Isolation and Characterization of Factors that Interact with Eukaryotic Transcriptional Enhancers

Thesis by

Keith D. Harshman

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1989

(Submitted December 21, 1988)

To my family

Acknowledgements

I would like to thank my advisor, Carl Parker, for his considerable patience and encouragement during the completion of the work described here. I would also like to thank Peter Dervan for his careful and constructive supervision of my first three years at Caltech. The support and encouragement of John Coutts, William Martin, Richard Jones and Bob and Linda Peterson prior to my arrival at Caltech is greatly appreciated. Friendships with Margaret Ferguson, Dick Bickle, Brian Feeney, Dave Beecher, Suzy Merrill and Sam Azlein, which began before coming to Caltech, and continued while here (over unfortunately long distances), were important reminders that there really was life 'on the outside'. Many thanks to Debora St-Claire, Tom Chesnutt, Melinda McCarthy and Stephanie Steele for making me realize there are many available paths after Caltech. Interactions with the Parker, Campbell and Dervan (Yo! Johnny Salami and Scotty!) labs- scientific and otherwise- have been enjoyable and rewarding. Joanne Topol, Kevin Sweder and Greg Wiederrecht deserve all the thanks I can possibly give (and more) for their friendship, advice, and remarkable ability to put things in the proper perspective. And finally, thanks to the boys at Hynes Meat Packing Company for good times on and off the job.

Abstract

The regulatory region controlling the transcription of the SV40 early genes provides a useful system in which to study the mechanisms of eukaryotic gene expression. The enhancer element of this region is composed of sequence motifs that can function independently or in combination to potentiate transcription in a wide range of cellular contexts. This thesis describes the purification and characterization of two proteins that specifically bind separate motifs.

An enhancer binding protein was identified in bovine thymus extracts and purified to homogeneity. This factor, designated EP2, was shown by chemical and enzymatic footprinting techniques to bind the core as well as two pseudo-core sequences. In a DNase I footprinting experiment, EP2 was unable to bind to a mutated core sequence that is incapable of activating transcription *in vivo*. EP-2 was shown to consist of a group of polypeptides, ranging in molecular weight from 34 kd to 43 kd, each of which has the ability to bind to the core and two pseudo-core sequences.

It was shown that the sequence bound by the mammalian transcription factor AP-1-- the AP-1 recognition element (ARE)-- was capable of activating transcription in yeast. The ARE is very similar in sequence to the GCN4 recognition element (GCRE), yet the ARE was shown to activate transcription in a *gcn4* yeast strain. A

protein present in yeast extracts, designated yAP-1, was shown to bind to the ARE and was purified to near homogeneity based on this ability. yAP-1 and AP-1 display remarkably similar biochemical and DNA binding characteristics. It was shown *in vitro* that yAP-1 can discriminate between the ARE and the GCRE, while GCN4 exhibits approximately equal affinities for the two elements.

The structural gene encoding yAP-1 was isolated and characterized. The DNA binding domain of the protein was localized to a sequence of 93 amino-acids in the amino-terminus. This sequence was shown to have significant homology with domains in c-JUN and GCN4, which have been ascribed the ability to bind DNA. The YAP1 gene was shown to be non-essential by gene disruption. DNA-affinity blot experiments performed using extracts from *YAP1* and *yap1* strains suggest the existence of a family of yeast genes encoding proteins that recognize the ARE.

Table of Contents

Introduction

The nature of the molecular mechanisms controlling the process of differential gene expression is one of the fundamental questions facing molecular biology today. This process controls many of the key events underlying important biological phenomena, such as organismal development, cellular differentiation, and neoplastic growth. The initiation of transcription represents one of the critical steps in the control of differential gene expression. The study of gene expression in bacteria has contributed greatly to the understanding of the mechanisms that control transcription initiation. Genetic studies of gene expression in bacteria first led to the proposition that genes can be divided into distinct functional regions-the structural region used as a template for the enzymatic synthesis of the RNA by RNA polymerase, and the promoter and regulatory regions that ensure that RNA synthesis initiates at the correct location and only when the correct environmental conditions prevail (Jacob and Monod, 1961). Extensive genetic, biochemical, and physical studies have subsequently revealed in great detail the DNA determinants in the bacterial promoter, the protein factors involved, as well as the biochemical steps required to initiate transcription (Miller and Reznikoff 1982; McClure, 1985; Helman and Chamberlin, 1988).

In *E. coli*, RNA polymerase is a multi-subunit enzyme which, in its active form contains five different peptide chains, α, β, β', and σ. This form of RNA polymerase, referred to as the holoenzyme, has the stoichiometry $\alpha_2\beta\beta'\sigma$ (Chamberlin, 1974). The σ subunit dissociates

relatively easily from the holoenzyme and in this dissociated form the polymerase is referred to as the core enzyme. The core enzyme can catalyze the formation of the inter-nucleotide phosphodiester bonds as well as the holoenzyme (Watson *et al.*, 1987). The σ subunit in the holoenzyme imparts to the enzyme the ability to recognize and bind the promoter with high affinity, while decreasing the enzyme's affinity for non-promoter sequences (Helmann and Chamberlin, 1988). Furthermore, the σ subunit may assist in DNA melting prior to transcription initiation (Simpson, 1979) and inhibit non-specific transcription (Tjian *et al.*, 1977). Evidence indicates that the σ subunit may make sequence-specific contacts with the promoter DNA when it is part of the holoenzyme (Simpson, 1979). Following transcription initiation, the σ subunit is no longer needed by the advancing polymerase and is released from the enzyme-nucleic acid complex (Hansen and McClure, 1980).

The promoters of nearly all *E. coli* genes contain two regions of highly conserved sequences as diagrammed in Fig. 1 (McClure, 1985). The exact positions of these two regions relative to the start site of transcription (defined as position +1) vary, but in most cases they are present in roughly the same position. The conserved region closest to the start site is typically 10 base pairs upstream of position +1 and is therefore called the -10 region (it is also referred to as the Pribnow box). The second conserved region is found centered approximately 35 base pairs upstream of the start site and is called

the -35 region. The consensus sequences formulated for the -10 region and -35 region are TATAAT and TTGACA, respectively.

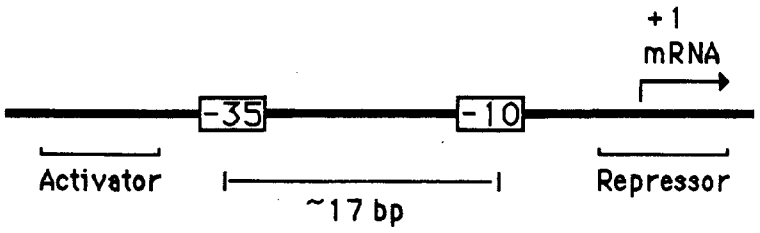


Figure 1. Schematic representation of a bacterial control region. Shown are the -10 and -35 elements and binding sites for activators and repressors. See text for details.

The importance of these two regions to *E. coli* promoter function is supported by both genetic and biochemical evidence. A majority of the mutations that significantly decrease transcription initiation change nucleotides in these conserved regions (Youderian *et al.*, 1983). Chemical and enzymatic footprinting experiments show important contacts are made between RNA polymerase and the -10 and -35 regions (Siebenlist *et al.*, 1980; Schmitz and Galas, 1979). Finally, the σ subunit can be UV cross-linked to the bacterial promoter close to the -10 region (Simpson, 1980). Although no actual *E. coli* promoters have -10 and -35 regions that correspond exactly to the consensus sequences, most differ by only a few nucleotides. In general, promoters that initiate transcription at low frequency (*i.e.*, "weak" promoters) contain -10 and -35 regions,

which differ from the consensus sequences more extensively than "strong" promoters. The degree to which a promoter diverges from the consensus sequence most likely represents an evolutionary pressure to optimize rather than maximize the expression of a given gene.

In order to survive, a bacterial cell must be able to adapt physiologically in response to environmental changes. To address this need, bacteria have developed the ability to regulate and change their pattern of gene expression in response to environmental signals. One way this is done is by using variant σ subunits, distinct from the predominant σ subunit, which interact only with groups of related genes (Losick and Pero, 1981). For example, there exists in *E. coli* a σ subunit called σ^{32} that promotes the transcription of a group of genes that allow the cell to survive the effects of growth at elevated temperatures (the so-called heat shock response; Grossman *et al.*, 1984). The -10 and -35 region sequence differs significantly from the consensus sequence in those genes that respond to alternative subunits. However, those genes that respond to the same σ subunit have -10 and -35 regions with highly homologous sequences (Hellman and Chamberlin, 1988).

In addition to holoenzyme, bacteria contain proteins that affect transcription initiation of specific genes by binding to DNA sequences found in or near the promoters of those genes (Miller and Reznikoff, 1980). In this way, a cell can regulate the expression of a gene by either regulating the presence of that gene-specific factor or the

ability of that factor to bind to the promoter. The effect of these regulatory proteins can be to either decrease or increase that rate of transcription initiation. Thus, proteins either inhibit the interaction of holoenzyme with the promoter, in which case they are called repressors, or facilitate that interaction, in which case they are called activators. The functional state of activator and repressor proteins can, in turn, be regulated by the presence in the cell of inducer and co-repressor molecules (Miller and Reznikoff, 1982). These molecules, also known as effectors, function by binding to the regulatory protein and activating or inactivating its ability to bind DNA.

The basic concepts of gene regulation that emerge from the studies of bacteria indicate the importance of protein-DNA interactions, protein-protein interactions, and protein-effector interactions in the control of transcription initiation. Although these basic concepts hold true for the control of transcription initiation in eukaryotes, the mechanistic details of these processes differ considerably and are significantly more complicated. This is not surprising considering the increased structural complexity displayed by eukaryotes over prokaryotes. *Saccharomyces cerevisiae* (baker's yeast) has proven to be a very convenient experimental system in which to study the processes of eukaryotic cells. Yeast display many of the basic structural and biological characteristics of higher eukaryotes, and yet, as microorganisms, offer many of the experimental advantages of prokaryotes. The relative ease with

which yeast can be genetically manipulated has made *S. cerevisiae* the most thoroughly dissected eukaryotic genetic system (Strathern *et al.*, 1982).

Unlike bacteria, where a single RNA polymerase is responsible for the transcription of protein coding and non-protein coding genes, the genome of yeast and other eukaryotes is transcribed by three different RNA polymerases (Sentenac and Hall, 1982). All three polymerases are large, multi-subunit enzymes composed of a number of different polypeptide chains. RNA polymerase I transcribes only the ribosomal RNA genes; RNA polymerase III transcribes the tRNA and 5S ribosomal RNA genes. RNA polymerase II transcribes the protein encoding genes and is therefore the primary polymerase involved in differential gene expression.

Yeast protein coding genes, like those in bacteria, consist of two regions: the control region, containing the promoter and regulatory elements that control transcription initiation, and the protein encoding structural region. The control region lies upstream of the structural region as diagrammed in Fig. 2. The initiation control regions-yeast promoters-are considerably larger than those in bacteria, extending hundreds of bases upstream of the transcriptional start site. Genetic analyses of the promoters of many yeast genes have allowed the identification of DNA sequences (called cis-acting elements), which regulate the transcription initiation of the gene. These cis-acting elements range in size from 10 to 30 base and can lie anywhere from 50 to 500 base pairs upstream from the

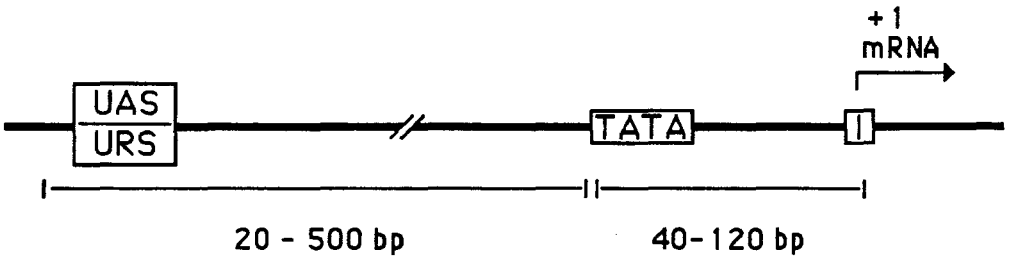


Figure 2. Schematic representation of a yeast control element. Shown are the relative positions of upstream activating (UAS) and repressing (URS) sequences, TATA box element, and the initiator element. See text for details.

start site. They are classified as either upstream activating sequences (UAS) or upstream repressing sequences (URS) on the basis of the effect of their presence in a promoter on the rate of initiation (Brent, 1985; Struhl, 1987). Unlike bacterial promoter elements, neither the orientation nor the exact position relative to the start site (within the 500 base pair limit) is important in order to influence transcriptional initiation (Guarente, 1987). Genes that respond to common environmental stimuli or that are involved in the same metabolic pathways contain upstream elements that are similar in DNA sequence (Arndt and Fink, 1986).

Not surprisingly, most UASs act as binding sites for protein factors (trans-acting factors). Protein binding at a UAS is nearly

always a requisite for transcriptional activation (Struhl, 1987). A number of models have been proposed to explain a factor's ability to activate transcription by binding at an UAS (Ptashne, 1988). One model proposes that activator binding alters the nearby chromatin structure in such a way that the transcription start site is more accessible to RNA polymerase. Another model proposes that RNA polymerase is attracted to the UAS-bound activator with high affinity, and then moves along the DNA to the transcriptional initiation site. A third model-one that has been receiving much experimental support recently-proposes the UAS-bound factor helps to recruit factors that bind at or near the start site via protein-protein interaction. The UAS is brought into the immediate vicinity of the binding sites for those other factors by looping out the intervening DNA. This model further proposes that transcriptional activators prossess surfaces that, by virtue of structure and/or chemical characteristics, interact with other proteins bound near the transcription start site and promoter elements (Ptashne, 1986). It is possible that more than one mechanism could be responsible for the effects produced by UASs.

Transcriptional repression, like activation, is thought to result from protein binding to an URS (Brent, 1985). The mechanism(s) by which these proteins act, however, is much less clear. The flexibility in positioning relative to the start site suggests that at least some do not function, as do bacterial repressors, by interfering with RNA polymerase binding. It has been proposed that they act without

affecting the ability of transcription activators to interact with polymerase and without interfering with their ability to bind DNA. This could be achieved by repressing the ability of the polymerase to travel from the UAS to the start site (Johnson and Herskowitz, 1985). It has also been proposed that repressors function by altering the chromatin structure such that the promoter regions become inaccessible to required transcription factors (Guarente, 1987). Again, more than one mechanism may be at work.

In addition to UASs and URSs, yeast promoters contain two other elements involved in transcription initiation. One of these elements is the TATA element (consensus sequence TATAA), which is similar in sequence to the bacterial -10 region and is required for initiation (Struhl, 1982). This element is found in the promoters of many different genes and is thought to play a general role in the transcription process. It does not have a strict position requirement, ranging in distance from about 40 to 120 base pairs upstream of the start site. A factor that binds to this element has been identified and is believed to function by interacting with polymerase (Buratowski *et al.*, 1988). The second element is the initiator element, which encompasses the start site and, while not affecting the level of transcription, is the primary determinant of the nucleotide at which transcription begins. It is not clear whether this element is the binding site for a separate factor or the DNA sequence preferred by RNA polymerase II for initiation (Burton *et al.*, 1986).

The structure and functions of the cells of higher eukaryotes are considerably more complicated than the uni-cellular yeasts. The mammalian genome, for example, is significantly larger than any of the yeasts and is composed of sequences of different classes: gene sequences transcribed by RNA polymerases, structural sequences necessary for DNA replication and chromosome sorting during cell division, and repetitive sequences whose function is a mystery (Igo-Kemenes *et al.*, 1982). Of even more importance is the fact that eukaryotic organisms are multi-cellular, being composed of different cell types and tissues with (sometimes vastly) different functions. Different cell types are characterized by the production of different proteins, requiring a system to accurately and precisely control the expression and, thus, the transcription initiation of the protein coding genes. The control regions of the protein coding genes of higher eukaryotes reflect the increased genomic complexity and specialized functions of these cells. This increased complexity has made the study of gene expression in higher eukaryotes a more difficult task.

Using a variety of genetic and molecular biological techniques, the structures of a number of control regions for mammalian protein encoding genes have been determined (Maniatis *et al.*, 1987). They exhibit many of the features of yeast control elements. Immediately upstream of the transcriptional start site in most mammalian promoters lies a highly conserved region with the consensus sequence $TATA^{A}/_{T}A^{A}/_{T}$. This "TATA" element is always located 25-30 base pairs upstream of the beginning of the transcribed region

and appears to define the position of the mRNA start site; mutations in this element produce 5' heterogeneity in the RNA transcripts (Benoist and Chambon, 1981). Further upstream from the start site is a region that contains one or more conserved sequences referred to as upstream elements. These sequences are generally 10-15 base pairs in size and can usually function in both their natural and inverse orientation. A wide variety of upstream elements have been found with different consensus sequences. Some upstream elements appear to be active in all, or nearly all, cell types at all times. Other upstream elements exhibit activity in a single or narrow range of cell types and/or in response to very specific environmental stimuli. These elements are generally required for efficient initiation of transcription (Dynan and Tjian, 1985). Deletion or mutation of these sequences results in promoter inactivation (Grosveld *et al.*, 1982; Dudler and Travers, 1984). There are cases, however, where upstream elements repress the initiation of transcription (Goodbourn *et al.*, 1986). The efficacy with which upstream elements potentiate transcription drops off steeply as their distance from the start site exceeds approximately 100 bases (Dynan and Tjian, 1985).

In addition to the TATA and upstream elements found in most promoters of higher eukaryotes, there are other control elements that profoundly affect the level of initiation by RNA polymerase II (Serfling *et al.*, 1985). These elements, called enhancers, are relatively complex and contain repeated sequences that are functionally independent (Ondek *et al.*, 1988). These sequences can

display unique cell type specificities and responsiveness to stimuli, which contribute to the overall character of the enhancer. Enhancers are in many ways similar to yeast UASs (Guarente, 1988). They can function in both orientations and at long and variable distances upstream of the mRNA start site. However, the distances over which they can influence transcription greatly exceed those of UASs-effects are seen when the elements are placed several thousand base pairs away from the start site. Additionally, enhancers need not be

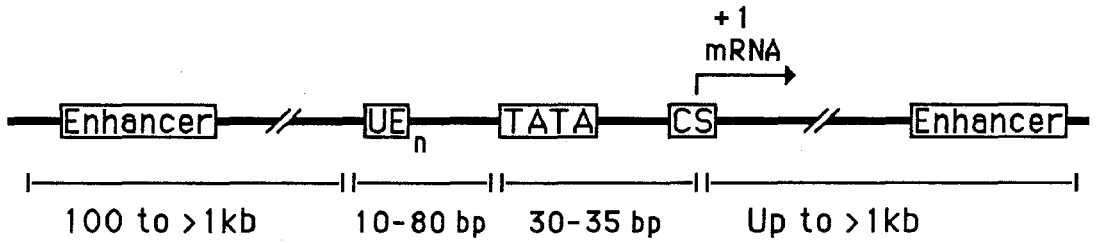


Figure 3. Schematic representation of a mammalian polymerase II control region. Shown are the relative positions of enhancer, upstream (UE), TATA, and cap site (CS) elements. See text for details.

positioned upstream of the start site in order to influence transcription-they can be placed within or downstream of the transcribed region.

Thus, the control regions of protein coding genes of higher eukaryotes are made up of a patchwork of both upstream elements and enhancers. A given gene's expression characteristics are

determined by the control elements contained within its regulatory region (Maniatis *et al.*, 1985). Those genes that respond to the same inductive signals contain the same regulatory sequences (Bienz and Pelham, 1986; Angel *et al.*, 1987).

The existence of sequence elements essential to transcription initiation in higher eukaryotes suggests, as in yeasts and prokaryotes, the presence of protein factors that control this process by binding to these elements. A large number of proteins, which specifically bind to upstream elements, enhancers, and the TATA box have been identified in protein extracts from animal tissues and tissue culture cells (Maniatis *et al.*, 1987; McKnight and Tjian, 1986; Guarente, 1988). Many of these factors have been extensively purified and shown to produce sequence-specific transcriptional enhancement when added to soluble cell extracts capable of accurate *in vitro* initiation (Parker and Topol, 1985; Sawadogo and Roeder, 1985; McKnight and Tjian, 1986; Lee *et al.*, 1987), thus verfying that they are bonafide transcription factors. The availability of highly purified factors along with *in vitro* transcription systems has allowed the investigation of the mechanistic and biochemical steps involved in transcription initiation (Parker and Topol, 1984; Reinberg and Roeder, 1980; Van Dyke *et al.*, 1988).

The rate at which the knowledge concerning the mechanisms of eukaryotic transcription has advanced has lagged behind that of prokaryotic transcription. The cloning of a number of genes that encode prokaryotic transcription factors (from both bacteria and

bacteriophage) has made possible a detailed analysis of the structural domains resposnsible for their biochemical functions (Ptashne, 1986). However, for technical reasons, only a small number of the genes encoding eukaryotic transcription factors have been cloned. Analysis of the functional domains of these proteins has revealed a number of conserved amino acid sequences, which appear to be involved in DNA binding. These sequences are different from those seen in bacterial DNA-binding proteins and presumably represent completely distinct structural motifs (Evans and Hollenberg, 1988). X-ray crystallographic studies are now possible because of the quantities of protein that can be produced using the cloned gene. Analyses of the amino acid composition of the domains responsible for transcriptional activation (separate from DNA binding) have been performed (Hope *et al.*, 1988). The results of these suggest amphipathic α-helices are responsible for the ability of at least some factors to activate transcription (Giniger and Ptashne, 1987).

Recent experiments have shown that cloned yeast transcription factors expressed in mammalian cells can activate transcription from a mammalian promoter containing the yeast UAS (Kakidani and Ptashne, 1988; Webster *et al.*, 1988). Similar experiments have shown that mammalian transcription factors can function in yeast (Metzger *et al.*, 1988). These results point to a remarkable degree of mechanistic conservation in eukaryotic transcription in addition to

the basic principles of protein-protein and protein-DNA interaction, which appear to apply across species boundaries.

The work described in this thesis addresses issues of both mechanistic diversity and conservation in the process of transcription initiation by RNA polymerase II. Each of the three chapters deals with aspects of proteins that bind DNA sequences in the regulatory regions of the mammalian tumor virus SV40. Double-stranded DNA viruses that infect animal cells, like SV40, rely heavily on the host enzymatic activities to express their genetic program. Additionally, a great deal is known about their genetics and molecular biology. Thus, these viruses provide a convenient and widely used experimental system in which to study mammalian transcription. The sequences in question, which are located in the enhancer of the early gene control region, were first identified on the basis of their ability to influence transcription from this promoter. They have since been found to be functional components of many cellular control regions. Chapter 1 describes the purification and characterization of a protein isolated from bovine thymus tissue, which binds to a closely related group of elements called the core and pseudo-core sequences. The range of sequences recognized by this factor is broader than that of similar factors isolated from other mammalian tissues. The experiments described in Chapter 2 demonstrate that a sequence, which mediates factor-dependent transcriptional activation in mammalian cells, is also capable of

activating transcription in yeast. The purification of a yeast nuclear protein, which binds to this sequence, is described as well as a comparison of the physical and DNA-binding characteristics of the yeast protein and a similar mammalian protein. Chapter 3 describes the molecular cloning of the gene encoding the yeast protein described in Chapter 2. The DNA and protein sequences of this yeast protein and related yeast and mammalian proteins are compared.

References

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

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

Bienz, M., and Pelham, H. (1986) *Cell* 45, 753-761.

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

Brent, R. (1985) *Cell* 49, 295-297.

Buratowski, S., Hahn, S., Sharp, P., and Guarente, L. (1988) *Nature* 334, 37-42.

Burton, A., Ortoalan, G., and Greenblatt, J. (1986) *EMBO J.* 5, 2923-2930.

Chamberlin, M. (1974) *Ann. Rev. Biochem.* 43, 721-775.

Chamberlin, M. (1976) in *RNA Polymerase* (R. Losick and M. Chamberlin, eds.) pp. 159-192 New York: Cold Spring Harbor Press.

Dynan, W., and Tjian, R. (1985) *Nature* 316, 774-778.

Dudler, R., and Travers, A. (1984) *Cell* 39, 391-398.

Evans, R., and Hollenberg, S. (1988) *Cell* 52, 1-3.

Goodbourn, S., Burstein, H., and Maniatis, T. (1986) *Cell* 45, 601-610.

Grossman, A., Erikson, V., and Gross, C. (1984) *Cell* 38, 383-390.

Grosveld, G., Rosenthal, A., and Flavell, R. (1982) *Nucl. Acid Res.* 10, 4951-4971.

Guarente, L. (1987) *Ann. Rev. Genet.* 21, 425-452.

Guarente, L. (1988) *Cell* 52, 303-305.

Hansen, U., Mclure, W. (1980) *J. Biol. Chem.* 255, 9564-9570.

Helman, J., and Chamberlin, M. (1988) *Ann. Rev. Biochem.* 57, 839-872.

Hope, I., and Struhl, K. (1985) *Cell* 43, 177-188.

Hope I., Mahadevan, S. and Struhl, K. (1988) *Nature* 333, 635-640.

Igo-Kemenes, K., Worz, W., and Zachau, H. (1982) *Ann. Rev. Biochem.* 51, 89-121.

Jacob, F., and Monod, J. (1961) *J. Mol. Biol.* 3, 318-356.

Johnson, A., and Herskowitz, I. (1985) *Cell* 42, 237-247.

Kadonaga, J., Carner, K., Masiarz, F., and Tjian, R. (1987) *Cell* 51, 1079-1090.

Kakidani, H., and Ptashne, M. (1988) *Cell* 52, 161-167.

Lee, W., Haslinger, A., Karin, M., and Tjian, R. (1987) *Nature* 325, 368-372.

Losick, R., and Pero, J. (1981) *Cell* 25, 582-584.

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

Metzger, D., White, J., and Chambon, P. (1988) *Nature* 334, 31-36.

McClure, W. (1985) *Ann. Rev. Biochem.* 54, 171-204

Ondek, B., Gross, L., and Herr, W. (1988) *Nature* 333, 40-45.

Pabo, C. (1983) *Nature* 301, 20-21.

Pabo, C., and Sauer, R. (1984) *Ann. Rev. Biochem.* 53, 293-321.

Parker, C., and Topol, J. (1984) *Cell* 37, 273-283.

Patshne, M. (1986) *A Genetic Switch* Cell Press, Palo Alto and Blackwell Publishing.

Ptashne, M. (1988) *Nature* 325, 683-689.

Reinberg, D., and Roeder, R. (1987) *J. Biol. Chem.* 262, 3310-3321.

Sawadogo, M., and Roeder, R. (1985) *Cell* 43, 165-175.

Schmitz, A., and Galas, D. (1979) *Nucleic Acid Res.* 6, 111-137.

Sentenac, A., and Hall, B. (1982) in *Molecular Biology of the Yeast Saccharomyces* (Strathern, J., Jones, E., and Broach, J., eds.) pp. 561-606 New York: Cold Spring Harbor Press.

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

Siebenlist, U., Simpson, R., and Gilbert, W. (1980) *Cell 20,* 269-281.

