reactions is indicated. The binding reactions were analyzed as described in Figure 3.

(D) Schematic representation of the chimeric CAT genes contained in the plasmids used in this study.

can bind to the core sequence that constitutes the collagenase AP-1 site. The SV40-AP-1 DNA-protein complex and the *HIS3* DNA-protein complex have very different gel mobilities, supporting our conclusion that each is derived from the binding of different proteins. At this point however, we cannot rule out the possibility that the two complexes contain the same protein. The different behavior of the complexes could be explained by differential modification of the protein or its interaction with different proteins.

#### The AP-1 Binding Site Can Serve as a UAS in *S. pombe*

Deletion of the AP-1 binding site of the SV40 promoter resulted in lower levels of expression in *S. pombe* (Figure 1B). This suggested to us that the AP-1-like factor may act as a *trans*-acting transcriptional regulator. In order to study this in more detail, we tested the ability of the AP-1-binding site to serve as a UAS element. A single copy of the oligonucleotide containing the collagenase AP-1 binding site was inserted immediately upstream of a TATA box sequence that was fused to the CAT coding region. Expression of CAT in cells transformed with this plasmid was compared with those transformed with the plasmid that lacked the insert. The presence of the Col-AP-1-binding site resulted in increased levels of CAT expression. A number of different transformants were analyzed in this way, and the average increase due to the Col-AP-1 insert

was 5-fold. A typical example is shown in Figure 5A. This difference was not due to differences in plasmid copy number in the transformants as determined by Southern analysis of transformed cell DNA (Figure 5B). We also tested the ability of the AP-1-binding site of the adenovirus E3 promoter (Hurst and Jones, 1987) to stimulate expression in a similar assay. Increased expression was observed, but in this case the increase was only about 2-fold. This result prompted us to compare the efficiency of binding of the *S. pombe* AP-1-like factor with the Col-AP-1- and the E3 AP-1-binding sites. Competitor retention assays showed that binding of the factor to SV-AP-1 oligomer was competed significantly less well by an excess of oligonucleotide containing the E3-binding site than one containing the collagenase site (Figure 5C). This difference in affinity of the two sites for the AP-1-like protein probably explains why the E3-binding site acts as a weaker UAS element.

#### The AP1-like Factor of *S. pombe* Is Sensitive to Phosphatase Treatment

The distinct mobility differences between the *S. pombe* and HeLa cell-derived complexes could be due to a number of reasons. For example, the AP-1 and AP-1-like factors might have a different size. Alternatively, the differences could reflect the consequences of posttranslational modification of the factors. If modification is involved, a



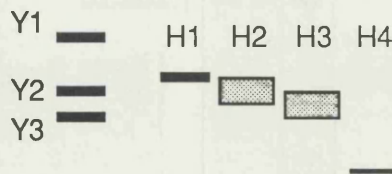
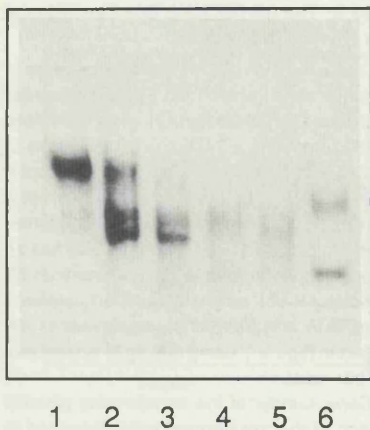


Figure 6. The *S. pombe* and HeLa AP-1 Factors Are Phosphorylated

Radiolabeled SV-AP-1 oligomer was incubated with 15  $\mu$ g of *S. pombe* nuclear extract (lanes 1–3) or 7  $\mu$ g of HeLa nuclear extract (lanes 4–6) that had been preincubated with 0 (lanes 1 and 6), 20 (lanes 2 and 4), or 40 (lanes 3 and 5) units of calf intestinal phosphatase for 10 min at room temperature. The binding reactions were analyzed as described in Figure 3. Also shown is a schematic of the yeast- and HeLa cell-derived protein complexes. Y1–Y3 are yeast-derived and H1–H4 are HeLa cell-derived; all complexes except H4 are competed in a specific fashion with an excess of unlabeled SV-AP-1 oligomer. H4 is due to nonspecific binding to the probe; its intensity is variable from reaction to reaction and is not competed in a specific fashion.

