J. Biosci., Vol. 14, Number 2, June 1989, pp. 189–202. © Printed in India.

Factors regulating the transcription of eukaryotic protein coding genes and their mechanism of action—A review

P. N. RANGARAJAN and G. PADMANABAN*
Department of Biochemistry, Indian Institute of Science, Bangalore 560 012, India

MS received 7 March 1989; revised 3 April 1989

Abstract. Protein factors play a crucial role in establishing gene-specific and cell-specific regulation of the process of transcription. These include general transcription factors which recognize TATA and CCAAT boxes and which form components of the RNA polymerase II system. Specific transcription factors interact with characteristic promoter elements of individual genes. Some of the examples are SP1, glucocorticoid receptor, GCN4, GAL4 and many others. Transcription factors have a DNA binding domain demarcated from the transcription activation domain. Some factors may have an additional ligand (small molecule) binding domain. Typical structural features such as helix-turn-helix motif, zinc finger and leucine zipper have been recognized in the DNA binding domain of the transcription factors. The acidic domain of the protein factors is involved in the transcription activation process. It appears that activation is the result of the combined action of several regulatory proteins binding at different regions of the promoter. Interaction between proteins bound to DNA but seperated by long stretches of nucleotides is facilitated by DNA bending. Functional specificity as well as diversity are feasible with a limited number of transcription factors through alterations in the architecture of interaction between a group of proteins bound to promoter elements.

Keywords. Transcription factors; promoters; Activation.

Introduction

In eukaryotes, multiple sequence-specific DNA-protein interactions occurring at distinct upstream regions regulate the initiation of transcription. Promoters of eukaryotic protein coding genes are characterized by the presence of a TATA box ($TATA^{A}_{T}A^{A}_{T}$)located 20–30 basepairs (bp) upstream of transcription initiation site. The TATA box present in eukaryotes is analogous to the Pribnow box (TATAAT) present in bacterial genes. It specifies the position where RNA synthesis is to begin and also determines which of the two DNA strands is to be transcribed (Corden *et al.,* 1980). The sequence $G^{C}_{T}CAATCT$ known as the CCAAT box is usually present 70–90 nucleotides upstream of the transcription initiation site and a family of transcription factors, referred to as the CCAAT binding proteins interact with these sequence elements (Chodosh *et al,* 1988). In addition, several regulatory sequences such as enhancers, silencers etc., located upstream of the TATA box and CCAAT box govern the efficiency of transcription. Enhancers have also been detected downstream of transcription initiation site. Purified RNA polymerase II initiates transcription randomly or at structures such as nicks, free ends etc., and thus

*To whom all the correspondence should be addressed.

Abbreviations used: bp, Basepairs; RAPs, RNA polymerase II associated proteins; NFI, nuclear factor I; HSTF, heat shock transcription factor; GRE, glucocorticoid regulatory element; UAS, upstream activation sequence; HNF, hepatocyte specific nuclear factor; TFIIIA, transcription factor IIIA; C/EBP, CCAAT box and enhancer binding protein; ATF, activating transcription factor.

behaves like the bacterial core polymerase. However, in the presence of crude nuclear extracts, RNA polymerase II transcription comes under strict promoter control (Lewis and Burgess, 1980). These crude nuclear extracts contain several auxillarly proteins referred to as transcription factors which interact with specific regulatory sequences and bring about accurate and efficient transcription. Several of these factors have been purified and biochemically characterized and are shown to promote specific and accurate initiation of transcription *in vitro*. A large number of transcription factors acting on a variety of eukaryotic promoters in a complex interactive manner have been identified. Some typical examples of transcription factors are listed in table 1. In this brief review, the regulation of RNA polymerase II transcription is discussed with selected examples.

General transcription factors

Factors that bind to common promoter elements such as TFIIA, TFIIB, TFIID, TFIIE and TFIIF isolated from HeLa cell extracts bind to TATA box and promote transcription in cell-free systems. Sopa *et al.* (1986) purified several RNA polymerase II associated proteins (RAPs) utilizing RNA polymerase II affinity columns. These include RAP30, RAP38 and RAP74. RAP30 is a functional component of TFIIF, whereas, RAP38 is required for transcription elongation. SP2 also purified from HeLa cell extracts, is yet another general transcription factor, known to promote transcription from several eukaryotic promoters such as SV40, adenovirus 2 major late promoter, human β-globin and avian sarcoma virus (Dynan and Tjian, 1983), Two factors, A and B are required for accurate initiation of transcription in. *Drosophila* cell extracts. Factor B binds near the TATA box and is compared to TFIIC and TFIID found in HeLa cell extracts (Parker and Topol, 1984a). A protein factor binding specifically to the TATA box of heat shock genes has been identified in *Drosophila.* It is postulated that all the TATA binding functions are governed by a family of proteins, each recognizing a unique sequence surrounding the TATA box core sequence (Wu, 1985).

