

# The isolation of transcription factors from $\lambda$ gt11 cDNA expression libraries: human steroid 5 $\alpha$ -reductase 1 has sequence-specific DNA binding activity

Kevin Gaston and Mike Fried

Eukaryotic Gene Organisation and Expression Laboratory, Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK

Received August 26, 1992; Revised and Accepted October 30, 1992

## ABSTRACT

**The Surf-1/Surf-2 bi-directional promoter contains binding sites for at least three transcription factors (Su1, Su2, and Su3). By screening a  $\lambda$ gt11 HeLa cell cDNA expression library with a concatenated Su2 factor binding site, we isolated a cDNA which encodes a protein with sequence-specific DNA binding activity. Gel retardation assays showed that the cloned factor binds specifically to the Su2 factor binding site present in the human Surf-1/Surf-2 promoter but not to an Su2 site containing mutated base pairs. Co-transfection experiments demonstrated that the cloned cDNA had little or no effect on the expression of a reporter gene under the control of multiple Su2 factor binding sites. Similarly a fusion protein in which the acidic activation domain of HSV VP16 was linked to the cloned factor had no effect, implying that the factor does not function as a DNA binding protein *in vivo*. DNA sequence analysis revealed that the cloned cDNA is identical to that of human steroid 5 $\alpha$ -reductase 1, an enzyme which converts testosterone to dihydrotestosterone. These results are discussed with respect to other putative transcription factors which have been isolated from cDNA expression libraries on the basis of their sequence-specific DNA binding activity.**

## INTRODUCTION

In higher eukaryotes genes are usually separated by many thousands of base pairs. The Surf-1 locus is unusual in that it contains at least 6 tightly clustered housekeeping genes: Surf-1 to Surf-6 (1). The mouse Surf-1 and Surf-2 genes are divergently transcribed from a bi-directional transcription unit in which their heterogeneous transcription start sites are separated by a maximum of 73bp (2). We have previously identified and characterised three factor binding sites within this promoter region (Su1, Su2 and Su3) which are required for accurate and efficient expression of both genes (3). These three factor binding sites are conserved between mouse and human and bind factors present in both mouse and human nuclear extracts (4). In order to understand the regulation of this class of cellular genes, we are

attempting to isolate and clone these factors. The Su1 factor binding site has sequence homology to the binding site of YY1 (also known as  $\delta$  and NF-E1) which has recently been cloned by three groups (5–7). The Su2 and Su3 factor binding sites have sequence homology to the binding site of the ETS family of DNA binding proteins (8).

The ability of proteins to bind to DNA in a sequence-specific manner has been used to clone a number of mammalian transcription factors from  $\lambda$ gt11 cDNA expression libraries, for example, NF $\kappa$ B (9) and CREB (10). In this communication, we report the isolation from a HeLa cell cDNA expression library of a cDNA which encodes a protein capable of specific binding to the human Su2 factor binding site. The encoded protein also binds to the equivalent factor binding site present in the mouse promoter, but not to factor binding sites containing mutations previously shown to prevent the binding of factors present in nuclear extracts (3). However, further experiments indicated that the cDNA product is unable to interact with the Su2 site *in vivo* and may not be a transcription factor.