likely candidate would be phosphorylation. In *S. cerevisiae*, the complex between the HSTF and its cognate binding site has a slower mobility following heat shock; this change in mobility appears to be due to increased phosphorylation (Sorger et al., 1987). In order to test this possibility, we pretreated the *S. pombe* and HeLa extracts with calf intestinal phosphatase prior to complex formation. Treatment of the yeast extract with 20 units of phosphatase resulted in a decrease of the normal complex Y1 and the appearance of two different complexes, Y2 and Y3, with increased mobility (Figure 6). With 40 units of phosphatase, Y1 virtually disappears and there is an increase in Y3 in relation to Y2. The simplest explanation for these results is that the *S. pombe* AP-1-like factor is phosphorylated, possibly at multiple sites, and that this phosphorylation induces some conformational change that alters mobility of the protein. The human AP-1 factor also appears to be phosphorylated. Treatment of the HeLa extract with 20 units of phosphatase results in the disappearance of the normal complex H1 (the other complex H4 obtained with the untreated extract is due to nonspecific binding, data not shown) and the appearance of a diffuse complex H2 that has an increased mobility. An even faster migrating, diffuse complex (H3) is obtained with 40 units of phosphatase. Although the H2 and H3 complexes do not migrate as a sharp band, it is interesting to note that the mobilities of Y2 and H2 and of Y3 and H3 are similar. These data strongly suggest that both the HeLa and *S. pombe* factors, or other proteins that may be associated with the complex, are phosphorylated and that the phosphorylation state is largely responsible for the difference in mobilities of the native HeLa and *S. pombe* complexes.

## Discussion

We have shown previously that the SV40 early gene promoter can function efficiently in *S. pombe* cells (Kaufer et al., 1985). In this paper, we show that transcription initiates from the same start site as that used in mammalian cells and that the majority of the enhancer sequences can be removed with little or no effect on expression levels. This is also the case in undifferentiated mammalian cells

where the SV40 promoter alone is as effective as the SV40 promoter plus enhancer (Gorman et al., 1985). The efficient expression of the SV40 promoter in *S. pombe* must be due to promoter elements other than those residing in the enhancer element. The preliminary deletion analysis we have done suggests that expression is affected to some extent by both the "P" element and the 21 bp repeat region. In mammalian cells the latter region has been shown to bind two transcription factors, Sp1 (Gidoni et al., 1985) and AP-2 (Mitchell et al., 1987). We are presently examining what factors in *S. pombe* interact with this 21 bp region.

In the SV40 promoter, the TATA element is 28–33 bp upstream of the major initiation site. In mammalian cells, it determines the site of initiation and its removal results in heterogeneous starts (Benoist and Chambon, 1981). Our demonstration that transcription initiates from the same site in *S. pombe* suggests that in these cells the TATA element is being recognized and is functioning in determining the position of initiation. In mammalian cells, the TATA box is recognized by a specific factor, and therefore it is likely that in *S. pombe* it also interacts with a TATA box binding factor. That initiation sites are determined by a similar mechanism in *S. pombe* and higher eukaryotes is also suggested by the finding that most *S. pombe* genes analyzed do contain TATA elements which are also positioned 25–30 bp upstream of the start site (Russell, 1983). In contrast however, TATA box elements in *S. cerevisiae* genes are located at variable distances from the start site, usually 60–120 bp upstream, and they appear not to direct transcription initiation at a fixed distance downstream (Struhl, 1987a). Indeed, mutational analysis of the *HIS3* promoter suggests that the signal for determining the position of initiation does not reside in the TATA element (Chen and Struhl, 1985). With respect to this aspect of transcription, therefore, it appears that there is greater similarity between the promoters of higher eukaryotes and *S. pombe* than with *S. cerevisiae*.