Several factors binding to the CCAAT box of various eukaryotic promoters have been identified. These include (i) CP 1, binding to the CCAAT sequence of human α-globin promoter and adenovirus major late promoter, (ii) CP2, binding to the CCAAT element of γ-fibrinogen promoter and (iii) the nuclear factor I(NFI), regulating transcription and replication in adenovirus. CP 1, CP2 and NF I appear to be distinct proteins in that each binds to its own recognition site with an affinity that is several orders of magnitude higher than that with which it binds to the recognition sequences of the other two proteins. These factors, categorized as CCAAT binding proteins, thus show specificity, at the same time perform diverse functions such as regulation of transcription and replication, perhaps through micro heterogeneity in their structure and heteromer formation (Chodosh *et al,* 1988).

Specific transcription factors

A variety of gene-specific and tissue-specific transcription factors have been identified and only a few examples will be discussed here. The factor SP1 purified from HeLa cell extracts binds specifically the sequence $\overset{G}{T}GGGCGG\overset{GGGC}{TAAT}$ and this

Table 1. Examples of eukaryotic transcription factors*.

Transcription factor	Species/tissue	Target genes
ADR1	Yeast	Alcohol dehydrogenase
AP-1	Human/HeLa	Collagenase
		Metallothionin II
		SV40 enhancer
AP-2	HeLa	Bovine papilloma virus enhancer
	Human/HeLa	Growth hormone
		Metallothionin IIA
		c-myc
		SV40 enhancer
AP-3	HeLa	SV40 enhancer
B factor	*Drosophila*	Actin 5c
		Histone H3
		Histone H4
CCAAT binding factor	Human	hsp70
	Sea urchin	Histone H2B-1
	Mouse	α2(I) collagen
Estrogen receptor	Chicken	Vitellogenin II
	Human	
	Rat	
GAL4	Yeast	GAL 1–10
		GAL 2
		GAL 7
		GAL 80
GCN4	Yeast	His 3
		His 4
Glucocorticoid receptor	Rat	Glucocorticoid receptor
		Growth hormone
		MMTV LTR
		trp oxygenase
	Rabbit	Uteroglobin
	Chicken	Vitellogenin II
		Lysozyme
	Human	Metallothionin II
HAP 1	Yeast	CYC 1
		CYC 7
HAP 2	Yeast	CYC 1
HSTF	*Drosophila*	hsp70
	Human	hsp70
	Yeast	hsp83
Matα1	Yeast	MFα1
		MFα2
		STE 3
NF-1	Mouse/chicken	MMTV LTR
	Mouse	α2(I) collagen
	Rat	Albumin
NFκβ	Human B cells	Ig enhancer
Octamer binding factor	HeLa	SV40 enhancer
Progesterone receptor	Chicken	Lysozyme
	Rabbit	MMTV LTR
		Uteroglobin
RC1	Yeast	CYC 1
RC2	Yeast	CYC 1

Table 1. (*Contd.*)

Transcription factor	Species/tissue	Target genes
SP1	HeLa	dhf reductase
		HIV-1 LTR
		Metallothionin IIA
		c-Ha-ras 1
		HSV-tk
		δ1 crystalline
		U2 Sn RNA
TFIIA	HeLa	AdML
TFIID	HeLa	AdML

*For a complete list of eukaryotic transcription factors and their recognition sequences refer Wingender (1988).

sequence is present in several eukaryotic promoters (Briggs *et al.,* 1986). Kadonaga *et al.* (1987) have shown that amino acids present in the C-terminal region of SP1 factor is responsible for the DNA binding. SP 1 is not a tissue-specific factor since it was detected in the nuclear extracts of several tissues.

A heat shock transcription factor (HSTF) binding to the heat shock responsive element (CTnnGAAnnTCnAG) has been identified in *Drosophila,* yeast and man (Widerrecht *et al,* 1987). HSTF has been purified and shown to stimulate transcription from hsp70 promoter *in vitro* (Parker and Topol, 1984b).

One of the well characterized transcription factors in higher eukaryotes is the glucocorticoid receptor that binds the glucocorticoid regulatory element (GRE) bearing the consensus sequence TGT/CCT. GRE is present in the promoters of several eukaryotic genes such as uteroglobin, metallothionin, chicken lysozyme etc. (Renkawitz *et al.,* 1984a). The binding sites for the progesterone receptor (Jost *et al.,* 1984) and estrogen receptor (Renkawitz *et al,* 1984b) are also known. In fact, it was reported that a 15bp sequence mediates both glucocorticoid and progeste rone induction of transcription suggesting that these proteins may bind similar sequences (Strahle *et al.,* 1987). The mechanism of activation of transcription by steroid hormone receptors is still not known. According to one view, the receptor is present in the cytosol in complex with a heat shock protein (hsp90) and steroid binding releases the receptor and allows its nuclear localization (Sanchez *et al,* 1985). Becker *et al.* (1986) argue that the ligand-free receptor itself can bind to target DNA sequences and steroid binding induces a conformational change in the receptor that alters its contacts with DNA. Recently, Godowski *et al.* (1988), have reported that transcriptional enhancement by glucocorticoid receptor requires hormone binding and the unliganded hormone binding region can inhibit all receptor activities other than hormone binding itself. They suggest that the unliganded hormone binding domain forms a complex with hsp90 and the formation of such a complex inhibits receptor function. Hormone binding derepresses receptor activity by altering the conformation of hormone binding domain and disrupting the receptor-hsp90 interactions.