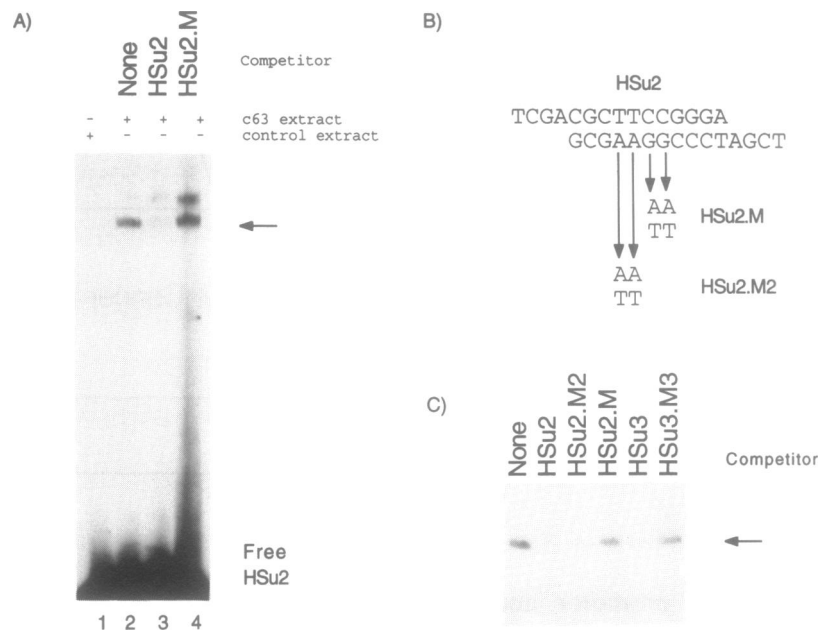
## MATERIALS AND METHODS

### Screening a cDNA expression library

A HeLa cell  $\lambda$ gt11 cDNA expression library (Stratagene) was screened with a concatenated oligonucleotide corresponding to the human Su2 factor binding site according to the method of Singh et al. (1989). The sequence of the oligonucleotide used for the screen is shown in Figure 1B. The bacterial strain used in the screen and in all subsequent work was *Escherichia coli* XL1-blue (Stratagene).

### Gel retardation assays

Crude bacterial cell extracts were prepared according to the method of Singh et al. (1989) from strains containing cDNA's cloned in the plasmid pBlueScript KS (–) (Stratagene). HeLa cell nuclear extracts were prepared exactly as previously described (3). Single-stranded synthetic oligonucleotides (100ng) were labelled with  $\gamma$ [<sup>32</sup>P]-ATP using T4 polynucleotide kinase. After



**Figure 1.** A gel retardation assay using bacterial cell extracts. (A) Crude bacterial extracts were incubated with a labelled oligonucleotide carrying a single copy of the human Su2 factor binding site (HSu2) (the sequence of which is shown in part B). Free and bound DNA were separated and visualised as described in the text. The source of the bacterial extract used in each case and the competitor oligonucleotide (200ng) present are as indicated in the Figure. (B) The sequence of the HSu2 and mutant HSu2 oligonucleotides. The HSu2.M and HSu2.M2 oligonucleotides have two base changes relative to the HSu2 sequence. (C) The same experiment as in (A) but performed using a number of double stranded competitor oligonucleotides; the HSu3 factor binding site has homology to the HSu2 sequence (4).

hybridisation to the corresponding oligonucleotide, unincorporated label was removed using a Sephadex G-50 spin column. Labelled double-stranded oligonucleotide (10,000cpm) was incubated with either bacterial cell extract in standard buffer (40mM Tris pH 8.0, 10mM MgCl<sub>2</sub>, 100mM KCl, 1mM DTT, and 0.1mg/ml bovine serum albumin), or HeLa cell nuclear extract in standard buffer supplemented with 1μg poly (dI-dC). After 20 minutes incubation at room temperature, free and bound DNA were separated on 6% polyacrylamide gels run in TBE as previously described (11, 12). The gels were dried onto Whatman 3MM paper and exposed to Kodak X-OMAT film at -70°C with an intensifying screen.

#### DNAs used in this study

The DNAs used in this study are shown diagrammatically in Figure 2. The acidic transcription activation domain (codons 410 to 490) of HSV VP16 was excised on a *Bam*HI-*Eco*RI fragment from plasmid pSD06a (13) and cloned into pBlueScript KS (-) carrying either c63 or c70. This generated the VP16 fusion proteins c63-VP16 and c70-VP16. These constructions, along with c63 and c70, were transferred on *Bam*HI-*Cl*ai fragments into the mammalian expression plasmid pMLVβplink (13) such that the expression of each protein was under the control of the human β-globin promoter activated by the MLV enhancer. The reporter plasmid was pCAT-Basic (Promega) containing three copies of either the wild-type human Su2 factor binding site (pCAT-HSu2) or three copies of a mutated Su2 site (pCAT-HSu2.M) cloned into the *Sal*I site upstream of the chloramphenicol acetyltransferase (CAT) reporter gene. The synthetic oligonucleotides used to create these plasmids were:

HSu2 5' *Sal*I-[CGCTTCCGGGA]×3 -*Sal*I  
and HSu2.M 5' *Sal*I-[CGCTTAAGGGA]×3 -*Sal*I.