A major finding of this paper is that in fission yeast there is a factor that very closely resembles the human factor AP-1. Both the *S. pombe* and human factors give the same footprint on the SV40 promoter and show similar binding

specificity when their affinity to mutant binding site templates is assessed; point mutations that decrease binding of the human AP-1 factor also decrease binding of the fission yeast factor. We also show that the AP-1-binding site can function as a UAS element in *S. pombe*, suggesting strongly that the yeast AP-1-like factor can act as a transcriptional activator. The 5-fold increase in expression we see is very similar to the increase observed in mammalian cells with the insertion of an AP-1-binding site upstream of a basal promoter (Lee et al., 1987b; Angel et al., 1987). Additionally, the E3 AP-1-binding site, which has a lower affinity for the *S. pombe* AP-1-like factor in vitro, served as a weaker UAS element. This correlation between activity and binding suggests that the increase observed is indeed due to the binding of the AP-1-like factor to the insert. Other studies (R. Jones and N. Jones, submitted) have shown that *S. pombe* cells contain a factor that has an identical binding specificity to the human transcription factor ATF (Lee et al., 1987c; Hurst and Jones, 1987). The consensus binding sites for human AP-1 and human ATF are very similar, the crucial difference being a single nucleotide (Hurst and Jones, 1987). It is striking therefore that despite the evolutionary distance between *S. pombe* and human cells, this subtle difference in the binding specificity of these two factors has been maintained.

*S. cerevisiae* cells also contain a nuclear factor that binds to the SV40 AP-1 site. This factor appears distinct from the well-characterized yeast activator encoded by *GCN4*, which regulates expression of a number of coordinately expressed amino acid biosynthetic genes (Hinnebusch et al., 1985; Hope and Struhl, 1985). The *GCN4* protein binds to a specific UAS element whose consensus sequence is very similar to the AP-1-binding site and to which purified AP-1 protein can also bind (Bohmann et al., 1987). However, both the collagenase and the SV40 AP-1 sites are more efficient competitors for human AP-1 than the *HIS3 GCN4* binding site. The situation is similar for the *S. pombe* and *S. cerevisiae* AP-1-like factors, but a greater stringency is shown as the *HIS3* UAS element is not an effective competitor even at high molar ratios. Therefore, in all three cell systems, both the SV40 and collagenase AP-1-binding sites have a significantly higher affinity for AP-1 and the AP-1-like factors than the *HIS3 GCN4* binding site.

In this paper, we identify by gel retention a complex resulting from the binding of a budding yeast protein to the *HIS3 GCN4* binding site. This complex shows a binding specificity that mirrors that expected for the *GCN4* protein (Hill et al., 1986). For the sake of further discussion, we will assume that this complex is due to *GCN4* protein although we have no immunological proof that this is the case. In contrast to the AP-1-like factor, this protein binds significantly better to the *HIS3* binding site than to the SV40 AP-1 site. Interestingly, however, it does bind well to the collagenase AP-1 site. This can be explained by considering the mutagenesis studies of Hill et al. (1986), who showed that the *HIS3 GCN4* binding site (TGACTCT) could be mutated to create a site (TGACTCA) with greater affinity for *GCN4* protein. The core sequence of the collagenase AP-1-binding site is identical to this high affinity site and therefore explains why *GCN4* protein binds to it.

It is also significant that the high affinity *GCN4* site is not found in any of the promoters under general amino acid control (Hill et al., 1986). A possible explanation is that such a site may also interact with the AP-1-like factor present in *S. cerevisiae* and therefore no longer be specific for *GCN4* protein binding.

Recent studies have shown that the cellular proto-oncogene *c-jun* encodes the human AP-1 transcription activator or at least a very close relative (Bohmann et al., 1987; Angel et al., 1988). Comparison of the *jun* sequences to known protein structures revealed a striking homology between the DNA-binding C-terminal region of *jun* and the DNA-binding domain of *GCN4* (Vogt et al., 1987). The homology does not extend beyond the binding domain, and large regions of the protein are unconserved. Since in this paper we show the existence of a second AP-1-like protein in yeast, it will be of interest to determine whether this protein has more extensive homology to mammalian AP-1.

We have also demonstrated here that the complexes formed by both the human AP-1 and the *S. pombe* AP-1-like factors are sensitive to phosphatase treatment; addition of alkaline phosphatase resulted in significant alteration in the mobilities of these complexes, and the difference in mobility of the yeast and HeLa-derived native complexes is largely negated by the phosphatase treatment. The simplest interpretation of this data is that the AP-1 and AP-1-like factors are phosphorylated. However, it is possible that the complexes contain multiple proteins and that the altered mobilities could result from loss or dephosphorylation of any one of these. Future studies should resolve this point and also reveal whether phosphorylation is critical for factor activity and regulation.