Simpson, R. (1979) *Cell* 18, 277-285.

Strathern, J., Jones, E., and Broach, J., eds. (1982) *The Molecular Biology of the Yeast Saccharomyces.* New York: Cold Spring Habor Press.

Struhl, K. (1982) *Proc. Natl. Acad. Sci. USA.* 79, 7385-7389.

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

Tjian, R., Losick, R., Pero, J., and Hinnebusch, A. (1977) *Eur. J. Biochem.* 74, 149-154.

Van Dyke, M., Roeder, R., and Sawadogo, M. (1988) *Science* 241, 1355-1338.

Webster, N., Rui Jin, J., Green, S., Hollis, M., Chambon, P. (1988) *Cell* 52, 169-178.

Wiederrecht, G., Seto, D., and Parker, C. (1988) *Cell* 54, 841-853.

Yousdarian, P., Bouvier, S., and Susskind, M. (1982) *Cell* 30, 843-853.

Chapter 1

Identification and Purification of a Bovine Thymus Nuclear Protein that Interacts with the SV40 Core and Psuedo-Core Enhancer Sequences

Identification and Purification of a Bovine Thymus Nuclear Protein that Interacts with the SV40 Core and Psuedo-Core Enhancer Sequences

Keith D. Harshman, Steve Sogo and Carl S. Parker
From the Division of Chemistry, California Institute of Technology, Pasadena, California 91125

Contribution #7369

(Submitted to *The Journal of Biological Chemistry*)

A bovine thymus nuclear protein designated enhancer protein 2 (EP2) has been identified and its DNA binding characteristics determined. Chemical and enzymatic footprinting identified as binding sites the SV40 core sequence as well as two pseudo-core sequences distal to the 72 base pair repeat as EP2 binding sites. A double point mutation in the core sequence, which abolishes enhancer dependent transcriptional activation *in-vivo* is not bound by EP2 as determined by DNase I footprinting . EP2 was purified to apparent homogeneity by sequence-specific DNA affinity chromatography. It consists of a group of polypeptides ranging in molecular weight from 34 kDa to 43 kDa. Following preparative SDS gel electrophoresis, the polypeptides were individually eluted, renatured and shown to produce identical DNase I footprints on each of the core and pseudo-core sequences present in the SV40 enhancer. This is in contrast to previously described enhancer binding factors, which interact with only single members of the SV40 core and psuedo-core cis-acting elements.

Enhancers are cis-acting DNA sequences that can stimulate transcription of an adjacent promoter from relatively great distances in an orientation independent fashion (1, 2). Although first identified as components of viral promoters (3, 4), they have since been found to be associated with many cellular genes (5-8). The most thoroughly characterized eukaryotic enhancer element is that found in the DNA tumor virus SV40 (9, 10). Detailed mutational analysis has shown this enhancer to be composed of a number of sequence motifs, which stimulate transcription synergistically (11). The individual elements or motifs display distinct cell-type specificity; the cell-type specificity of the entire enhancer is the sum of the specificities of the individual motifs (12-14). Although the exact mechanism(s) by which enhancers stimulate transcription is unclear, the results of *in vivo* and *in vitro* experiments have shown that the elements act as binding sites for trans-acting factors and that enhancer activity is dependent on these trans-acting factors (15, 16). A number of the proteins that bind to various sequence elements found in the SV40 and other enhancers have been identified and purified on the basis of *in vitro* binding studies (17, 18).

The motif within the 72 base pair repeat of the SV40 enhancer, referred to as the core sequence, is also found in a number of other viral and mammalian enhancers (19). This sequence, TGTGGA/TA/TA/TG, has been shown to be important for SV40

enhancer function (11) and exhibits enhancer activity in a wide range of cell types (12, 14). Interestingly, two different enhancer elements nearly identical to the core sequence exhibit distinct ranges of cell-type specific activities, which overlap that exhibited by the core (12). Proteins identified in extracts from animal tissue, and tissue culture cells that bind to the core sequence, exhibit no binding affinity for these two 'psuedo-core' sequences (20, 21). Likewise, proteins have been isolated from tissue culture cells, which bind specifically to a given pseudo-core sequence but not to the core or the other pseudo-core sequence (32).

We will describe the identification, characterization and purification of a group of polypeptides from bovine thymus nuclear extracts that bind to the core sequence within the 72 base pair repeat and the upstream two pseudo-core sequences. After extensive purification using sequence-specific oligonucleotide chromatography, these binding activities have been assigned to a set of polypeptides ranging in molecular weight from 25 to 46 kDa. Interestingly, each of these polypeptides appears to be able to bind to both of the upstream pseudo-core sequences as well as the core sequence within the 72 base pair repeat.

MATERIALS AND METHODS

Bovine Thymus Extract Preparation-Freshly obtained bovine thymus glands were transported on ice from the slaughter house, and either used immediately or stored at -80°C. All steps of the

extract procedure were performed at 4°C. The bovine thymus (1 Kg) tissue was blended at low speed in 3 l of A buffer [15 mM KCl, 10 mM Hepes (pH 7.6), 2 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 1 mM PMSF (Sigma)] for 1 min, followed by 30 sec at medium speed and 30 sec at high speed. The nuclei were then collected by centrifugation at 10,000 rpm for 10 min in a Beckman JA10 rotor. The nuclear pellet was resuspended in 200 ml of A buffer plus 10 mM 1,10-phenanthroline (Sigma) and benzamidine hydrochloride (Sigma) by blending at low speed for 10 sec. The nuclei were lysed by the addition of 310 ml (1/10 volume) of 4 M ammonium sulfate (pH 7.9). At this stage diisopropyl-fluorophosphate (Sigma) was added to 0.4 mM. The DNA was sedimented from the lysate by centrifugation at 35,000 rpm for 90 min in a Beckman 45 Ti rotor. The clear supernatant was collected and the protein precipitated by the addition of 0.25 g ammonium sulfate per ml of supernatant. After mixing for 30 min, the precipitated proteins were collected by centrifugation at 15,000 rpm for 15 min in a Beckman JA10 rotor. The supernatant was discarded and the pellet was resuspended in 150 ml C buffer [10% glycerol, 50 mM Hepes (pH 7.6), 40 mM KCl, 0.1 mM EDTA, 0.1 mM DTT]. The protein suspension was subsequently dialyzed against C buffer for 8 hr. The insoluble protein after dialysis was removed by centrifugation at 10,000 rpm for 10 min in a Beckman JA10 rotor. The supernatant was either stored at -80°C or immediately subjected to the chromatographic techniques described below.

Chromatographic Purification of EP2-The standard chromatography buffer was HGED, which consists of 25 mM Hepes (pH 7.6), 10% (v/v) glycerol, 0.1 mM EDTA, 1 mM dithiothreitol. KCl was added and its molarity indicated as follows: 0.1 HGKED is 0.1 M KCl in the otherwise standard buffer. The protein concentrations were determined by the Bradford procedure with BSA as the standard (23). DEAE-cellulose (Whatman DE52), and BioRex 70 (Bio-Rad) resins were prepared according to the manufacturer's recommendations. Heparin-Sepharose was prepared following the procedure of Davidson *et al.* (24).

A 300 ml DE52 column was packed and equilibrated with 0.1 HGKED. 150 ml of bovine thymus nuclear extract was loaded onto the column and washed with 900 ml 0.1 HGKED. The proteins that flowed through the DE52 column were passed directly onto a 90 ml heparin-Sepharose column equilibrated with 0.1 HGKED. Following a 300 ml 0.1 HGKED wash, the heparin-Sepharose bound proteins were eluted with a 5-column volume linear salt gradient from 0.1 to 0.8 HGKED. Protein-containing fractions were dialyzed *vs.* 0.1 HGKED. In this and subsequent steps the activity was assayed by DNase I footprinting. Active fractions were pooled and loaded onto a 30 ml BioRex 70 column. Bound protein was eluted with 0.7 HGKED. The protein sample was then loaded onto a column (900mm x 30mm) containing 500 ml AcA44 gel filtration resin equilibrated with 0.7 HGKED. The protein was chromatographed in the same buffer. The active fractions were pooled, dialyzed against 0.25 HGKED and slowly

loaded onto a 0.8 ml EP2 oligonucleotide-Sepharose column equilibrated with the same buffer. The column was prepared according to the procedure of Wiederrecht *et al.* (25). The sequence of the synthetic oligonucleotide used in the column preparation was:

strand 1 5' - GATCCGTAGGTGTGGAAAGTCCCCA - 3'

strand 2 5' - GATCTGGGGACTTTCCACACCTACG - 3'

After loading, the resin was washed with 50 column volumes of 0.27 HGKED. The bound protein was eluted from the resin with a 5 column volume linear salt gradient from 0.27 to 1.5 M HGKED. The active fractions were pooled, dialyzed, and subjected to one additional round of affinity chromatography. The active fractions were pooled, dialyzed against C buffer, and stored at -80°C.

DNase I Footprinting Reactions-The general DNase I footprinting reaction was as follows: reactions, in a final total volume of 50 µl, contained 1-3 ng of labeled fragment, 25 µg/ml freshly denatured salmon sperm DNA, 5 mM $MgCl_2$ and the indicated volume of EP2. After a 20 min 0°C binding reaction, DNase I (Cooper) was added to 50 µg/ml. After 30 sec at 0°C, the DNase I digestion was terminated by the addition of 100 µl of a termination buffer [1% sarkosyl (w.v), 0.1 M Tris (pH 7.8), 0.1 M NaCl, 0.01 M EDTA, 25 µg/ml salmon sperm DNA, 100 µg/ml proteinase K]. The samples were incubated at 37°C for 15 min, followed by a 90°C incubation for 2 min. The samples were phenol extracted, ethanol precipitated, washed, and dried under vacuum. Samples were

electrophoresed according to the procedures of Maxam and Gilbert (26).

MPE·Fe(II) Footprinting Reactions-MPE·Fe(II) footprinting reactions were conducted as follows: reactions were carried out in 10 mM Tris (pH 7.4) and 50 mM NaCl in a final reaction volume of 30 µl. Each reaction contained 20 µl purified EP2, 1-3 ng end-labeled DNA fragment (12,000-15,000 cpm) and 25 µg/ml ng freshly denatured salmon sperm DNA. After a 10 min binding reaction at 22°C, 1.5 µl 100 mM MPE·Fe(II) was added and incubated at 22°C for 3 min. Cleavage was initiated by addition of 1 µl 40 mM dithiothreitol and continued for 10 min at 22°C. Cleavage was stopped by the addition of 1 µl 0.6 M EDTA. After butanol extraction, 100 µl of termination buffer was added and sample treatment continued as described for DNase I footprinting reactions.

Methylation (Dimethyl Sulfate) Protection Reactions-- Binding reactions were 5mM $MgCl_2$ in a final volume of 25µl. Reactions contained 20 µl of EP2 and were conducted for 5 min. DMS was added to a final concentration of 5 mM. The mixture was incubated at 22°C for 5 min. The DMS reaction was terminated by addition of 6 µl stop solution (1.5 M sodium acetate (pH 7.0), 1.0 M mercaptoethanol, 100 µg/ml tRNA), followed by the addition of 50 ml of sarkosyl solution [1% sarkosyl (w/v), 0.1 M Tris (pH 7.8)], 0.1 M NaCl, and 0.01 M EDTA). After phenol and ether extractions, the DNA was precipitated with ethanol. The DNA pellet was resuspended in 20 µl TE (10 mM Tris (pH 7.8), 1 mM EDTA). Base cleavage was

accomplished by addition of NaOH to a final concentration of 0.15 M and heating at 90°C for 20 min. The DNA was precipitated with ethanol, dried and electrophoresed according to the procedures of Maxam and Gilbert (26).

SDS Gel Electrophoresis and Renaturation-Fractions containing EP2 binding activity after the second round of affinity chromatography were combined. Protein was precipitated by the addition of 3 volumes of acetone followed by incubation for 30 min in a dry ice-ethanol bath. The protein was pelleted by centrifugation at 15,000 rpm (Beckman JA20 rotor) for 30 min at 0°C. The pellet was vacuum dried and then resuspended in an SDS protein gel loading buffer. The sample was applied to a 10% SDS polyacrylamide gel (27) and electrophoresed at 250 volts for 3 hr at 4°C. The gel was then soaked for 5-10 min in 0.1 M KCl. At the appearance of the protein-SDS-potassium precipitate, the bands were excised from the gel and placed into small dialysis bags along with 700 ml SDS gel running buffer. The protein was electroeluted from the gel slice for 3 hr at 250 volts. The sample was then dialyzed against 0.05 HGKED for 3 hr to remove glycine. The protein was precipitated as before and washed twice with ice cold 80% acetone to remove residual SDS. The pellet was vacuum dried and resuspended in 300 µl C buffer containing 6 M guanidine hydrochloride. The sample was dialyzed overnight against C buffer to remove the guanidine hydrochloride, allowing protein renaturation.

RESULTS

Identification and Partial Purification of Enhancer Binding Proteins-Bovine thymus was selected as the initial starting material for the identification and purification of enhancer binding proteins because of its relative low cost and abundance. It was not, however, intrinsically obvious that bovine thymus would be a good source for enhancer binding proteins. An extract was prepared from freshly obtained bovine thymus following the protocol described in the Materials and Methods. A direct analysis of the crude extract failed to reveal the presence of proteins that could specifically bind to the SV40 enhancer as determined by DNase I footprinting experiments. The extract was applied to a DEAE-cellulose column and the flow through applied directly to a heparin-Sepharose column and the bound proteins were eluted with a linear salt gradient. At this point it is possible to see (by DNase I footprinting) specific interactions in two regions of the SV40 enhancer (Fig. 1, panel A). These two regions are approximately 20 base pairs in size and are positioned over the 5' and 3' ends of the SV40 72 base pair repeat. The binding activity that elutes first from the heparin-Sepharose column was designated EP2. This protein binds to the distal end of the SV40 72 base pair repeat relative to the 21 base pair repeats and the origin of replication. Included in this binding site is the core consensus sequence (27). The other binding activity elutes at a higher salt concentration from the heparin-Sepharose resin and was designated

EP1. EP1 binds to a 20 base pair region of the SV40 enhancer proximal to the 21 base pair repeats, which contain the recognition sequence for the mammalian transcription factor AP-1 (28). EP2 was subsequently purified to apparent homogeneity by AcA44 chromatography and sequence-specific DNA-affinity chromatography as shown in Fig. 1 and described in detail below. A flow diagram describing the purification of EP2 is shown in Fig. 2.

Detailed Analysis of EP2 Interactions with the SV40 72 Base Pair Repeat-EP2 binding to the SV40 72 base pair repeat has been examined in detail. DNase I protection (37) on the bottom strand as a result of EP2 binding lies between residues 163 and 180 (Fig. 3A, lane 8). Protection on the top strand is observed between residues 163 and 175 (Fig. 3B, lane 8). It is interesting to note that the core sequence is situated asymmetrically to the distal side of the DNase I footprint. The core sequence is also protected in DNase I footprinting experiments conducted with EP2 preparations on DNA templates derived from the IgG gene heavy chain enhancer (5) and the polyoma virus enhancer (40) (data not shown). DMS protection studies (38) on the EP2 sequence reveal two enhanced guanine cleavages at 163 and 164 and one reduced guanine cleavage at 165 of the bottom strand (Fig. 3A, lane 10). Mutations of residues 163 and 164 have little effect on enhancer function *in vivo* (27). EP2 interactions determined by DMS (38) on the top strand consist of two strongly suppressed methylations at guanines 172 and 173 and a weakly enhanced guanine reaction at residue 168 (Fig. 3B, lane 10).

The two guanine residues at positions 172 and 173 are critical for *in vivo* enhancer activity as shown by Herr and Clark (29).

A third reagent, MPE (methidium propyl-EDTA) was used to determine more precisely the boundaries of the EP2 recognition sequence. MPE's relatively small size allows it to bind and cleave DNA nearer to the sites of protein-DNA interaction than DNase I (30, 31). MPE will also cleave DNA with less sequence-specificity than DNase I, so one obtains a very uniform cleavage ladder, allowing a more precise determination of the boundaries of protein binding. MPE footprints with partially purified EP2 were performed on both strands of the SV40 enhancer and are shown in Fig. 3 (panel A, lane 3; panel B, lane 2). A small region of protection is observed over nine residues on the bottom strand between positions 167 and 175, inclusive. On the top strand, an eight base pair region of protection is observed between residues 165 and 172. As expected, this small protected region includes the core consensus sequence (27).

EP2 was also observed to protect a region outside of the 72 base pair repeats, between bases -263 and -289. The protected sequences can be seen in the column profiles of Fig. 1 as well as in Figs. 4, 5 and 6. Contained within this region of DNase I protection are two sequences-GCGGAAAG and GTGGAATG-which are very similar to the core sequence. These EP2 sites correspond to the GT-IIA and GT-IIC sites bound by proteins found in HeLa Cell and BJA-B cell extracts (32). A systematic mutagenesis screen through the SV40 enhancer by Chambon and colleagues has shown that all of the

EP2 sites are important for enhancer activity *in vivo*. The relative positions of all the EP2 binding sites determined in our footprint assays are shown in Fig. 3C.

A Double-Point Mutation in the Core Sequence Prevents EP2 Binding-Herr and Clarke (29) have shown that a double point mutation in the core consensus sequence (dpm 6) greatly reduces enhancer function *in vivo*. If EP2 is a protein that in some way mediates enhancer function, then one might predict that such a double point mutation would reduce EP2 binding. Shown in Fig. 4 is a titration of partially purified EP2 on the wild-type SV40 enhancer (lanes 1-3) and on the double point mutation (lanes 6-9). It is evident that EP2 is unable to bind to the sequence containing the point mutations at levels that fully protect the wild-type sequence. It is noteworthy that the two alterations in dpm 6 correspond to guanine residues 172 and 173 (wild-type GTGGAAAG to dpm 6 GTCCAAAG). These residues were shown by DMS footprints to make close contact to EP2 (Fig. 3B, lane 10) and are also included within the MPE footprint (Fig. 3). Evidence showing the *in vivo* importance of guanine 172 for SV40 enhancer function has been described by Weiher *et al.* (27). In this case, mutation of guanine 172 to adenine completely abolished enhancer activity. As mentioned in the previous section, EP2 binding sites exist outside of the 72 base pair repeat; these sites serve as an internal control for EP2 binding activity for the dpm 26 template used in the experiment shown in Fig. 4.

Purification of EP2 to Apparent Homogeneity-The EP2 binding activity was purified to near homogeneity using a combination of gel filtration and affinity chromatography. Gel filtration was performed using an AcA44 resin. The chromatographic profile of EP2 activity on this resin is shown in Fig. 2B. The activity profile was bimodal with a small amount of EP2 activity eluting in the void volume of the column while greater than 80% of the EP2 activity was well included within the column. The included activity has an apparent molecular weight between 50 and 30 kDa (fractions 32 through 43). These fractions were pooled and concentrated on BioRex 70 prior to DNA-affinity chromatography. (The low molecular weight DNase activity that elutes between fractions 44 through 68 was discarded.)

A DNA-affinity resin was constructed by concatamerizing a synthetic oligonucleotide containing the core consensus sequence and coupling the concatamers to cyanogen bromide-activated Sepharose (25). The active fractions from the BioRex 70 concentration step were slowly applied to the column. Next, the column was washed extensively with buffer at a salt concentration high enough to elute the majority of non-specifically bound proteins from the resin, while the core and pseudo-core (EP2-site) specific binding protein(s) remain absorbed. The tightly bound proteins were eluted with a salt gradient from 0.27 to 1.5 M KCl. To complete the purification, it was necessary to repeat the affinity chromatography step to remove minor contaminating proteins. DNase I footprint assays of the fractions eluted from the first passage over the affinity resin are

shown in Fig. 2C. The binding activity elutes between approximately 0.35 M KCl and 1.1 M KCl. DNase I footprint assays of fractions from the second passage are shown in Fig. 5A. Analysis of the protein composition of the active fractions by SDS polyacrylamide gel electrophoresis is shown in Fig. 5B. Several polypeptides are observed, including a major species at 39 kDa, and two other species with molecular weights of 43 kDa and 41 kDa. Additionally, a series of evenly spaced polypeptides ranging in molecular weight from 34 to 38 kDa are observed.

To identify which of these polypeptides is responsible for the specific binding activity observed, we performed a preparative SDS protein gel purification step. The protein bands were visualized by KCl staining, individually excised and electrophoretically eluted from each gel slice. The eluted proteins were further denatured with guanidine-HCl and renatured by slow dialysis (25). Each eluted and renatured protein was subsequently analyzed on an SDS protein gel to measure recovery and purity (Fig. 6, panel B; note that the 70 kDa band present in each lane is BSA added after electroelution to aid in protein recovery during the renaturation procedure). In this particular experiment two bands from the 34-38 kDa cluster as well as the 43 kDa band were not efficiently recovered. The binding activity from each of the isolated polypeptides is shown in Fig. 6A. Footprinting reactions were conducted with 5 µl and 40 µl from each fraction. Footprints were observed with each of those polypeptides recovered in good yield from the gel. In other experiments, the 43

kDa polypeptide has been shown to bind to the core sequence in a fashion identical to the other polypeptides (data not shown). As seen in Fig. 6A, each of the gel-eluted, renatured polypeptides is capable of interacting with the core sequence and both of the pseudo-core sequences upstream of the 72 base pair repeat. The boundaries of the footprints are identical for each polypeptide and correspond to the boundaries observed when partially and affinity purified EP2 preparations were used in the footprinting reactions (Figs. 4 and 5).

DISCUSSION

The SV40 enhancer has proven to be an amalgam of sequence motifs, which can act in homologous or heterologous combinations to stimulate transcription (13, 33). The cell-type specificity of this enhancer has been shown to be the sum of the specificities exhibited by each of the individual elements, which constitute the enhancer (12, 13). These sequence motifs are thought to potentiate transcription by functioning as binding sites for trans-acting factors (15). Interestingly, a number of these sequence elements act as binding sites for multiple trans-acting factors (17). A number of these factors have been purified (17); some of these purified factors have been shown to enhance *in vitro* transcription in a sequence-dependent fashion (21, 28). Here we have described the identification and purification of a set of polypeptides from bovine thymus tissue, which bind to sequences within the SV40 enhancer as

well as other viral and mammalian enhancers. We have designated this activity as enhancer protein 2 or EP2. A comparison of the sequences bound by EP2 preparations in the SV40 enhancer (Fig. 3A and B, Figs. 5 and 6) as well as the IgG heavy chain enhancer and the polyoma virus enhancer, reveal the common presence of the core sequence and sequences nearly or exactly homologous to the core. Further support for the assignment of the core nucleotides as the EP2 recognition sequence comes from the results of DMS and MPE footprinting studies. The core sequence lies symmetrically within the boundaries of both the DMS and MPE footprints produced by EP2 (Fig. 3A, B). Nucleotide residues included in the DNase I and DMS footprints on the core sequence include some to the 3' side of the core sequence, TGTGGAAAG, defined on the basis of point mutation and sequence comparison (27). These residues, however, are included in the larger functionally defined core sequence, TGTGGAAAGTCCCCA (34).

Our results suggest that EP2 is a protein, which is distinct from other recently identified proteins that bind to the SV40 core and pseudo-core motifs. The HeLa cell derived core-binding protein, AP-3 (21) differs from EP2 in that it is incapable of binding to the pseudo-core motifs in the SV40 enhancer. Similarly, the rat liver-derived factor C/EBP (originally referred to as EBP20) has been shown to bind with high affinity to the SV40 core element, but not to the pseudo-core enhancer elements (20). EP2 and C/EBP produce significantly different DMS footprints over the core motif as well.

Neither TEF-1, which binds to the distal pseudo-core (GTIIC; 35) nor GTIIA, which binds to the proximal pseudo-core (GTIIA; 32) bind the neighboring core or pseudo-core elements. This does not, however, preclude the possibility that any or all of these proteins are genetically related to EP2. The fact that these proteins recognize DNA sequences bound by EP2 certainly suggests some degree of amino acid and/or nucleic acid homology. The cloning of the genes that encode these proteins (the gene encoding C/EBP has recently been cloned; 36) and a comparison of the DNA and protein sequences will provide the definitive answer to these questions.

We presently have no direct evidence showing an EP2-mediated effect on transcription in thymus tissue. The involvement of this factor in transcriptional regulation is suggested by the results of DNase I and MPE footprinting experiments (Figs. 3 and 5). EP2 interacts specifically with sequences that have been shown by mutational analysis to be important for SV40 enhancer function (11, 27, 29). Additionally, multimerized copies of the core sequence and the promoter-distal pseudo-core stimulate the transcription of a recorder gene in transient expression assays (12). Involvement in transcription regulation is further suggested by the results of the methylation protection experiments shown in Fig. 3. Guanine residues 173 and 172, which show a decreased reactivity toward DMS in the presence of EP2 (Fig. 3B), suggest an important contact between these nucleotide residues and the protein, and require maximal SV40 enhancer function *in vivo* (11). Similarly, guanine

residues 163, 164 and 165 show a change in reactivity toward DMS due to EP2 binding and have been shown to be important for *in vivo* SV40 enhancer function by analysis of mutations (11). Finally, Fig. 4 shows that a double point mutation in the core sequence which abolishes SV40 enhancer activity *in vivo* (29) fails to be bound by EP2. Studies are currently in progress in this laboratory to determine directly whether EP2 can stimulate transcription in a core-sequence-dependent fashion from RNA polymerase II promoters *in vitro*.

Following chromatographic purification, which included two cycles over a core-sequence DNA-Sepharose affinity column, EP2 activity identity could by assigned to a small cluster of polypeptides--a major polypeptide at 41 kDa, two somewhat less abundant polypeptides at 43 and 39 kDa, and a still less abundant set of polypeptides with evenly spaced molecular weights ranging from 38 to 34 kDa (Fig. 5B). Gel elution/renaturation experiments show that each of these polypeptides can bind the core and the pseudo-core sequences (Fig. 6). Furthermore, the individually gel eluted and renatured polypeptides produce identical DNase I footprints over these DNA sequences. These results, in addition to their identical chromatographic behavior, suggest thatthese polypeptides are closely related and are perhaps the products of the proteolytic breakdown of a single larger protein. It is interesting to note the small amount of EP2 activity in the high molecular weight fraction that eluted in the void volume of AcA44 gel filtration resin. Assuming that this EP2

activity is not an aggregate of lower molecular species, this could be the parent protein of the lower molecular weight polypeptides visible in the affinity purified fractions. Alternatively, the high and low molecular weight activities may represent two or more different gene products (intact or degraded), which bind to the core sequence.

Sequence-specific DNA affinity chromatography has facilitated the homogeneous purification from bovine thymus nuclear extracts of a set of polypeptides, which bind to the core and pseudo-core elements of the SV40 enhancer. Homogeneous preparations of this binding activity will allow the determination of a partial amino acid sequence and preparation of antibodies against the protein(s); both endeavors are currently being undertaken in this laboratory. The antibodies will be useful in studying the localization and potential modifications that might occur to the sequence-specific binding proteins. The partial amino acid sequence will also facilitate cloning the gene for the factor by allowing synthesis of appropriate oligonucleotide probes to screen a recombinant DNA library.