The c70 cDNA was obtained from the HeLa cell λgt11 cDNA library by PCR using specific oligonucleotide primers based on the published sequence of the human steroid 5α-reductase 1 cDNA (14).

#### Cell culture, transfections and CAT assays

HeLa cells were grown in 10% fetal calf serum (2×10<sup>6</sup> cells per 14cm petri dish) before transfection with 20μg of total plasmid DNA by electroporation (460Volts-960uF). After 48hrs, CAT activity was determined as described by Gorman et al. (1982)(15) and quantified by scintillation counting after three extractions with xylene (16). The β-galactosidase expressing plasmid pRSV-βgal (5μg) was included in each transfection as an internal control of transfection efficiency and the CAT expressing plasmid pMLV-CAT (5μg) was used as a positive control for the CAT assay.

#### DNA sequence analysis

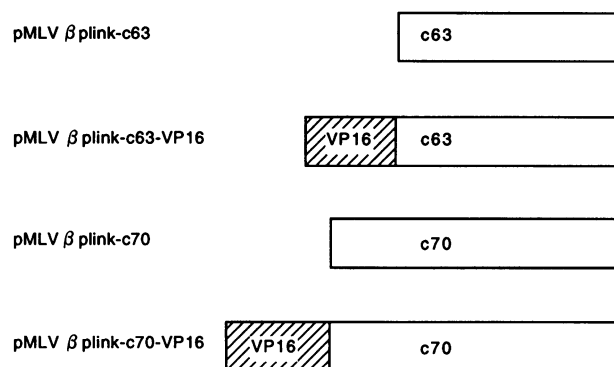
The DNA sequences of the c63 and c70 cDNA's and the other constructions used in this work were determined by the dideoxy-chain elongation method (17) using a Sequenase kit (USB) according to the manufacturer's instructions. Both strands of the templates were sequenced using specific oligonucleotide primers.

## RESULTS

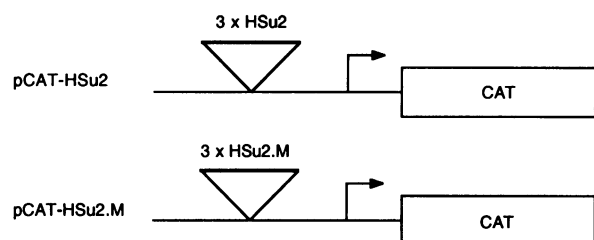
#### The isolation of a clone with sequence-specific DNA binding activity

A λgt11 HeLa cell cDNA expression library was screened with a <sup>32</sup>P-labelled concatenated oligonucleotide carrying the human Su2 factor binding site. From an initial screen of 500,000 plaques, 6 positive clones were identified. Only one of these clones (c63) bound the multimerised Su2 factor binding site after secondary

## A) Effectors



## B) Reporters



**Figure 2.** Constructs used for transfection experiments. (A) The construction of the pMLV $\beta$ plink series of plasmids is described in the text. The hatched box represents the HSV VP16 transcription activation domain, open boxes represent the c63 and c70 open reading frames. (B) The reporter plasmids used in this study were derivatives of pCAT-basic and contained three copies of the human Su2 factor binding site (pCAT-HSu2) or three copies of the mutant factor binding site, HSu2.M (pCAT-HSu2.M). The arrow represents the transcription start point.

and tertiary screens and gave a strong positive signal. *In vivo* excision was used to recover this clone in the pBlueScript KS (–) plasmid. Restriction enzyme analysis showed that the c63 cDNA was 1.7 kb in length.