An exciting possibility is that the yeast system may provide a convenient means for the identification and isolation of cDNAs that encode mammalian transcription factors and for their subsequent genetic analysis. Using standard genetic methods, it should be possible to isolate yeast mutants that are defective in the genes encoding factors such as the AP-1-like and ATF-like factors that appear to be closely related to mammalian counterparts. Functional complementation of such mutants with human cDNA libraries should then be possible. Such a strategy would depend, of course, on the human factor being able to function in yeast cells and interact with the yeast transcriptional machinery. The demonstration that yeast activators can function in human cells makes this a distinct possibility. Such a complementation strategy has already been used successfully to clone the human homolog of the fission yeast *cdc2<sup>+</sup>* gene (Lee and Nurse, 1987).

## Experimental Procedures

### Plasmid Constructions

The plasmid pSABCAT was constructed by cloning the HindIII–BamHI fragment from pSV2CAT (Gorman et al., 1982) that contains the CAT coding sequence and SV40 3' regulatory sequences into the plasmid pSAB1 (Durkacz et al., 1985). pSMCAT was constructed by cloning the 323 bp PvuII–HindIII fragment containing the SV40 promoter–enhancer region upstream of CAT in pSABCAT. SMD1CAT was prepared by cloning the BglII–HindIII fragment of pA<sub>10</sub>CAT2 (Gorman et al., 1982) into pSABCAT; this fragment contains the sequences of the SV40 promoter region to be the SphI site at –198. SMD2CAT was prepared

by cloning into SABCAT the HincII–HindIII fragment from the plasmid p21Tg (a kind gift from C. Goding); this fragment contains the promoter sequences up to position –150. SMd13CAT was prepared by cloning into SABCAT the NcoI–HindIII fragment from p21Tg, which contains promoter sequences up to position –107.

The plasmids pYT3, pYT3CP, and pYT3E3 were constructed by inserting specific oligonucleotide sequences upstream of the CAT coding sequences into the plasmid pU4E2CAT. This plasmid was constructed by cloning the 1.95 kb XhoI–XbaI fragment of pE2A-E-CAT (Murthy et al., 1985) into pUC19 and subsequent subcloning of a BamHI–SphI fragment from this intermediate into the BamHI–SphI sites of pURA4. The E2 promoter sequences present in pU4E2CAT were removed by restricting with Sall and BamHI and replaced with oligonucleotides containing the adenovirus E3 TATA region from +10 to –37 (pYT3) and, additionally, sequences from –79 to –103 of the E3 promoter (pYT3E3) or an oligonucleotide containing the core AP-1 binding site of the human collagenase promoter (pYT3CP).

#### Extract Preparation

HeLa cell extract was prepared as previously described (Hurst and Jones, 1987). Total extract from *S. cerevisiae* was a gift from P. Sorger (Sorger et al., 1987). Extract from *S. pombe* was prepared as follows: The *S. pombe* strain Ade 704 h<sup>+</sup> was grown to a density of 6 × 10<sup>6</sup> cells/ml in minimal medium containing 200 mg/l adenine, histidine, leucine, uracil, and 0.5% glucose. Yeast spheroplasts were then prepared as previously described (Beach and Nurse, 1981), except that Novozym 234 was removed by repeated washing in ice-cold 1.2 M sorbitol, 10 mM Tris-HCl (pH 7.6), and 1 mM PMSF. Nuclei were prepared from spheroplasts as described by Wiederrrecht et al. (1987), except that ammonium sulphate was added at 0.3 g/ml at the final extract precipitation.

#### DNA Preparation and CAT Assays

Twenty milliliters of a particular clone were grown to a density of 3–5 × 10<sup>6</sup> cells/ml in minimal medium and harvested by centrifugation. Total DNA was prepared from half of the cells according to Durkacz et al. (1986). After digestion with BamHI and Sall, the samples were Southern blotted and probed with <sup>32</sup>P-labeled CAT DNA to determine plasmid copy number. All CAT band intensities were normalized with respect to the hybridization signal for the single copy *cdc2*<sup>+</sup> gene.