A plethora of factors regulating RNA polymerase II transcription have been identified in yeast. These include GAL4, GCN4, HAP1, HAP2, MATα2 and ADR1 gene products. The GAL4 protein regulates the transcription of GAL 1 gene by binding to an upstream activation sequence (UAS) called UAS_G. The first 74 N-

terminal amino acids of GAL4 protein are involved in DNA binding, while amino acid residues 148–196 and 768–881 are involved in transcription activation (Ma and Ptashne, 1987a).

The GCN4 protein promotes transcription of several co-regulated genes in response to amino acid starvation (general amino acid control). GCN4 protein binds a 9bp palindrome, ATGAC/GTCAT, which is also recognized by the mammalian transcription factor AP1 as well as the jun oncoprotein (Rausher *et al,* 1988). The DNA binding domain of GCN4 protein lies in the 60 C-terminal amino acid residues. The C-terminus also has a 19 amino acid long acidic domain that is implicated in transcription activation function (Hope and Struhl, 1986).

The iso-1-cytochrome C gene(CYCl) of yeast contains two UASs known as UAS1 and UAS2 which interact with proteins coded by HAP1, HAP2 and HAP3 loci (Guarente and Hoar, 1984). HAP 1 product binds at UAS 1, whereas UAS2 is activated by the combined action of HAP2 and HAP3 proteins. HAP1 protein also binds the UAS of CYC7 gene with equal affinity. However, there is no sequence homology between UASl of CYC1 gene and the UAS of CYC7 gene (Pfeifer *et al,* 1987). Thus HAPl protein is a rare example of a transcription factor binding two unrelated sequences. Other transcription factors identified in yeast include the MATα2 protein (Johnson and Herskowitz, 1985) and the ADR1 gene product (Hartshore *et al.,* 1986). The latter is required for the transcriptional activation of alcohol dehydrogenase gene.

Several transcriptional factors expressed in selective tissues have been identified; A liver specific transcription factor referred to as NFI-L, binding to a sequence similar to that recognized by NFI (TGGCA) has recently been identified (Paonessa *et al.,* 1988). Similarly, Courtois *et al* (1988), purified a 88 kDa protein binding to the α_1 -antitrypsin gene promoter and this is identified as hepatocyte specific nuclear factor (HNF1). A pitutary-specific factor called Pit-1, regulating the transcription of prolactin and growth hormone genes in the lactotrophs and somatotrophs of pitutary gland has been reported (Nelson *et al,* 1988). An erythroid-specific factor called ERYF-1, binding to the sequence AGATAA of chicken globin genes was shown to play a prominent role in the initial establishment of erythroid cell lineage (Evans *et al,* 1988). Other tissue-specific factors include the octamer binding protein of immunoglobulin genes expressed in B cells and that expressed in sea urchin testis (Barberis *et al.,* 1987). Several factors claimed to be tissue-specific, were later found to induce transcription in other tissues as well For example, the factor NF- *κβ*, binding to Ig*κ* -enhancer and expressed only in *κ*-producing cells was later shown to be induced by phorbol esters in HeLa cells as well (Sen and Baltimore, 1986).

Structure of the DNA binding domain of transcription activators

Despite several unique characteristics of their own, transcription factors share some common structural features so that certain generalizations can be made. Majority of these factors are sequence-specific DNA binding proteins and contain two major functional domains, namely, the DNA binding domain and the transcription activation domain. Proteins like steroid hormone receptors possess an additional domain for ligand binding. In addition, several of these proteins interact with other proteins of the transcription machinery through their carboxy terminii. Many of the transcription factors show considerable homology in their DNA binding domains

and studies so far have led to the identification of 3 major DNA binding motifs, namely, the helix-turn-helix motif, the zinc finger motif and the leucine zipper.

Helix-turn-helix motif

This motif was first identified in the DNA binding domain of prokaryotic regulatory proteins such as λ-repressor, cro protein, *Escherichia coli* Gal repressor, Lac repressor and Trp repressor (Pabo and Sauer, 1984). A similar motif was later identified in the yeast MATα2 protein and certain homeotic products of *Drosophila* (Sheperd *et al.,* 1984). This motif is characterized by the presence of two α-helices, one of which called the helix3, lies within the major groove of DNA while the other lies across. Helix3 plays a major role in DNA recognition and mutation of amino acids in helix3 leads to weak promoter recognition. Amino acids that form the hinge region between the two α-helices are highly conserved among all these proteins. The helix-turn-helix motif lies in the amino terminus while the carboxy terminal amino acids provide sites for the dimerization of two monomers such that the axis of 2-fold symmetry of the dimer is coincident with that of DNA. These proteins dock with operators in the B-form. Hydrogen bonds and van der Waals forces are involved in the interactions between the amino acid side chains and edges of bases in the major groove of DNA. These, together with the backbone and electrostatic interactions stabilize DNA-protein complexes (Ptashne *et al.,* 1982) (figure 1).