Crude bacterial cell extracts made from bacteria carrying the c63 plasmid and from bacteria carrying pBlueScript KS (–) with no insert, were tested for sequence-specific DNA binding activity using a gel retardation assay. A labelled double-stranded oligonucleotide carrying a single copy of the human Su2 factor binding site was incubated with crude cell extract for 20 minutes at room temperature. Free and bound DNA were then separated on 6% polyacrylamide gels as previously described (3) and viewed by autoradiography (Figure 1). As can be seen from the figure, crude extracts from bacteria containing the vector plasmid with no insert contain no Su2 binding activity (Figure 1A lane 1). In contrast, extracts from bacteria containing the plasmid carrying the c63 clone produced a retarded band characteristic of sequence-specific DNA binding activity (Figure 1A lane 2). The addition of non-specific competitor (1 $\mu$ g poly (dI-dC)) had no effect on the retarded band. In contrast, the upper band in

**Table 1.** A comparison of the DNA binding specificities of HeLa cell nuclear extract and c63 containing bacterial extract

Competitor oligonucleotide	c63 bacterial cell extract	HeLa cell nuclear extract
HSu1	+/-	+
HSu2	+	+
HSu2.M	-	-
HSu2.M2	+	-
HSu3	+	+
HSu3.M3	-	-

(+) competes, (-) fails to compete, (+/-) partially competes

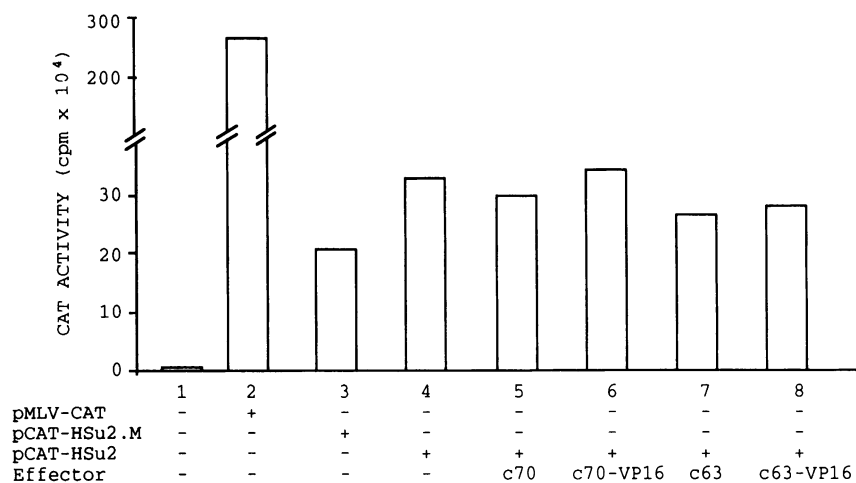
Figure 1A varied in intensity between experiments and was abolished by the addition of poly (dI-dC).

To examine the DNA binding specificity of the protein encoded by the c63 cDNA, bacterial extract was pre-incubated with a 1000-fold excess of competitor oligonucleotide carrying either the HSu2 factor binding site or a mutated HSu2 site (HSu2.M) which does not bind factors present in HeLa cell nuclear extracts (these oligonucleotides are shown in Figure 1B). The wild-type HSu2 site effectively competed for the DNA binding activity present in the bacterial extract (Figure 1A lane 3), whereas, the mutant site failed to compete (Figure 1A lane 4). This shows that the DNA binding activity present in the c63 extract is specific for the Su2 factor binding site. Competition experiments with a variety of oligonucleotide sequences, including the wild-type and mutant Su3 factor binding sites, showed that the isolated factor has similar DNA binding specificity to the Su2 factor present in HeLa cell nuclear extracts (Figure 1C). However, a mutant HSu2 site which does not compete for the Su2 binding activity present in HeLa cell nuclear extracts (HSu2.M2) does compete for the Su2 binding activity present in the bacterial extract (Figure 1C). The results of these competition assays are summarised in Table 1. These data show that the DNA binding specificity of the c63 encoded protein is not identical to that present in nuclear extracts.