The remaining cells were resuspended to 50 μl of 250 mM Tris-HCl (pH 7.5), 1 mM PMSF and broken by vortexing with acid washed glass beads for 30 sec. The beads were washed with 3 × 0.5 ml of the same buffer and cell debris removed by centrifugation. Samples of these extracts were incubated at 37°C in the presence of <sup>14</sup>C-chloramphenicol and acetyl CoA (Gorman et al., 1982).

#### DNAase I Footprinting

Footprinting with uninfected HeLa cell and *S. pombe* extracts were carried out as previously described (Hurst and Jones, 1987). Each reaction contained 0.8 μg of end-labeled DNA probe and 2 μg of poly (dl:dC).

#### Gel Retention Assay

Binding reactions between extracts and end-labeled probes were carried out in a final volume of 25 μl in 20 mM HEPES-KOH (pH 8.0), 100 mM NaCl, 2 mM EDTA, 5 mM MgCl<sub>2</sub>, 7 mM β-mercaptoethanol, 1 mM PMSF, and 10% glycerol. In addition to probe (0.1–0.2 ng) and crude extract, the reactions also contained 2 μg poly (dl:dC) (Pharmacia) as nonspecific DNA competitor. Binding reactions were carried out at 20°C for 20 min and loaded directly onto 6%–8% polyacrylamide gels and run in 0.5× TBE at 200 V.

#### Primer Extension

An oligonucleotide (5'-GTGACCTATATGGTGGCA-3') complementary to the CAT coding sequence (3'-CACTGGATATACACCGT-5') was <sup>32</sup>P-labeled with T4 polynucleotide kinase and γ-<sup>32</sup>[PATP] (Amersham 5000 Ci/mmol) (Maniatis et al., 1982). Four nanograms of labeled oligonucleotide was annealed with 6 μg of total RNA, prepared by the method of Kaufer et al. (1985), in 10 μl of 10 mM Tris-HCl (pH 8), 1 mM EDTA, and 250 mM KCl for 1 hr at 50°C. Extension was initiated by addition of 25 μl of 20 mM Tris-HCl (pH 8.7), 10 mM MgCl<sub>2</sub>, 5 mM DTT, and 0.25 mM of each dNTP and 10 units of Super RT (Anglian). After

1 hr at 37°C, the reactions were ethanol precipitated and analyzed on a 6% acrylamide gel containing 7 M urea.

#### Phosphatase Reactions

Ten micrograms of *S. pombe* nuclear extract or 5 μg of HeLa nuclear extract were incubated at room temperature for 10 min with or without the addition of 20 or 40 units of calf intestinal phosphatase (Boehringer Mannheim). The reactions were in a total volume of 5 μl in the gel retention buffer additionally containing 10 mM MgCl<sub>2</sub> and 2 mM ZnCl<sub>2</sub>. The treated extracts were subsequently assayed by gel retention as described above except that the retention buffer also contained 5 mM sodium molybdate.

#### Oligonucleotides

The oligonucleotides were purified and annealed as previously described (Hurst and Jones, 1987). The sequence of the oligonucleotides used in this study are:

HIS-168	5'-3'	TCGACGGATCCAGCGGATGACGCTTTTTTTGCATG
	3'-5'	GCCTAGGTCGCCTACTGCGAAAAAACC
HIS-wt	5'-3'	TCGACGGATCCAGCGGATGACTCTTTTTTTGCATG
	3'-5'	GCCTAGGTCGCCTACTGAGAAAAAACC
Col-mut1	5'-3'	TCGACGGATCCTTGAGGAGTCAGCCGCATG
	3'-5'	GCCTAGGAACTCCTCAGTCGGC
Col-mut2	5'-3'	TCGACGGATCCTTGATGAGTCCTCCGCATG
	3'-5'	GCCTAGGAACTACTCAGGAGGC
Col-AP-1	5'-3'	TCGACATCGATTTGATGAGTCAGCCGCATG
	3'-5'	GTAGCTAACTACTCAGTCGGC
SV40-AP-1	5'-3'	GATCCATCTCAATTAGTCAGCAAG
	3'-5'	GTAGAGTTAATCAGTCGTTCTAG

The sequences in the E3-AP-1 and nonsense oligomers were as previously described (Hurst and Jones, 1987).

