REVIEW ARTICLE

Eukaryotic transcription factors

David S. LATCHMAN

Medical Molecular Biology Unit, Department of Biochemistry, University College and Middlesex School of Medicine, The Windeyer Building, Cleveland Street, London W1P 6DB, U.K.

INTRODUCTION

It is now well established that the primary control of gene expression lies at the level of gene transcription, many genes being transcribed only in particular tissues where their protein products are required [1,2]. Genes which are regulated in parallel in response to a particular inducing signal or in a particular tissue have been shown to contain common DNA sequence elements which are often but not always located up-stream of the start site of transcription [3–6]. These elements play an essential role in the specific expression pattern of such genes. Thus their destruction by mutation results in the abolition of the specific pattern of gene expression whilst their transfer to a marker gene confers this pattern of regulation on the heterologous gene [7–10]. Similarly, other short DNA sequences such as the TATA box which are common to many genes with different expression patterns are essential for the basic process of transcription itself [11,12].

It has been believed for some time that such DNA sequence elements would act by binding specific regulatory proteins. These transcription factors would interact with each other and the RNA polymerase enzyme itself in order to modulate transcription. The study of such transcription factors has been very difficult however, principally because they are present in very small amounts. Hence, even when they were purified, only limited amounts were obtained which were insufficient for detailed biochemical study of the functional properties of the proteins.

In the last few years however this difficulty has been overcome by the cloning of the genes encoding these factors. This has been achieved in two ways. In some cases the factor has been purified to homogeneity and the small amounts obtained used to obtain a partial amino acid sequence. In turn, this is used to predict a set of redundant oligonucleotides containing all the possible DNA sequences capable of encoding this region of the protein. These oligonucleotides are then used to screen a cDNA library prepared from a cell type which expresses the factor. This approach has been used, for example, to isolate cDNA clones for the general transcription factor Sp1 [13,14] and the CAAT box binding factor CTF/NF1 [15,16].

The alternative approach to cloning the genes encoding these factors is based on the fact that the specific DNA sequence to which these factors bind is often known. This sequence can therefore be used to screen a cDNA library in which the eukaryotic inserts are expressed as proteins in the bacterial cell. If the library is screened with a labelled probe containing the binding site of a particular factor, clones containing the mRNA for the factor and expressing it as a protein will bind the labelled probe and can hence be readily identified (see [17,18] for examples of the successful use of this approach).

The successful isolation of the genes encoding many eukaryotic transcription factors by these means has resulted in an explosion of information on these factors (for general reviews see [19–21]). Rather than consider each factor individually, this review will consider the general properties that such a factor must have and illustrate how these are achieved in particular factors.

Clearly the first essential feature that such factors require is the ability to bind to DNA in a sequence specific manner. Following such binding, the factor must interact with other factors or with the RNA polymerase itself in order to influence transcription either positively or negatively. Finally, in the case of factors which regulate transcription in response to a particular stimulus or in a particular tissue, some means must exist to regulate the synthesis or activity of the factor so that it is only active in the correct situation.

These three characteristics, namely DNA binding, effect on transcription and control of activity are discussed in the subsequent sections.

DNA BINDING

Once the gene encoding a particular transcription factor has been cloned, the region of the protein containing the DNA binding domain can be readily identified. This is achieved by expressing portions of the gene as protein fragments either in bacteria [14,22] or by coupled *in vitro* transcription and translation [23,24] and then testing each of these protein fragments for their ability to bind to DNA in a sequence specific manner. Such studies have identified several structural elements common to different transcription factors which are responsible for sequence-specific binding and these will be discussed in turn (for reviews see [25,26]).

The zinc finger

One of the first transcription factors to be purified and cloned was TFIIIA which plays a critical role in the transcription of the 5 S ribosomal RNA genes by RNA polymerase III [27]. The DNA binding region of this factor contains nine repeats of a 30 amino acid sequence of the form Tyr/Phe-Xaa-Cys-Xaa-Cys-Xaa_{2,4}-Cys-Xaa₃-Phe-Xaa₅-Leu-Xaa₂-His-Xaa_{3,4}-His-Xaa₅ where Xaa is a variable amino acid [28]. Each of these repeats therefore contains two invariant pairs of cysteine and histidine residues which co-ordinate a single atom of zinc. This results in a finger-like structure (Fig. 1a) in which the conserved phenylalanine and leucine residues and several basic residues in the finger project from the surface of the protein. The tips of these fingers make direct contact with the major groove of the DNA, alternate fingers binding on opposite sides of the helix ([29], Fig. 1b).

Multiple examples of this zinc finger motif have subsequently been identified in a number of transcription factors for genes transcribed by RNA polymerase II including Sp1 [14], the *Drosophila* kruppel protein [30], the yeast ADRI protein [31] and many others [26,32–34]. Interestingly, a single mutation in one of the zinc finger motifs of the kruppel protein which replaces a cysteine by a serine that cannot bind zinc, results in a mutant fly whose appearance is exactly identical to that produced by complete deletion of the gene [30]. Hence the ability to bind zinc is essential for DNA binding activity and therefore for the functioning of the protein as a transcription factor.

A similar zinc binding motif is also found in the DNA binding

Fig. 1. Schematic diagram of the cysteine-histidine zinc finger (a) and the interaction of the fingers with DNA (b)

Each successive finger enters the DNA from opposite sides of the helix (from [29] by permission of Professor Sir Aaron Klug and Elsevier Publications).

regions of members of the steroid/thyroid hormone receptor family whose members bind specific steroid hormones and either activate or repress gene expression by binding to specific DNA sequences in target genes [10,35,36]. In this case however the binding region consists of two fingers each of which contains four cysteine residues co-ordinating the zinc atom rather than two each of cysteine and histidine and it also lacks the conserved phenylalanine and leucine residues found in the other type of finger ([37,38]; Fig. 2a). In addition, the two-finger element is present only once in each receptor, as opposed to the multiple fingers ranging from two to 37 found in genes having cysteine-histidine fingers. The two types of finger may not therefore be evolutionarily related [39].