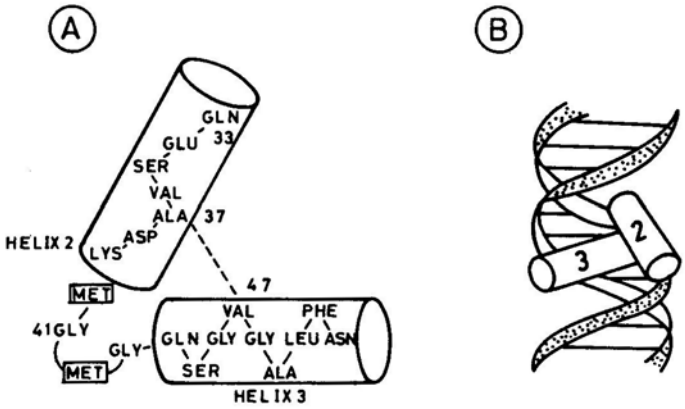

Figure 1. (A) The helix-turn-helix motif of λ-repressor. Amino acids 33(GLN), 37(ALA), 41(GLY) and 47(VAL) are highly conserved among all the prokaryotic repressor proteins. In addition, glycine at position 41 is usually flanked by hydrophobic amino acids. The angle between the helix2 and helix3 is controlled by the contact between the side chains of alanine at position 37 and valine at position 47. (**B**) Interaction of λ-repressor with the operator DNA.

Zinc finger motif

The zinc finger motif was first identified in the DNA binding domain of transcription factor IIIA(TFIIIA) which activates transcription from promoters recognized by

RNA polymerase III. Similar motifs were later discovered in the ADR1 gene product of yeast, the gag gene product of retroviruses, ElA gene product of adenovirus, large T antigen of papova viruses, the SP1 factor and the product of *Drosophila* developmental regulatory loci like kruppel and serendipity (Berg, 1986; Rosenber, 1986). The sex determining region of human Y chromosome was shown to encode a finger protein (Page *et al.,* 1987).

TFIIIA is a 40 kDa polypeptide with a 30 kDa aminoterminal domain retaining DNA binding activity and a 10 kDa carboxy terminal domain involved in transcription activation (Smith, 1984). Zinc (7-11 atoms) are complexed with each TFIIIA molecule. The DNA binding domain of TFIIIA consists of 9 homologous subunits termed zinc fingers. Each zinc finger is composed of 30 amino acids and contain the sequence of the form Cys-X_{2-5}-Cys-X_{12}-His-$X_{2,3}$-His, where X may be any amino acid. Each of the 30 amino acid domains of TFIIIA forms a loop or finger of 12 amino acids, the base of which is formed by Zn^{2+} bound by Cys and His residues (figure 2). The amino acids in the loop bind DNA with each finger interacting with one half turn of the helix (Rhodes and Klug, 1986). Analysis of the crystal structure of a double stranded nonadeoxynucleotide that corresponds to the tightest binding part of the sequence recognized by TFIIIA revealed that it has an A-type rather than a B-type conformation (McCall *et αl.,* 1986). TFIIIA molecule binds to the internal control region of the 5S rRNA gene between the nucleotides +40 and +90 and this interaction is stabilized by the sequential binding of two other protein factors called B and C. Each of the 9 finger domains of TFIIIA interacts with about half a DNA period and thus the 9 domains cover the entire internal control region. An advantage of the multiple fingers is that the stable transcriptional complex, once formed, can sustain several rounds of transcription, during which process the factor remains bound to the gene. As the RNA polymerase III passes through the gene, the multifingered protein releases the fingers bound ahead of the progressing polymerase but remains bound to the DNA by its remaining fingers (Miller *et al,* 1985).

When genomic libraries of different animals were screened with a probe coding

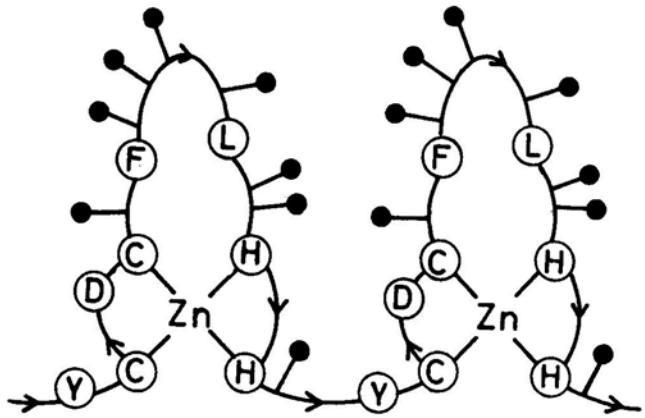

Figure 2. The zinc finger motif of Xenopus TFIIIA. Only two of the 9 repetitive motifs are shown. (●), Most probable DNA binding side chains. F, phenyl alanine; L, leucine; C, cysteine; H, histidine; D, aspartic acid; Y, tyrosine.

for the finger domain, several copies of homologous DNA sequences were detected suggesting that several finger proteins exist in eukaryotes and these are well conserved during evolution (Schutz *et al.,* 1986).

Leucine zipper

Landschulz *et al.* (1988) identified a new motif termed leucine zipper in the CCAAT box and enhancer binding protein (C/EBP) of rat liver nuclei. A computer search revealed that a similar motif is present in several other DNA binding proteins such as mouse c-myc, human N-myc, human L-myc, v-jun, v-fos and GCN4. The region of homology is characterized by a periodic repetition of leucine residues at every seventh position over a distance covering 8 helical turns. The leucine repeats in these proteins can form an α-helix with the leucine side chains projecting out from the helix at regular intervals (figure 3A). It was proposed that the leucine side chains from one protein molecule interdigitate with those on the other forming a zipper that holds the two molecules together and this dimerization facilitates appropriate positioning of amino acid side chains of these proteins into the grooves of the double helical DNA (figure 3B). Thus the leucine zipper is not directly involved in DNA binding but aligns the interacting protein molecules alongside the DNA such that atomic interactions can take place between the amino acid side chains and the bp of DNA.