### Sequence analysis of c63 and the isolation of a full length cDNA

DNA sequence analysis revealed that the c63 cDNA contained an open reading frame with the potential to encode a 159 amino acid protein fused to the N-terminal 37 amino acids of  $\beta$ -galactosidase. A comparison of the sequence of this protein to those present in the EMBL database was performed using the FASTA program (18). The c63 encoded protein is identical to the C-terminal 159 amino acids of a protein previously identified as human steroid 5 $\alpha$ -reductase 1 (14). Furthermore, the nucleotide sequence of the c63 clone proved to be identical to that of the steroid 5 $\alpha$ -reductase 1 cDNA.

The full length steroid 5 $\alpha$ -reductase 1 protein contains 259 amino acids and is encoded by a 2.2 kb mRNA. A full length cDNA clone (c70) encoding the steroid 5 $\alpha$ -reductase 1 protein was obtained using PCR from the HeLa cell cDNA library as described in the Methods. Crude extracts of bacterial cells containing pBlueScript carrying the full length c70 clone also contained Su2 binding activity (data not shown). Computer analysis of the protein sequence of steroid 5 $\alpha$ -reductase 1 revealed no motifs characteristic of known DNA binding proteins (for example, basic region-helix-loop-helix, zinc finger, or homeo-domain motifs).



**Figure 3.** The expression of pCAT-basic reporter plasmids in HeLa cells. The graph shows the levels of CAT activity found in HeLa cell extracts 48hrs after transfection with the plasmids indicated. Enzyme activity has been normalised with respect to transfection efficiency and in each case values represent the average of three independent experiments.

### The effect of the cloned factor on transcription *in vivo*

HeLa cells were transfected with a reporter plasmid in which three copies of the human Su2 factor binding site (HSu2) or three copies of a mutant site (HSu2.M) were placed upstream of the CAT gene: plasmids pCAT-HSu2 and pCAT-HSu2.M respectively. The levels of CAT activity present in cell extracts made 48 hours after transfection are shown in Figure 3. The presence of the wild-type HSu2 sites resulted in slightly higher levels of CAT activity than those obtained in the presence of the mutant sites (Figure 3 lanes 3 and 4). Presumably, this increase in CAT activity is due to endogenous Su2 binding factors which activate transcription of the CAT gene.

In order to determine whether the c70 cDNA (steroid 5 $\alpha$ -reductase 1) encoded a transcription factor, we co-expressed both the full length c70 protein and the truncated c63 protein with the CAT reporter plasmid containing three copies of the wild-type Su2 factor binding site. As can be seen from Figure 3 (lanes 5 and 7), the presence of either the c70 or c63 protein had little or no effect on the expression of the CAT gene showing that these proteins do not act as strong activators of transcription.

The c70 protein could be a repressor of transcription or require a co-activator. To discount these possibilities, we fused the c70 and c63 proteins to the strong transcription activation domain of HSV VP16 (19). When the VP16 activation domain is fused to heterologous DNA binding domains the resulting hybrid proteins act as sequence-specific transcription activator proteins in both mammalian and yeast cells (20). However, co-expression of the fusion proteins, c70-VP16 and c63-VP16, with the pCAT-HSu2 reporter plasmid did not increase CAT activity in cell extracts (Figure 3, lanes 6 and 8). Therefore, the c70 protein and the truncated c63 protein both fail to act as efficient DNA binding proteins in intact cells.

## DISCUSSION

By screening a  $\lambda$ gt11 HeLa cell cDNA expression library with a multimerised Su2 factor binding site, we isolated a clone which encodes sequence-specific DNA binding activity. The cloned factor binds to a single copy of the human Su2 factor binding

site but not to a Su2 site in which two base pairs have been mutated (Figure 1). Analysis of the DNA sequence of the cloned factor showed that the encoded protein had no features characteristic of a sequence-specific DNA binding protein but was identical to a protein previously identified as the enzyme steroid 5 $\alpha$ -reductase 1. There are at least three possible explanations for the sequence-specific binding activity of steroid 5 $\alpha$ -reductase 1. First, steroid 5 $\alpha$ -reductase 1 may be solely a transcription factor. Second, steroid 5 $\alpha$ -reductase 1 may be a bi-functional protein: other proteins which function both as enzymes and regulatory factors have been found, for example, the bacterial protein BirA which binds DNA (21) and the IRE-BP/aconitase protein which binds RNA (22). Finally, the sequence-specific DNA binding activity of steroid 5 $\alpha$ -reductase 1 might be detected *in vitro* by the sensitive gel retardation assay (11, 12) but might not occur in intact cells owing to either the weak nature of the interaction or the intracellular localisation of the protein away from the DNA.