The existence of different related DNA binding regions in the steroid receptors which bind to distinct but related sequences ([10,35]; Fig. 2b) has allowed detailed study of the features in the protein which are important for sequence specific DNA binding [40–42]. Thus the oestrogen receptor and the glucocorticoid receptor bind to related but distinct palindromic sequences in their target genes (Fig. 2b). Alteration of two amino acids in the *N*-terminal finger of the oestrogen receptor to those found in the same region of the glucocorticoid receptor results in a receptor which can bind to the DNA sequence normally bound by the oestrogen receptor and hence switch on oestrogen responsive genes even though the other 775 amino acids of the protein are derived from the glucocorticoid receptor. Similarly, in the case of the oestrogen receptor and the thyroid hormone receptor, which bind the same palindromic DNA sequence but in which the spacing between the two halves of the palindrome differs (Fig. 2b) alteration of five amino acids in the second finger allows an oestrogen receptor to recognize thyroid hormone response elements in the DNA [43,44]. Hence, whilst the first finger plays a critical role in determining the precise palindromic sequence

Fig. 2. DNA binding domain (a) and DNA recognition sequences (b) of the steroid/thyroid hormone receptors

(a) Schematic diagram of the pair of cysteine fingers found in the steroid/thyroid hormone receptor family. Regions of the finger critical for determining the palindromic DNA sequence recognized by the receptor (A) and the optimal spacing of the two halves of the palindromic sequence (B) are indicated. (b) Relationship of the palindromic sequences recognized by the receptors for various hormones. N indicates that any base can be present at this position, a dash indicates that no base is present, the gap having been introduced to align the sequence with the other sequences.

to which the receptor binds, the second finger determines the optimal spacing between the two halves of the palindrome for binding to occur.

Although best analysed in the steroid/thyroid receptor gene family similar multi-cysteine fingers are found in other transcription factors such as the yeast transcription factor GAL 4 [45,46] and the adenovirus E1A protein [47].

The helix–turn–helix motif

A number of homeotic mutations which dramatically affect the development of the fruit fly *Drosophila melanogaster* have been described. The genes in which these mutations map are likely therefore to play a critical role in the development of the fly and are known as homeotic genes [48,49]. When these genes were isolated they were found to contain a common 60-amino acid region which was highly conserved between different genes [48–51]. This region, known as the homeobox or homeodomain [52,53], was subsequently shown to be critical for the DNA binding ability of these proteins [22,54]. Structure predictions of this region indicated that it could form a helix–turn–helix motif in which an α -helical region is followed by a β -turn and then another α -helical region [55,56]; Fig. 3). This structure, which was originally predicted on the basis of homology to bacterial repressor proteins, has now been directly confirmed by n.m.r. spectroscopic studies of the Antennapedia homeodomain [57]. In the bacterial repressors the first helical motif lies across the major groove of the DNA while the second helix, known as the recognition helix, lies partly within the major groove and makes sequence specific contact with the DNA [58]. In agreement with a similar role for the second helix in sequence specific DNA recognition by the homeodomain, a mutation at residue 9 of this helix which results in substitution of the serine present in the Prd protein by the glutamine present in the Ftz protein confers on the

Fig. 3. Schematic diagram of the helix–turn–helix motif in the homeobox

Prd protein the ability to bind to Ftz binding sites [59]. Similar dramatic effects of other alterations at this position have also been demonstrated [59,60].

Following its original identification in the *Drosophila* homeotic genes, similar homeoboxes containing the helix–turn–helix motif have also been identified in the yeast mating type transcriptional regulatory proteins [61,62], in a variety of amphibian and mammalian proteins [63] and in plants such as *Antirrhinum* [64].

More recently another class of regulatory proteins has been identified in which the homeobox forms one part of a larger conserved domain known as the POU domain that also includes another, POU-specific region [65,66,67]. These proteins, which include the octamer binding proteins Oct-1 and Oct-2 [68,69], the pituitary specific protein Pit-1 [70] and the nematode gene *unc-86* [71], all use the POU (Pit-Oct-Unc) domain to bind to DNA. The relative contribution of the homeodomain and the POU-specific domain differs between the different proteins however, with the homeodomain being sufficient for sequence specific binding of Pit-1 [72] whilst both the homeodomain and the POU-specific domain are necessary in the case of Oct-1 [73].

The leucine zipper and the basic DNA binding domain

As discussed above, the study of motifs common to numerous transcription factors has led to the identification of the role of these motifs in DNA binding. Another element found in several transcription factors such as the liver specific transcription factor C/EBP, the yeast factor GCN4 and the proto-oncogene proteins Myc, Fos and Jun is the leucine zipper [74,75]. In this structure leucine residues occur every seven amino acids in an α -helical structure such that the leucines occur every two turns on the same side of the helix.

Rather than acting directly as a DNA-binding motif, however, the zipper facilitates the dimerization of the protein by interdigitation of two leucine-containing helices on different molecules. In turn such dimerization results in the correct protein structure for DNA binding by the adjacent region which in C/EBP, Fos and Jun is a highly basic region, distinct from those discussed so far, that can interact directly with the acidic DNA ([76–79]; Fig. 4).

Both the Fos and Jun proteins bind to sequences in DNA

Fig. 4. Leucine zipper and DNA binding domain

Schematic diagram indicating how dimerization of two molecules of the C/EBP transcription factor via the leucine (L) zipper motif aligns the adjacent DNA binding domain in the correct orientation to bind DNA.

known as AP-1 sites which mediate gene induction following phorbol ester treatment [80,81]. Interestingly however, whereas the Jun protein can bind specifically to this sequence as a homeodimer [82], the Fos protein can only do so after formation of a heterodimer with Jun [83]. This difference is directly due to a difference in the leucine zipper motif of the two proteins, which prevents Fos homodimer formation. Thus substitution of the Fos leucine zipper region with that of Jun allows the chimaeric protein to bind to DNA through the basic region of Fos [84]. The requirement for dimer formation of these proteins prior to DNA binding thus introduces another potential regulatory point in the control of gene expression.