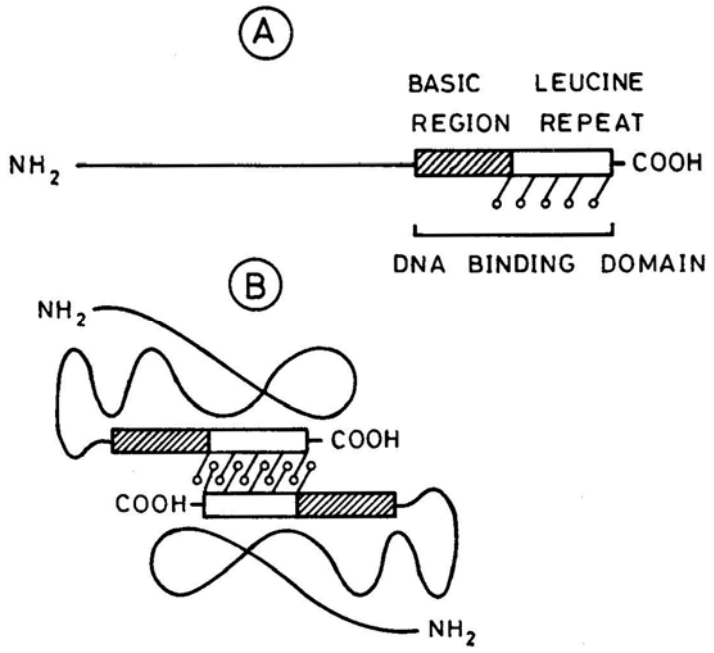

Figure 3. Leucine zipper of the C/EBP. **(A)** Location of the leucine repeat within the C/EBP polypeptide. The DNA binding domain lies beyond the leucine zipper domain towards the aminoterminus of the protein (striped region). **(B)** A hypothetical model of the C/EBP dimer formed by the interdigitation of the leucine zipper domains in an antiparallel conformation.

Mechanism of transcription activation

Role of acidic amino acids in transcription activation

Analysis of the transcription activation domain of certain transcription factors revealed that a group of acidic amino acids conferring a net negative charge may play an important role in activation of transcription. It was observed that the transactivation domain of GAL4 protein is composed of several acidic amino acids and mutational analysis confirmed that the negative charge is essential for transcription activation function (Gill and Ptashne, 1987). The activation domain of GCN4 protein also bears an acidic character but the amino acid sequence in this region bears no homology to that of GAL4 protein (Ma and Ptashne, 1987b). Although acidic amino acids seem to impart transcription activation function to these proteins, there is poor correlation between the strength of activation and the number of acidic amino acid residues. Single amino acid changes reduce GCN4 activation potential without much change in acidic character. Giniger and Ptashne (1987) constructed a gene encoding a 15 amino acid long polypeptide that is capable of forming an α-helix, one surface of which consists of negatively charged amino acids, while the other bears hydrophobic amino acids. When attached to a DNA binding domain, this peptide was found to activate transcription in yeast. Thus higher level of structural determinants such as an amphipathic helix with charged amino acids on one face, may be more important than the actual amino acid sequence, in transcription activation.

The transcription activation domain of gene regulatory proteins need not always be acidic as shown in the case of glucocorticoid receptor, which consists of two well characterized enhancement domains, enh1 and enh2. The enh2 domain is very similar to the activation domain of GAL4 protein in containing several acidic amino acids and the distribution of negative charge is also similar between the two proteins. However, the enh1 domain is positively charged, containing 6 acidic and 17 basic amino acid residues between amino acids 440 and 525 and it can function independantly of enh2 domain. It is suggested that these enhancement domains might display distinct activities in different cells or they might operate by entirely different mechanisms (Godowski *et al.,* 1988).

Hybrid gene regulatory proteins

One of the current approaches for the understanding of the mechanism of trans cription activation involves construction of hybrid regulatory proteins. It was shown that a chimeric protein consisting of the DNA binding domain of estrogen receptor and transcription activation domain of either GCN4 or GAL4 protein can activate a promoter containing estrogen responsive element (Webster *et al.,* 1988). Similarly, a hybrid protein containing the DNA binding domain of *E. coli* lexA repressor and the trans-activation domain of GAL4 protein was able to activate transcription from a promoter containing lexA operator. In yet another study, the DNA binding domain of progesterone receptor was replaced by the corresponding region of glucocorticoid receptor and this hybrid protein specifically binds the glucocorticoid regulatory element and leads to progesterone dependant activation of heterologous genes (Green and Chambon, 1987). Recently, Hollis *et al.* (1988),

constructed a hybrid phage 434 repressor in which the DNA recognition helix was replaced by the corresponding region of phage P22 repressor. This chimeric protein binds specifically the P22 operator. Further, a dimer comprising of the 434 repressor monomer and the chimeric 434-P22 repressor monomer specifically recognizes a chimeric operator that lacks the usual 2-fold symmetry, suggesting that combination of two DNA binding proteins may form a novel protein that recognizes a new DNA sequence.