ACKNOWLEDGMENTS

We thank Winship Herr for the gift of the dpm 6 and 1 x 72 SV40 clones and Peter Dervan for the gift of MPE. We are indebted to the members of the Parker lab for their valuable comments on this manuscript. We also are grateful to Phoebe Ray for her expert preparation of the manuscript. CSP is a Rita Allen Foundation Scholar. This research was supported in part by NIH Biomedical

Research Support Grant RR07003.

REFERENCES

1. Gluzman, Y. (ed.) (1985) *Eukaryotic Transcription - The Role of cis and trans-Acting Elements in Initiation,* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

2. Wasylyk, B. (1986) in *Maximizing Gene Expression* (Reznikoff, W., and Gold, L., eds), pp. 79-99, Butterworths, Boston.

3. Banerji, J., Rusconi, S., and Schaffner, W. (1981) *Cell* 27, 299-308.

4. Moreau, P., Ritten, B., Wasylyk, B., Everett, R., Gaub, M. P., and Chambon, P. (1981) Nucl. Acids Res. 9, 6047-6068.

5. Banerji, J., Olson, L., and Schaffner, W. (1983) Cell 33, 729-740.

6. Gilles, S. D., Morrison, S. L., Oi, V. T., and Tonegawa, S. (1983) Cell 33, 717-728.

7. Queen, C., and Baltimore, D. (1983) Cell 33, 741-748.

8. Walker, M. D., Edlund, T., Boulet, A. M., and Rutter, W. J. (1983) Nature, 306, 557-561.

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

10. Khoury, G., and Gruss, P. (1983) Cell 33, 313-314.

11. Zenke, M., Grundstrom, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A., and Chambon, P. (1986) EMBO J. 5, 387-397.

12. Schirm, S., Jiricny, J., and Schaffner, W. (1987) Genes and Develop. 1, 65-74.

13. Fromental, C., Kanno, M., Nomiyama, H., and Chambon, P. (1988)

Cell 54, 943-953.

14. Ondek, B., Shepard, A., and Herr, W. (1987) EMBO J. 6, 1017-1025.
15. Scholer, H. R., and Gruss, P. (1984) Cell 36, 403-411.
16. Sassone-Corsi, P., Dougherty, J., Wasylyk, B., and Chambon, P. (1984) Proc. Natl. Acad. Sci. USA 81, 308-312.
17. Wildeman, A. (1988) Biochem. Cell Biol. 66, 567-577.
18. Jones, N., Rigby, P., and Ziff, E. (1988) Genes and Develop. 2, 267-281.
19. Sassone-Corsi, P., and Borrelli, E. (1986) Trends Genet. 2, 215-219.
20. Johnson, P., Landshulz, W., Graves, B., and McKnight, S. (1987) Genes and Develop. 1, 133-146.
21. Mitchell, P., Wang, C., and Tjian, R. (1987) Cell 50, 847-861.
22. Chiu, R., Imagawa, M., Imbra, R., Bockoven, J., and Karin, M. (1987) Nature 329, 648-651.
23. Bradford, M. M. (1976) Anal. Biochem. 72, 248-254.
24. Davidson, B. L., Leighton, T., and Rabinowitz, J. C. (1979) J. Biol. Chem. 254, 9220-9226.
25. Wiederrecht, G., Shuey, D. J., Kibbe, W. A., and Parker, C. S. (1987) Cell 48, 507-515.
26. Maxam, A., and Gilbert, W. (1980) Meth. Enzymol. 65, 499-560.
27. Weiher, H., Konig, M., and Gruss, P. (1983) Science 219, 626-631.
28. Lee, W., Mitchell, P., and Tjian, R. (1987) Cell 49, 741-752.
29. Herr, W., and Clarke, J. (1986) Cell 45, 461-470.
30. Van Dyke, M., and Dervan, P. (1983) Nucl. Acids Res. 11, 5555-

5567.

31. Sawadogo, M., and Roeder, R. (1985) Cell 43, 165-175.
32. Xiao, J., Davidson, I., Ferrandon, D., Rosales, R., Vigneron, M., Macchi, M., Ruffenach, F., and Chambon, P. (1987) EMBO J. 6, 3005-3013.
33. Ondek, B., Gloss, L., and Herr, W. (1988) Nature 333, 40-45.
33. Hager, D., and Burgess, K. (1980) Anal. Biochem. 109, 76-86.
34. Herr, W., and Gluzman, Y. (1985) Nature 313, 717-714.
35. Davidson, I., Xiao, J., Rosales, R., Staub, A., and Chambon, P. (1988) Cell 54, 931-942.
36. Landschulz, W., Johnson, P., Adashi, E., Graves, B., and McKnight, S. (1988) Genes and Develop. 2, 786-800.
37. Galas, P., and Schmitz, A. (1978) Nucl. Acids Res. 5, 3157-3170.
38. Siebenlist, U., and Gilbert, W. (1980) Proc. Natl. Acad. Sci. USA 77, 122-126.
39. Herbomel, P., Bourachot, B., and Yaniv, M. (1984) Cell 39, 653-662.

FIGURE LEGENDS

FIG. 1. Chromatographic purification of EP2. (A) Heparin-Sepharose gradient elution of DE52 flow-through material. The protein elution profile is shown along with the results of the footprint analysis (Materials and Methods) using 40 µl of the indicated heparin-sSpharose column fraction. Lane C contained no added protein. The fragment used was derived from the 1 x 72. The bracketed fractions (21-23) were pooled, concentrated by step elution and BioRex 70 and subsequently fractionated by gel filtration on AcA44. (B) Gel filtration on AcA44. The protein elution profile and footprint analysis using 40 µl of the indicated fraction are shown. The DNA fragment used was that described in (A) of this figure. Lane i is the input to the column. The bracketed fractions (32-44) were pooled and chromatographed on the sequence-specific oligonucleotide column. (C) EP2 oligonucleotide sepharose gradient elution. Shown is the protein elution profile and results of footprint assay using 20 µl of the indicated fractions from the first pass over the oligonucleotide sepharose column. The DNA fragment used was the 1 x 72 fragment described in (A) of this figure. The bracketed fractions (12-26) were pooled and chromatographed again on an EP2 oligonucleotide-Sepharose column (see Fig. 5 and Materials and Methods).

FIG. 2. Chromatography scheme for the purification of EP2 from bovine thymus nuclear extracts.

FIG. 3. Analysis of EP2 binding to the SV40 72 base pair

repeat. (A) MPE, DNase I and DMS footprints on the bottom strand of the SV40 72 base pair repeat. Footprint reactions were conducted as described under Materials and Methods. The fragment used was derived from the Xho plasmid diagrammed in (C) of this figure. Lanes 3, 8 and 10 correspond to the addition of 20 µl of partially purified EP2 to the footprint reaction; lanes 4, 7 and 9 correspond to no added protein. G and R correspond to chemical cleavage reactions specific for guanine and purine residues, respectively. (B) MPE, DNase I and DMS footprints on the top strand of the SV40 72 base pair repeat. Lanes 2, 8 and 10 correspond to the addition of 20 ml of partially purified EP2 to the footprint reaction; lanes 1, 7 and 9 correspond to no added protein. R and G are as previously described.

FIG. 4. EP2 footprint titration on wild-type (1 x 72) and mutant (dpm6) core elements. Lanes 1-3 contain 0, 5, 10 and 20 µl, respectively, of partially purified EP2 added to a footprinting reaction with the wild-type core containing fragment (1 x 72). Lanes 4-7 contain 0, 5, 10 and 20 µl, respectively, of partially purified EP2 in footprinting reactions with a template derived from the mutated core sequence double point mutant, containing plasmid dpm 6. R and G correspond to a product of the chemical cleavage reaction specific for purines and guanines, respectively. Also shown are the wild-type (1 x 72) and mutant (dpm6) core sequences along with a diagram of the fragment and the locations of the EP2 binding sites.

FIG. 5. Sequence-specific oligonucleotide chromatography of EP2. (A) Fractions containing EP2 binding activity from the first

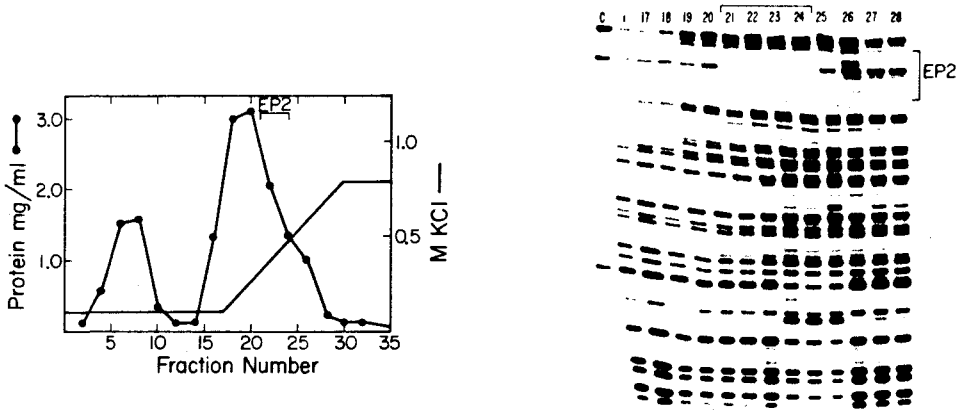
cycle over the DNA affinity column were pooled and gradient eluted again from the same oligonucleotide resin column. Shown is the footprint analysis using 20 µl of the indicated gradient fraction. Lane s contains no added protein; lane i contains the column input; lanes R and G correspond to the chemical cleavage reactions specific for purines and guanines, respectively. (B) Polypeptide composition of the sample assayed in (A). Assays consisted of 20 µl of the indicated gradient fraction; M corresponds to protein gel markers with masses of 160 kDa, 90 kDa and 40 kDa.

FIG. 6. Renaturation of EP2 binding activity. (A) Footprint analysis on gel-eluted, denatured and renatured protein samples (Materials and Methods). Lane 1: no added protein; lane 2: 20 µl protein gel input; lanes 3 and 4: 5 and 40 µl of sample obtained from 36 kDa polypeptide cluster; lanes 5 and 6 contain 5 and 40 µl of sample obtained from the 39 kDa polypeptide; lanes 7 and 8 contain 5 and 40 l of the sample obtained from the 41 kDa polypeptide. R and G correspond to products of the chemical sequencing reactions specific for purines and guanines, respectively. (B) Polypeptide analysis of gel-eluted, denatured and renatured protein samples. Lanes 1-3 each contain 40 µl of sample obtained from the individual bands. The large band at 70 kDa is BSA that was added during the renaturation procedure to aid in protein recovery. Protein gel markers are as in Fig. 5B.

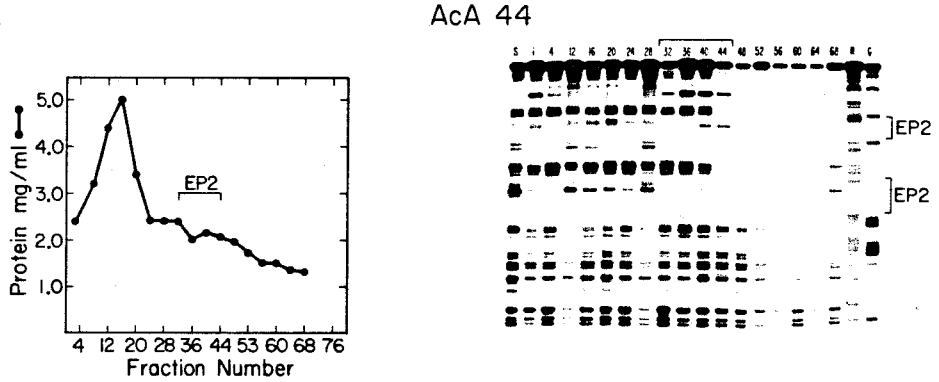
Fig. 1

EP2 Chromatography

A. Heparin Sepharose



B. AcA 44



C. Oligonucleotide Sepharose

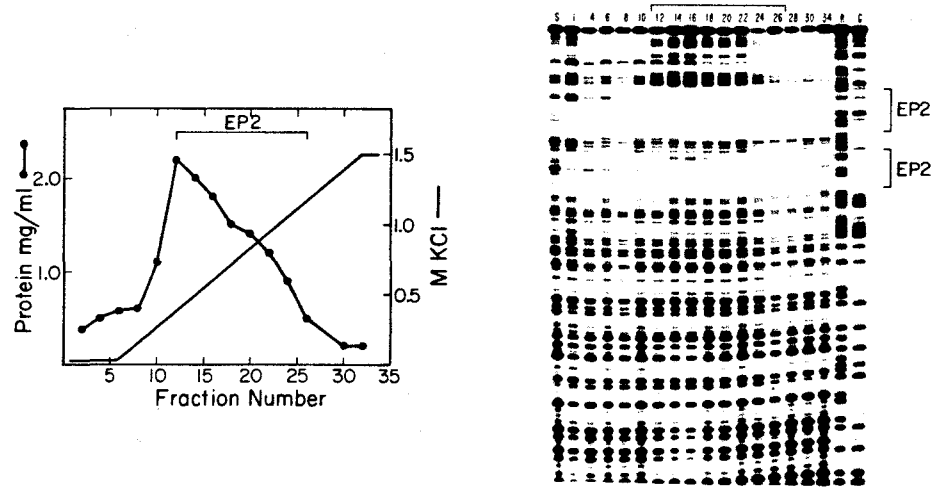


Fig 2

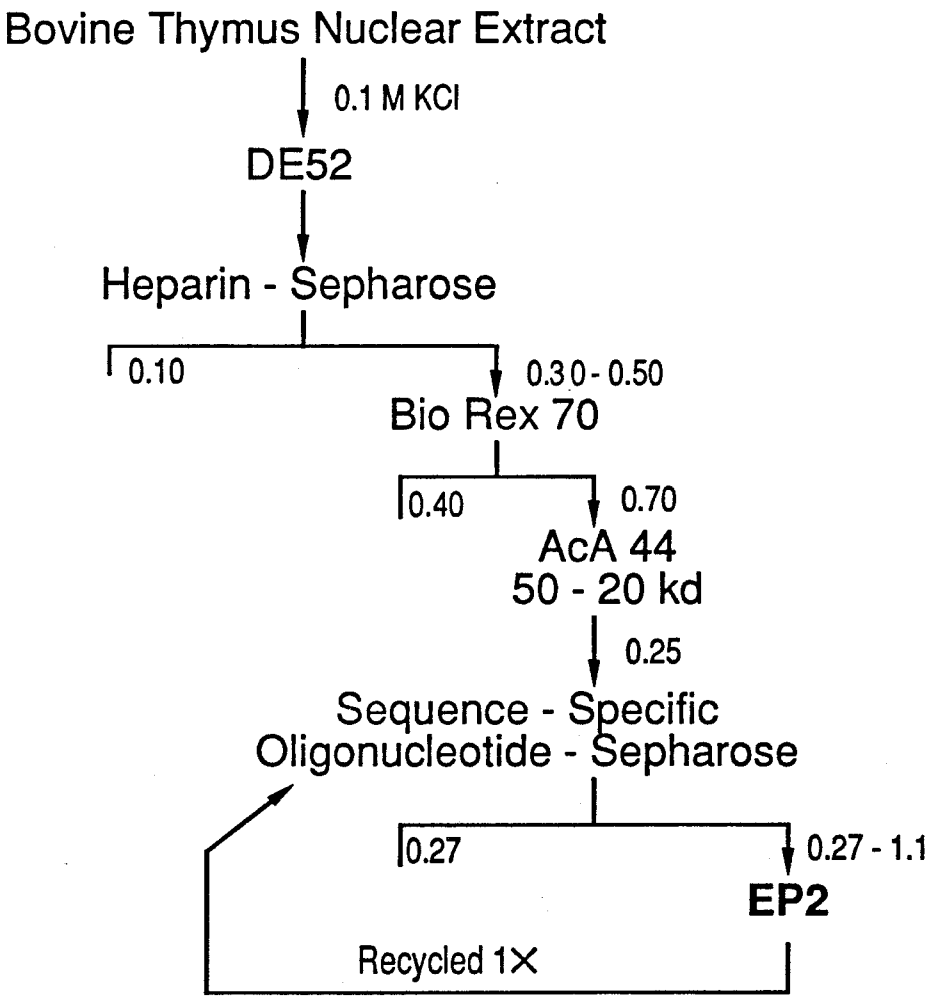
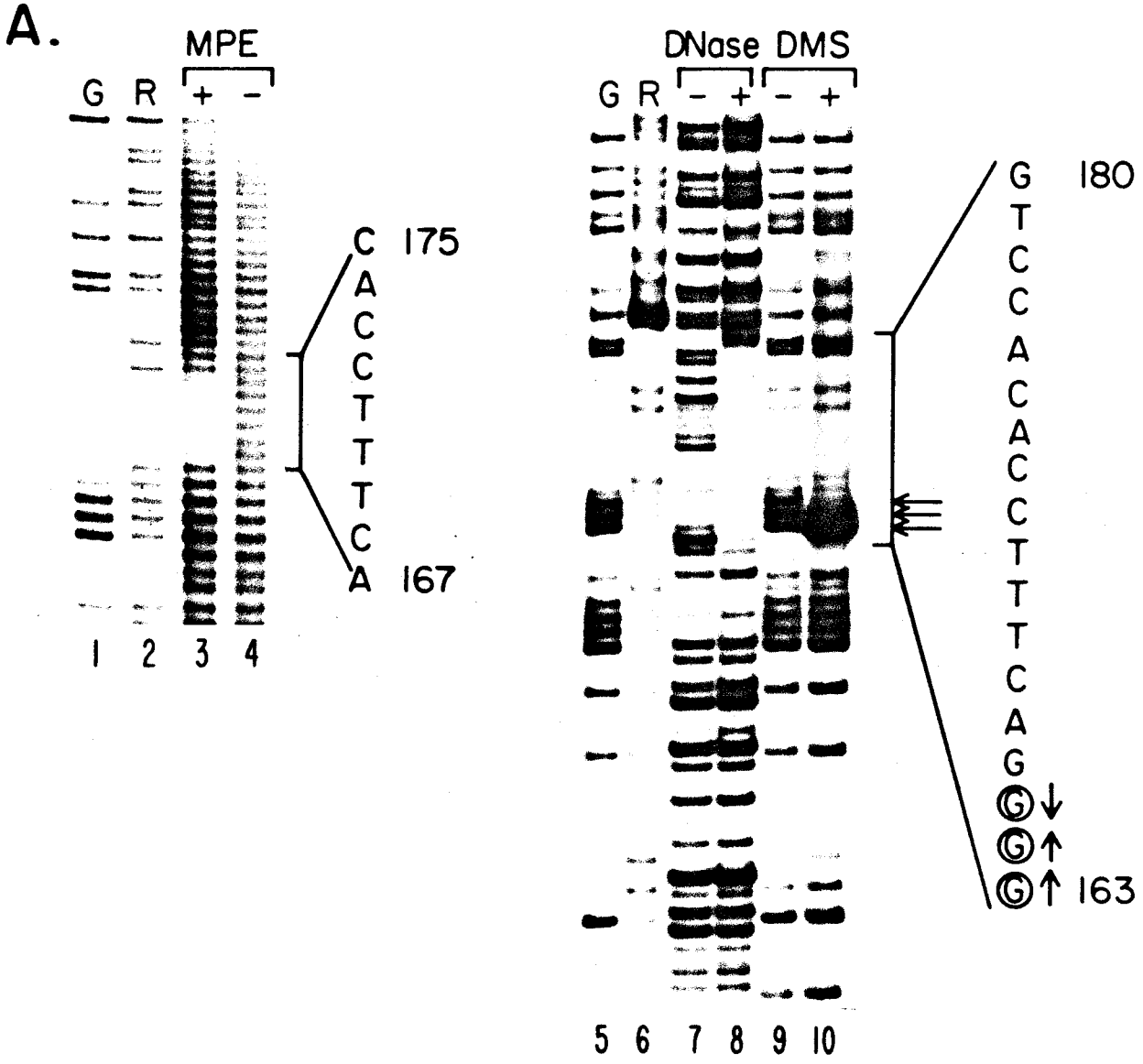
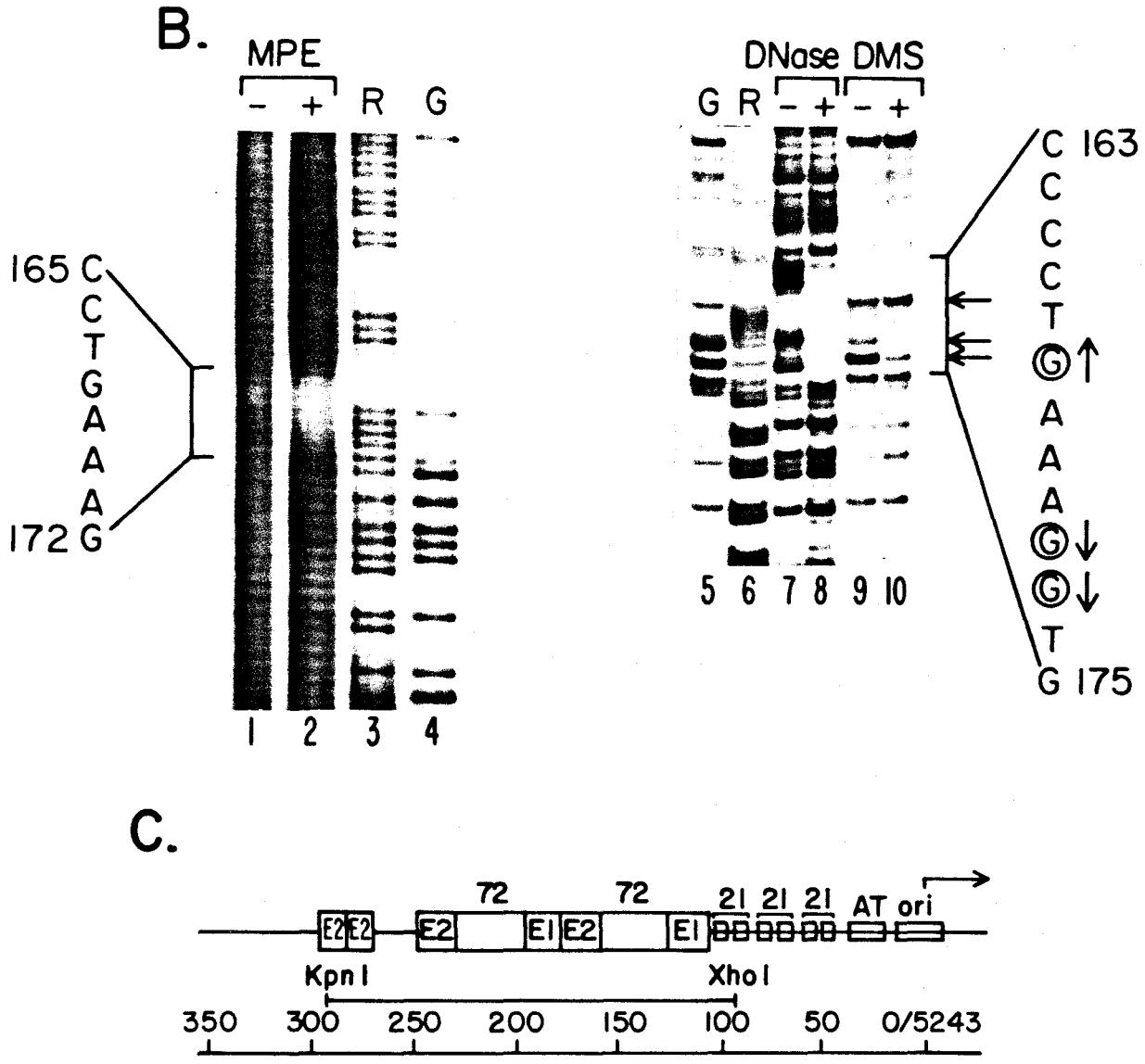
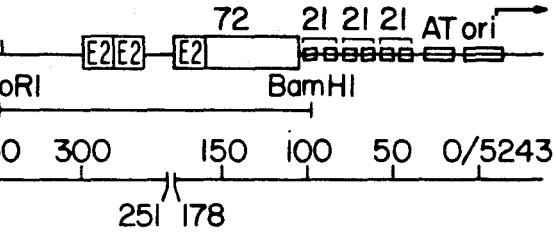

Bovine Thymus Nuclear Extract
0.1 M KCl
DE52
Heparin - Sepharose
0.10
0.30 - 0.50
Bio Rex 70
0.40
0.70
AcA 44
50 - 20 kd
0.25
Sequence - Specific
Oligonucleotide - Sepharose
0.27
0.27 - 1.1
EP2
Recycled 1X

Fig 3a


A.
MPE
G R + -
C 175
A
C
C
T
T
T
C
A 167
1 2 3 4
DNase DMS
G R - + - +
G 180
T
C
C
A
C
A
C
C
T
T
T
C
A
G
G↓
G↑
G↑ 163
5 6 7 8 9 10

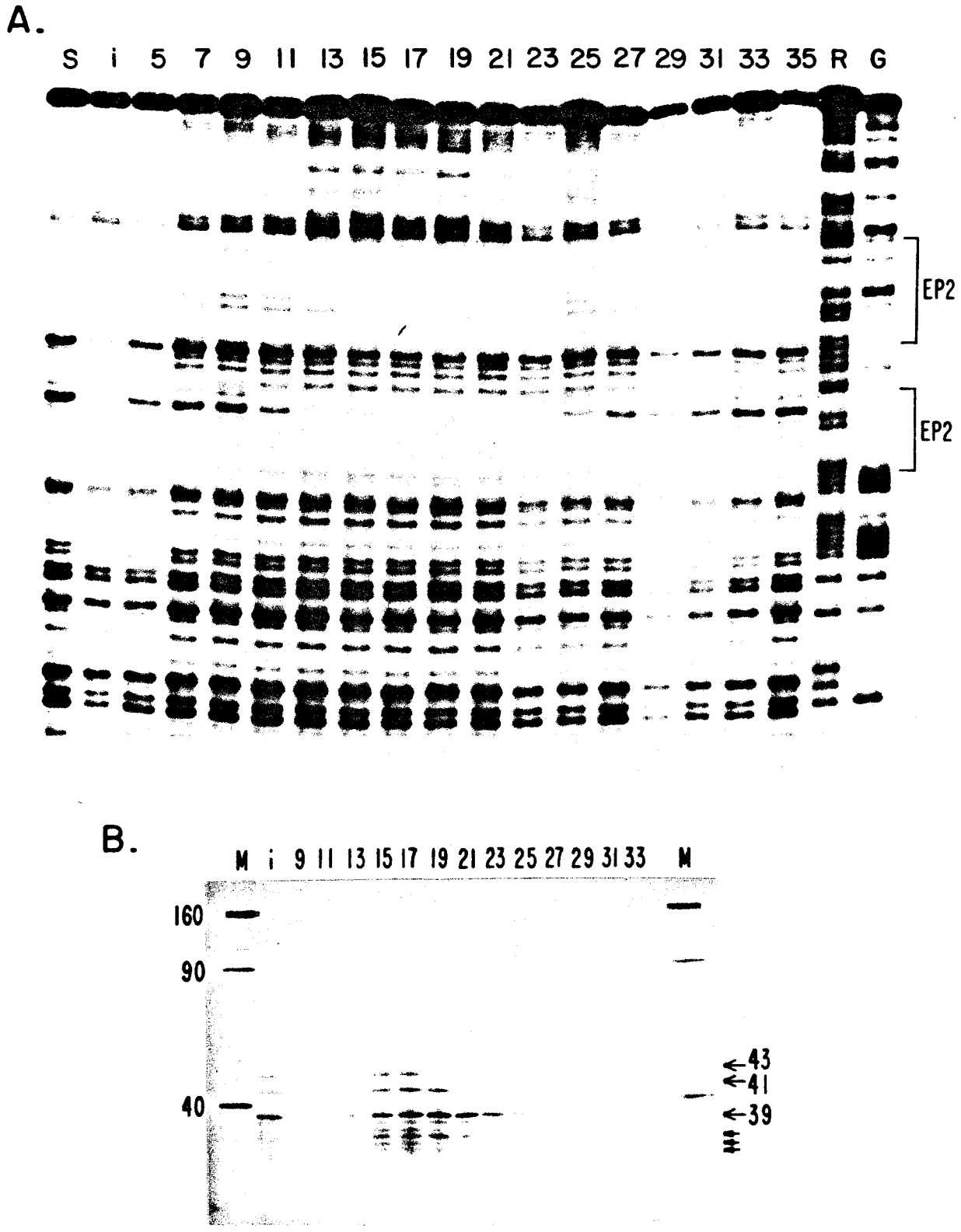

B.
MPE
- + R G
165 C
C
T
G
A
A
A
172 G
1 2 3 4
DNase DMS
G R - + - +
5 6 7 8 9 10
C 163
C
C
C
T
G
A
A
A
G
G
T
G 175
C.
72
72
21 21 21
AT ori
E2 E2 E2 E1 E2 E1
Kpn I
Xho I
350 300 250 200 150 100 50 0/5243