Although originally identified in leucine zipper-containing proteins, the basic DNA-binding domain has also been identified by homology comparisons [85] in a number of other transcriptional regulatory proteins including two proteins, E12 and E47, which bind to the immunoglobulin enhancer [86], the muscle regulatory protein MyoD1 [87] and the *Drosophila* daughterless protein [88]. In these cases, the basic domain is associated with an adjacent region that can form a helix–loop–helix structure in which two amphipathic helices (containing all the charged amino acids on one side of the helix) are separated by an intervening non-helical loop [86,89]. Although originally thought to be the DNA-binding domain of these proteins, this helix–loop–helix motif is now believed to play a similar role to the leucine zipper in mediating protein dimerization and facilitating DNA binding by the adjacent basic DNA-binding motif [85,89]. In agreement with this, deletions or mutations in the basic region of MyoD1 do not prevent dimerization but abolish its ability to bind to DNA. Interestingly, DNA binding of deleted MyoD1 lacking the basic DNA binding domain can be restored by substitution of the basic domain from the E12 protein [90]. Such substitution does not allow the hybrid protein to activate muscle specific gene expression however, suggesting that the basic region of MyoD1 contains elements mediating both DNA binding and the activation of muscle-specific genes [90].

Hence both the leucine zipper and the helix–loop–helix motif act by causing dimerization, allowing DNA binding by the adjacent basic domain. Interestingly, the *myc* oncogene proteins contain both a leucine zipper and a helix–loop–helix motif adjacent to the basic DNA region [81,86] suggesting that proteins

Fig. 5. Domains of the glucocorticoid receptor protein

The regions mediating DNA binding, transcriptional activation and hormone binding are indicated.

containing this DNA binding region may form a related family comprising three sub-families having either a leucine zipper, a helix-loop-helix motif or both [85].

Other DNA-binding motifs

As further transcription factor genes are analysed other DNA binding regions distinct from those discussed above are being defined. Thus the DNA binding regions of transcription factors such as AP2 [23], the serum response factors [24], the CAAT box binding protein CTF/NF1 [91] and the yeast transcription factors HAP2 and HAP3 [92,93] are distinct from the known motifs and from each other and may therefore be founder members of new families of DNA binding motifs.

It is clear therefore that a number of different structures exist which can mediate sequence-specific DNA binding. Several of these are common to a number of different transcription factors, with differences in the precise amino acid sequence of the motif in each factor controlling the precise DNA sequence to which it binds and hence the target genes for the factor.

EFFECT ON TRANSCRIPTION

Although binding to DNA is obviously a necessary prerequisite for a factor to affect transcription, it is not in itself sufficient. Thus following binding the factor must interact with other factors or the RNA polymerase itself in order to modulate transcription. Although such an interaction very often results in the activation of transcription, a number of cases have now been described in which factor binding can result in transcriptional repression. Activation and repression of transcription will therefore be discussed in turn.

Activation

In order to identify the regions of transcription factors which mediate transcriptional activation, a number of domain swapping experiments have been carried out in which different regions of one factor have been combined with the DNA-binding region of another factor allowing the effect on transcription of each region of the protein to be assayed. These studies have led to the identification of specific regions within transcription factors which can activate transcription following DNA binding [42,94-97]. These activation domains are normally distinct from the region which mediates DNA binding, indicating that transcription factors have a modular structure with different regions of the protein having distinct, independent activities. This is well

illustrated in the case of the members of the steroid/thyroid hormone receptor family [35] where distinct regions of the protein mediate DNA binding, transcriptional activation and binding of the hormone (Fig. 5). An extreme example of such modularity is provided by the herpes simplex virus virion protein VP16. This protein possesses a strong activating region but lacks any DNA binding activity and must therefore form a complex with the cellular DNA binding protein Oct-1, in order that DNA binding and consequent transcriptional activation can occur [97-99]. Hence, in this case, DNA binding and transcriptional activation domains are located on different proteins.

When the activation regions of a number of transcription factors were compared, it was found that, although they did not share amino acid homology, they possessed a high proportion of acidic amino acids [100,101]. These acidic amino acids were arranged in such a way that they formed an amphipathic α -helix in which all the negative charges were displayed along one surface of the helix. In agreement with the critical role of this structure in transcriptional activation, a peptide which can form an acidic amphipathic helix can activate transcription when linked to the DNA binding domain of the yeast transcription factor GAL4 whereas the same amino acids placed in a random order could not do so ([102]; Fig. 6).

Fig. 6. Activation of transcription can be mediated by an amphipathic helix linked to a truncated GAL4 protein lacking the activation domain but not by the same amino acids arranged so that the negatively charged amino acids are randomly distributed

Fig. 7. Schematic diagram of the interaction of the transcriptional activator ATF with the TATA box binding protein TFIID to create a stable transcription complex and activate transcription

Although this acidic activation domain is common to a number of transcription factors from yeast to man, other non-acidic activation domains have also been described. Thus the activation domain of Sp1 contains a glutamine rich region [103,104], whilst that of CTF/NF1 is very proline rich [91]. The finding of similar glutamine- or proline-rich regions in other transcription factors [68,69,92,93,105] suggests that activation domains of this type are not confined to a single protein.

It is likely that the different activation domains act by interacting with other protein factors in order to facilitate transcription. Although this may occur by direct interaction with the RNA polymerase itself [101], at least in the case of the acidic activation domain it seems more likely that its effect is mediated via the TATA box binding factor TFIID [106]. Thus binding of yeast GAL4 or the mammalian transcription factor ATF to their specific binding sites in regulated promoters has been shown to change the conformation of already bound TFIID so that, instead of contacting the TATA box alone, it contacts both the TATA box and the transcriptional start site [107,108]. In turn such altered binding facilitates the binding of other factors such as TFIIC, TFIIE and the RNA polymerase itself into the stable transcriptional complex [109] necessary for transcription to occur. The altered conformation of TFIID is directly dependent upon the activation domain of GAL4 or ATF since the binding of truncated factors lacking the activation domain does not produce the change in TFIID binding [107].