Protein-protein interactions

The binding sites of gene regulatory proteins often lie several bp upstream of transcription start site. Thus arises the question as to how do these proteins, binding far upstream regions regulate transcription? A glimpse at the various eukaryotic genes reveals that transcription activation, often, is the result of the combined action of several regulatory proteins binding at different regions of the promoter. It is suggested that these proteins interact with one another leading to the formation of a DNA loop thereby bringing the upstream sites adjascent to the transcription start site. The evidences for such protein-protein interactions are, infact, overwhelming. For example, the carboxy terminal amino acids of TFIIIA are essential for the correct initiation of transcription and it is suggested that this region is involved in the binding of other components of the transcription machinery (Harrison, 1986). Similarly, the promoters of heat shock genes contain multiple copies of heat shock elements and it is proposed that the HSTF binds at these sites in a cooperative fashion and the binding of the first HSTF molecule causes a bend in DNA that facilitates the binding of the second HSTF molecule, whose binding results in additional conformation changes in the contacts of DNA to HSTF, triggering transcription (Shuey and Parker, 1986). In amphibian oocytes, heat shock element as well as the CCAAT box are essential for the efficient binding of HSTF to hsp70 promoter suggesting that the CCAAT box binding protein may interact with HSTF, thereby enhancing the affinity of the latter for the promoter (Beinz, 1986). The recognition sequences of SP1 factor are often found near the binding sites of other transcription factors such as AP1 or NF1 and it is believed that these factors may act in conjunction with one another and modulate transcrip tion (Kadonaga *et al.,* 1987). Similar observations have been made in prokaryotes as well. For example, it was shown in the case of λ-repressor, which binds at the right operator (O_R) of the λ-phage promoting the transcription of genes involved in lysogeny, that mutations in the amino terminal domain seriously disrupt the transcription activation function but the DNA binding function remains unaffected. It was suggested that the amino terminus represents the positive control domain and the repressor interacts with RNA polymerase through its amino terminal amino acids (Ptashne, 1986). It has recently been reported that the yeast GAL4 protein can activate transcription synergistically in conjunction with several mammalian transcription factors such as activating transcription factor (ATF), upstream stimulating factor and glucocorticoid receptor (Kakidani and Ptashne, 1988; Webster *et al.,* 1988). It is believed that the interaction between any two transcription activators may be mediated by a common target protein such as TFIID, a TATA binding protein. Infact, it has been shown by Horikoshi *et al.* (1988), that ATF binds to multiple upstream elements of adenovirus E4 promoter

and alters the promoter interactions of TFIID and these interactions in turn facilitate promoter recognition by RNA polymerase II and the subsequent formation of pre initiation complex.

It is to be mentioned here that motifs such as the leucine zipper play a very important role in facilitating protein-protein interactions among different transcription activators. For example, it has been shown recently that the leucine zipper stabilizes the interaction between fos and jun proteins and the fos–jun heteromer binds to DNA more tightly than either protein alone (Kouzarides and Ziff, 1988; Sassone-corsi *et al.,* 1988). Thus, different transcription activators may interact with one another through motifs such as the leucine zipper and such interactions may lead to several combinations of dimers. A large number of such heteromers may indeed be involved in the regulation of transcription of eukaryotic gene families.

Squelching

Gill and Ptashne (1988) observed that the yeast GAL4 protein, when expressed at high levels, inhibits transcription of certain genes that lack GAL4 binding sites. This inhibitory effect, known as squelching, is independant of DNA binding domain, but involves the transcription activation domain. A GAL4 protein, lacking the DNA binding domain inhibits, whereas, a mutant lacking the activator domain does not inhibit. GAL4 derivatives possessing stronger activating domains inhibit more efficiently. It is suggested that squelching may be the result of the interaction between the activating domain of transcription factor and a target protein such as TFIID. When the transcription factor binds to DNA, its activating domain is brought into the immediate vicinity of the gene, where it interacts with a specific target protein leading to the formation of an active transcription complex. The target protein is present in limiting amounts and an activator protein may interact with the former even without binding to DNA and as a result, the transcription is blocked because of the nonavailability of the target protein. Thus, overproduction of a transcription activator would repress transcription from promoters lacking the activator binding sites and at very high concentrations, promoters bearing the binding sites would also be inhibited. These results suggest that the rate of transcription depends on the concentration of activators in the cell and the activator concentration should be regulated such that the activating domains are exposed only when required. The following examples indicate that such a regulation may indeed be operating *in vivo.* In yeast, the negative regulatory protein GAL80 binds to the activating domain of GAL4 in the absence of galactose, thereby blocking transcription. In higher eukaryotes, steroid hormone receptors are complexed with hsp90 in the absence of steroid hormone. Several transcription factors are turned over rapidly, whereas in others, the activating domain is functional only when phosphorylated and regulation is effected at the level of phosphorylation (Ptashne, 1988).

