COMMENTARY

The role of DNA-binding proteins in differentiation and transformation

E. WHITELAW

Sir William Dunn School of Pathology, University of Oxford, South Parks Rd, Oxford OX1 3RE, UK

Introduction

Ten years ago we were unsure how closely the transcriptional regulation in eukaryotes mimicked that found in prokaryotes. Today it seems that the mechanisms are reassuringly similar. This may not be a very great surprise but it is certainly a great delight. Ptashne and others, working with the bacteriophage λ have shown in simple molecular terms how two DNA-binding proteins – the λ repressor and cro – could determine its mode of growth (for review, see Ptashne *et al.* 1980). These two proteins compete for binding at an operator. If the λ repressor is successful, transcription of lytic functions is repressed and lysogenic functions simultaneously activated, more repressor is produced and the phage enters stable lysogeny. If however the cro protein reaches the operator first, the balance swings the other way. Transcription of the λ repressor is prevented and free expression of the lytic proteins is allowed. So changes in the concentrations of these proteins have a far-reaching impact on λ's lifestyle. A variety of prokaryotic gene systems have been studied and clear patterns are emerging. The binding sites tend to be palindromes, recognized by transcription factors that bind as dimers. Transcription or repression can depend upon a very fine balance of concentrations and protein–protein interactions that favour an 'opening' or 'closing' of the promoter region.

Over the last few years a great deal of information has become available that emphasizes the importance of DNA-binding proteins in the differentiation of higher organisms through their actions on specific sets of genes. Furthermore, it is now becoming clear that protein–protein interactions, i.e. interactions between DNA-binding proteins, are also important. In this review, I will attempt to outline the types of experiments that have brought this to light and to discuss in detail a few examples from different systems that suggest that this will turn out to be a general mechanism.

Homeobox proteins are transcription factors

First, let us consider the homeobox proteins. This family of proteins is known to play a vital role in the development of multicellular organisms (for review see Scott and Carroll, 1987; Dressler, 1989; Wright *et al.* 1989). They are characterized by a highly conserved 'homeodomain' about 60 amino acids in length. Homeodomains, first recognized in *Drosophila*, have now been found in a wide variety of eukaryotic organisms ranging from yeast to human, and more than 80 such proteins have been identified. Originally the homeobox proteins were thought of as the key to segmentation in flies and vertebrates. However, more recently, they have assumed a much broader role in cell commitment in the early embryo. Almost every one of the homeobox genes in *Drosophila* is expressed in a characteristic subset of embryonic cells, and evidence from *in situ* hybridization suggests that each cell type may contain a unique combination of homeobox gene products. It is thought that these permutations of gene expression play a crucial part in establishing diverse pathways of morphogenesis.

The possibility that homeodomains might be sequence-specific DNA-binding proteins that directly influence transcription came from the realization that they shared considerable sequence similarity with the yeast mating type proteins α1 and **a**2 (Shepard *et al.* 1984). These yeast proteins were already known to be similar to various prokaryotic repressors that interact with DNA (Laughon and Scott, 1984). Furthermore, the recent discovery that three mammalian transcription factors that are known to be DNA-binding proteins contain homeobox sequences provides another clue as to how homeobox proteins control gene expression. The genes encoding these three factors, Oct-1, Oct-2 and Pit-1/GHF, are closely related and represent a new subfamily of homeobox genes (Herr *et al.* 1988).

Oct-1 is a ubiquitous protein that binds a conserved eight-nucleotide sequence called the octamer: ATTTGCAT. Oct-2 is a lymphoid-specific protein that binds the same sequence. This sequence is associated with promoter and enhancer elements of many genes and has been shown to play an important role in the transcriptional activation of these genes. Direct proof that these proteins are sequence-specific transcription factors has been obtained by adding purified proteins to *in vitro* transcription reactions (Scheidereit *et al.* 1987).