IX72 (WT) T G T G G A A A G T
dpm 6 T G T C C A A A G T

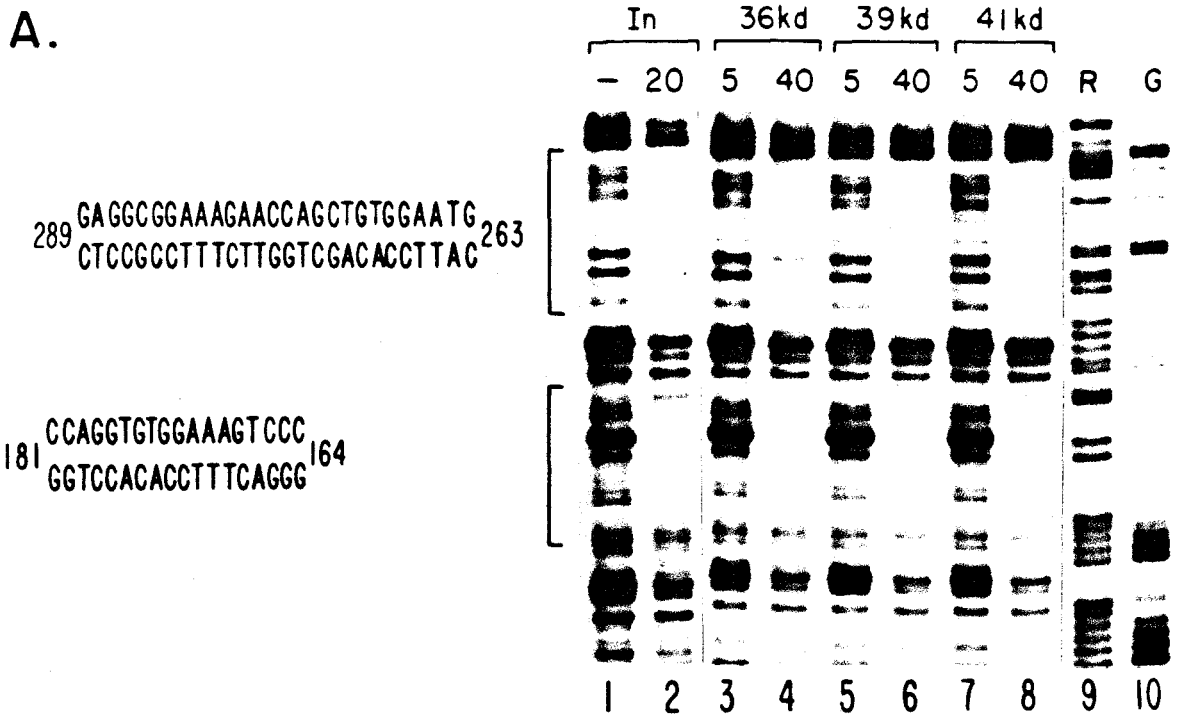
72
21 21 21
AT ori
E2 E2
E2
EcoRI
BamHI
350 300 150 100 50 0/5243
251 178


IX72
dpm 6
– 5 20 R G
– 5 10 20 R G
1 2 3 4 5 6 7 8 9 10 11

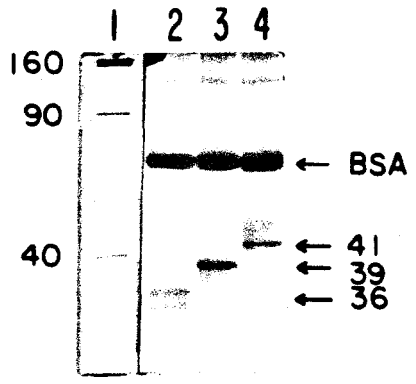
Fig 5


A.
S i 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 R G
EP2
EP2
B.
M i 9 11 13 15 17 19 21 23 25 27 29 31 33 M
160
90
40
43
41
39

A.


In
36kd
39kd
41kd
– 20 5 40 5 40 5 40 R G
289 GAGGCGGAAAGAACCAGCTGTGGAATG 263
CTCCGCCTTTCTTGGTCGACACCTTAC
181 CCAGGTGTGGAAAGTCCC 164
GGTCCACACCTTTCAGGG
1 2 3 4 5 6 7 8 9 10

B.


1 2 3 4
160
90
40
BSA
41
39
36

Chapter 2

Transcriptional Activation by the SV40 AP-1 Recognition Element in Yeast is Mediated by a Factor Similar to AP-1 that is Distinct from GCN4

Transcriptional Activation by the SV40 AP-1 Recognition Element in Yeast is Mediated by a Factor Similar to AP-1 that is Distinct from GCN4

Keith D. Harshman, W. Scott Moye-Rowley, and Carl S. Parker
Division of Chemistry
California Institute of Technology
Pasadena, California 91125

Running Title: Yeast Transcription Factor yAP-1

Contribution # 7754

[Published in *Cell* 53, 321-330 (1988)]

Abstract

The consensus recognition element for the mammalian transcription factor AP-1 is very similar to that of the transcriptional activator GCN4. In this report it is shown that the AP-1 recognition element (ARE) found in the SV40 enhancer can activate transcription from a heterologous promoter in *S. cerevisiae*. This activation, however, is not dependent on the presence of GCN4 as evidenced by ARE-dependent transcription in a *gcn4* yeast strain. A previously unknown yeast transcription factor that is probably responsible for this activation was identified and highly purified. The yeast factor, designated yAP-1, shares remarkably similar biochemical and DNA binding characteristics with mammalian AP-1. These data suggest that the yeast and mammalian AP-1s are evolutionarily conserved and perhaps functionally related. Also noteworthy is that GCN4 can bind to a GCN4 recognition element (GCRE) and the ARE with approximately equal affinities; yAP-1, however, has a much lower affinity for the GCRE than the ARE, suggesting that yAP-1 can discriminate between these elements *in vivo*.

Introduction

The initiation of transcription is a primary control point in gene regulation. This control is mediated in part by specific interactions between trans-acting protein factors and their cis-acting target DNA sequences. A number of general and gene-specific cis-acting regulatory elements have been identified by a combination of both genetic and molecular analyses (Guarente, 1984; Serfling et al.,1985; Maniatis et al.,1987). Genetic and, more recently, biochemical studies have led to the identification and purification of several trans-acting protein factors that interact with these regulatory elements (Guarente, 1987; McKnight and Tjian, 1987; Struhl 1987a).

A well characterized example of a gene specific trans-activator in yeast is GCN4 (Hinnebusch and Fink, 1983; Lucchini et al., 1984). GCN4 is known to be required for the transcriptional activation of many of the genes involved in the biosynthesis of amino acids (Jones and Fink, 1982). The cellular levels of GCN4 are believed to increase in response to amino acid starvation, resulting in the elevated transcription of a set of genes whose products are necessary for amino acid synthesis (Hope and Struhl, 1985). GCN4 binds to the GCN4 recognition element (GCRE) which contains the consensus sequence: ATGACTCAT (Hope and Struhl, 1985; Arndt and Fink, 1986; Hill et al., 1986). This sequence bears a remarkable similarity to the consensus sequence recognized by a recently identified mammalian activator protein isolated from HeLa cells called activator protein -1

(AP-1; Lee et al., 1987a). The AP-1 recognition element (ARE) has been found to consist of the following consensus sequence: C/GTGACTC/AA (Angel et al., 1987; Lee et al., 1987b). The ARE has the same central pentameric sequence as the GCRE: TGACT. AREs have been found in the SV40 enhancer and in the human metallothionein II promoter as well as in the promoters of other mammalian genes (Angel et al., 1987; Lee et al., 1987b). These observations suggested that a DNA binding motif has been highly conserved from yeast to man, but is used by different transcriptional activators with distinct biological roles (Struhl, 1987b). These activators may represent a new class of DNA binding proteins with a conserved structural motif analogous to those that contain either the helix-turn-helix (Pabo and Sauer, 1984) or zinc finger (Miller et al., 1985; Evans and Hollenberg, 1988) motifs.

During an analysis of DNA binding proteins present in yeast nuclear extracts, we identified a factor that binds to the SV40 ARE. This factor is present in normal and *gcn4* deletion mutant strains, demonstrating that this protein is not GCN4. Here we report on the purification and biochemical characterization of this previously unknown factor. We demonstrate that a mammalian ARE can activate transcription in both wild type and in *gcn4* deletion mutant yeast strains. We further show that the yeast factor and AP-1 have nearly identical DNase I, MPE, and DMS footprints. These data provide evidence that the factor we have identified is the yeast homolog of the mammalian transcription factor AP-1, and therefore

designate this factor as the yeast activator protein-1, or yAP-1.

One of the more interesting aspects of the DNA binding properties of yAP-1 is that it binds very poorly to the GCRE, despite the fact that the GCRE is almost identical to the ARE. GCN4, however, can bind to both the GCRE and the ARE with approximately equal affinities. Taken together, these observations suggest that yAP-1 can discriminate between the GCRE and the ARE *in vivo*, whereas GCN4 cannot.

Results

Saccharomyces cerevisiae Contains an ARE Binding Activity Distinct from GCN4

Fractions obtained by heparin Sepharose chromatography of a yeast nuclear extract were examined for the presence of DNA binding activities that would specifically interact with the SV40 enhancer element. An activity that specifically bound to the ARE of the SV40 enhancer was identified by DNase I footprinting (data not shown). The ARE has been shown to be sufficient for transcriptional activation both *in vivo* (Angel et al., 1987) and *in vitro* (Lee et al., 1987a) in mammalian systems. As noted previously (Struhl, 1987b), the DNA sequence of the ARE is very similar to that of the GCRE. This raised the possibility that the observed yeast ARE binding activity was due to the presence of GCN4 in the heparin Sepharose column

fractions. To test this possibility, an extract was prepared from a yeast *gcn4* deletion mutant strain (9995-8D). This extract contained a binding activity which, upon purification, had the same ARE binding characteristics as the strain containing the wild-type GCN4 gene (Figure 1, compare lanes 2 and 4). These data demonstrate that *S. cerevisiae* contains a DNA binding activity distinct from GCN4 that binds to a mammalian ARE.

The ARE Activates Transcription *In Vivo*

To evaluate the effectiveness of the ARE as a yeast upstream activating sequence (UAS), the resident UAS present in a yeast *CYC1:* β-galactosidase fusion gene (Guarente and Ptashne,1981) was replaced with single or multiple copies of a 26 base pair oligonucleotide containing the SV40 ARE (Figure 2). Similarly, constructions were made in which a *HIS3* GCRE (Hope and Struhl, 1985) was used to replace the resident UAS. The fusion genes were transformed into a yeast strain that contains a normal GCN4 gene (CG219) and a yeast strain that has had the GCN4 gene deleted (9995-8D). The fusion genes were present on both high copy number 2μm plasmids and single copy centromere containing (CEN) plasmids. The relative β-galactosidase activities of the different constructions were assayed and used as a measure of the transcriptional activity dependent upon the newly introduced AREs and GCREs. The results of these analyses are shown in Figure 2. In

the wild type GCN4 strain (CG219), a single copy of the ARE will stimulate transcription between 60-90 fold (depending upon the orientation) when present on a 2μm plasmid. A single copy of the GCRE, under the same circumstances, will stimulate transcription approximately 100-fold. Increasing the number of AREs from one to four further increases the transcription level 2-fold. As expected, the GCRE stimulates transcription significantly less in the *gcn4* mutant strain (9995-8D). A modest 5-6 fold stimulation above the basal activity level was observed. ARE dependent transcription remained high in the mutant strain, however, with a 35-70 fold stimulation above basal activity (see Figure 2 for further details).

The observations just described demonstrate that the ARE can function as a yeast UAS in the absence of GCN4. ARE dependent transcription is also seen with single copy CEN plasmid based constructions. Single copies of the ARE stimulate transcription approximately 20-fold in the presence of GCN4 and 7-10 fold in the absence of GCN4. Increasing the number of AREs to four results in a 4-fold increase in transcription over that seen with a single ARE (see Figure 2). These results are in agreement with similar experiments demonstrating ARE-dependent transcription in mammalian cells (Angel et al., 1987; Lee et al., 1987). It is clear from all of these experiments that the ARE is capable of activating transcription in yeast and in man.

Purification and Biochemical Characterization of yAP-

yAP-1 was highly purified by following the procedures previously described for the yeast heat shock transcription factor (HSTF, Wiederrecht, et al., 1987). The final step in the procedure consisted of sequence-specific DNA-affinity chromatography. The DNA affinity resin employed was made by coupling the ligated ARE oligonucleotide, shown in Figure 2, to Sepharose. ARE binding activity present in the column fractions was determined by DNase I footprinting using a DNA fragment that contained the SV40 enhancer element. The results obtained on fractions derived from one cycle through the DNA affinity column are shown in Figure 3a. Two AREs are bound by yAP-1 with differing affinities. The site to which yAP-1 binds with higher affinity lies within the 72 base pair repeat between nucleotides -181 and -198 (numbering according to Tooze, 1980). The other site lies just distal to the 72 base pair repeat, between nucleotides -251 and -267. The locations and relative yAP-1 affinities for the two sites are in agreement with the data reported for mammalian AP-1 (Lee et al., 1987b).

The polypeptide components of these same fractions were examined by SDS polyacrylamide gel electrophoresis and the results are shown in Figure 3b. A number of strongly silver staining polypeptides are present in the yAP-1 active fractions after one round of affinity chromatography. Recycling the yAP-1 activity multiple times through the affinity column does not result in any

significant further purification. Thus, it was not possible to clearly demonstrate the molecular weight of yAP-1 by conventional SDS polyacrylamide gel electrophoresis and silver staining. To determine the molecular weight of yAP-1, alternative approaches were employed.

A sample of affinity purified yAP-1 was subjected to SDS polyacrylamide gel electrophoresis and then transferred to nitrocellulose under non-denaturing conditions (see Experimental Procedures for details) . The nitrocellulose sheet was probed with a ligated ARE oligonucleotide radioactively labeled with g-^{32}P ATP and T4 polynucleotide kinase. After several high salt washes, specific binding of the oligonucleotide by polypeptides immobilized on the nitrocellulose was observed by autoradiography (Miskimins et al., 1985). Control experiments were performed in which the yeast HSTF was probed with a radioactively labeled ARE and yAP-1 was probed with a radiolabeled heat shock element (HSE; Figure 4). These experiments demonstrated that specific binding was observed only between a given protein and its corresponding recognition element (see Figure 4 for details). As shown in Figure 5a (lanes 2 and 3), nine polypeptides that specifically interact with the ligated ARE are routinely observed in the yAP-1 preparations derived from a single pass through the affinity resin . The polypeptides range in molecular weight form 90 to 25kD and do not correspond to any of the major polypeptides found by silver staining.

The polypeptides identified with the DNA probes were further

implicated in specific ARE binding by experiments that employed monoclonal antibodies. Mouse monoclonal antibodies were generated against preparations of affinity purified yAP-1. These monoclonal antibodies fell into two groups: A and B. Group A antibodies reacted with a single 70kD protein, presumably the same 70kD protein that was observed by silver staining. The group B antibodies reacted with a set of 7 polypeptides ranging in molecular weight from 90kD to approximately 35kD (Figure 5a, lane 1). The group B antibodies were determined to be yAP-1-specific by their ability to immunoprecipitate yAP-1 activity from yeast extracts (Figure 5b, lane 3). Group A antibodies as well as anti-HSTF monoclonal antibodies were unable to immunoprecipitate yAP-1 from yeast extracts (data not shown).

A comparison between the yAP-1 polypeptides observed using the Group B antibodies and those that specifically bound the ARE DNA probe is shown in Figure 5a, lanes 1 and 2. A striking similarity in the pattern of polypeptides identified by both methods is observed. Each polypeptide recognized by the anti-yAP-1 monoclonal antibody also bound specifically to the ARE oligonucleotide. Three polypeptides identified with the DNA probe (47kD, 26kD, and 24kD) are not recognized by the anti-yAP-1 antibody. This difference is most likely the result of proteolysis that specifically removes the epitope recognized by the antibody but leaves the DNA binding domain of the protein intact.

Additional evidence that the polypeptides identified by the

antibody and DNA probes are responsible for specific ARE binding is shown by the results of a polyacrylamide gel elution-renaturation experiment (Hager and Burgess,1980; Wiederrecht et al.,1987). After SDS polyacrylamide gel electrophoresis, the polypeptide components of affinity-purified yAP-1 were visualized by KCl staining and excised from the gel (molecular weight standards were utilized to determine the approximate molecular weights of the excised polypeptides). After electroelution, the polypeptides were renatured and assayed for DNA binding activity by DNase I footprinting (see Experimental Procedures for complete details). Polypeptides of 160kD, 90kD, 70kD, as well as two groups of polypeptides between 40-50kD and 30-40kD were examined and the results shown in Figure 5c. The 160kD and 70kD polypeptides do not bind to the ARE (Figure 5c, lanes 1 and 2), whereas the 90kD polypeptide and the two groups of polypeptides between 40-50kD and 30-40kD were able to specifically bind to the ARE (Figure 5c, lanes 3-8).

The group of ARE-binding polypeptides between 30 and 50kD is reminiscent of those found in human AP-1 preparations of Lee et al., (1987) and Angel et al., (1987). To investigate this observation further a detailed comparison between the molecular weights of the ARE-binding activities present in affinity purified preparations of AP-1 and yAP-1 was made. A remarkably similar pattern of polypeptides capable of binding to the ARE is observed in affinity purified preparations of human AP-1 and yAP-1 immobilized on nitrocellulose. As shown in Figure 5a , six polypeptides ranging in

molecular weight from 47kD to 35kD are observed both for yAP-1 and human AP-1 (lanes 2 and 3, respectively). It is noteworthy that no 90kD polypeptide is observed in any of the human AP-1 preparations, although it is possible that human AP-1 is proteolyzed during its isolation. In addition to this difference, monoclonal antibodies against yAP-1 have not been observed to react with human AP-1 (data not shown).

Comparison of yAP-1 and AP-1 Binding to the SV40 ARE

Because of the similarities between the molecular weights and functional characteristics of yAP-1 and AP-1, a detailed comparison of their DNA binding properties was carried out. Affinity purified preparations of yAP-1 and AP-1 were used in the DNA binding experiments. No contaminating GCN4 activity was detected by antibodies specific for GCN4 (data not shown) in the yAP-1 preparations used in the DNA binding experiments shown in Figure 6. The footprinting reactions employed a DNA fragment that contains a synthetic SV40 ARE (as described in Figure 2) inserted into the polylinker site of pUC19.

yAP-1 and AP-1 protect the same 18 nucleotides of the non-coding strand and the same 20 nucleotides of the coding strand from DNase I cleavage (Figure 6a and b, lanes 2 and 3). One minor difference between the two proteins is that AP-1 binding produces a DNase I hypersensitive site on the 3' boundary of the coding strand

footprint that is not observed with yAP-1. Methidiumpropyl-EDTA (MPE; Sawadogo and Roeder, 1985) footprints also demonstrate that both proteins bind to DNA in a nearly identical fashion (see Figure 6a and b, lanes 8 and 9).

Dimethyl sulfate (DMS; Shuey and Parker, 1986) footprints of both yAP-1 and AP-1 show that guanine residue -185 is increased in DMS reactivity whereas guanine -189 is reduced in DMS reactivity. yAP-1 binding greatly increases the reactivity of adenine -192 and greatly decreases the reactivity of guanine -187, while the effect of AP-1 binding on these residues is qualitatively the same, but quantitatively reduced (Figure 6a and b, lanes 5 and 6, respectively). yAP-1 decreases the reactivity of guanine residue -184; this effect is not observed with AP-1. A summary of the DNaseI, MPE, and DMS footprinting data is presented in Figure 6c.

The ARE and GCRE are Recognized with Equal Affinity by GCN4 but not by yAP-1

On the basis of the sequence homology between the ARE and the GCRE, one might expect yAP-1 and GCN4 to bind to both sites with similar affinities. To test this hypothesis directly, DNase I footprinting studies were carried out using affinity purified yAP-1 and partially purified GCN4. The DNA fragment used contained both an ARE and a GCRE, the sequences of which are shown in Figure 2. Figure 7 shows the results of the addition of increasing amounts of

GCN4 (lanes 2-7) and yAP-1 (lanes 9 to 14) to the binding reactions. Both the ARE and GCRE are protected from DNase I cleavage at approximately the same concentrations of GCN4. yAP-1, however, exhibits a significantly greater affinity for the ARE than for the GCRE. The addition of 1 binding unit (see Legend to Figure 4 for unit definition) of yAP-1 results in approximately 50% protection of the ARE (lane 11) while the GCRE shows no protection (lane 12). The GCRE is approximately 50% protected only after 20 binding units of yAP-1 are added to the reaction (lane 15). Thus, GCN4 has approximately equal affinity for the GCRE and the ARE while yAP-1 shows an approximately 20-fold higher affinity for the ARE.

It is important to stress that these experiments do not address the issue of the equilibrium DNA binding constants of GCN4 and yAP-1. This is because the precise concentration of active DNA binding proteins present in both the GCN4 and yAP-1 preparations was not determined. Until equilibrium binding constants are measured the possibility cannot be ruled out that the relatively lower affinity of yAP-1 for the GCRE is sufficient for specific binding of yAP-1 to the GCRE in the cell.

Discussion

Within the SV40 enhancer are a number of separate cis-acting elements, which activate transcription in a variety of different contexts and cell types (Zenke et al., 1986; Schirm et al., 1987). One

such element, the ARE, is also found in the promoter regions of several genes whose transcription is induced by phorbol-esters (Angel et al., 1987; Lee et al., 1987a). Examples of these genes include the human metallothionein II, collagenase, and interleukin 2 genes (Angel et al., 1987). Synthetic oligonucleotides containing this sequence element confer basal level as well as phorbol-ester-inducible transcription on heterologous promoters (Angel et al., 1987; Lee et al., 1987a). The mammalian ARE is almost identical to the consensus sequence bound by the yeast transcriptional activator, GCN4 . In this report we have demonstrated that the ARE is capable of activating transcription in yeast *in vivo* in the absence of GCN4. Thus, a previously unknown transcription factor exists in yeast that binds to the ARE and activates transcription.

The mammalian protein that binds to the ARE, designated AP-1, has been purified from HeLa cells and found to consist of a set of polypeptides that include a major component of 47kD (Lee et al., 1987a). The yeast counterpart to AP-1, designated yAP-1, has been extensively purified by sequence-specific DNA-affinity chromatography as described in this report. Silver-stained SDS polyacrylamide gels containing affinity purified yAP-1 revealed the presence of several bands that consistently co-purified with yAP-1 activity. None of the readily visualized silver stained bands possessed ARE binding activity as determined by blotting and renaturation experiments. This observation points out the potential risks in assigning molecular weights for DNA binding proteins simply

on the basis of silver staining. The reason that yAP-1 is not readily visualized by silver staining is not known, however, comparable amounts of other DNA binding proteins (based on DNA binding activity) that have been examined in this laboratory are stained by silver, suggesting that yAP-1 is not efficiently bound by silver.

Monoclonal antibodies raised against affinity purified yAP-1 that specifically immunoprecipitated yAP-1 activity identified several polypeptides that did not correspond to those seen by silver staining. These polypeptides consisted of a 90kD species and several smaller components that may be proteolytic breakdown products from the 90kD polypeptide, or perhaps distinct forms of yAP-1. Renaturation and blotting experiments further demonstrated that the polypeptides recognized by the monoclonal antibodies were responsible for the ARE binding activity and therefore constituted yAP-1.

AP-1 purified from HeLa cells contains polypeptides that are nearly identical in molecular weight and number to the smaller yAP-1 components. yAP-1 preparations consistently possessed a 90kD species that was not observed in any of the human AP-1 preparations (this report; Angel et al., 1987; Lee et al., 1987b). One possible explanation for this difference is that the HeLa cell derived AP-1 is proteolyzed from a larger component that is closer in molecular weight to yAP-1. A second possibility is that although yAP-1 and AP-1 share DNA binding domains (and perhaps other activity domains), a domain that is present in the 90 kD form of yAP-

1 is absent from AP-1. Along these lines, it is intriguing that the smaller polypeptides present in the yAP-1 preparations are almost identical in size and number to those present in human AP-1, leading to the speculation that both mammals and yeast may possess a family of related proteins that can bind to the ARE. In yeast it is clear that at least two members of such a family exist, GCN4 and yAP-1.

When the DNA binding properties of yAP-1 and AP-1 are compared, further similarities are observed. Footprint analyses employing both DNase I and MPE reveal that yAP-1 and AP-1 protect identical sequences on both strands of the DNA double helix. The purine residues implicated by DMS footprinting as important for DNA recognition are nearly identical for both proteins. All of these observations lead to the conclusion that yAP-1 and AP-1 are closely related. A comparison of the DNA binding properties of GCN4 with those of yAP-1 demonstrated a significant difference between the two proteins. GCN4 is capable of binding to the GCRE and the ARE with almost identical affinities. This observation is not surprising considering that the GCRE and ARE contain the same pentameric consensus sequence. yAP-1, however, gave the unexpected result of binding to the GCRE with an approximate 20-fold lower affinity than to the ARE. These observations suggest that yAP-1 may be capable of discriminating between the GCRE and the ARE *in vivo*. One caveat to this interpretation is that we have not measured the equilibrium DNA binding constants for yAP-1 or GCN4. Therefore, it is possible

that yAP-1 may in fact possess a sufficiently high DNA binding constant for the GCRE to interact with this element in the cell. This would of course require that the cellular levels of yAP-1 be sufficiently high to allow for GCRE binding.