#### Acknowledgments

We thank Viesturs Simanis for his generous provision of advice and materials, Peter Sorger for budding yeast extract, Helen Hurst for HeLa cell extract and advice, Iain Goldsmith for synthesis of the oligonucleotides and Tom Dooley for helpful discussion. S. M. is a fellow of the C.S.I.C. (Consejo Superior de Investigaciones Científicas, Spain). We finally thank Clare Middlemiss and Fiona Gallivan for preparing the manuscript.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 31, 1988.

#### References

- Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R. J., Rahmsdorf, H. J., Jonat, C., Herrlich, P., and Karin, M. (1987). Phorbol ester-inducible genes contain a common *cis* element recognized by a TPA-modulated *trans*-acting factor. *Cell* 49, 729–739.
- Angel, P., Allegretto, E., Okino, S., Hattori, K., Boyle, W. J., Hunter, T., and Karin, M. (1988). Oncogene *jun* encodes a sequence-specific *trans*-activator similar to AP1. *Nature* 332, 166–171.
- Arndt, K. T., Stiles, C., and Fink, G. R. (1987). Multiple globin regulators control HIS4 transcription in yeast. *Science* 237, 874–880.
- Beach, D., and Nurse, P. (1981). High frequency transformation of the fission yeast *S. pombe*. *Nature* 200, 140–142.
- Benoist, C., and Chambon, P. (1981). In vivo sequence requirements for the SV40 early promoter region. *Nature* 290, 304–310.
- Bohmann, D., Bos, T. J., Admon, A., Tetsuji, N., Vogt, P. K., and Tjian, R. (1987). Human proto-oncogene *c-jun* encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. *Science* 238, 1386–1392.
- Chen, W., and Struhl, K. (1985). Yeast mRNA initiation sites are deter-