Hence specific regions of transcription factors can activate transcription following DNA binding, by altering the conformation of another bound factor and facilitating the assembly of a stable transcriptional complex (Fig. 7).

Repression

Although the majority of transcription factors act in a positive manner, a number of cases have now been described in which a transcription factor exerts an inhibiting effect on transcription

and several possible mechanisms by which this may be achieved have been described ([110]; Fig. 8).

The simplest means of achieving repression is seen in the β -interferon promoter where the binding of two positively acting factors is necessary for gene activation. Another factor acts negatively by binding to this region of DNA and simply preventing the positively acting factors from binding (Fig. 8a). In response to viral infection, the negative factor is inactivated, allowing the positively acting factors to bind and transcription occurs [111,112]. In a related phenomenon known as squelching [113] (Fig. 8b), a strong *trans*-activating protein (B) binds all the molecules of a general transcription factor (A) in solution. This prevents binding of the general transcription factor to other genes which contain binding sites for this factor but lack binding sites for the *trans*-activator, resulting in their repression. Although such squelching can be produced experimentally, it is unclear however whether it actually occurs naturally.

In addition to inhibiting DNA binding, a negative factor can also act by interfering with the activation of transcription mediated by a bound factor in a phenomenon known as quenching (Fig. 8c). Thus the negatively acting yeast protein GAL80 prevents gene activation by the GAL4 protein by binding to it and masking the activation domain of GAL4. In response to the presence of galactose however, GAL80 dissociates allowing GAL4 to activate the genes required for the metabolism of galactose [114,115]. A related example in which the inhibitory factor binds directly to the DNA but to a site distinct from that binding the positive factor is provided by the yeast mating type α 2 protein. This protein binds to a site adjacent to that bound by the MCM1 transcriptional activator protein and prevents it from activating the α -specific genes [116].

In all these cases the negative factor exerts its inhibiting effect by neutralizing the action of a positively acting factor by preventing either its DNA binding or its activation of transcription. In principle however, it is possible that some factors

Fig. 8. Repression of transcription by (a) competition for binding, (b) sequestration, (c) quenching of activity or (d) direct repression

Note that in (b) unlike the other cases, the B protein acts as a strong transactivator for genes containing its binding site. However, in genes which lack its binding site but require protein A for activity, protein B can act as a repressor by binding to protein A and removing it from these genes.

may have an inherently negative effect and may directly inhibit transcription possibly via a discrete domain analogous to the activation domains (Fig. 8d). Although a number of cases where this may be occurring have been described, in no case has the possibility of the neutralization of a positive factor been eliminated and an inherent negative action unequivocally demonstrated [110]. Nonetheless even if their action can only be mediated by neutralizing a positively acting factor, it is clear that negative acting factors play an important role in transcriptional regulation.

CONTROL OF ACTIVITY

The primary role of many transcription factors is to activate particular genes in specific tissues or in response to a particular signal such as a steroid hormone. Clearly therefore some means must exist of ensuring that such a factor is active only in the correct tissue or becomes active only in response to the signal. This is achieved in different cases either by controlling the synthesis of the protein so that it is made only when necessary (Fig. 9a) or by regulating the activity of the protein so that pre-existing protein becomes activated when required (Figs. 9b–9d). These two methods will be discussed in turn.

Regulation of synthesis

Many regulatory proteins are synthesized in only a limited range of cell types. Thus the octamer binding protein Oct-2, which is involved in the stimulation of immunoglobulin gene expression in B cells, is found in these cells but is absent from cell types such as HeLa cells which do not express the immunoglobulin genes [69, 117]. The expression of the gene encoding Oct-2 in HeLa cells results in the transcriptional activation of

Fig. 9. Activation of transcription factors by (a) new synthesis, (b) ligand binding, (c) dissociation of an inhibitory protein or (d) protein modification

immunoglobulin genes introduced into these cells [118]. Hence the synthesis of Oct-2 only in specific cell types allows it to activate specifically the expression of particular genes in such cells.

In many cases, such regulated synthesis of a transcription factor is achieved by regulating the transcription of the gene encoding the factor. Thus like the Oct-2 protein, the Oct-2 mRNA is absent from HeLa cells [118–120], suggesting the existence of transcriptional control mechanisms acting on the gene encoding Oct-2, whilst the gene encoding the liver specific factor C/EBP has been directly shown to be transcribed only in the liver by using a nuclear run-on assay [121].

It is clear however that such regulation of transcription only sets the problem of gene regulation one stage further back, requiring mechanisms to activate the transcription of the gene encoding the transcription factor itself. It is not surprising therefore that the synthesis of specific transcription factors is often modulated by post-transcriptional control mechanisms. Such post-transcriptional regulation is observed for example in the case of the yeast factor GCN4 which activates the genes encoding the enzymes of amino acid biosynthesis in response to amino acid starvation. In this case the increased synthesis of GCN4 following amino acid starvation is mediated by increased translation of its specific mRNA [122]. This is achieved by inhibiting the use of translational initiation codons upstream of the correct start site for the translation of GCN4 protein. This results in increased use of the GCN4 initiation codon and production of GCN4 in response to amino acid starvation.

Regulation of transcription factor biosynthesis can also be achieved at the level of RNA splicing [123–124]. Thus two alternatively spliced forms of the thyroid hormone receptor exist, one of which lacks the hormone binding domain [123]. This

protein can therefore bind to the same DNA sequences as the hormone binding receptor but cannot respond to the hormone and activate transcription. It therefore acts as a dominant repressor of gene activation mediated by the normal receptor in response to hormone binding. A similar repression of gene expression is also produced by the *v-erbA* oncogene of avian erythroblastosis virus which similarly encodes a truncated thyroid hormone receptor lacking the hormone-binding domain [125,126].

Regulation of activity

Given the obvious advantages of regulating transcription factor synthesis at a post-transcriptional level, it is not surprising that in other transcription factors this process has been taken further with such factors being present in all tissues and becoming activated post-translationally in response to a particular signal or in a specific tissue.