Cytochrome P-450 gene as a model system

In this laboratory, detailed investigations have been carried out using rat liver cytochrome P-450 genes as a model system to study regulation of eukaryotic gene

transcription. Prototype chemicals such as phenobarbitone and 3-methylcholanthrene enhance the transcription of cytochrome P-450 b/e and c/d genes by 20–50-fold in the whole animal after a single injection (Ashwanikumar and Padmanaban, 1980; Ravishankar and Padmanaban, 1985). An interesting finding has been that heme, the prosthetic group of cytochrome P-450, is a positive modulator of the transcription of this gene (Ravishankar and Padmanaban, 1983; Sathyabhama *et al*., 1986; Bhat and Padmanaban, 1988). Studies also reveal that cycloheximide inhibits phenobarbitone or 3-methylcholanthrene mediated activation of the transcription of the respective P-450 genes implicating the possible involvement of a drug-induced transcription factor (Bhat *et al.,* 1987). In support of this, nitrocellulose filter binding, gel retardation, foot print assays and south western blot analysis of the interaction between the upstream region of the P-450 b/e gene (Rangarajan *et al.,* 1987) and nuclear extracts, reveal drug- and heme-modulated binding of a transcription factor(s) to the upstream sites. A major protein involved has a molecular weight of 85 kDa. The binding of the transcription factor(s) to the upstream region correlates very well with the pattern of activation of transcription under different conditions of treatment (Rangarajan and Padmanaban, 1989). The system offers excellent scope to dissect the various processes associated with the phenomenon of transcription.

Prospects

Eukaryotic transcription, once considered an impenetrable black box, is now amenable for detailed analysis and significant information has already been accumulated. The generation of cell-free transcription systems responding to tissue and gene specific factors will help in understanding the process of selectivity in transcription activation. Negative interactions are as important as the positive ones, since the process of inactivation of genes and reasons for their silence need to be understood. The involvement of higher order structures of the genome in transcription regulation is still an open question. The main problem is to seggregate the cause from effect. Despite the complexity of chromatin structure, it is interesting to note that transcriptional activation state in a variety of systems is reflected in the functional consequence of the interaction between naked DNA and the transcription factors involved. Another interesting fact is the possible functional diversity of the transcription factors, which while manifesting specificity in one plan of architecture arising out of interaction with a set of proteins may show a different specificity in another combination or even assume the function of a replication factor. Finally, the regulatory potential of the different facets of transcription, namely, initiation, elongation and termination have to be understood and integrated.

Acknowledgement

Studies from the authors' laboratory are carried out in a project funded by the Department of Science and Technology, New Delhi.