ARE dependent transcription is not inducible by 3-aminotriazole (K. Harshman and S. Moye-Rowley, unpublished results), an amino acid analog that induces the expression of GCRE responsive genes (Penn et al., 1983). Thus it is unlikely that yAP-1 is directly involved in general amino acid control. It is possible that yAP-1 can bind to the GCRE *in vivo* and provide a certain basal level of promoter activity. Indeed, our own data show that there is a significant GCRE-dependent basal activity in the *gcn4* deletion strain that is perhaps mediated by yAP-1 binding to this element. The elements controlling basal expression of *HIS3* may reside within a specific promoter proximal GCRE (Struhl and Hill, 1987). However, the basal elements involved in *HIS4* expression are distinct from the GCRE (Arndt et al., 1987). It is most likely that yAP-1's primary role is to facilitate the expression of genes distinct from those involved in the control of general amino acid biosynthesis. An interesting speculation is that GCN4 may play a role in the expression of genes that possess an ARE. This is a reasonable possibility because GCN4 can bind with relatively high affinity to the ARE. Studies addressing these issues are currently in progress.

It is not known at the present time how many genes in yeast possess an ARE. Several genes have been found that have very good

matches to the ARE consensus sequence. For example, *PUT2* and *HOM3* have a potential ARE located in the promoter region upstream from the start point of transcription (S. Moye-Rowley and K. Harshman, unpublished observations). However, in none of these cases is it known whether this element is functional. It will be very interesting for future work to determine which genes require an ARE and whether these genes have any physiological relationship to each other.

The protein product of the avian oncogene *v-jun* has been shown to have a significant amino acid homology with the DNA binding domain of GCN4 (Vogt et al., 1987). More recently, it was demonstrated that the homologous unit present in the v-jun protein can substitute for the DNA binding domain of GCN4 (Struhl, 1987b). Using a *v-jun* clone, Tjian and co-workers have isolated the human proto-oncogene *c-jun* (Bohmann et al., 1987). The DNA sequence of *c-jun* has revealed a significant homology to the GCN4 DNA binding domain. Indeed, *c-jun* expressed in *E. coli* produces a protein that will bind to the ARE. This strongly suggests that *c-jun* is the AP-1 gene. These observations, combined with the current finding of yAP-1, bring to light a new and interesting class of trans-activating proteins. All of these activators bind to a DNA element that has the same central pentameric consensus sequence. GCN4 can bind efficiently to both the ARE and the GCRE whereas yAP-1 is capable of distinguishing between the ARE and the GCRE. An analysis of the primary amino acid sequence of the DNA binding domains of this

class of activators reveals no homology with any of the known DNA binding motifs such as helix-turn-helix or the hypothetical zinc fingers (Pabo and Sauer, 1984; Evans and Hollenberg, 1988). It will be interesting to isolate the gene for yAP-1 and determine its amino acid sequence. It is anticipated that there will be considerable amino acid homology between the DNA binding domain of yAP-1 and the other activators of this class. Indeed, one can further speculate that the differences between the sequences present in the DNA binding domains of yAP-1 and the other members of this group might lead to certain predictions as to which amino acids are responsible for the ability of yAP-1 to discriminate between the GCRE and the ARE.

Experimental Procedures

Yeast Strains and Yeast Nuclear Extract Preparation

Large scale yeast extracts were prepared using the protease-deficient strain EJ926 as previously described (Wiederrecht, et al., 1987). Additionally, gcn4 yeast extracts were prepared using the yeast strain 9995-8D (MATa ura3-52 leu2-3,-112 gcn4-D1; the generous gift of Drs. Kim Arndt and Gerald Fink). This strain and CG219 (MATa ura3-52) were used in the β-galactosidase activity assays.

HeLa Cell Nuclear Extract Preparation and Chromatographic Procedures

Nine liters of HeLa cell suspension cultures were grown in modified Eagles medium supplemented with 5% calf serum and harvested during the exponential growth period (Hare et al., 1980). The cells were collected by centrifugation (4500 x g; 5 min) and the loose cellular pellet was resuspended in 200 ml buffer A: 10 mM Hepes (pH 7.6), 15 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA (all steps were performed at 4°C unless otherwise noted). The cells were collected by centrifugation (18,000 x g; 5 min), resuspended in 130 ml of buffer A with 1 mM PMSF (Sigma), and allowed to swell on ice for 20

min. The cells were broken by homogenization (15-20 strokes) in a large dounce homogenizer. The homogenate was subjected to centrifugation (18,000 x g; 10 min) and the nuclear pellet resuspended in 60 ml of buffer A plus 3.75 ml of 4 M KCl and 1 mM PMSF. The solution was divided equally into Beckman 45 Ti polycarbonate tubes and 4 M ammonium sulfate was added to a final concentration 0.36M. After gentle mixing for one hour the tubes were subjected to centrifugation for 1 hr at 100,000 x g. The supernatant was collected and the protein precipitated by the addition of 0.33 g solid ammonium sulfate per ml of supernatant. The precipitated protein was collected by centrifugation (18,000 x g; 10 min) and resuspended in 15 ml of a buffer containing 25 mM Hepes (pH 7.6), 10% (v/v) glycerol, 50 mM KCl, 0.1 mM EDTA, and 0.2% (v/v) Triton X-100 (abbreviated as 0.05 HGKET where 0.05 is the molarity of KCl). The nuclear extract was dialyzed overnight against the same buffer and either used immediately or stored at -80°C.

The purification of yeast yAP-1 and human AP-1 was performed exactly as described for the yeast HSTF by Wiederrecht et al. (1987). The affinity resin used in the sequence-specific DNA - affinity chromatography was made according to the procedure of Wiederrecht et al. (1987) using the ARE oligonucleotide described in Figure 2.

Preparation of GCN4

GCN4 was isolated from *E.coli* strain AR68 harboring the yeast GCN4 gene on the plasmid pAB100. Both the plasmid and the bacterial strain were the kind gifts of Drs. Kim Arndt and Gerald Fink. GCN4 was partially purified through a phosphocellulose column as previously described (Arndt and Fink, 1986).

Footprint Reactions

Construction of Plasmids and Preparation of Labeled Fragments

pENH1x72, which contains the enhancer region of SV40 including a single copy of the 72 base pair repeat inserted into the polylinker of pUC13, was the kind gift of Dr. Winship Herr. pARE was constructed by inserting the ARE oligonucleotide shown in Figure 2 into the Bam HI site of pUC19. pARE/GCRE was constructed by inserting the GCRE oligonucleotide shown in Figure 2 into the Pst I site of pARE. All fragments were labeled with T4 polynucleotide kinase and γ-^{32}P-ATP. pENH1x72 was labeled at the Bam H1 site present in the polylinker, and subsequently cleaved with Eco RI also in the polylinker. Both pARE and pARE/GCN4 were labeled at the Eco RI site present in the polylinker and the cut with Pvu II when the non-coding strand was analyzed. When the coding strand was analyzed,

the DNA was labeled at the Hind III site in the polylinker and then cut with Pvu II.

DNase I and DMS Methylation Footprints

The DNase I footprints employed poly dI-dC as the carrier and were performed as previously described (Wiederrecht et al., 1987). The DMS footprinting reactions were carried out as described by Shuey and Parker (1986), omitting the gel isolation and elution steps.

Methidiumpropyl-EDTA (MPE) Footprints

The MPE reactions were performed in a final volume of 25 µl in 10 mM Tris (pH 7.4), 50 mM NaCl, 10 mg/ml poly dI-dC, and yAP-1 or AP-1 as indicated in the legend to Figure 6. After incubation for 20 min at 0°C, a solution containing 100 mM MPE and 200 mM $Fe(NH4)_2(SO4)_2$ was added to a final concentration of 25 mM MPE. Following a 3 min incubation at 22°C, the MPE-Fe(II) cleavage reaction was initiated by the addition of DTT to a final concentration of 2 mM. After further incubation for 15 min at 22°C, the reaction was terminated by first adding EDTA to 20 mM, and then 100 µl of footprint termination buffer as described (Parker and Topol, 1984).

Preparation of Anti-yAP-1 and Anti-GCN4 Antibodies

Anti-yAP-1 monoclonal antibodies were generated from two mice as follows: the mice received two intraperitoneal injections at two week intervals of 25 mg affinity purified yAP-1 in MDL and TDM emulsion. Two additional boosts of 25 mg in PBS (6 mM sodium phosphate pH 7.5, 3 mM potassium chloride, 140 mM sodium chloride) were made at ten day intervals. Myeloma-spleen fusions were analyzed for the presence of reacting antibody by the Western blotting procedure (Towbin et al., 1979) using affinity purified yAP-1. The monoclonal supernatants obtained were used undiluted. Anti-GCN4 polyclonal sera was the generous gift of Dr. Alan Hinnebusch. The polyclonal sera was diluted 1:1000 in PBS containing 0.02% sodium azide and 0.3% bovine serum albumin.

SDS Polyacrylamide Gel Electrophoresis, Western Blotting, DNA Probing of Blotted Proteins, and Renaturation Experiments

SDS polyacrylamide gel electrophoresis was performed as previously described (Laemmli, 1970). Western blotting was performed as described (Towbin et al., 1979); after blocking and incubating the nitrocellulose filter with poly- or monoclonal sera, bound antibodies were visualized with alkaline-phosphatase conjugated goat anti-mouse IgG (BioRad) according to the recommended procedure of the

supplier.

The probing of nitrocellulose-bound proteins using radiolabeled, ligated oligonucleotides was performed as follows: proteins were subjected to SDS polyacrylamide gel electrophoresis. All subsequent steps were performed at 4° C. The proteins were transferred to nitrocellulose by electroblotting for 6 hr at 100 mAmps in a buffer containing 25 mM Tris (pH8.3) and 190 mM glycine. The nitrocellulose filters were then blocked for 1 hr with 2.5% (w/v) non-fat dry milk in 0.05 HGKE. The filter was incubated for 1 hr in a solution containing $2x10^5$ cpm per ml radiolabeled oligonucleotide and 30 mg/ml sonicated double-stranded salmon sperm DNA. The filter was then washed twice with 0.1 HGKET for 5 min and once with 0.3 HGKET for 10 min. The probe specifically bound by the immobilized polypeptides was visualized by autoradiography.

Renaturation of yAP-1 after elution from polyacrylamide gels was performed as previously described (Wiederrecht et al., 1987).

Immunoprecipitation of yAP-1

Staph A cells (100 µl) were mixed with 20 µl of goat anti-mouse antibody and 400 µl of anti-yAP-1 antibody for 2 hrs at 4°C. The Staph A cell-antibody conjugate was pelleted by brief centrifugation and washed three times with 300 µl of buffer IP-1: 50mM Tris (pH 7.5), 2% (v/v) Triton X-100, 0.4 M NaCl, 0.1mM EDTA. Following

resuspension on the Staph A cell-antibody conjugate in 40 µl 0.05 HGKET, 200 µl of yeast nuclear extract was added and the resulting mixture incubated for 3 hrs at 4°C. The Staph A cell-antibody conjugate was washed twice with IP-1 and once with IP-2: 100mM Tris (pH 7.5), 2M Urea, 200mM NaCl, 0.5% Tween 20. The bound proteins were twice eluted from the Staph A cells by incubation in 30 µl of 6M guanidine hydrochloride in 0.05 HGKET for 10 min at 22°C. After brief centrifugation, the supernatant was removed and the denatured proteins renatured by overnight dialysis against 0.05 HGKET at 4°C. The presence of active yAP-1 was determined by DNase I footprinting on the pENH1x72 DNA fragment.

β-Galactosidase Activity Assays

Plasmids used in the assays were constructed by standard techniques (Maniatis et al., 1982). The Xho I-Sal I fragment of pLG669Z (Guarente and Ptashne, 1981), which contains the *CYC1* wild type UASs, was replaced with Bgl II linkers. ARE and GCRE oligonucleotides (See Figure 2 for sequences) were inserted into this Bgl II site and the resulting plasmids characterized by restriction mapping. Fusion gene constructions on plasmids containing a centromere based origin of replication were made by transferring the Sal I -Sst I fragment from the pLG669Z derived plasmids to Sal I/Sst I digested YCpSalZ, a derivative of pSEYC102 (Emr et al., 1986). Yeast strains CG219 and 9995-8D were transformed with these

plasmids using the lithium acetate procedure of Ito et al. (1983) . Yeast cultures were grown from single colonies in minimal media to an approximate $O.D._{600}$ of 1.0, at which time they were harvested by brief centrifugation, washed with 5 ml of H_2O and then resuspended at 50 $O.D._{600}$/ml in a buffer containing 0.1 M Tris (pH 7.5) and 0.05% (v/v) Triton X-100 (Miozzari et al., 1978). The cells were permeabilized by freezing at -80°C followed by thawing on ice. Reactions containing 800 µl of Z buffer (Miller, 1972) and 200 µl of 4 mg/ml o-nitro-phenyl-b-D-galactoside were equilibrated at 28°C for three minutes; the reactions were initiated by the addition of 200 µl of cell suspension. After a sufficient yellow color had developed, the reaction was stopped by the addition of 0.5 ml 1 M Na_2CO_3 and the time of the reaction noted. The samples were placed on ice until all of the reactions were completed, at which time the cells were removed by centrifugation and the A_{420} of the supernatants determined. The protein concentration of the individual cell suspensions was determined by first mixing 100 µl of cell suspension with 900 µl of H_2O and 0.5 ml of 3 N NaOH and heating at 90°C for 5 min. The solution was then cooled on ice and 0.5 ml of 2.5% $CuSO_4$ was added with thorough mixing. After incubating for 10 min at room temperature, the precipitate that forms was removed by brief centrifugation at room temperature and the A_{555} of the supernatant determined. Standard reactions were performed using bovine serum albumin solutions of known concentrations. The β-galactosidase units are defined as: $(2.3\times10^{-7})(A_{420}/v)(t)^{-1}(m)$, where v is the

volume of cell suspension used in ml, t is the time of the reaction in minutes, and m is the protein concentration of the cell suspension in mg/ml. The values shown represent an average of two determinations, each conducted with a separate transformant.

Acknowledgements

We wish to thank Drs. Kim Arndt and Gerald Fink for the gift of the *E. coli* strain AR68 harboring the yeast GCN4 gene on the plasmid pAB100 as well as the *gcn4* strain 9995-8D. We also thank Dr. Alan Hinnenbush for the gift of the anti-GCN4 polyclonal antibody sera. We thank Dr. Winship Herr for the plasmid pENH1X72 and Dr. Peter Dervan for the gift of MPE. We greatly appreciate the efforts of Susan Ou of the Caltech Monoclonal Core Facility for her expert preparation of the anti-yAP-1 monoclonal antibodies used in these experiments and Dr. Giuseppe Attardi for advice on HeLa cell culture techniques. We found the assistance and advice of Dr. Greg Wiederrecht invaluable during the purification of the yeast factor. We thank the members of the Parker lab and Dr. Barbara Wold for helpful comments on the manuscript. We thank Phoebe Ray for aid in the preparation of this manuscript. W.S. Moye-Rowley is the recipient of postdoctoral fellowship 2901 from the American Cancer Society. This research was supported by a grant from the ACS (NP-604) to C.S.P. C.S.P. is a Rita Allen Foundation Scholar.

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.

Arndt, K., and Fink, G. R. (1986). GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5' TGACTC 3' sequences. Proc. Natl. Acad. Sci. USA 83, 8516-8520.

Arndt, K. A., Styles, C., and Fink, G. R. (1987). Multiple global regulators control *HIS4* transcription in yeast. Science 237, 874-880.

Bohmann, O., Bos, T. J., Admon, A., Nishimura, T., 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.

Emr, S. D., Vassarotti, A., Garrett, J., Geller, B. C., Takeda, M., and Douglas, M. G. (1986). The amino terminus of the yeast F1-ATPase b-subunit precursor functions as a mitochondrial import signal. J. Cell Biol. 102, 523-533.

Evans, R. M., and Hollenberg, S. M. (1988). Zinc fingers: guilt by association. Cell 52, 1-3.

Guarente, L., and Ptashne, M. (1981). Fusion of Escherichia coli *lac Z* to the cytochrome c gene of Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 78, 2199-2203.

Guarente, L. (1984). Yeast promoters: positive and negative elements. Cell 36, 799-800.

Guarente, L. (1987). Regulatory proteins in yeast. Ann. Rev. Genet. 21, 425-452.

Hagar, D., and Burgess, K. (1980). Elution of proteins from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity. Anal. Biochem. 109, 76-86.

Hare, J. F., Ching, E., and Attardi, G. (1980). Isolation, subunit composition, and site of synthesis of human cytochrome c oxidase. Biochemistry 19, 2023-2030.

Hill, D. E., Hope, I. A. Macke, 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., and Fink, G. R. (1983). Positive regulation in the general amino acid control of Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 80, 5374-5378.

Hope, I. A., and Struhl, K. (1985). GCN4 protein, synthesized *in vitro*, binds *his3* regulatory sequences: implications for general control of amino acid biosynthetic genes in yeast. Cell 43, 177-188.

Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983). Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. 153, 163-168.

Jones, E. W., and Fink, G. R. (1982). Regulation of amino acid biosynthesis in yeast. In Strathern, J. H., Jones, E. W., and Broach, J. R., Eds. The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 181-299.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head bacteriophage T4. Nature 227, 680-685.

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.

Lucchini, G., Hinnebusch, A. G., Chen, C., and Fink, G. R. (1984). Positive regulatory interactions of the *HIS4* gene of Saccharomyces cerevisiae. Mol. Cell. Biol. 4, 1326-1333.

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.

McKnight, S., and Tjian, R. (1986). Transcriptional selectivity of viral genes in mammalian cells. Cell 46, 795-805.

Miller, J., McLachlan, A.D., and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. 4, 1609-1614.

Miller, J. H. (1972). Experiment in Molecular Genetics (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Miozzari, G. F., Niederberger, P., and Hutter, R. (1978) Permeabilization of microorganisms by Triton X-100. Anal. Biochem. 90, 220-233.

Miskimins, W. K., Roberts, M. P., McClelland, A and Ruddle, F. H. (1985). Use of a protein-blotting procedure and a specific DNA probe to identify nuclear proteins that recognize the promoter region of the transfer in receptor gene. Proc. Natl. Acad. Sci. USA 82, 6741-6744.

Pabo, C. O., and Sauer, R. T. (1984). Protein-DNA recognition. Ann. Rev. Biochem. 53, 293-321.

Parker, C. S., and Topol, J. (1984) A Drosophilia RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. Cell 46, 357-369.

Penn, M. D, Galgoci, B., and Greer, H. (1983). Identification of AAS genes and their regulatory role in general control of amino acid biosynthesis in yeast. Proc. Natl. Acad. Sci. USA 80, 2704-2708.

Sawadogo, M., and Roeder, R. G. (1985). Interaction of a gene-specific transcription factor with the adenovirus major late promoter

upstream of the TATA box region. Cell 43, 165-175.

Schirm, S., Jiricny, J., and Schaffner, W. (1987). The SV40 enhancer can be dissected into multiple segments, each with a different cell type specificity. Genes Develop. 1, 65-74.

Serfling, E., Jasin, M., and Schaffner, W. (1985). Enhancers and eukaryotic gene transcription. Trends Genet. 1, 224-230.

Shuey, D. J., and Parker, C. S. (1986) Binding of Drosophilia heat-shock gene transcription factor to the *hsp70* promoter. J. Biol. Chem. 261, 7934-7940.

Struhl, K. (1987a). Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. Cell 49, 295-297.

Struhl, K. (1987b). The DNA-binding domain of the *jun* oncoprotein and the yeast GCN4 transcriptional activator protein are functionally homologous. Cell 50, 841-846.

Struhl, K., and Hill, D. E. (1987). Two related regulatory sequences are required for maximal induction of Saccharomyces cerevisiae *his3* transcription. Mol. Cell. Biol. 7, 104-110.

Tooze, J., Ed. (1980). DNA Tumor Viruses. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Towbin, H., Stalhelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA 76, 4350-4354.

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 carboxyl-terminal region of a protein coded for by the oncogene *jun*. Proc. Natl. Acad. Sci. USA 84, 3316-3319.

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.

Zenke, M., Grundstrom, T., Mattes, H., Wintzerith, M., Schatz, C., Wildeman, A. G., and Chambon, P. (1986). Multiple sequence motifs are involved in SV40 enhancer function. EMBO J. 5, 387-397.

Figure Legends

Figure 1. Footprint of yAP-1 Isolated from *GCN4* and *gcn4* Yeast Strains.

Footprint reactions (20 µl final volume) contained 2.0 ng pENH1x72 DNA fragment and protein as indicated. Lanes 1 and 3: no protein; lane 2: 10 µl affinity purified yAP-1 isolated from yeast strain 9995-8D (*gcn4*); lane 4: 5 µl affinity purified yAP-1 isolated from yeast strain EJ926 (*GCN4*).

Figure 2. ARE- and GCRE-*CYC1* :β-Galactosidase Gene Fusion Constructions and Resulting β-Galactosidase Activity Levels in Yeast.

The ARE and GCRE oligonucleotides shown were inserted into the upstream Bgl II site of a *CYC1:lacZ* fusion gene promoter as described in the Experimental Procedures. The ARE oligonucleotide sequence was derived from the 72 base pair repeat of SV40; the GCRE oligonucleotide sequence was derived from the promoter proximal GCRE of the yeast *HIS3* gene. The constructions chosen for activity assays are shown. Insertion of the oligonucleotide with the sequence orientation shown is indicated by a rightward pointing arrow; the opposite orientation is indicated by a leftward pointing arrow. *GCN4* (CG219) and *gcn4* (9995-8D) yeast strains were transformed with

the ARE and GCRE constructions as well as constructions containing no UAS and the wild type *CYC1* UAS. β-galactosidase activity was quantitated for cultures grown from single colonies as described in the Experimental Procedures. The average value of two such determinations is presented in the table.

[a]2μm based origin of replication.

[b]Centromere based origin of replication.

[c]Four AREs with undetermined orientation were inserted into the Bgl II site.

Figure 3. yAP-1 DNA Affinity Chromatography Elution Profiles.

(a) DNase I footprint analysis of fractions eluted from the DNA affinity resin. Lane A: column flow through; lane B: 0.32 M KCl step fraction; numbered lanes: sequential 2.0 M KCl step fractions. Footprint reactions (20 μl final volume) contained 5 μl of the indicated column fraction and 2 ng pENH1x72 DNA fragment. The SV40 coordinates and positions of the yAP-1 binding sites relative to other SV40 promoter elements are shown.

(b) Silver-stained SDS polyacrylamide gel analysis of affinity chromatography fractions. Lane M: protein gel markers with molecular weights as indicated. Other lanes are as marked in (a); 5 μl of each fraction was analyzed.

Figure 4. DNA Probe Analysis of Affinity Purified yAP-1 and Yeast HSTF.

Samples containing 30 binding units yAP-1 (lanes 1 and 3) and HSTF (lanes 2 and 4) were subjected to SDS polyacrylamide gel electrophoresis, transferred to nitrocellulose and probed with radiolabeled, ligated HSE (lanes 1 and 2) or ARE (lanes 3 and 4) as described in the Experimental Procedures. One binding unit is defined as the amount of factor required to completely protect from DNase I cleavage 10 fmol of a DNA fragment containing a single binding site for the protein in question. The molecular weights of the polypeptides identified are given in kilodaltons. The 130kD polypeptide identified as HSTF using the HSE probe is identical to that identified by Wiederrecht et al. (manuscript in preparation).

Figure 5. Molecular Weight Determination of yAP-1

(a) Antibody and DNA probe analysis of affinity purified yAP-1 and AP-1. Samples containing 30 binding units of yAP-1 (lanes 1 and 2) and AP-1 (lane 3) were subjected to SDS-PAGE, transferred to nitrocellulose, and probed with either radioactively labeled, ligated ARE (lanes 2 and 3) or anti-yAP-1 antibody (lane 1) as described in Experimental Procedures. The molecular weights of the polypeptides identified are shown in kilodaltons.

(b) Immunoprecipitation of yAP-1 with anti-yAP-1 monoclonal

antibodies. DNase I footprint reactions (30 µl final volume) contained 2 ng of pENH1x72 and protein as indicated. Lanes 1 and 5: no added protein; lane 2: 5 binding units affinity purified yAP-1; lane 3: 20 µl of proteins precipitated from yeast extract by anti-yAP-1 antibodies; lane 4: 20 µl of proteins precipitated from yeast extract with monoclonal antibody raised against the Drosophila *Shaker* gene peptide III as a control.

(c) Renaturation of yAP-1 binding activity. Footprint reactions (50 µl final volume) contained 2 ng pENH1x72 DNA fragment and the specified volume of gel-eluted, renatured polypeptides of the indicated molecular weight. Lane 1: 40 µl, 160 kD: lane 2: 40 µl, 70kD; lanes 3 and 4: 20 µl and 40 µl, respectively, 90 kD; lanes 5 and 6: 20 µl and 40 µl, 40-50 kD; lanes 7 and 8: 20 µl and 40 µl, 30-40 kD.

Figure 6. DNase I, DMS, and MPE Footprints of yAP-1 and AP-1 on the SV40 ARE

Footprint reactions (final volume 20 µl) contained 5 binding units affinity purified yAP-1 or AP-1 and either 2.5 ng (DNase I) or 5 ng (DMS and MPE) of pARE DNA fragment .

(a) Footprints on the SV40 ARE coding strand. Lanes 1, 2, and 3: DNase I footprints of no protein, yAP-1, and AP-1, respectively; lanes 4, 5, and 6: DMS footprints of no protein, yAP-1 , and AP-1, respectively; lanes 7, 8, and 9: MPE footprints of no protein, yAP-1,

and AP-1, respectively.

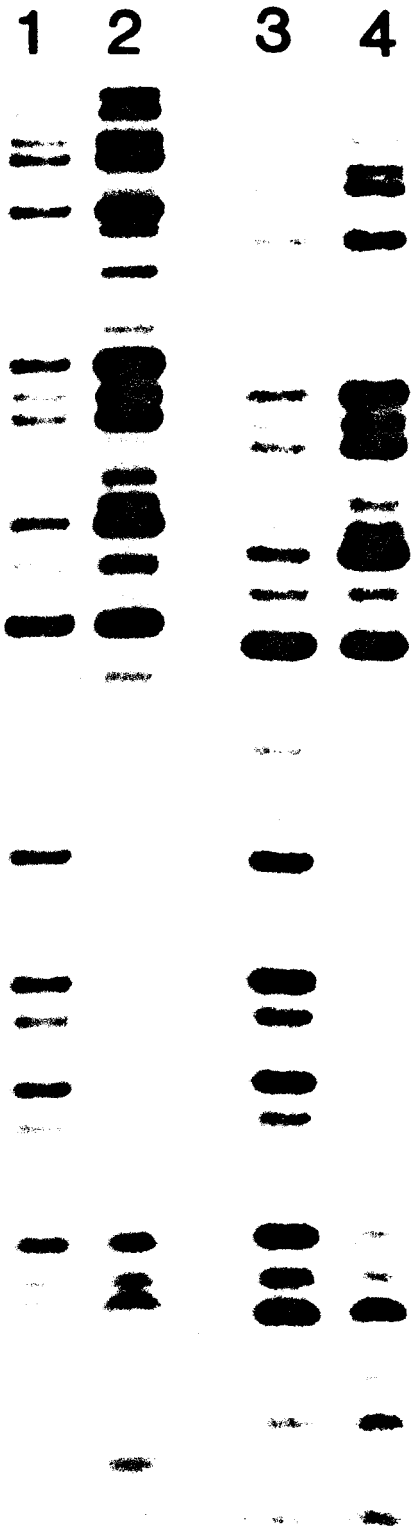
(b) Footprints on the SV40 ARE non-coding strand. Lanes are as marked in (a).