A simple example of such modulation occurs in the yeast transcription factor ACE1 which activates transcription of the metallothionein gene in response to copper. In this case the protein undergoes a conformational change in the presence of copper which allows it to bind to regulatory sites in the metallothionein gene and activate transcription ([127]; Fig. 9b).

A similar dependence on an activating ligand is also observed in mammalian cells in members of the steroid/thyroid hormone receptor family. Thus these receptors must bind the appropriate steroid hormone in order to activate transcription of target genes, and a region at the C-terminus of each receptor which binds the appropriate hormone has been identified ([10,36]; see Fig. 5). It was thought therefore that binding of the hormone to the receptor activates its ability to bind to DNA and thereby switch on transcription. It is now clear however that although *in vivo* the receptors bind to DNA only in the presence of hormone, *in vitro* they can do so even in its absence [128,129]. This suggests that *in vivo* such DNA binding is prevented by some other anchorage protein, with the hormone acting to release the receptor from this factor, thus allowing it to bind DNA. In agreement with this, several steroid receptors have been shown to be associated with the 90 kDa heat-inducible protein hsp90 prior to hormone treatment [130,131], the receptor dissociating upon hormone addition [132,133].

Hence in this case factor activation is mediated not by a ligand-induced change in conformation but by disruption of an

inhibitory protein-protein interaction (Fig. 9c). A similar disruption of an interaction with an inhibitory protein is also responsible for the activation of the GAL4 protein in response to galactose [114,115] and the NF kappa B protein in response to phorbol ester treatment of T cells or HeLa cells [134–136]. Moreover, this mechanism of activation is not confined to factors which mediate gene activation in response to specific substances but is also observed in factors which mediate the activation of specific genes in a particular tissue. Thus the transcription factor MyoD1 plays a critical role in the activation of muscle-specific gene expression as myoblasts differentiate into myotubes [137]. This activation occurs not because of a rise in the level of MyoD1 during this differentiation event but rather because of a fall in the level of an inhibitory protein Id which associates with MyoD1 and prevents it binding to DNA [138]. Interestingly Id, like MyoD1, contains a helix-loop-helix motif which mediates dimerization but lacks the adjacent basic DNA binding domain (see above). It may therefore act by dimerizing with MyoD1 and preventing its binding to DNA, as occurs with truncated forms of MyoD1 lacking the DNA binding domain [90].

In addition to protein-protein interaction, activation of transcription factors can also be achieved by protein modification, providing a direct means of activating a particular factor in response to a specific signal (Fig. 9d). One example of this is provided by the CREB transcription factor which mediates the activation of several cellular genes following cyclic AMP treatment [139]. Thus cyclic AMP is known to stimulate the protein kinase A enzyme [140]; in turn this enzyme phosphorylates CREB [141] stimulating the activity of a transcriptional activation domain adjacent to the site of phosphorylation [142,143]. Hence stimulation of gene expression by cyclic AMP is directly mediated via its stimulation of protein kinase A and the consequent phosphorylation of CREB. Similar activation by phosphorylation is also involved in the activation of the yeast heat shock transcription factor in response to heat [144,145] and the activation of the NF kappa B protein by phorbol esters [134–136]. In this latter case however it is the inhibitory protein associated with NF kappa B (see above) which is phosphorylated causing it to dissociate from the NF kappa B protein allowing NF kappa B to bind to DNA and activate transcription [146].

Although phosphorylation is obviously an important means of achieving activation by protein modification, transcription

Table 1. Transcription factor domains

Domain	Role	Factors containing domain	Comments
Homeobox	DNA binding	Numerous <i>Drosophila</i> homeotic genes, related genes in other organisms	DNA binding mediated via helix-turn-helix motif
POU	DNA binding	Mammalian Oct-1, Oct-2, Pit-1, nematode unc-86	Related to homeodomain
Cysteine-histidine zinc finger	DNA binding	TFIIIA, kruppel, Sp1 etc	Multiple copies of finger motif
Cysteine-cysteine zinc finger	DNA binding	Steroid/thyroid hormone receptor family	Single pairs of fingers, related motifs in adenovirus E1A and yeast GAL4 etc.
Basic element	DNA binding	C/EBP, <i>c-fos</i> , <i>c-jun</i> , GCN4	Often found in association with leucine zipper, helix-loop-helix motif or both
Leucine zipper	Protein dimerization	C/EBP, <i>c-fos</i> , <i>c-jun</i> , GCN4, <i>c-myc</i>	Mediates dimerization which is essential for DNA binding by adjacent domain
Helix-loop-helix	Protein dimerization	<i>c-myc</i> , <i>Drosophila</i> daughterless MyoD, E12, E47	Mediates dimerization which is essential for DNA binding by adjacent domain
Amphipathic acidic α -helix	Gene activation	Yeast GCN4, GAL4, steroid/thyroid receptors	Probably interacts directly with TFIID
Glutamine-rich region	Gene activation	Sp1	Related regions in Oct-1, Oct-2, AP-2 etc.
Proline-rich region	Gene activation	CTF/NF1	Related regions in AP-2, <i>c-jun</i> , Oct-2

factors are also modified in other ways, for example by glycosylation [147], and activation could therefore take place by alteration of the levels of these modifications also.

A variety of mechanisms involving both increased synthesis and protein activation by ligand binding, protein modification or disruption of protein-protein interaction (Fig. 9) therefore exist to allow specific factors to become active in response to a particular signal or in a particular cell type.

CONCLUSIONS

It should be clear from the above that the cloning of the genes encoding many eukaryotic transcription factors has allowed the identification of a number of motifs which mediate DNA binding, protein dimerization or gene activation and these are listed in Table 1. Similarly in several cases the critical amino acids for sequence-specific DNA binding have been identified and the manner in which some transcription factors become activated in a particular cell type or in response to a specific signal has been determined. What remains to be achieved is to link this information with detailed protein structural information. This will then allow an understanding, for example, of how single amino acid changes affect the DNA sequence which a factor recognizes, how an acidic domain mediates activation and how phosphorylation or ligand binding can affect the activity of a transcription factor. In this manner it will eventually be possible to understand precisely how transcription factors interact with DNA, with each other and with the RNA polymerase itself in order to regulate transcription.