The steroid 5 $\alpha$ -reductase 1 cDNA was originally cloned using an expression cloning system (23). Briefly, random cDNA pools were injected into *Xenopus* oocytes and, after overnight incubation, 5 $\alpha$ -reductase activity was assayed in cell free extracts. By repeated rounds of screening, a single cDNA was obtained which, when injected into oocytes, produced increased steroid 5 $\alpha$ -reductase activity. The cDNA cloned in this way was presumed to encode the enzyme steroid 5 $\alpha$ -reductase although the protein product of the cDNA had not been directly demonstrated to have steroid 5 $\alpha$ -reductase activity.

We reasoned that the steroid 5 $\alpha$ -reductase 1 cDNA cloned by Andersson et al. (1989), and subsequently selected from the HeLa cDNA library on the basis of DNA binding activity, could in fact be a transcription factor responsible for positive regulation of the steroid 5 $\alpha$ -reductase gene and possibly the Surf-1 and Surf-2 genes. However, co-transfection of steroid 5 $\alpha$ -reductase 1 and a reporter plasmid, in which the chloramphenicol acetyltransferase gene is under the control of multiple Su2 factor binding sites, demonstrated that steroid 5 $\alpha$ -reductase 1 had no effect on the level of CAT activity (Figure 3). In order to determine whether steroid 5 $\alpha$ -reductase 1 or the C-terminal region of steroid 5 $\alpha$ -reductase 1 could act as a DNA binding

protein in intact cells, we fused the acidic activation domain of HSV VP16 to the N-terminus of full length steroid 5 $\alpha$ -reductase 1 (c70-VP16) or the truncated c63 protein (c63-VP16). Co-expression of these fusion proteins with the reporter plasmid also had no effect on CAT activity, implying that the neither the steroid 5 $\alpha$ -reductase 1 protein nor the isolated C-terminal region of the protein can function as a DNA binding domain *in vivo*. The possibility remains that steroid 5 $\alpha$ -reductase 1 is a DNA binding protein but that its activity may be restricted to a particular tissue or cell type. Furthermore, like many other transcription factors, for example, myc and max (24), it may require a partner in order to bring about changes in gene expression. Another possibility is that the DNA binding activity of steroid 5 $\alpha$ -reductase 1 is masked by other proteins present in the cell which prevent DNA binding or nuclear localisation. Finally, it might be that the DNA binding activity of steroid 5 $\alpha$ -reductase 1 is only seen *in vitro*.

Our isolation of steroid 5 $\alpha$ -reductase 1 may not be first time that an enzyme has been cloned using this technique. Ray and Miller (1989) cloned a factor, which they called MBP1, from a HeLa  $\lambda$ gt11 cDNA library (25). The MBP1 factor binds to an oligonucleotide carrying a sequence present in the human myc promoter as judged by gel retardation assay and DNase I footprinting. Furthermore, MBP1 was shown to repress transcription from a reporter plasmid in which the expression of the CAT gene was under the control of the human myc promoter. However, our comparison of the nucleotide sequence of the MBP1 clone with that of the human alpha enolase cDNA (26) shows that the two sequences are 97% identical. These data show that the MBP1 clone is most probably identical to human alpha enolase and suggest that either the DNA binding and transcription repression activity of this protein is fortuitous or, less likely, the cytoplasmic enzyme alpha enolase (like BirA) is a bi-functional protein which acts as both an enzyme and a repressor of transcription. These results, and ours with steroid 5 $\alpha$ -reductase 1, demonstrate a possible problem in the use of DNA binding capability alone as a method for the isolation of transcription factors from cDNA expression libraries. Clearly *in vitro* DNA binding activity must be accompanied by a demonstration of function *in vivo* before the product of a cloned cDNA can be classified as a transcription factor.