(c) Diagrammatic summary of yAP-1 and AP-1 footprints. The nucleotide sequence protected from DNase I cleavage by yAP-1 and AP-1 is shown. The nucleotides protected from MPE cleavage by the two proteins are contained within the boxes. The purine residues which show affected DMS reactivity due to protein binding are underlined. Effects due to AP-1 are indicated by open arrows and yAP-1 by solid arrows; increased reactivity is indicated by "up" arrows and decreased reactivity by "down" arrows.

Figure 7. Footprint Titrations of GCN4 and yAP-1 on a Fragment Containing Both the GCRE and ARE.

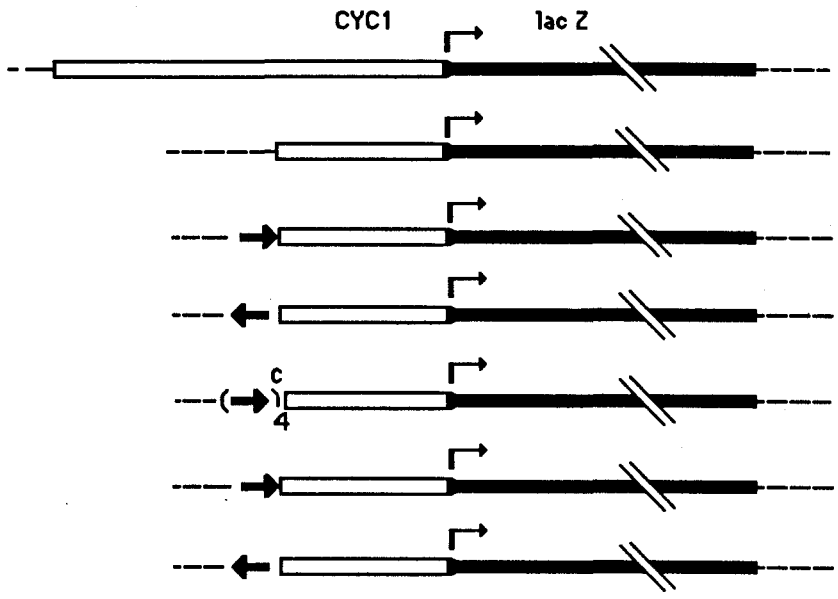
Footprint reactions (final volume 50 µl) contained 2.5 ng of the pARE/GCRE DNA fragment and the indicated amount of partially purified GCN4 or affinity purified yAP-1. Lanes 1, 8, and 15: no added protein; lanes 2-7: 0.2, 0.5, 1, 2, 4, and 10 binding units of GCN4, respectively; lanes 9-15: 0.2, 0.5, 1, 5, 10, and 20 binding units of yAP-1, respectively; lanes Y and R: chemical sequencing reactions specific for pyrimidines and purines, respectively. The positions of the ARE and the GCRE are labeled. Visible in lanes 5-7 is a footprint 3' to the labeled GCRE over a sequence in the *lacZ* gene of pARE/GCRE which has a weak homology to the GCRE.

Fig 1



ARE Oligonucleotide: 5'-GATCCATGGTTGCTGACTAATTGA
GTACCAACGACTGATTAACTCTAG-5'

GCRE Oligonucleotide: 5'-GATCCGGATGACTCTTTTTTTA
GCCTACTGAGAAAAAAATCTAG-5'



	β-Galactosidase Activity (nmol/min/mg)			
	CG219 (GCN4)		9995-8D (gcn4)	
UAS	2μm [a]	CEN [b]	2μm	CEN
CYC1	15	---	13	---
---	1.1	0.09	2.1	0.2
ARE	62	2.0	66	1.4
ARE	94	1.7	104	2.0
ARE	131	8.3	145	7.9
GCRE	94	---	9.2	---
GCRE	113	---	13.5	---

Fig 3

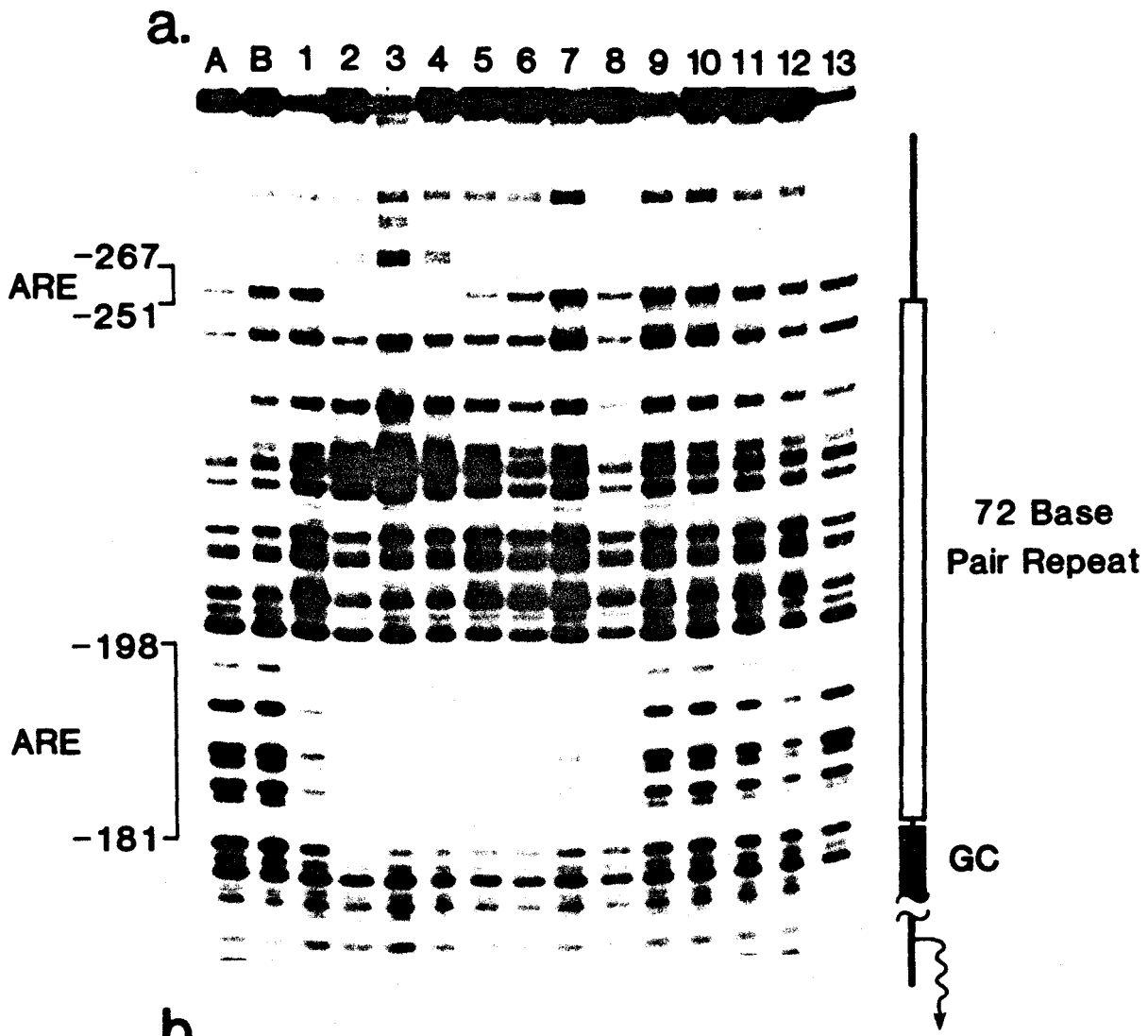

a.
A B 1 2 3 4 5 6 7 8 9 10 11 12 13
-267
ARE
-251
-198
ARE
-181
72 Base
Pair Repeat
GC
b.
M A B 1 2 3 4 5 7 9 11 13 M
165/155
94
90
67
60
43
40
30

Fig 4

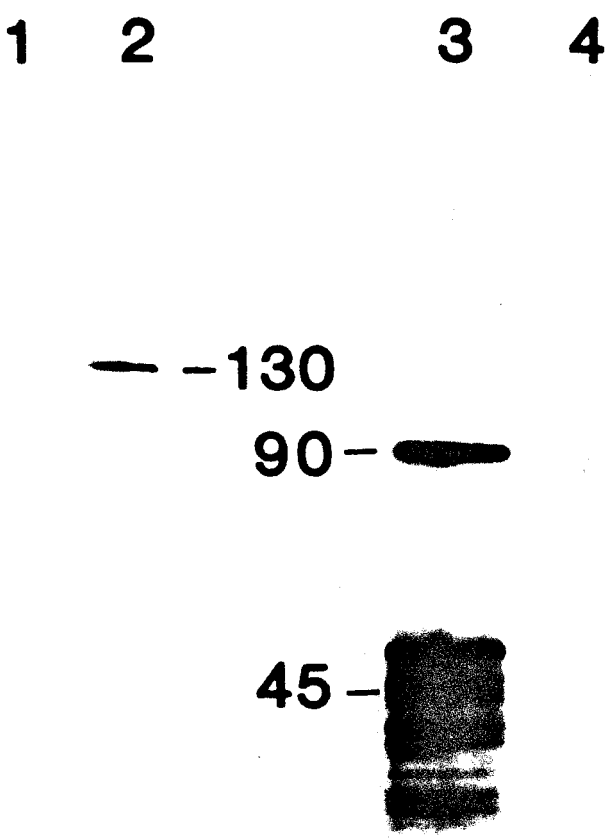


Fig 5

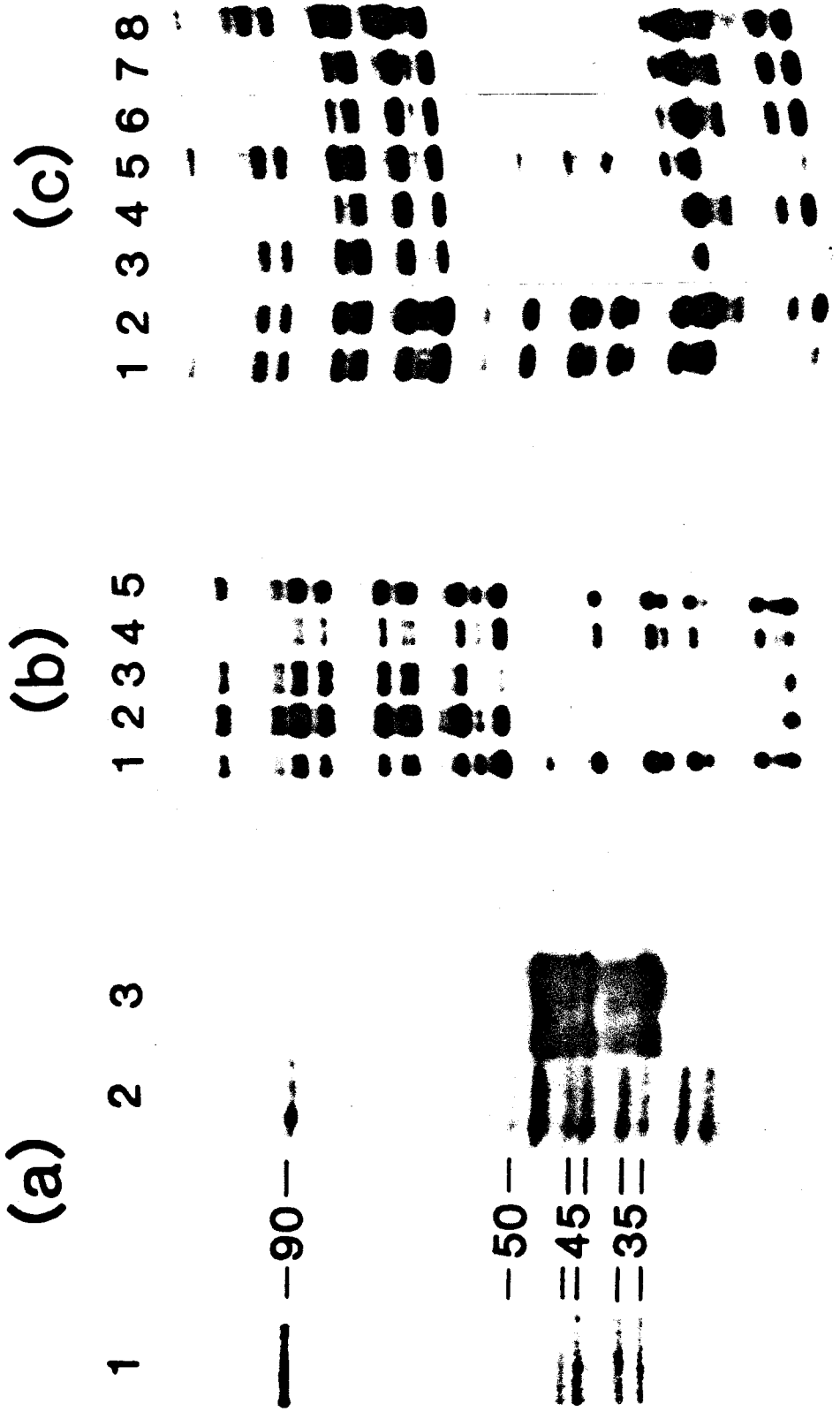
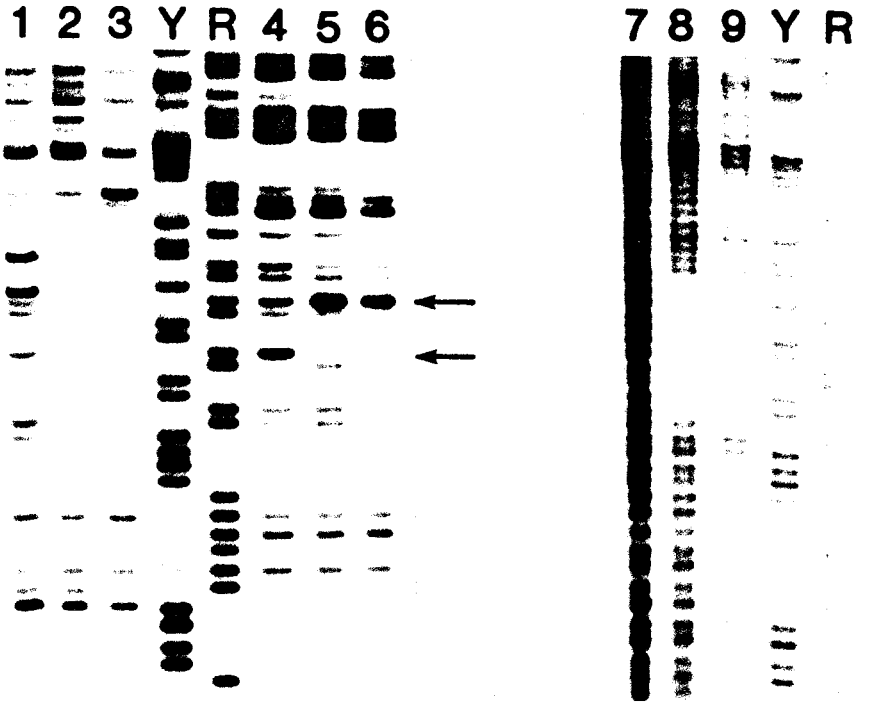


Fig 6a,b

(a)



(b)

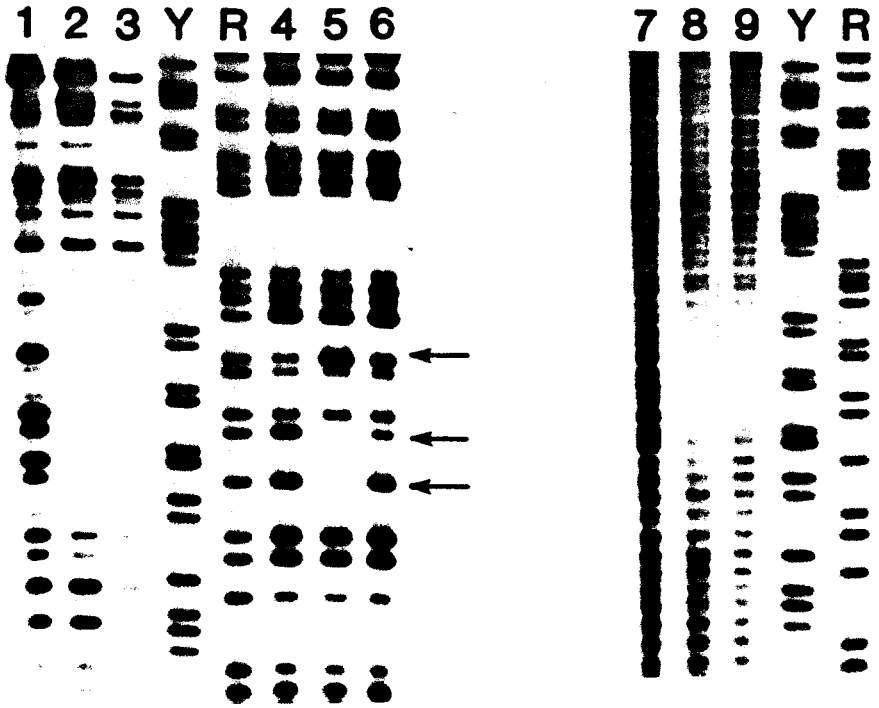
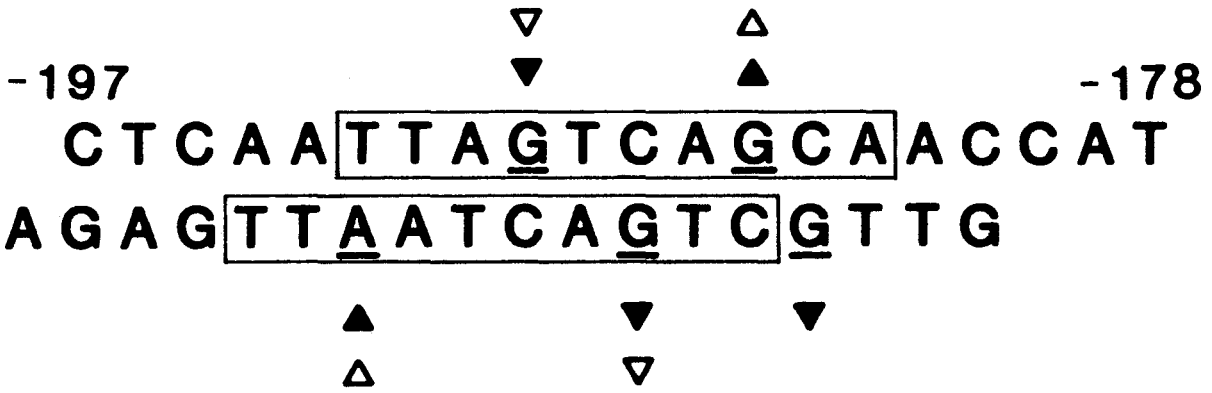
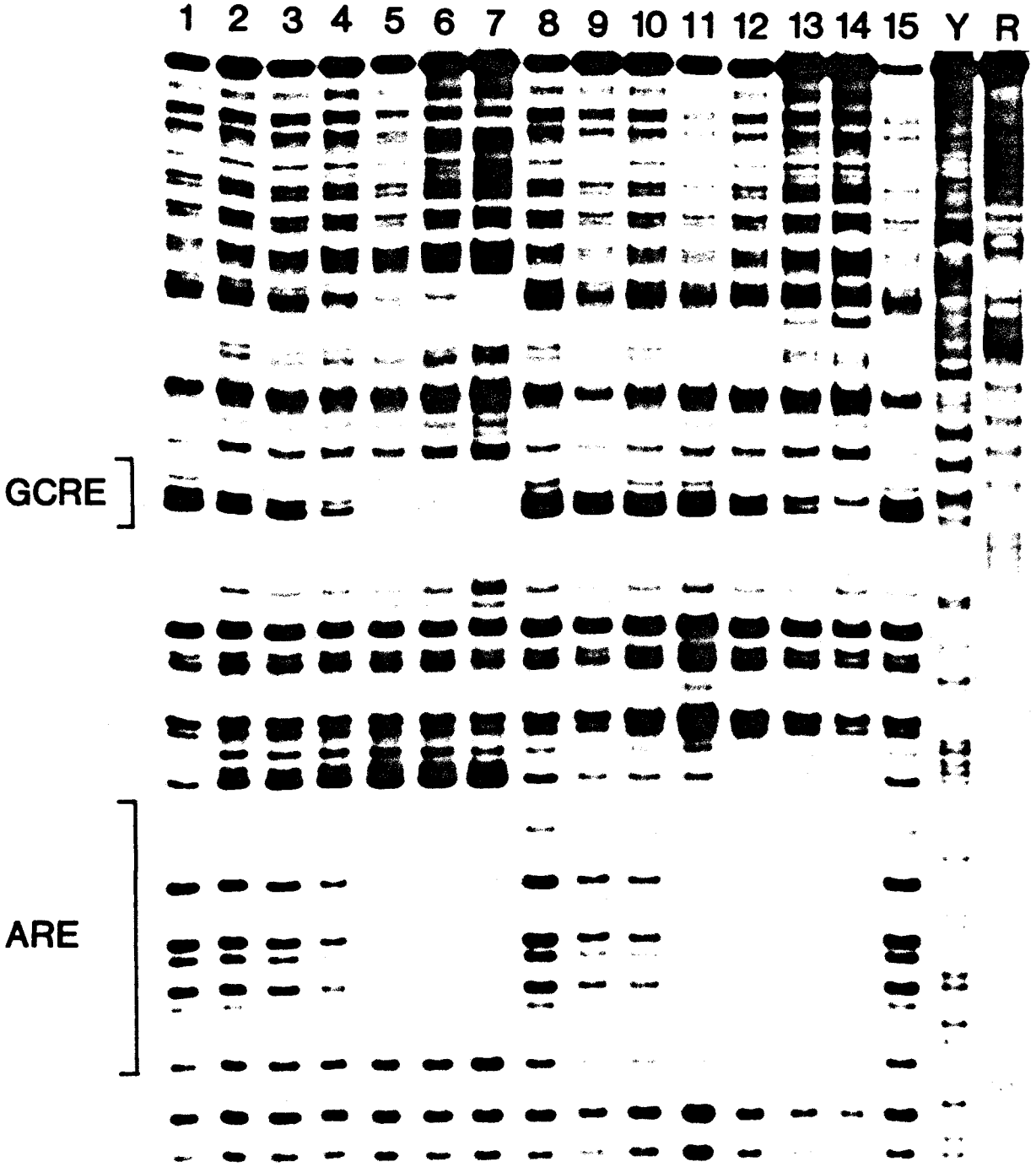


Fig 6c




1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 Y R
GCRE
ARE

Chapter 3

Isolation and Characterization of the Yeast *YAP1* Gene

Yeast *YAP1* encodes a novel form of the jun-family of transcriptional activator proteins

W. Scott Moye-Rowley, Keith D. Harshman, and Carl S. Parker
Division of Chemistry, 147-75
California Institute of Technology
Pasadena, California 91125

[*Key Words*: yeast AP-1; transcription factor; DNA-binding domain; *jun*-family]

Running Title: Yeast Gene for yAP-1

(Submitted to *Genes and Development*)

The jun-family of transcriptional activators includes mammalian AP-1 well as the yeast regulatory protein GCN4. Recently, an additional transcriptional activator has been found in yeast that recognizes the TGACTCA sequence element common in GCN4/AP-1 sites. This factor was designated yAP-1. The structural gene for yAP-1 has now been isolated and characterized. The deduced amino acid sequence predicts a protein of 650 residues, considerably larger than GCN4 or c-jun. The amino terminus of yAP-1 is homologous to the carboxy terminal DNA-binding domains of GCN4 and c-jun. Disruption of the *YAP1* gene demonstrates this gene is non-essential but required for ARE (AP-1 recognition element)-dependent transcriptional activation. DNA-affinity blots of proteins from *yap1* cells suggest the presence of additional TGACTCA-binding proteins, other than GCN4 and yAP-1. Furthermore, expression of at least one of these related DNA-binding proteins appears to be under control of yAP-1.

The use of a single DNA-binding, site which can be recognized by multiple transcription factors, is emerging as a common theme in eukaryotic transcriptional regulation. There are at least 3 different types of CCAAT-binding proteins from humans which recognize distinct subsets of the central CCAAT sequence (Graves et al. 1986, Jones et al. 1987, Dorn et al. 1987; Chodosh et al. 1988a). The SV40 "core" or TGTGGAAAG site can be recognized by the AP3 factor as well as C/EBP (Mitchell et al. 1987; Landschulz et al. 1988b; Jones et al. 1988). The use of a common binding site allows for multiple transcriptional effects mediated through a single recognition element.

This motif of multiple factors/single binding site introduces a further level of complexity to the analysis of purified DNA-binding proteins. It is possible that factors purified on the basis of binding may, in fact, consist of distinct proteins recognizing the same site. An example of this situation is mammalian transcription factor AP-1 (Lee et al. 1987a).

AP-1 binds to sites containing the sequence TGACTCA (Angel et al. 1987; Lee et al. 1987b). Purified AP-1 can stimulate *in vitro* transcription of the SV40 early promoter and human metallothionein IIA gene (Lee et al. 1987a). In addition, AP-1 binding sites were found to mediate phorbol ester-inducibility through a heterologous promoter (Angel et al. 1987; Lee et al. 1987b).

Recent work has indicated that the protein product of the *c-jun* proto-oncogene has DNA-binding properties which are identical to those of AP-1 (Bohmann et al. 1987; Angel et al. 1988). Detailed analysis of the polypeptides from HeLa nuclear extracts, purified by use of an AP-1 recognition element (ARE) DNA-affinity column, indicated there are several different proteins capable of binding the ARE (Rauscher et al. 1988). The deduced protein sequence of a serum-induced protein from mouse cells (jun-B) shows homology with v-jun (Ryder et al. 1988). These findings suggest that a family of related proteins exist that will bind to DNA sites containing the conserved pentameric element.

Members of the jun-family have also been identified in yeast. Both *c-jun* and *v-jun* as well as *jun-B* have homology at the C-terminus of their deduced protein sequences with the DNA-binding domain of the yeast, positively-acting, transcription factor GCN4 (Bohmann et al. 1987; Ryder et al. 1988; Vogt et al. 1987). The GCN4 recognition element (GCRE: GTGACTCAC) is also bound by c-jun (Bohmann et al. 1987). Additional functional homology comes from the observation that a lexA-v-jun fusion protein can bind to the GCRE upstream from the *HIS3* promoter and activate transcription of that gene (Struhl 1987).