REFERENCES

- Darnell, J. E. (1982) *Nature (London)* **297**, 365–371
- Latchman, D. S. (1990) *Gene Regulation: A Eukaryotic Perspective*, Unwin Hyman, London
- Davidson, E. J., Jacobs, H. T. & Britten, R. J. (1983) *Nature (London)* **301**, 468–470
- Jones, N. C., Rigby, P. W. J. & Ziff, E. B. (1988) *Genes Dev.* **2**, 267–281
- Maniatis, T., Goodbourn, S. & Fischer, J. A. (1987) *Science* **236**, 1237–1245
- Hatzopoulos, A. K., Schlokot, U. & Gruss, P. (1988) in *Transcription and Splicing* (Hames, B. D. & Glover, D. M., eds.), pp. 43–96, IRL Press, Oxford
- Pelham, H. R. B. (1982) *Cell* **30**, 517–528
- Robins, D. M., Paek, I., Seeburg, P. H. & Axel, R. (1982) *Cell* **29**, 623–631
- Wirth, T., Staudt, L. & Baltimore, D. (1987) *Nature (London)* **329**, 174–178
- Beato, M. (1989) *Cell* **56**, 335–344
- Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383
- McKnight, S. & Tjian, R. (1986) *Cell* **46**, 795–805
- Kadonga, J. T. & Tjian, R. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5889–5893
- Kadonga, J. T., Carner, K. R., Masiarz, F. R. & Tjian, R. (1987) *Cell* **51**, 1079–1090
- Rosenfeld, P. J. & Kelley, T. J. (1986) *J. Biol. Chem.* **261**, 1398–1408
- Santoro, C., Mermod, N., Andrews, P. C. & Tjian, R. (1988) *Nature (London)* **334**, 218–224
- Singh, H., Le Bowitz, J. H., Baldwin, A. S. & Sharp, P. A. (1988) *Cell* **52**, 413–419
- Vinson, C. R., La Marco, K. L., Johnson, P. F., Landschulz, W. H. & McKnight, S. L. (1988) *Genes Dev.* **2**, 801–806
- Johnson, P. F. & McKnight, S. L. (1989) *Annu. Rev. Biochem.* **58**, 799–839
- La Thangue, N. B. & Rigby, P. W. J. (1988) in *Transcription and Splicing* (Hames, B. D. & Glover, D. M., eds.), pp. 3–42, IRL Press, Oxford
- Mitchell, P. J. & Tjian, R. (1989) *Science* **245**, 371–378
- Muller, M., Afrolter, M., Leupin, W., Otting, G., Wuthrich, K. & Gehring, U. J. (1988) *EMBO J.* **7**, 4299–4304
- Williams, T., Admon, A., Luscher, B. & Tjian, R. (1988) *Genes Dev.* **2**, 1557–1569
- Norman, C., Runswick, M., Pollock, R. & Treisman, R. (1988) *Cell* **55**, 989–1003
- Schleif, R. (1988) *Science* **241**, 1182–1187
- Struhl, K. (1989) *Trends Biochem. Sci.* **14**, 137–140
- Sakonju, S. & Brown, D. D. (1982) *Cell* **31**, 395–405
- Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4**, 1609–1614
- Klug, A. & Rhodes, D. (1987) *Trends Biochem. Sci.* **12**, 464–469
- Redemann, N., Gaul, U. & Jackle, H. (1988) *Nature (London)* **332**, 90–92
- Parraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T. & Klevitt, R. E. (1988) *Science* **241**, 1489–1492
- Ruiz-Altaba, A., Perry-O'Keefe, H. & Melton, D. A. (1987) *EMBO J.* **6**, 3065–3070
- Page, D. C., Mosher, R., Simpson, E. M., Fisher, E. M. C., Mardon, G., Pollack, J., McGillivray, B., DeLa Chappelle, A. & Braun, L. G. (1981) *Cell* **51**, 1091–1104
- Evans, R. M. & Hollenberg, S. M. (1988) *Cell* **52**, 1–3
- Evans, R. M. (1988) *Science* **240**, 889–895
- Yamamoto, K. R. (1985) *Annu. Rev. Genet.* **19**, 209–252
- Hollenberg, S. M., Giguere, V., Segui, P. & Evans, R. M. (1987) *Cell* **49**, 39–46
- Freedman, L. P., Luisi, B. F., Korszin, Z. R., Basuappa, R., Sigler, P. B. & Yamamoto, K. R. (1988) *Nature (London)* **334**, 543–546
- Frankel, A. D. & Pabo, C. O. (1988) *Cell* **53**, 775
- Berg, J. M. (1989) *Cell* **57**, 1065–1068
- Green, S. & Chambon, P. (1987) *Nature (London)* **325**, 75–78
- Green, S. & Chambon, P. (1988) *Trends Genet.* **4**, 309–314
- Danielsen, M., Hinck, L. & Ringold, G. M. (1989) *Cell* **57**, 1131–1138
- Umesono, K. & Evans, R. M. (1989) *Cell* **57**, 1139–1146
- Giniger, E., Varnum, S. M. & Ptashne, M. (1985) *Cell* **40**, 767–774
- Ma, J. & Ptashne, M. (1987) *Cell* **48**, 847–853
- Moran, E. & Mathews, M. B. (1987) *Cell* **48**, 177–178
- Ingham, P. W. (1988) *Nature (London)* **335**, 25–34
- Scott, M. P. & Carroll, S. B. (1987) *Cell* **51**, 689–698
- Gehring, W. J. (1987) *Science* **236**, 1245–1252
- Scott, M. P., Tamkun, J. W. & Hartzell, G. W. (1989) *Biochim. Biophys. Acta* **989**, 25–48
- McGinnis, W., Levine, M., Hafer, E., Kuriowa, A. & Gehring, W. J. (1984) *Nature (London)* **308**, 428–433
- Scott, M. P. & Weiner, A. J. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4115–4119
- Mihara, H. & Kaiser, E. T. (1988) *Science* **242**, 925–927
- Laughon, A. & Scott, M. P. (1984) *Nature (London)* **310**, 25–31
- Tsonis, P. A., Carperas, V. & Shihahan, T. (1988) *Biochem. Biophys. Res. Commun.* **157**, 100–105
- Quian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J. & Wuthrich, K. (1989) *Cell* **59**, 573–580
- Ptashne, M. (1986) *A Genetic Switch*, Cell Press and Blackwell Scientific Publications, Cambridge and Palo Alto
- Treisman, J., Gonczy, P., Vashishtha, M., Harris, E. & Desplan, C. (1989) *Cell* **59**, 553–562
- Hanes, S. P. & Brent, R. (1989) *Cell* **57**, 1275–1283
- Nasmyth, K. A. (1982) *Annu. Rev. Genet.* **16**, 493–500
- Shepherd, J. C. W., McGinnis, W., Carrasco, A. E., De Robertis, E. M. & Gehring, W. J. (1984) *Nature (London)* **310**, 70–71
- Akam, M. (1989) *Cell* **57**, 347–349
- Sommer, H., Belfran, J.-P., Huijser, P., Pape, H., Lonig, W.-E., Saedler, H. & Schwarz-Sommer, Z. (1990) *EMBO J.* **9**, 615–622
- Levine, M. & Hoey, T. (1988) *Cell* **55**, 537–540
- Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D., Sharp, P. A., Ingraham, J. A., Rosenfeld, M. G., Finney, M., Ruvkun, G. & Horvitz, H. R. (1988) *Genes Dev.* **2**, 1513–1516
- He, X., Treacy, M. N., Simmons, D. M., Ingraham, H. A., Swanson, L. S. & Rosenfeld, M. G. (1989) *Nature (London)* **340**, 35–42
- Sturm, R. A., Das, G. & Herr, W. (1988) *Genes Dev.* **2**, 1582–1599
- Clerc, R. G., Corcoran, L. M., Le Bowitz, J. H., Baltimore, D. & Sharp, P. A. (1988) *Genes Dev.* **2**, 1570–1581
- Ingraham, H. A., Chen, R., Mangalam, H. J., Elsholtz, H. P., Flynn, S. C., Linn, C. R., Simmons, D. M., Swanson, L. & Rosenfeld, M. G. (1988) *Cell* **55**, 519–529
- Finney, M., Ruvkun, G. & Horvitz, H. R. (1988) *Cell* **55**, 757–769
- Theill, L. E., Castrillo, J.-L., Wu, D. & Karin, M. (1989) *Nature (London)* **342**, 945–948
- Sturm, R. A. & Herr, W. (1988) *Nature (London)* **336**, 601–604

74. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240**, 1759–1764
75. Abel, T. & Maniatis, T. (1989) *Nature (London)* **341**, 24–25
76. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1989) *Science* **243**, 1681–1688
77. Gentz, R., Rauscher, F. J., Abate, C. & Curran, T. (1989) *Science* **243**, 1695–1699
78. Turner, R. & Tjian, R. (1989) *Science* **243**, 1689–1694
79. Kouzarides, T. & Ziff, E. (1988) *Nature (London)* **336**, 646–651
80. Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R. J., Ramsdorf, H. J., Jonat, C., Herrlich, P. & Karin, M. (1987) *Cell* **49**, 729–739
81. Lee, W., Mitchell, P. & Tjian, R. (1987) *Cell* **49**, 741–752
82. Bos, T. J., Bohmann, D., Tsuchie, H., Tjian, R. & Vogt, P. K. (1988) *Cell* **52**, 705–712
83. Rauscher, F. J., Cohen, D. R., Curran, T., Bos, T. J., Bogt, P. K., Bohmann, D., Tjian, R. & Franza, B. R. (1988) *Science* **240**, 1010–1016
84. Neuber, M., Adamkiewicz, J., Hunter, J. B. & Muller, R. (1989) *Nature (London)* **341**, 243–245
85. Prendergast, G. C. & Ziff, E. B. (1989) *Nature (London)* **341**, 392
86. Murre, C., McCaw, P. S. & Baltimore, D. (1989) *Cell* **56**, 777–783
87. Tapscott, S. J., Davis, R. L., Thayer, M. J., Cheng, P.-F., Weintraub, J. & Lassar, A. B. (1988) *Science* **242**, 405–411
88. Caudy, M., Vassin, J., Brand, M., Ruma, R., Jan, L. Y. & Jan, Y. N. (1988) *Cell* **55**, 1061–1067
89. Murre, C., McCaw, P. S., Vaessin, H., Candy, M., Jan, L. Y., Jan, Y. N., Cabera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., Weintraub, H. & Baltimore, D. (1989) *Cell* **58**, 537–544
90. Davis, R. L., Cheng, P.-F., Lassar, A. B. & Weintraub, H. (1990) *Cell* **60**, 733–746
91. Mermod, N., O'Neill, E. A., Kelley, T. J. & Tjian, R. (1989) *Cell* **58**, 741–753
92. Pinkham, J. L., Olesen, J. T. & Guarente, L. (1987) *Mol. Cell. Biol.* **7**, 578–587
93. Hahn, S., Pinkham, J., Wei, R., Miller, R. & Guarente, L. (1988) *Mol. Cell. Biol.* **8**, 655–663
94. Godowski, P. J., Picard, D. & Yamamoto, K. (1988) *Science* **241**, 812–816
95. Hollenberg, S. M. & Evans, R. M. (1988) *Cell* **55**, 899–906
96. Hope, I. A. & Struhl, K. (1986) *Cell* **46**, 885–894
97. Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. (1988) *Nature (London)* **335**, 563–564
98. O'Hare, P. & Hayward, G. S. (1987) *J. Virol.* **61**, 190–199
99. Preston, C. M., Frame, M. C. & Campbell, M. E. M. (1988) *Cell* **52**, 425–434
100. Ptashne, M. (1988) *Nature (London)* **335**, 683–689
101. Sigler, P. B. (1988) *Nature (London)* **333**, 210–212
102. Giniger, E. & Ptashne, M. (1987) *Nature (London)* **330**, 670–672
103. Courey, A. J. & Tjian, R. (1988) *Cell* **55**, 887–898
104. Courey, A. J., Holtzman, D. A., Jackson, S. P. & Tjian, R. (1989) *Cell* **59**, 827–836
105. Bohmann, D., Box, T. J., Adman, A., Nishimura, T., Vogt, P. K. & Tjian, R. (1987) *Science* **238**, 1386–1392
106. Sawadogo, M. & Roeder, R. G. (1985) *Cell* **43**, 165–175
107. Horikoshi, M., Carey, M. F., Kakidani, H. & Roeder, R. G. (1988) *Cell* **54**, 665–669
108. Horikoshi, M., Hai, T., Lin, Y.-S., Green, M. R. & Roeder, R. G. (1988) *Cell* **54**, 1033–1042
109. Brown, D. D. (1984) *Cell* **37**, 359–365
110. Levine, M. & Manley, J. L. (1989) *Cell* **59**, 405–408
111. Goodbourn, S., Burstein, H. & Maniatis, T. (1986) *Cell* **45**, 601–610