## ACKNOWLEDGEMENTS

We wish to thank Steve Dalton and Richard Treisman for plasmids pSD06 and pMLV $\beta$ plink, David Bentley for pointing out the similarity between MBP1 and enolase, and Clive Dickson, Ken Garson, Sheela Jayaraman, Mike Owen, and Malcolm Parker for comments on the manuscript.

## REFERENCES

- Huxley, C., and Fried, M. (1990) *Mol. Cell. Biol.*, **10**, 605–614.
- Williams, T. J., and Fried, M. (1986) *Mol. Cell. Biol.*, **6**, 4558–4569.
- Lennard, A., and Fried, M. (1991) *Mol. Cell. Biol.*, **11**, 1281–1294.
- Lennard, A., Gaston, K., and Fried, M. (Manuscript in preparation).
- Shi, Y., Seto, E., Chang, L-S., and Shenk, T. (1991) *Cell*, **67**, 377–388.
- Park, K., and Atchison, M. L. (1991) *Proc. Natl. Acad. Sci.*, **88**, 9804–9808.
- Hariharan, N., Kelley, D. E., and Perry, R. P. (1991) *Proc. Natl. Acad. Sci.*, **88**, 9799–9803.
- Karim, F. D., Urness, L. D., Thummel, C. S., Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C., Maki, R. A., Gunther, C. V., Nye, J. A., and Graves, B. J. (1990) *Genes Dev.*, **4**, 1451–1453.
- Singh, H., LeBowitz, J. H., Baldwin, A. S., and Sharp, P. A. (1989) *Cell*, **52**, 514–523.
- Hoeffler, J. P., Meyer, T. E., Yun, Y., Jameson, J. L., and Habener, J. F. (1988) *Science*, **242**, 1430–1433.
- Garner, M. M., and Revzin, A. (1981) *Nucl. Acids Res.*, **9**, 3047–3060.
- Fried, M.G., and Crothers, D.M. (1981) *Nucl. Acids Res.*, **9**, 6505–6525.
- Dalton, S., and Treisman, R. (1992) *Cell*, **68**, 597–612.
- Andersson, S., and Russell, D. W. (1990) *Proc. Natl. Acad. Sci.*, **87**, 3640–3644.
- Gorman, C. M., Moffat, L. F., and Howard, B. H. (1982) *Mol. Cell. Biol.*, **2**, 1044–1051.
- Seed, B., and Sheen, J-Y. (1988) *Gene*, **67**, 271–277.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci.*, **74**, 5463–5467.
- Pearson, W. R. (1990) *Methods Enzymol.*, **183**, 63–98.
- Trizenberg, S. J., Kingsbury, R. C., McKnight, S. L. (1988) *Genes Dev.*, **2**, 718–729.
- Ma, J., and Ptashne, M. (1988) *Cell*, **55**, 443–446.
- Cronan, J. E. Jr. (1989) *Cell*, **58**, 427–429.
- Kaptain, S., Downey, W. E., Tang, C., Philpott, C., Haile, D., Orloff, D. G., Harford, J. B., Rouault, T. A., and Klausner, R. D. (1991) *Proc. Natl. Acad. Sci.*, **88**, 10109–10113.
- Andersson, S., Bishop, R. W., and Russell, D. W. (1989) *J. Biol. Chem.*, **264**, 16249–16255.
- Blackwood, E. M., and Eisenman, R. N. (1991) *Science*, **251**, 1211–1217.
- Ray, R., and Miller, D. M. (1991) *Mol. Cell. Biol.*, **11**, 2154–2161.
- Giallongo, A., Feo, S., Moore, R., Croce, C. M. and Showe, L. C. (1986) *Proc. Natl. Acad. Sci.*, **83**, 6741–6745.