We have reported the isolation and characterization of an additional factor, distinct from GCN4, which is present in yeast cells and that binds the ARE and activates transcription (Harshman et al. 1988). This protein was named yAP-1 by virtue of the identical

footprint formed on the ARE by this yeast factor and human AP-1. Analysis of the molecular weights of the polypeptide components present in the yeast AP-1 preparations revealed a major species of approximately 90 kD as well as several smaller proteins. The presence of lower molecular weight proteins might be due to proteolysis of the larger species or to the presence of a family of proteins that can recognize the same conserved sequence analogous to the situation in mammalian cells.

To more fully address this and related issues we have cloned the gene that encodes the 90 kD species of yAP-1. This gene has been designated *YAP1*. The structure of the deduced yAP-1 protein is quite different from that of c-jun or GCN4. While all three proteins have homology within their DNA-binding domains, the location of this region is carboxy terminal for both c-jun and GCN4, but amino terminal for yAP-1. Disruption of the gene that encodes yAP-1 is not lethal but results in the loss of yAP-1 expression as well as that of several other proteins that can bind selectively to the ARE and GCRE. These findings suggest that yAP-1 is a member of a family of transcription factors which are inter-related at the level of regulation as well as recognizing similar sequence motifs.

Results

Isolation of the YAP1 Gene

We utilized monoclonal antibodies generated against yAP-1 protein (Harshman et al. 1988) in order to clone the *YAP1* gene from a λgt11 expression library constructed from yeast genomic DNA

(Young and Davis 1983). The screening of approximately 10^6 phage (of which 30% were recombinant) yielded 8 positive signals. Five of these clones reproducibly reacted with the monoclonals and were taken to plaque purity. Phage DNA preparations of these recombinants were made and cleaved with EcoRI. Two different classes of phage were observed on the basis of the EcoRI digestion products (Figure 1). One class, a representative of which we designated λAC2, contained 2 EcoRI fragments of 1.9 kb and 0.22 kb. The other, represented by λAC5, also contained 2 EcoRI fragments but of 2.5 kb and 0.8 kb. All these EcoRI fragments were subcloned individually into pUC19 (Yanisch-Perron et al. 1985) for further analysis.

Restriction mapping suggested and Southern blotting confirmed that the 1.9 kb and 2.5 kb fragments were homologous. Western analysis of protein extracts from E. coli strains bearing the pUC19 subclones of the 1.9 kb and 2.5 kb fragments indicated that immunoreactive yAP-1 was being produced in these cells (data not shown). Both the 1.9 kb and 2.5 kb fragments were found to hybridize to a single 6.1 kb EcoRI fragment present in the yeast genome (Figure 6B). Further Southern blotting (data not shown) indicated that the 2.5 kb fragment extends from an authentic yeast genomic EcoRI site to a random shear endpoint 2.5 kb upstream. The 800 bp fragment is an authentic genomic EcoRI fragment and lies next to the 2.5 kb fragment in the genome.

DNA Sequence of the YAP1 Gene

Western analyses of bacterially-produced proteins encoded by subclones of the 2.5 kb fragment suggested a large portion of the yAP-1 coding sequence was contained within this region (unpublished observations). The DNA sequence of this fragment was determined. The sequence predicts one long open reading frame of 650 amino acid residues with no introns. The deduced molecular weight of this protein would be 72.5 kD in contrast to the observed molecular weight of 90 kD from SDS polyacrylamide gels (Harshman et al. 1988). This discrepancy between the observed and predicted molecular weights has been seen for several yeast transcription factors including GCN4 (Hope and Struhl 1985), RAP1 (Shore and Nasmyth 1987) and HSTF (heat shock transcription factor: Wiederrecht et al. 1988). The cause of this difference is unknown but may reflect some unusual structural features in these proteins.

While yAP-1 is predicted from the deduced amino acid sequence to be an acidic protein, the charge distribution indicates that the amino terminus is basic but the carboxy terminal region of the protein has an acidic nature. As we show later, the amino terminus of yAP-1 corresponds to the DNA binding domain.

S1 nuclease analysis of the 5' end of the *YAP1* mRNA was performed to localize the startpoints of *YAP1* gene transcription (Figure 3). The major 5' terminus is designated position +1 with several other minor startpoints at positions -7, +10, +15 and +38. Remarkably, all these startpoints map upstream of an ATG at

position +81 which is not predicted to be included in the *YAP1* coding sequence. The ATG is followed 15 bp downstream by a TAG terminator in frame. Since eukaryotic ribosomes generally initiate at the first ATG encountered in an mRNA (Kozak 1984), this ATG upstream from the *YAP1* ORF would be expected to have an effect on downstream *YAP1* expression. While we have no evidence of a role for the short upstream ORF, other yeast transcription factor genes, such as *GCN4* (Hinnebusch 1984) and *PPR1* (Kammerer 1984), have been found to have short potential coding sequences upstream of the main ORF encoding the transcription factor. It will be of interest to determine the type of role the upstream ORF has in gene expression of *YAP1*.

yAP1 is homologous to GCN4 and jun

Previous work has demonstrated that the recognition sequences for yAP-1, AP-1 and GCN4 are very closely related (Bohmann et al. 1987; Harshman et al. 1988). We compared our deduced amino acid sequence for yAP-1 to those for c-jun and GCN4 to ascertain if the known DNA-binding regions of the GCN4 (Hope and Struhl 1986) and c-jun (Bohmann et al. 1987) proteins were conserved in yAP-1. Figure 4A shows a computer generated alignment of the DNA-binding domains of GCN4 and c-jun with a homologous region of the amino terminus of yAP-1. The region of homology of yAP-1 contains 36% and 27% sequence identities with GCN4 and c-jun, respectively. Over the same stretch of amino acids, c-jun and GCN4 have 30% sequence identities. This strong sequence conservation prompted us to test the ability of a restriction fragment of *YAP1* encoding the c-jun/GCN4 homology region to produce yAP-1 DNA-binding activity.

Identification of the YAP1 DNA-Binding Domain

The production of high levels of yAP-1 activity in E. coli was accomplished by placing fragments of the *YAP1* gene under control of the strong, heat-regulated λPL promoter contained in the vector pOTSV (Shatzman and Rosenberg 1987). Initial experiments utilized extracts prepared from bacteria carrying a pOTSV derivative containing a BamHI/EcoRI fragment which expresses 90% of the full-length yAP-1 protein. This construction was designated pPL-*YAP1*. A BamHI/HpaI fragment encoding 93 amino acids of the

amino terminus of yAP-1 was also inserted into pOTSV to form pBH1. Crude extracts were prepared from appropriate E. coli transformants bearing these two plasmids and assayed for yAP-1 binding activity. Control experiments established that E. coli cells contained no specific ARE-binding activity (data not shown).

DNAse I footprinting demonstrates that the protection of the ARE seen in yAP-1 samples from yeast (Figure 5, lane 2) or from the bacterially produced protein (Figure 5, lanes 3-7) are indistinguishable. This result confirms we have isolated the gene encoding the yAP-1 protein. In addition, the ability of the extract from the pBH1 transformants to produce a wild-type yAP-1 footprint indicates that the DNA-binding domain of yAP-1 is encoded by this fragment and is located between amino acid residues 63 to 156. A comparison of the relative positioning of the DNA-binding domains of GCN4, c-jun, and yAP-1 is shown in Figure 4B.

The DNA template used for footprint analysis also contains a GCRE which is labeled in Figure 5. The yAP-1 sample purified from yeast gives a partial footprint on this site while neither of the bacterially derived extracts show any detectable protection. This is consistent with the belief that there exist, in yeast, other proteins that are capable of binding to sequences related to the TGACT repeat core of the ARE and GCRE. A large body of evidence has accumulated implicating the presence of multiple factors capable of recognizing the ARE in mammalian cells (Rauscher et al., 1988). In

order to examine the possibility that an analogous situation existed in yeast, we constructed a yeast strain which lacked a functional *YAP1* gene.

Characterization of a Yeast yap1 Mutant

We employed the method of one-step gene disruption (Rothstein, 1983) in order to generate a mutant containing a non-functional yAP-1 protein. The 1.7 kb *HIS3* BamHI fragment was inserted into the *YAP1* gene between amino acid residues 156 and 157. The resulting fragment was introduced into a homozygous *his3* diploid strain as shown in Figure 6A and *HIS+* transformants were selected. Correct integration was verified by Southern blot analysis (Figure 6B). Selected heterozygotes were sporulated and the resulting tetrads dissected by standard techniques (Sherman et al., 1979).

Analysis of spores of representative tetrads indicated that *HIS+:his-* segregants were recovered in a 2:2 fashion, thus indicating that the *yap1::HIS3* insertion is not a lethal mutation. We could detect no gross phenotypic change in segregants bearing the insertion mutant. Direct examination of the mutant yAP-1 protein was next undertaken. An extract prepared from a *yap1::HIS3* segregant was chromatographed through a heparin-agarose column and assayed for GCRE and ARE binding activity by DNAse I footprinting. A control extract was prepared using an *YAP1, his3* segregant from the same tetrad. HSTF activity was also assayed in both extracts as a control DNA-binding protein.

Fractionation of a yeast nuclear extract over heparin agarose results in the yAP-1 and HSTF DNA-binding activities being eluted in the 0.8 M KCl step (Harshman et al. 1988; Wiederrecht et al. 1987). We recovered equal levels of HSE binding activity in the *yap1* and *YAP1* extracts (not shown). DNA-affinity blotting (Miskimins et al. 1985) was performed to determine the type of polypeptide chains in the two extracts capable of recognizing the GCRE and ARE (Figure 7).

A DNA-affinity blot of the heparin agarose fraction from the *YAP1* strain, with the ARE as the DNA probe, detects the characteristic set of bands, starting with full-length yAP-1 at 90 kD as well as a set of proteins from 50 kD to 30 kD (Harshman et al. 1988). However, the fraction from the *yap1* strain gives a distinctly different result. The 90 kD band is entirely absent while the lower molecular weight species are reduced in intensity or absent. Remarkably, a strong signal at approximately 45 kD remains. It is unlikely that this binding activity is due to a novel yAP-1 fusion protein formed by the *HIS3* insertion as the deduced molecular weight from the known DNA sequence (Struhl 1985) of the fusion junction predicts a primary translation product of only 27 kD.

A distinct set of signals is observed when the GCRE is used to probe a parallel DNA-affinity blot. Strong binding is seen from polypeptides of 69, 65 and 55 kD in the *YAP1* extract. Only the 55 kD signal is seen in the yap1 extract. GCN4 protein is unlikely to be detected in this assay as the cells from which the extracts were

made were grown under conditions which repress GCN4 synthesis (Thireos et al. 1984; Hinnebusch 1984). The 55 kD polypeptide is larger than the observed molecular weight for GCN4 (Hope and Struhl 1985; Arndt and Fink 1986) and may represent some other TGACT-recognizing protein. The 69 kD band which is reduced in intensity and the 65 kD band which is not seen in the *yap1* mutant are probably not forms of yAP-1 as they are not detected by the ARE. Functional yAP-1 may be required for the expression of the 69 and 65 kD proteins but further experiments are required to confirm this.

Transcriptional Activation by Cloned YAP1

Further evidence that the gene we isolated encodes yAP-1 comes from an analysis of the ability of the cloned gene to stimulate ARE-dependent transcription in an *yap1* mutant background. We employed *CYC1-lacZ* gene fusions in which β-galactosidase activity is dependent upon the presence of a functional ARE. Expression of *YAP1* was placed under control of the inducible galactose promoter (Johnston and Davis 1984). The results are shown in Table 1.

In a wild-type *YAP1* cell, 130 units/mg of β-galactosidase are produced when glucose is present as the carbon source while 70 units/mg are generated during growth on galactose. When an isogenic *yap1::HIS3* (see experimental details) is assayed under the same growth conditions, 0.2 and 0.11 units are formed during glucose and galactose growth, respectively. Introduction of a *GAL1-YAP1* gene fusion into this mutant background leads to no change in

the level of β-galactosidase activity during glucose growth. However, induction of the *GAL1* promoter by growth on galactose leads to a 1000-fold elevation in ARE-dependent β-galactosidase activity. These data support the conclusion that the gene we have isolated indeed encodes the positively-acting transcription factor yAP-1. Disruption of the coding sequence of this gene leads to an inability to activate transcription through the ARE while expression of the clone under *GAL* promoter control can correct this defect.

Discussion

YAP1 Encodes Transcription Factor yAP-1

We have employed monoclonal antibodies directed against the yAP-1 protein (Harshman et al. 1988) to isolate from a λgt11 library (Young and Davis 1983) the yeast gene encoding this factor. Three independent lines of evidence lead us to conclude we have cloned the *YAP1* gene: (i) the footprint of the bacterially expressed protein is identical to that of authentic yeast factor; (ii) the DNA sequence of the DNA-binding region of the clone has homology with the analogous region from GCN4 and c-jun; (iii) disruption of the *YAP1* ORF prevents ARE-dependent transcriptional activation.

An unexpected result is the loss of ARE-dependent activation upon disruption of the *YAP1* gene (Table 1). In these cells, the *GCN4* gene is still intact. Previous work has demonstrated that GCN4 binds as well to the ARE as to a GCRE from the *HIS3* promoter (Harshman et al. 1988) and that transcriptional activation through the GCRE is dependent on GCN4 binding (Hill et al. 1986). These data support the idea that GCN4 should be able to bind to the ARE and activate transcription. With this in mind, it is surprising that the level of β-galactosidase activity produced in the *GCN4, yap1* cell from the *ARE-CYC1-lacZ* gene fusion is no higher than the same fusion lacking an upstream activation site (unpublished results). This suggests that authentic yeast GCN4 protein can discriminate between the ARE and yeast GCRE. Alternatively, yAP-1 may be involved in expression of the *GCN4* gene.

yAP-1: A Representative of a New Class of jun-Type Transcriptional Activator?

Our original observations (Harshman et al. 1988) of the AP-1-like footprint on and transcriptional activation by the ARE in yeast led us to believe we had identified the yeast AP-1 factor. However, the isolation of a human *c-jun* genomic clone which predicts a primary translation product of only 340 residues (Bohmann et al. 1987; Angel et al. 1988) and comparison to our yAP-1 sequence indicates yAP-1 is a protein with related properties to c-jun but quite different in structure. yAP-1 recognizes the same sequence as c-jun but has its DNA binding domain located at the amino terminus of the protein while the c-jun DNA-binding domain is at the carboxy terminus (Figure 4B). In this regard, GCN4 protein structure is much closer to that of c-jun with both proteins having their DNA-binding domains at the carboxy end.

Even though the DNA-binding domain of AP-1 and c-jun/GCN4 are located at different ends of the respective proteins, the primary sequences of these regions are homologous (Figure 4A). Within this region of homology, a recently described potential structure is conserved between c-jun, GCN4 and yAP-1, the "leucine zipper" (Landschulz et al. 1988). This structure is characterized by the repeat -(leu-6 residues)-, repeating at least 4 times. Although, at present, there is no experimental evidence to support the importance of this repeating pattern of leucines, the availability of the cloned genes will permit this hypothesis to be tested.

Existence of a jun-type Family of Proteins in Yeast

Compelling evidence has emerged indicating that a family of proteins recognizing the ARE exists (Rauscher et al. 1988) in mammalian cells. Southern blots using a *c-jun* probe (Bohmann et al. 1987) detect several different fragments indicating the presence of other genes related to *c-jun*. DNA sequencing of a mouse cDNA encoding the serum-stimulated jun-B protein demonstrates this factor has homology with v-jun although this protein is not mouse c-jun (Ryder et al. 1988).

Yeast contains no less than a two member ARE-binding family, GCN4 and yAP-1. The availability of the *YAP1* gene in cloned form combined with the manipulations possible with yeast genetics, have allowed us to construct a genetic background depleted of yAP-1 activity. DNA-affinity blots of protein from the disrupted *yap1* strain suggest that another ARE-binding protein of 45 kD is present in yeast. In addition, the GCRE detects three more distinct proteins of 69, 65 and 55 kD. A recent report demonstrates the presence of an 80 kD ARE-binding protein in mammalian cells as well as the previously observed set of proteins migrating from 39 to 47 kD (Rauscher et al. 1988). Experiments are underway to test the relationship of the 90 kD yAP-1 factor with factors in mammalian cells as well as to characterize further the novel ARE and GCRE binding activities we have detected.

The presence in yeast cells of multiple TGACT-recognizing proteins may provide an explanation as to the non-essentiality of

the *YAP1* gene. Loss of the yAP-1 protein may be compensated for by other proteins with overlapping DNA-binding specificity. This compensation would have to be a subtle effect since ARE-dependent β–galactosidase activity remains low in an *yap1* strain. It will be of interest to examine the phenotype of an *yap1, gcn4* double mutant to assess the effect of loss of the two known TGACT-recognizing proteins on yeast physiology.

A number of examples have appeared demonstrating that the transcriptional machinery is more similar between yeast and mammals than previously thought. Proteins analogous to mammalian CAAT factors (Chodosh et al. 1988b) as well as a TATA-binding protein (Buratowski et al 1988; Cavallini et al. 1988) have been detected in yeast. This conservation suggests that the basic eukaryotic process of transcription by RNA polymerase II may be studied in yeast to yield results extrapolable to higher cells. Bearing in mind the functional conservation between yeast and mammalian transcription, the analysis of transcription factors binding the ARE and GCRE in yeast may further the understanding of the role played in mammalian cells by c-jun and related factors.

Materials and Methods

Bacterial and Yeast Strains

All M13 subclones and most plasmids were constructed and propagated in JM101 (Messing et al. 1981). DH5 (Hanahan 1985) was used as the host for construction of the *GAL* fusion vector. MM294 (*cI*$^+$) was employed to construct the fusion between the λpL promoter and *YAP1* in pOTSV while AR68 (*cI857, htpR*) was used to permit heat inducible expression as described (Shatzman and Rosenberg 1987).

Yeast strains SEY6210 (*MATα, leu2-3,-112, ura3-52, his3-Δ200, trp1-Δ901, lys2-801, suc2-Δ9, Mel*$^-$) and SEY6211 (*Mata, leu2-3,-112, ura3-52, his3-Δ200, trp1-Δ901, ade2-101, Mel*$^-$) were obtained from S. Emr, California Institute of Technology. The diploid formed by crossing these two strains was designated SEY6210.5. This diploid strain was employed for the initial gene disruption experiments. In order to avoid any spurious results due to effects of the backgrounds of the two strains used to construct the diploid, we also disrupted the *YAP1* gene in SEY6210, thus generating an isogenic pair of *YAP1:yap1* strains. The *yap1* strain was designated SM9. Standard yeast genetic techniques were employed (Sherman et al, 1979). Plasmids were introduced into yeast cells by the method of Beggs (Beggs, 1978) or by a modification of the Li-acetate technique (Ito et al, 1983 , R. Aroian, California Institute of Technology).

DNA Manipulations

DNA fragments were cloned into plasmid and M13 vectors by standard techniques (Maniatis et al. 1982). λDNA was prepared by the use of LambdaSorb (Promega Biotec). The *YAP1* gene disruption plasmid, pSM25, was constructed by inserting a BglII linker (New England Biolabs, CAGATCTG) into the HpaI site of the 2.5 Kb EcoRI fragment subcloned in pUC19. The 1.7 kb BamHI *HIS3* fragment was then inserted into this BglII site. pSM25 was cleaved with EcoRI prior to yeast transformation.

The *GAL-YAP1* gene fusion was constructed by attaching BglII linkers to the 1.9 kb EcoRI fragment of *YAP1* and inserting this fragment into the BamHI site of pSEYC68-GAL (Emr et al, 1986) to form pSEYC68-GALR1.9 . Next, a BamHI/SalI fragment containing the complete carboxy terminus was inserted into BamHI/SalI cleaved pSEYC68-GALR1.9 to reconstitute the 3' end of *YAP1*. This construct was named pGAL19. Transformants bearing pGAL19 could be selected on the basis of the *URA3* gene in the pSEYC68 backbone (Emr et al, 1986). The *ARE-lacZ* fusion gene was carried on a *LEU2* containing derivative of pLGΔBS (Harshman et al, 1988). The StuI/SalI fragment of pLGΔBS was replaced by the 2.0 kb HpaI/SalI LEU2 fragment to form pCS10. Two ARE oligonucleotides were inserted into the BglII site to construct pSM38. pCS10 was constructed by C. Schoenherr of this laboratory. Transformants containing both pGAL19 and pSM38 could be detected by the Ura^+, Leu^+ phenotype.

Bacterial Production of yAP-1

Appropriate AR68 transformants were grown at 30° in 2XYT media and 100 μg/ml ampicillin until the A_{600}=0.6. An equal volume of 65° media was added and the cultures placed at 42° for 80 min. After harvesting, the pellet was resuspended at 2ml B-buffer (0.1 M Tris, pH 7.5, 0.2 M KCl, 10% glycerol, 10 mM B-mercaptoethanol) per gram cells. PMSF was added to 1 mM and the cell suspension digested with 750 μg/ml lysozyme for 30 min at 4°. The cells were then sonicated 6 times for 10 sec each time at 350 watts on ice. Debris was removed by centrifugation. Solid ammonium sulfate was added to 0.38 gm/ml original volume with mixing for 30 min. The precipitated proteins were pelleted, resuspended in 2 volumes per gram original cell weight of 0.05 HGKE buffer (Harshman et al. 1988) and dialyzed against the same buffer overnight. This extract could be used directly for footprinting.

S1 Mapping

Yeast RNA was prepared from SEY6210.5 by the method of Hinnebusch and Fink (Hinnebusch and Fink 1983). This RNA was chromatographed on oligo d(T) cellulose to enrich for poly(A)+ RNA as described (Maniatis et al. 1982). The 5' end-labeled probe was prepared by kinasing an oligonucleotide corresponding to +183 to +203 in *YAP1*. This labeled oligonucleotide was annealed to an M13mp18 clone of the 500 bp EcoRI/BamHI fragment from the 5' end of *YAP1*. The primer was extended with cold dNTPs and the

Klenow fragment as described (Greene 1987). The radiolabeled strand was isolated from a strand separation gel (Maxam and Gilbert 1980). Hybridization and S1 nuclease treatment were performed as in (Moye et al. 1985).

Yeast Biochemistry

Selected segregants were grown in a 10 l New Brunswick fermenter in yeast minimal media (Sherman et al. 1979) supplemented with 1% yeast extract. Yeast nuclear extracts were made and heparin-agarose columns run as in (Wiederrecht et al. 1987). The footprinting template used contains an ARE as well as a GCRE and is derived from pARE/GCRE (Harshman et al. 1988). DNA-affinity blots were performed as described (Harshman et al. 1988) using radiolabeled and ligated oligonucleotides as probes.

Other Methods

All gene disruptions and segregants were verified by Southern blotting (Maniatis et al. 1982). DNA sequencing employed the dideoxy method of Sanger (Sanger et al. 1977) with the use of (α-^{35}S) dATP as the label (Biggin et al. 1983). The λgt11 library was screened using standard protocols (Mierendorf et al. 1987) . β-Galactosidase assays were performed as before (Harshman et al. 1988).

ACKNOWLEDGEMENTS

We thank C. Schoenherr and P. Herman for plasmids and discussions. We are indebted to Dr. G. Wiederrecht for his assistance during the course of this work. We thank Drs. Z.-Y. Zhang-Keck, K. Hughes and G. Wiederrecht for their critical review of this manuscript. W.S.M. is a Postdoctoral Fellow of the American Cancer Society (PF2901). This research was supported by a grant from the American Cancer Society (NP604) to C.S.P. C.S.P. is a Rita Allen Foundation Scholar.

FIGURE LEGENDS

Figure 1. Restriction Map of *YAP1* Genomic Region and λgt11 subclones. The arrow indicates the bounds and polarity of the yAP-1 coding sequence. Restriction sites are indicated as: B, BamHI; E, native yeast genomic EcoRI site; *E, synthetic EcoRI linker; H, HpaI; K, KpnI.

Figure 2. DNA Sequence of the *YAP1* Gene. The numbering is relative to position +1 being the major 5' end for *YAP1* mRNA. The one letter code is used for the amino acid sequence.

Figure 3. S1 Nuclease Mapping of the 5' End of the *YAP1* mRNA. The same *YAP1*-specific oligonucleotide used to synthesize the S1 probe was also employed to generate a dideoxy sequence ladder from the M13 template used for probe synthesis. Lanes 1-4: Dideoxy G,A,C,T reactions on M13 mp18:RI/Bam 500 clone, Lane 5: S1 digestion of probe annealed to yeast tRNA, Lane 6: S1 digestion of probe annealed to yeast poly(A)$^+$ RNA. The major 5' terminus (position +1) is indicated.

Figure 4. (A) Amino Acid Sequence Alignment of c-jun, GCN4 and yAP-1. Only the region corresponding to the DNA-binding domains of the respective protein is shown. The positions corresponding to the repeating pattern of leucines are designated with asterisks. This alignment represents the best fit for the three proteins. Amino acid identities are indicated by upper-case letters and two dots while conservative replacements are shown by a single dot. The one letter amino acid code is used throughout. The

numbers refer to the amino acid position in each factor where the alignment stops.

(B) Location of DNA-Binding Domains of c-jun, GCN4 and yAP-1. Representation of the position of the region encoding the DNA-binding domain relative to the rest of the protein sequence. The numbers refer to the length of the respective factor. N, amino terminus of the protein; C, carboxy terminus; DNA, DNA-binding domain.

Figure 5. DNAase I Footprint Comparison of Yeast and Bacterially-Produced yAP-1. The ARE and GCRE are denoted by brackets. Footprint reactions were performed as described (Wiederrecht et al. 1987) using the indicated amount of extract. Lane 1: no protein, Lane 2: 10 µl of affinity-purified yAP-1, Lanes 3-5: 1,5,10 µl of crude extract from AR68/pPL-*YAP1*, Lanes 6-7: 1,5 µl of crude extract from AR68/pBH1.

Figure 6. Gene Disruption of *YAP1*

(A) Schematic showing the integration of the *HIS3* insertion mutant into the wild-type 6.1 kb EcoRI fragment. Restriction sites are indicated as in Figure 1. ATG, start codon for *YAP1* ORF; TAA, termination codon.

(B) Southern blot of yeast strains used in disrupting *YAP1*. All DNA preparations (Winston et al. 1983) were cleaved with EcoRI. The blot was probed with the nick-translated 2.5 kb EcoRI fragment. The sizes of the detected fragments are shown. Lanes 3-6 represent DNA preparations from each segregant of the same tetrad. Lane 1:

SEY6210.5, Lane 2:SM81 (heterozygous *YAP1* disruption in SEY6210.5), Lane 3: *YAP1, his3* segregant, Lane 4: *yap1::HIS3, his3* segregant, Lane 5: *yap1::HIS3, his3* segregant, Lane 6: *YAP1, his3 segregant.*

Figure 7. DNA-Affinity Blot of *YAP1* and *yap1* Heparin-Agarose Fractions. Equal amounts of protein derived from either the *YAP1* or *yap1* strain were loaded onto a 10% SDS-polyacrylamide gel and processed for a DNA-affinity blot as described (Harshman et al. 1988). Lanes 1 and 2 were probed with the ARE while lanes 3 and 4 were probed with the GCRE. Lanes 1 and 3 represent protein from the *YAP1* cell while lanes 2 and 4 contain protein from the *yap1* strain. The molecular weight standards (Rainbow Markers, Amersham) are indicated.