- mined primarily by specific sequences not by distance from the TATA element. *EMBO J.* **4**, 3273–3280.
- Durkacz, B., Beach, D., Hayles, J., and Nurse, P. (1985). The fission yeast cell cycle control gene *cdc2*: structure of the *cdc2* region. *Mol. Gen. Genet.* **201**, 543–545.
- Fromm, M., and Berg, P. (1982). Deletion mapping of DNA regions required for SV40 early promoter functions in vivo. *J. Mol. Appl. Genet.* **1**, 457–481.
- Fromm, M., and Berg, P. (1983). Simian virus 40 early- and late-region promoter functions are enhanced by the 72-base-pair repeat inserted at distant locations and inverted orientations. *Mol. Cell. Biol.* **3**, 991–999.
- Gidoni, D., Dynan, W. S., and Tjian, R. (1984). Multiple specific contacts between mammalian transcription factor and its cognate promoters. *Nature* **312**, 409–413.
- Gorman, C. M., Moffat, L., and Howard, B. H. (1982). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* **2**, 1044–1051.
- Gorman, C. M., Rigby, P. W. J., and Lane, D. P. (1985). Negative regulation of viral enhancers in undifferentiated embryonic stem cells. *Cell* **42**, 519–526.
- Gosh, P. K., Lebowitz, P., Frisque, R. J., and Gluzman, Y. (1981). Identification of a promoter component involved in positioning the 5' termini of the SV40 early mRNAs. *Proc. Natl. Acad. Sci. USA* **78**, 100–104.
- Hill, D. E., Hope, I. A., Mackie, J. P., and Struhl, K. (1986). Saturation mutagenesis of the yeast *his3* regulatory site: requirements for transcriptional induction and for binding by *GCN4* activator protein. *Science* **234**, 451–457.
- Hinnebusch, A. G., Lucchini, G., and Fink, G. R. (1985). A synthetic *HIS4* regulatory element confers general amino acid control of the cytochrome *c* gene (*CYC1*) of yeast. *Proc. Natl. Acad. Sci. USA* **82**, 498–502.
- Hope, I. A., and Struhl, K. (1985). *GCN4* protein, synthesized in vitro, binds *HIS3* regulatory sequences: implications for general control of amino acid biosynthesis genes in yeast. *Cell* **43**, 177–188.
- Hurst, H. C., and Jones, N. C. (1987). Identification of factors that interact with the E1A-inducible adenovirus E3 promoter. *Genes and Development* **1**, 1132–1146.
- Jones, N. C., Rigby, P. W. J., and Ziff, E. B. (1988). Transacting protein factors and the regulation of eukaryotic transcription: lessons from studies on DNA tumour viruses. *Genes and Development* **2**, 267–281.
- Kakidani, H., and Ptashne, M. (1988). *GAL4* activates gene expression in mammalian cells. *Cell* **52**, 161–167.
- Kaufman, N. F., Simanis, V., and Nurse, P. (1985). Fission yeast *Schizosaccharomyces pombe* correctly excises a mammalian RNA transcript intervening sequence. *Nature* **318**, 78–80.
- Lee, M. G., and Nurse, P. (1987). Complementation used to clone a human homologue of the fission yeast cell cycle control gene *cdc2*. *Nature* **327**, 31–35.
- Lee, W., Haslinger, A., Karin, M., and Tjian, R. (1987a). Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. *Nature* **325**, 368–372.
- Lee, W., Mitchell, P., and Tjian, R. (1987b). Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell* **49**, 741–752.
- Lee, K. A. W., Hai, T. Y., Siva Raman, L., Thimmappaya, B., Hurst, H. C., Jones, N. C., and Green, M. R. (1987c). A cellular transcription factor ATF activates transcription of multiple E1A-inducible adenovirus early promoters. *Proc. Natl. Acad. Sci. USA* **84**, 8355–8359.
- Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).
- Maniatis, T., Goodbourn, S., and Fischer, J. A. (1987). Regulation of inducible and tissue-specific gene expression. *Science* **236**, 1237–1244.
- 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.
- Murthy, S. C. S., Bhat, G. P., and Thimmappaya, B. (1985). Adenovirus E1A early promoter: transcriptional control elements and induction by the viral pre-early E1A gene, which appears to be sequence independent. *Proc. Natl. Acad. Sci. USA* **82**, 2230–2234.
- Pelham, H. (1985). Activation of heat-shock genes in eukaryotes. *Trends Genet.* **1**, 31–35.
- Russell, P. R. (1983). Evolutionary divergence of the mRNA transcription initiation mechanism in yeast. *Nature* **307**, 167–169.
- Russell, P., and Nurse, P. (1986). *cdc25<sup>+</sup>* functions as an inducer in the mitotic control of fission yeast. *Cell* **45**, 145–153.
- Sentenac, A. (1985). Eukaryotic RNA polymerases. *CRC Crit. Rev. Biochem.* **18**, 31–91.
- Serfling, E., Jasin, M., and Schaffner, W. (1985). Enhancers and eukaryotic gene transcription. *Trends Genet.* **1**, 224–230.
- Sorger, P. K., Lewis, M. J., and Pelham, H. R. B. (1987). Heat shock factor is regulated differently in yeast and HeLa cells. *Nature* **329**, 81–84.
- Struhl, K. (1987a). Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. *Cell* **49**, 295–297.
- Struhl, K. (1987b). The DNA-binding domains of the *jun* oncoprotein and the yeast *GCN4* transcriptional activator proteins are functionally homologous. *Cell* **50**, 841–846.
- Vogt, P. K., Bos, T. J., and Doolittle, R. F. (1987). Homology between the DNA-binding domain of the *GCN4* regulatory protein of yeast and the carboxy-terminal region of a protein encoded for by the oncogene *jun*. *Proc. Natl. Acad. Sci. USA* **84**, 3316–3319.
- Webster, N., Jin, J. R., Green, S., Hollis, M., and Chambon, P. (1988). The yeast *UAS<sub>G</sub>* is a transcriptional enhancer in human HeLa cells in the presence of the *GAL4 trans*-activator. *Cell* **52**, 169–178.
- Wiederrecht, G., Shuey, D. J., Kibbe, W. A., and Parker, C. S. (1987). The *Saccharomyces* and *Drosophila* heat shock transcription factors are identical in size and DNA binding properties. *Cell* **48**, 507–515.