References

Ashwanikumar and Padmanaban, G. (1980) *J. Biol. Chem.,* **225,** 522.
Barberis, A., Superati-furga, G. and Busslinger, M. (1987) *Cell,* **50,** 347.
Becker, P. B., Gloss, B., Schmid, W., Strahle, U. and Schutz, G. (1986) *Nature* (*London*) **324**,686
Beinz, M. (1986) *Cell,* **46,** 1037.
Bhat, G. J. and Padmanaban, G. (1988) *Arch. Biochem. Biophys.,* **264,** 584.
Bhat, G. J., Rangarajan, P. N. and Padmanaban, G. (1987) *Biochem. Biophvs. Res Commun* **148,** 1118
Berg, J. (1986) *Science,* 485, 82.
Briggs, M. R., Kadonaga, J. T., Bells, S. P. and Tjian, R. (1986) *Science,* **234,** 47.
Chodosh, L. A., Baldwin, A. S., Carthew, R. W. and Sharp, P. A. (1988) *Cell,* **53,** 11.
Corden, J., Wasylyk, B., Buchwalder, A., Sassone-corsi, P., Kedinger, D. and Chambon P (1980) *Science* **209,** 1406.
Courtois, G., Baumhueter, S. and Crabtree, G. (1988) *Proc. Natl. Acad. Sci. USA,* **85,** 7937
Dynan, W. S. and Tjian, R. (1983) *Cell,* **32,** 669.
Evans, T., Reitman, M. and Felsenfield, G. (1988) *Proc. Natl. Acad. Sci. USA,* **85,** 5976.
Gill, G. and Ptashne, M. (1987) *Cell,* **51,** 121.
Gill, G. and Ptashne, M. (1988) *Nature* (*London*), **334,** 721.
Giniger, E. and Ptashne, M. (1987) *Nature* (*London*), **330,** 670.
Green, S. and Chambon, P. (1987) *Mature* (*London*), **325,** 670.
Godowski, P. J., Picard, D. and Yamamoto, K. R. (1988) *Science,* **241**, 812.
Guarente, L. and Hoar, E. (1984) *Proc. Natl. Acad. Sci. USA,* **81,** 7860.
Harrison, S. C. (1986) *Nature* (*London*), **322**, 597.
Hartshore, T. A., Blumberg, H. and Young, E. T. (1986) *Nature (London),* **320,** 283.
Hope, I. A. and Struhl, K. (1986) *Cell,* **46,** 885.
Hollis, M., Valenzuela, D., Pioli, D., Wharton, R. and Ptashne, M. (1988) *Proc. Natl. Acad. Sci. USA,* **85,** 5834.
Horikoshi, M., Carey, M. F., Kakidani, H. and Roeder, R. G. (1986) *Cell,* **54,** 665.
Johnson, A. D. and Hesrkowitz, I. (1985) *Cell,* **42,** 237.
Jost, J. P., Seldran, M. and Geisser, M. (1984) *Proc. Natl. Acad. Sci. USA,* **81,** 429.
Kadonaga, J. T., Carner, K. R., Masiarz, F. R. and Tjian, R. (1987) *Cell,* **51,** 1079.
Kakidani, H. and Ptashne, M. (1988) *Cell,* **52,** 161.
Kouzarides, T. and Ziff, E. (1988) *Nature* (*London*), **336**, 646.
Landschulz, W. H., Johnson, P. F. a n d McKnight, S. L. (1988) *Science,* **240**, 1759.
Lewis, M. K. and Burgess, R. R. (1980) *J. Biol. Chem.,* **255,** 4928.
Ma, J. and Ptashne, M. (1987a) *Cell,* **48,** 847.
Ma, J. and Ptashne, M. (1987b) *Cell,* **51,** 113.
McCall, M., Brown, T., Hunter, W. N. and Kennard, O. (1986) *Nature* (*London*), **322,** 661.
Miller, J., McLachlan, A. D. and Klug, A. (1985) *EMBO. J.,* **4,** 1609.
Nelson, C, Albert, V. R., Elshltz, H. P . , Laslie, I., Lu, W. and Rosenfield, M. G. (1988) *Science,* **239**,1400.
Pabo, C. O. and Sauer, R. T. (1984) *Annu. Rev. Biochem.,* **53,** 293.
Page, D. C, Mosher, R., Simpson, E. M., Fisher, E. M., Mardon, G. and Pollack, J. (1987) *Cell,* **51**, 1091.
Paonessa, G., Gounari, F., Frank, R. and Cortese, R. (1988) *EMBO. J.,* **7,** 3115.
Parker, C. S. and Topol, J. (1984a) *Cell,* **36**, 357.
Parker, C. S. and Topol, J. (1984b) *Cell,* **37,** 273.
Pfiefer, K., Prezant, T. and Guanrente, L. (1987) *Cell,* **49,** 19.
Ptashne, M., Johnson, A. D. and Pabo, C. O. (1982) *Sci. Am.,* **247**,106.
Ptashne, M. (1986) *Nature* (*London*), **322,** 697.
Ptashne, M. (1988) *Nature* (*London*), **335,** 683.
Ravishankar, H. and Padmanaban, G. (1983) *Arch. Biochem. Biophys.,* **225,** 16.
Ravishankar, H. and Padmanaban, *G.* (1985) *J. Biol. Chem.,* **260,** 1588.
Rangarajan, P. N., Ravishankar, H. and Padmanaban, G. (1987) *Biochem. Biophys. Res. Commmun,* **144**, 258.
Rangarajan, P. N. and Padmanaban, G. (1989) *Proc. Natl. Acad. Sci.* USA **86,** 3963.
Rausher, F. J. III., Sambucetti, L. C, Curran, T., Distel, R. J. and Spiegelman, B. M. (1988) *Cell,* **52**,471.
Renkawitz, R., Schutz, G., Vonder Ahe, D. and Beato, M. (1984a) *Cell,* **37,** 503.
Renkawitz, R, Schutz, G., Vonder Ahe, D. and Beato, M. (1984b) *Cell,* **39,** 2771.
Rhodes, D. and Klug, A. (1986) *Cell,* **46**, 123.

Rosenber, V. B. (1986) *Nature* (*London*), **319,** 336.
Sanchez, E. R., Toft, D. O., Schleisinger, M. T. and Pratt, W. B. (1985) *J. Biol. Chem.,* **260,** 12398.
Sassone-corsi, P., Ransone, C. J., Lamph, W. W. and Verma, I. M. (1988) *Nature* (*London*), **336,** 692.
Satyabhama, S., Seelan, R. S. and Padmanaban, G. (1986) *Biochemistry,* **25,** 4508.
Sheperd, J. C. W., McGinnis, W., Carrasco, A. W., DeRobertis, E. M. and Gehring, W. J. (1984) *Nature* (*London*), **310,** 70.
Schutz, R., Aicher, W., Gaul, W., Cote, S., Driess, A., Maier, D., Seifert, E. and Jackie, H. (1986) *Cell,* **47,** 1025.
Shuey, D. J. and Parker, C. S. (1986) *J. Biol. Chem.,* **261,** 7934.
Sen, R. and Baltimore, D. (1986) *Cell,* **47,** 921.
Smith, D. R. (1984) *Cell,* **37,** 645.
Sopa, M., Carthew, R. W. and Greenblatt, J. (1986) *EMBO. J.,* **55,** 2923.
Strahle, U., Klock, G. and Schutz, G. (1987) *Proc. Natl. Acad. Sci. USA,* **84,** 7871.
Webster, N., Jin, J. R., Green, S., Hollis, M. and Chambon, P. (1988) *Cell,* **52,** 169.
Widerrecht, G, Shuey, D. J., Kibble, W. N. and Parker, C. S. (1987) *Cell,* **48,** 507.
Wu, C. (1985) *Nature* (*London*), **317,** 84.
Wingender, E. (1988) *Nucleic Acids Res.,* **16,** 1879.