REFERENCES

Angel, P., M. Imagawa, R. Chiu, B. Stein, R.J. Imbra, H.J. Rahmsdorf, C. Jonat, P. Herrlich, and M. Karin. 1987. Phorbol ester-inducible genes contain a common *cis* element recognized by a TPA-modulated *trans*-acting factor. *Cell* **49**: 729-739.

Angel, P., E. A. Allegretto, S. T. Okino, K. Hattori, W. J. Boyle, T. Hunter, and M. Karin. 1988 Oncogene *jun* encodes a sequence-specific *trans*-activator similar to AP-1. *Nature* **332**: 166-171.

Arndt, K. and G. Fink. 1986. GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5'TGACT3' sequences. *Proc. Natl. Acad. Sci. U.S.A.* **83**: 8516-8520.

Beggs, J.D. 1978. Transformation of yeast by a replicating hybrid plasmid. *Nature* 275: 682-688.

Biggin, M., T. Gibson, and G. Hong. 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. U.S.A.* **80**: 3963-3965.

Bohmann, D., T.J. Bos, A. Admon, T. Nishimura, P.K. Vogt, and R. Tjian. 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.

Buratowski, S. , S. Hahn, P.A. Sharp, and L. Guarente. 1988. Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature* **334**: 37-42.

Cavallini, B., J. Huet, J.-L. Plassat, A. Sentenac, J.-M. Egly, and P. Chambon. 1988. A yeast activity can substitute for the HeLa cell TATA box factor. *Nature* **334**: 77-80.

Chodosh, L.A., A.S. Baldwin, R.W. Carthew, and P.A. Sharp. 1988a. Human CCAAT-binding proteins have heterologous subunits. *Cell* **53**: 11-24

Chodosh, L.A., J. Olesen, S. Hahn, A.S. Baldwin, L. Guarente, and P.A. Sharp. 1988b. A yeast and human CCAAT-binding protein have Heterologous subunits that are functionally interchangeable. *Cell* **53**: 25-35.

Dorn, A., J. Bollekens, A. Staub, C. Benoist, and D. Mathis. 1987. A multiplicity of CCAAT box-binding proteins. *Cell* **50**: 863-872.

Emr, S., Vassarotti, A., Garrett, J., Geller, B., Takeda, M., and Douglas, M. 1986. The amino terminus of the yeast F_1-ATPase β-subunit functions as a mitochondrial import signal. *J. Cell Biol.* 102; 523-533.

Graves, B.J., P.F. Johnson, and S.L. McKnight. 1987. Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. *Cell* **44**: 565-576.

Greene J.M. 1987. S1 analysis of mRNA using single-stranded DNA probes. In *Current Protocols in Molecular Biology*, F.M. Ausubel, R. Brent, R. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds.(John Wiley and Sons, New York): pp 4.6.1.-4.6.9.

Hanahan, D. 1985. Techniques for Transformation of E. coli. In *DNA Cloning: Volume I*, D.M. Glover, ed. (IRL press, Oxford, England): pp109-135.

Harshman, K.D., W.S. Moye-Rowley, and C.S. Parker. 1988. Transcriptional activation by the SV40 AP-1 recognition element in yeast is mediated by a factor similar to AP-1 that is distinct from GCN4. *Cell* **53**: 321-330.

Hill D.E., I.A. Hope, J.P. Macke,, and K. Struhl. 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. and G. Fink. 1983. Repeated DNA sequences upstream from *HIS1* also occur at several other co-regulated genes in Saccharomyces cerevisiae. J. *Biol. Chem.* **258**: 5238-5247.

Hinnebusch, A.G. 1984. Evidence for translational regulation of the activator of general amino acid control in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **81**: 6442-6446.

Hope,I. and K. Struhl. 1986. Functional dissection of eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* **46**: 885-894.

Ito, H., Y. Fukuda, K. Murata, and A. Kimura. 1983. Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* **153**: 163-168.

Johnston, M., and R. Davis. 1984 Sequences that regulate the divergent *GAL1-GAL10* promoter in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **4**: 1440-1448.

Jones, K. A., J. T. Kadonaga, P. J. Rosenfeld, T. J. Kelly, and R. Tjian. 1987 A cellular DNA-binding protein that activates eukaroytic transcription and DNA replication. *Cell* **48**: 79-89.

Jones, N.C., P.W.J. Rigby, and E.B. Ziff 1988. *Trans*-acting protein factors and the regulation of eukaryotic transcription: lessons from studies of DNA tumor viruses. *Genes Dev.* **2**: 267-281.

Kammerer, B., A. Guyonvarch, and J.D. Hubert 1984. Yeast regulatory gene *PPR1*. *J. Mol. Biol.* **180**: 239-250.

Kozak, M. 1984. Selection of initiation sites by eukaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. *Nucl. Acids Res.* **12**: 3873-3893.

Landschulz, W.H., P.F. Johnson, and S.L. McKnight. 1988a. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* **240**: 1759-1764.

Landschulz, W., P. Johnson, E. Adashi, B. Graves, and S. McKnight. 1988b. Isolation of a recombinant copy of the gene encoding C/EBP. *Genes Dev.* **2**: 786-800.

Lee, W. A. Haslinger, M. Karin, and R. Tjian. 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., P. Mitchell and R. Tjian. 1987b. Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell* **49**: 741-752.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982 *Molecular Cloning: A Laboratory Manual.* (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory)

Maxam, A. and W. Gilbert. 1980. Sequencing end labeled DNA. *Meth. Enzymol.* **65**: 499-560.

Messing, J., R. Crea, and P. Seeburg. 1981. A system for shotgun DNA sequencing. *Nucl. Acids Res.* **9**: 309-321.

Mierendorf, R. C. Percy, and R. Young. 1987. Gene isolation by screening λgt11 libraries with antibodies. *Meth. Enzymol.* **152**: 458-469.

Miskimins, W., M. Roberts, A. McClellend, and F. Ruddle. 1985. Use of a protein-blotting procedure and a specific DNA probe to identify nuclear proteins that recognize the promoter region of the transferrin receptor gene. *Proc. Natl. Acad. Sci. U.S.A.* **82**: 6741-6744.

Mitchell, P.J., C. Wang, and R. Tjian. 1987. Positive and negative regulation of transcription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T-antigen. *Cell* **50**: 847-861.

Moye, W.S., N. Amuro, J.K.M. Rao, and H. Zalkin. 1985. Nucleotide sequence of yeast *GDH1* encoding NADP-dependent glutamate dehydrogenase. *J. Biol. Chem.* **260**: 8502-8508.

Rauscher,III,F.J., D.R. Cohen, T. Curran, T.J. Bos, P.K. Vogt, D Bohmann, R. Tjian, and B.R. Franza, Jr. 1988. Fos-associated protein p39 is the product of the *jun* proto-oncogene. *Science* **240**: 1010-1016.

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

Ryder, K., L.F. Lau, and D. Nathans. 1988. A gene activated by growth factors is related to the oncogene *v-jun*. *Proc. Natl. Acad. Sci. U.S.A.* **85**: 1487-1491.

Shore, D. and K. Nasmyth. 1987. Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. *Cell* **51**: 721-732.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**: 5463-5467

Shatzman, A. and M. Rosenberg. 1987 Expression, identification and characterization of recombinant gene products in Escherichia coli. *Meth. Enzymol.* **152**: 661-673.

Sherman, F., G. Fink, and J. Hicks. 1979 *Methods in Yeast Genetics*. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Struhl, K. 1985. Nucleotide sequence and transcriptional mapping of the yeast *PET56-HIS3-DED1* gene region. *Nucl. Acids Res.* **13**: 8587-8601.

Struhl, K. 1987. The DNA-binding domains of jun oncoprotein and the yeast GCN4 transcriptional activator protein are functionally homologous. *Cell* **50**: 841-846.

Thireos, G., M.D. Penn, and H. Greer. 1984. 5' untranslated sequences are required for the translational control of a yeast regulatory gene. *Proc. Natl. Acad. Sci. U.S.A.* **81**: 5096-5100.

Wiederrecht, G., D.J. Shuey, W.A. Kibbe, and C.S. Parker. 1987. The Saccharomyces and Drosophila heat shock transcription factors are identical in size and DNA binding properties. *Cell* **48**: 507-515.

Wiederrecht, G., D. Seto, and C.S. Parker. 1988. Isolation of the gene encoding the Saccharomyces cerevisiae heat shock transcription factor. *Cell* **54**: 841-853.

Winston, F., F. Chumley, and G. Fink. 1983. Eviction and transplacement of mutant genes in yeast. *Meth. Enzymol.* **65**: 211-228.

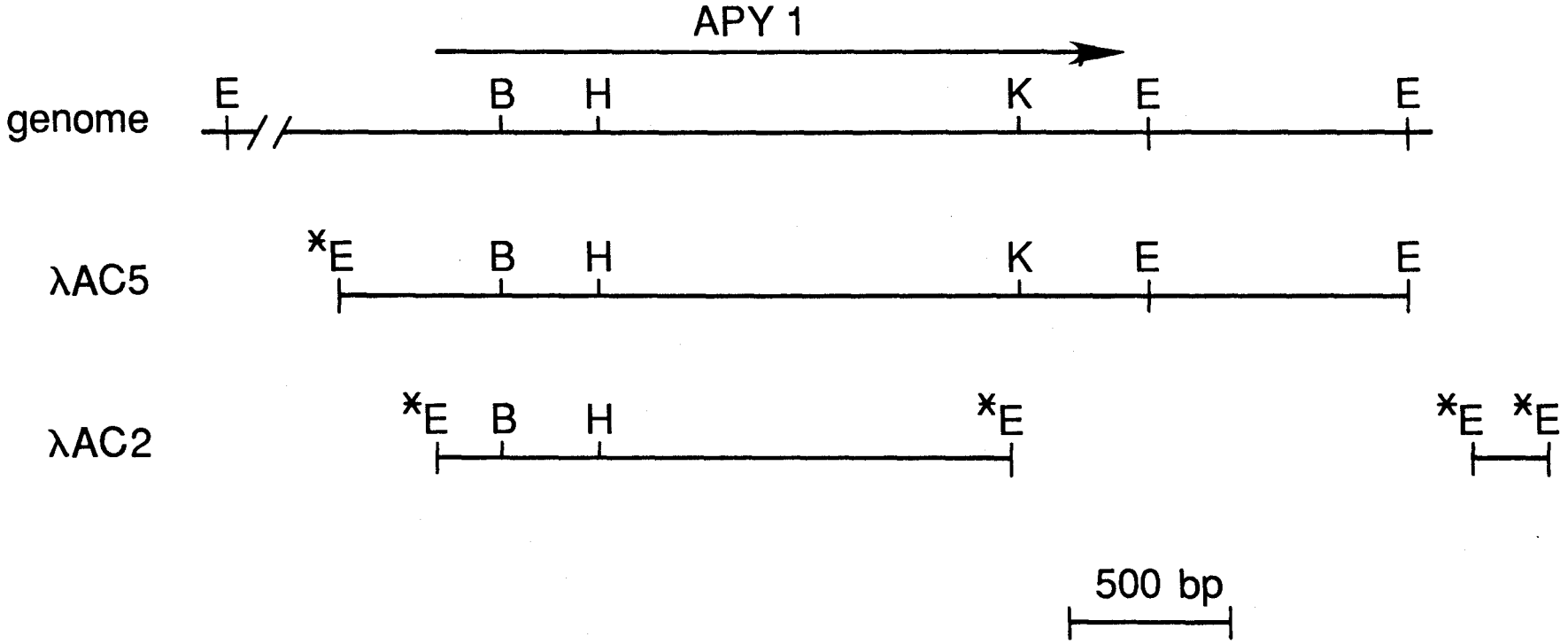
Yanisch-Perron, C., J. Viera, and J. Messing. 1985. Improved M13 phage cloning vectors and strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103-119.

Young, R., and R. Davis. 1983. Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad. Sci. U.S.A.* **80**: 1194-1198.

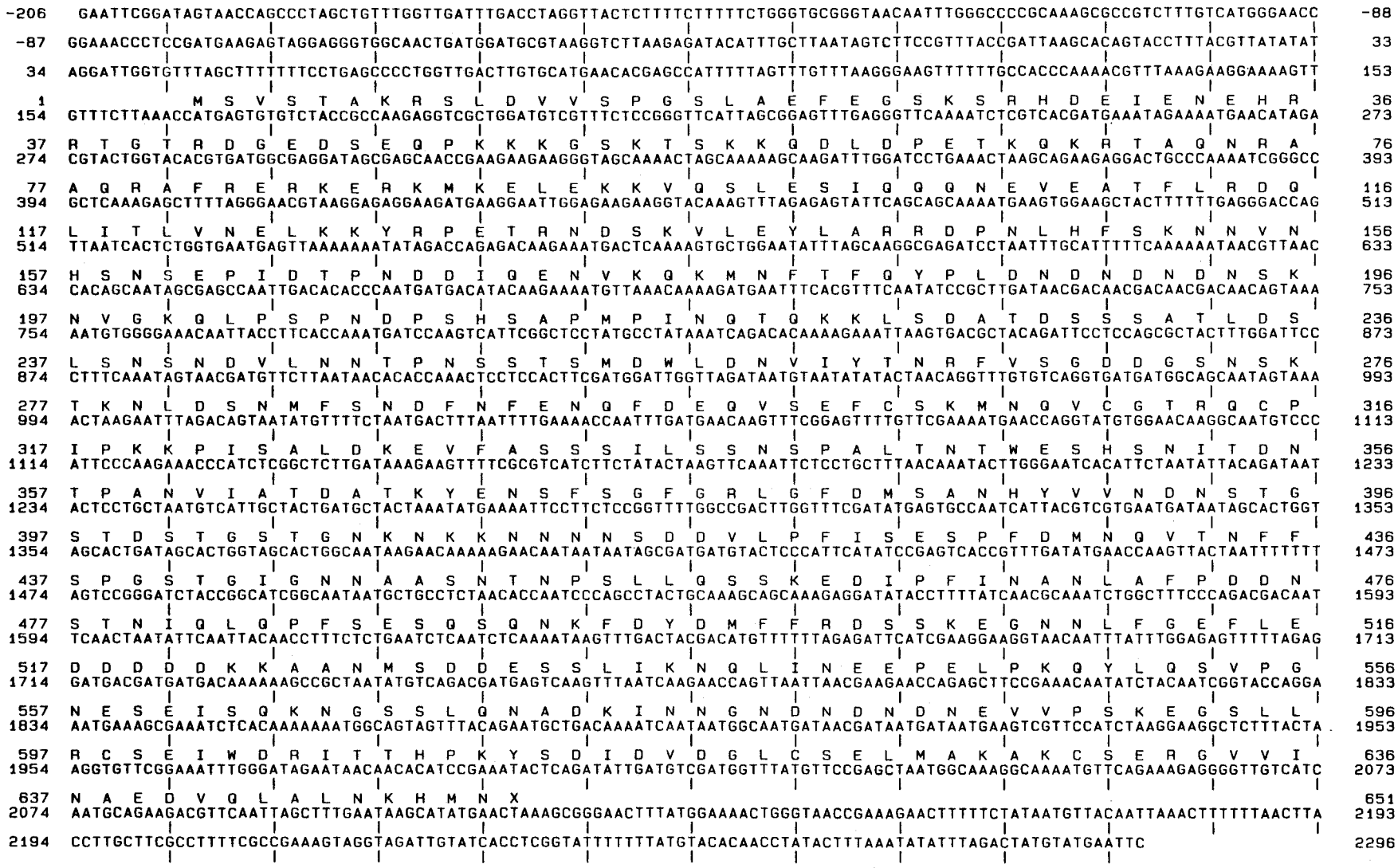
Table 1. Transcriptional Activation in Yeast by Cloned *YAP1*.

	β-Galactosidase Activity	
YAP1 Allele	glucose	galactose
YAP1	130	70
yap1::HIS3	0.2	0.11
GAL-YAP1, yap1::HIS3	0.15	103

The transformed yeast strains were grown with either 2% glucose or 3% galactose as the carbon source. β-galactosidase activities are expressed as nmol/min mg protein and were performed as in (Harshman et al. 1988). The *YAP1* alleles are described in the text.

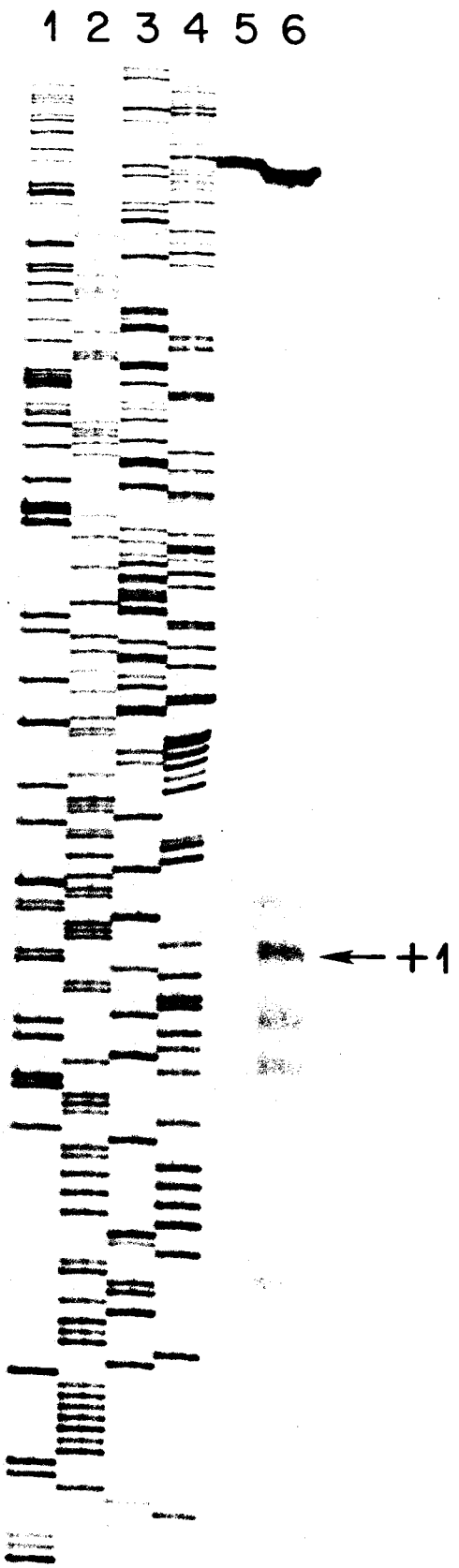


140
Fig 1



141
Fig 2

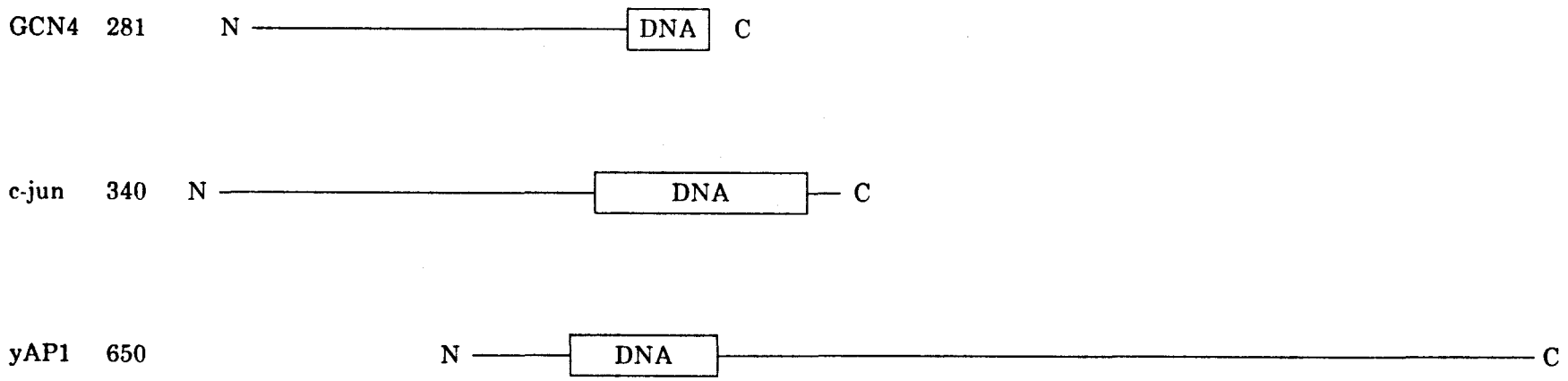
Fig 3



A

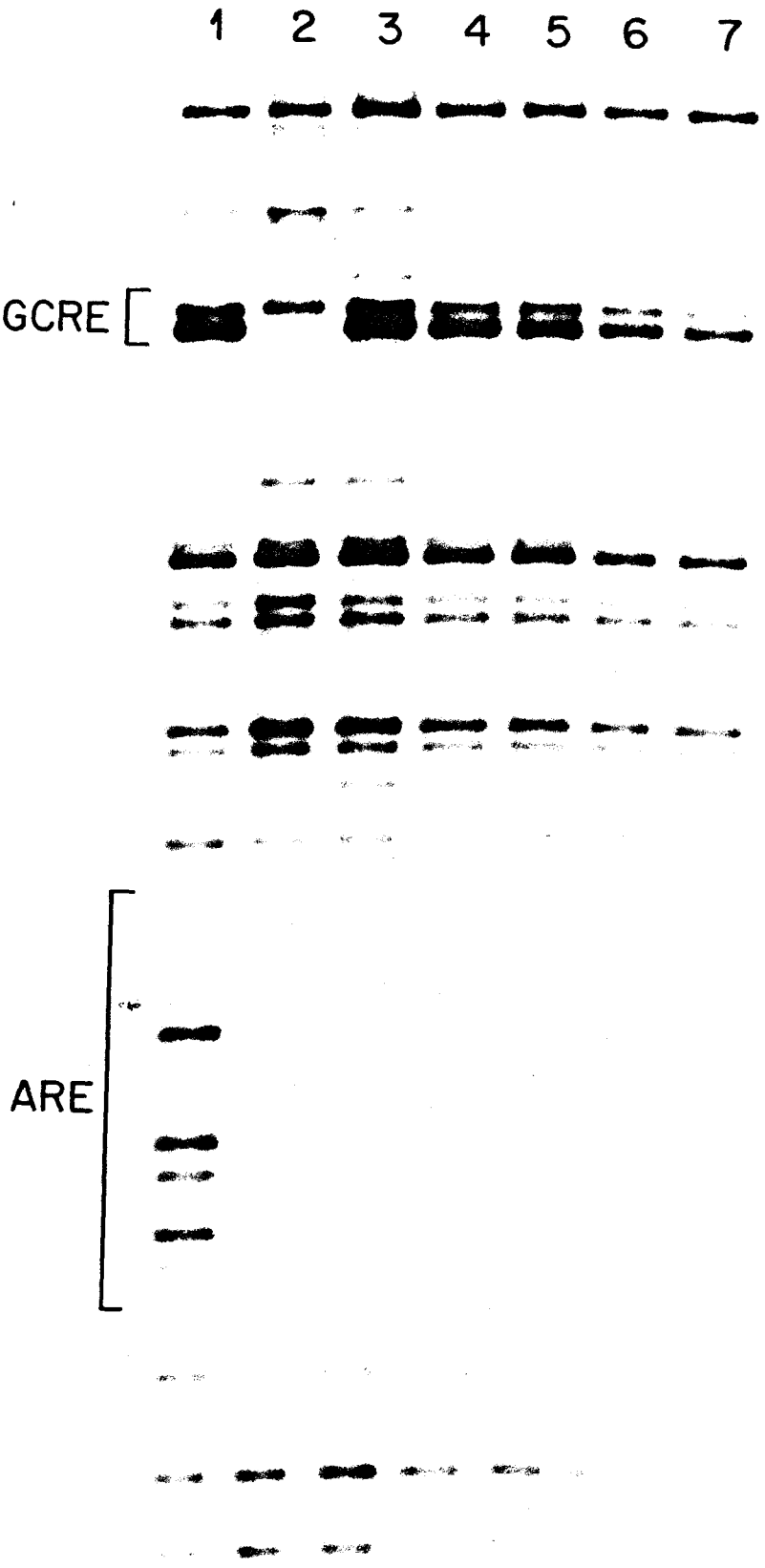
```
                                          *       *       *       *       *
yAP1 dldpetkqkrtaqNraAqrafReRKerkmkeLEkKVqsLesiqqqneveatfLrdqlitlvn 122
GCN4 pessdpaalkrarNteAarrsRaRKlqrmkqLEdKVeeLlsknyhlenevarLkklvger   281
     . .    . .  .:  :  . : :: ...  ::.::..: . . .       :.. . .  .
CJUN sqerikaerkrmrNriAaskcRkRKleriarLEeKVktLkaqnselastanmLreqvaqlkq 118
```

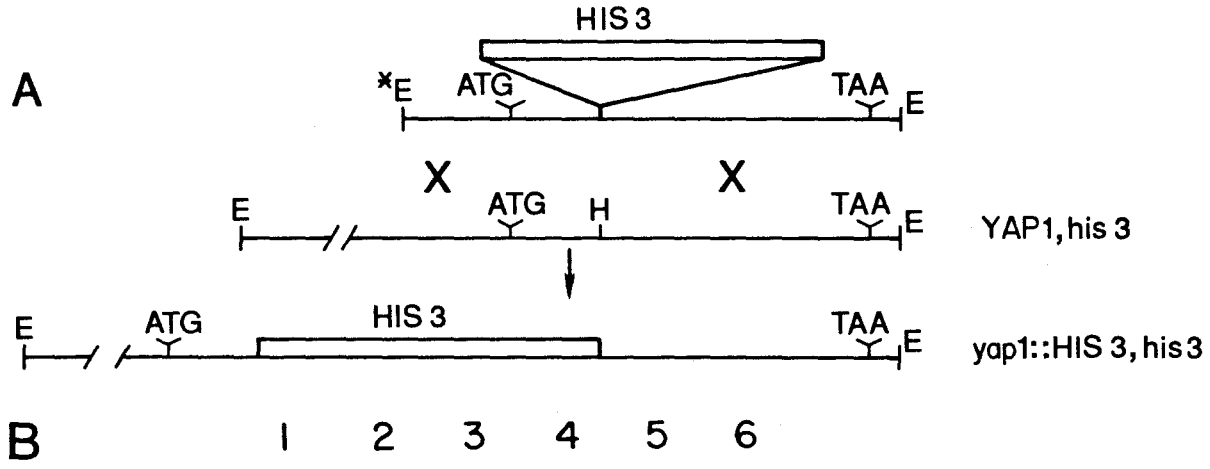
B



143
Fig 4

Fig 5




A
HIS 3
*E
ATG
TAA
E
X
X
E
ATG
H
TAA
E
YAP1, his 3
E
ATG
HIS 3
TAA
E
yap1::HIS 3, his 3
B
1
2
3
4
5
6


7.8 Kb
6.1 Kb

Fig 7