112. Zinn, K. & Maniatis, T. (1986) *Cell* **45**, 611–618
113. Gill, G. & Ptashne, M. (1988) *Nature (London)* **334**, 721–724
114. Lohr, D. & Hopper, J. E. (1985) *Nucleic Acids Res.* **13**, 8409–8423
115. Johnston, S. A., Salmeron, J. M. & Pincher, S. E. (1987) *Cell* **50**, 143–146
116. Keleher, C. A., Coutte, C. & Johnson, A. D. (1988) *Cell* **53**, 927–936
117. Scheidereit, C., Heguy, A. & Roeder, R. G. (1987) *Cell* **51**, 783–793
118. Muller, M. M., Ruppert, S., Schaffner, W. & Mathias, P. (1988) *Nature (London)* **336**, 544–551
119. Scheidereit, C., Cromlish, J. A., Gerster, T., Kawamaki, K., Balmaceda, G.-C., Currie, R. A. & Roeder, R. G. (1988) *Nature (London)* **336**, 551–557
120. Staudt, L. M., Clerc, R. G., Singh, H., Le Bowitz, J. H., Sharp, P. A. & Baltimore, D. (1988) *Science* **241**, 577–580
121. Xanthopoulos, K. G., Mirkovitch, J., Decker, T., Kuo, C. F. & Darnell, J. E. Jr. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4117–4121
122. Fink, G. R. (1986) *Cell* **45**, 155–156
123. Koenig, R. G., Lazar, M. A., Hodin, R. A., Brent, G. A., Larsen, P. R., Chin, W. W. & Moore, D. D. (1989) *Nature (London)* **337**, 659–661
124. Larosa, G. J. & Gudas, L. J. (1989) *Mol. Cell. Biol.* **8**, 3906–3917
125. Sap, J., Munoz, A., Damm, K., Goldberg, Y., Ghysdael, J., Leutz, A., Berg, H. & Vennstrom, B. (1986) *Nature (London)* **324**, 635–640
126. Sap, J., Munoz, A., Schmitt, A., Stunnenberg, H. & Vennstrom, B. (1989) *Nature (London)* **340**, 242–244
127. Furst, P., Hu, S., Hackett, R. & Hamer, D. (1988) *Cell* **55**, 705–717
128. Becker, P. B., Ruppert, S. & Schutz, G. (1987) *Cell* **51**, 435–443
129. Wilmann, T. & Beato, M. (1986) *Nature (London)* **324**, 688–691
130. Catelli, M. G., Binart, N., Jung-Testas, I., Renou, J. M., Baulieu, E. E., Foramisco, J. R. & Welch, W. J. (1985) *EMBO J.* **4**, 3131–3135
131. Sanchez, E. R., Toft, D. O., Schlesinger, M. J. & Pratt, W. B. (1985) *J. Biol. Chem.* **260**, 12398–12401
132. Sanchez, E. R., Meschini, S., Tienrungroj, W., Schlesinger, M. J., Toft, D. O. & Pratt, W. B. (1987) *J. Biol. Chem.* **262**, 6986–6991
133. Howard, K. J. & Distelhorst, C. W. (1988) *J. Biol. Chem.* **263**, 3474–3481
134. Baeuerle, P. A. & Baltimore, D. (1988) *Cell* **53**, 211–217
135. Baeuerle, P. A. & Baltimore, D. (1988) *Science* **242**, 540–546
136. Lenardo, M. J. & Baltimore, D. (1989) *Cell* **58**, 227–229
137. Davis, R. L., Weintraub, H. & Lassar, A. B. (1987) *Cell* **51**, 987–1000
138. Benezra, R., Davis, R. L., Lockshen, D., Turner, D. L. & Weintraub, H. (1990) *Cell* **61**, 49–59
139. Montminy, M. R. & Bilezikjian, L. M. (1987) *Nature (London)* **328**, 175–178
140. Nigg, E. A., Hilz, H., Eppenberger, H. M. & Dutly, F. (1985) *EMBO J.* **4**, 2801–2806
141. Gonzalez, G. A. & Montminy, M. R. (1989) *Cell* **59**, 675–680
142. Yamamoto, K. K., Gonzalez, G. A., Biggs, W. H. & Montminy, M. R. (1988) *Nature (London)* **334**, 494–498
143. Yamamoto, K. R., Gonzalez, G. A., Menzel, P., Rivier, J. & Montminy, M. R. (1990) *Cell* **60**, 611–617
144. Sorger, P. K., Lewis, M. J. & Pelham, H. R. B. (1987) *Nature (London)* **329**, 81–84
145. Sorger, P. K. & Pelham, H. R. B. (1988) *Cell* **54**, 855–864
146. Ghosh, S. & Baltimore, D. (1990) *Nature (London)* **344**, 677–682
147. Jackson, S. P. & Tjian, R. (1988) *Cell* **55**, 125–133