Pit-1 (or GHF-1) is a transcription factor involved in

Journal of Cell Science 94, 169–173 (1989)

the pituitary-specific expression of two related hormones, growth hormone (GH) and prolactin (PRL). It provides a stunning example of how a transcription factor, confined to a single tissue, can activate specified genes, in that tissue alone. In this case the GH and PRL genes are marked out since they contain the 9 base pair consensus $\begin{matrix}TT\\AA\end{matrix}$TATNCAT. Pit-1 recognizes this sequence and these genes are expressed (Nelson *et al.* 1988). The GH and PRL promoters can be transcriptionally activated *in vitro* by adding purified Pit-1 protein to HeLa nuclear extract (Nelson *et al.* 1988). The recent isolation of the cDNAs that encode Pit-1 and GHF-1 has led to the discovery that they both encode the same 291 amino acid protein and that their expression is restricted to the two cell types of the anterior pituitary that express the GH and PRL genes. Expression of Pit-1 in heterologous cells results in the activation of cotransfected growth hormone and prolactin promoters (Ingraham *et al.* 1988). The protein contains a homeodomain with 35 % amino acid identity to the *Drosophila* prd homeobox and 57 % and 55 % identity to Oct-1 and Oct-2, respectively (Herr *et al.* 1988).

The homeodomain proteins have long been thought of as factors that directly activate or repress the transcription of batteries of other genes and it is now becoming clear that the control is mediated directly by DNA-binding in promoter regions. For example, we know that the *Drosophila* Even-skipped (Eve), Engrailed (En) and Antennapedia (Antp) homeodomains bind DNA in a sequence-specific manner (Hoey and Levine, 1988; Desplan *et al.* 1988; Müller *et al.* 1988). Indeed, the homeobox proteins Eve and Fushi tarazu (Ftz) bind with high affinity to specific sites within the promoter of the segmentation gene engrailed (En). Eve and Ftz are known to participate in the activation of the engrailed gene. However, there is as yet little information about how the binding of these proteins at their recognition sequences activates or represses gene expression.

All homeodomains appear to contain a helix-turn-helix motif similar to regions previously shown in prokaryotic repressors (e.g. λ repressor) to interact with DNA (Otting *et al.* 1988). In prokaryotic repressors it is known that the sequence specificity is afforded by the second of these two helices (close to the carboxy terminus). It is this part of the homeodomain that is most conserved between different proteins. This suggests that many different homeoproteins are capable of binding to similar DNA target sequences. In fact, the recent experiments show that a number of different homeobox proteins can bind the same binding sites at the 5′ end of the En and Eve genes, lending support to the idea that regulatory interactions among homeobox genes involve competition of different proteins for the same *cis*-regulatory sequences (Hoey and Levine, 1988). In this model, different proteins, some perhaps activators and others repressors, 'compete' for binding to a specific site and the 'on/off' state of the target gene depends on which protein has the highest affinity and also on the intracellular concentration of the various proteins. This type of mechanism is analogous to the regulatory interactions of λ repressor and cro proteins discussed above.

Some oncogenes are transcriptional activators

Until recently, transcriptional regulators had not been shown to harbour oncogenic potential. The first definitive example of a transcription factor gene inducing cancer is the oncogene *jun*. Over the past year it has become evident that Jun is an important cellular control element that interacts with other proteins and with DNA, to affect transcription and induce oncogenic transformation.

The replication-defective retrovirus avian sarcoma virus 17 (ASV17), isolated from a chicken sarcoma (Maki *et al.* 1987), causes fibrosarcomas in chickens and oncogenic transformation in avian embryonic fibroblasts in culture. The oncogenic potential of ASV17 is due to the presence of the *jun* gene (Ball *et al.* 1989). The first clue as to the mechanism by which Jun triggered oncogenesis began with the realization that it shared 44 % homology with the well-characterized yeast transcriptional activator GCN4 (Vogt *et al.* 1987). The homology is restricted to the carboxy-terminal 70 amino acids, which corresponds to the DNA binding domain of both proteins. Furthermore, the Jun protein itself can function as GCN4 in yeast, despite the lack of homology in the amino-terminal two-thirds of the molecules (Struhl, 1988). This meant that there was both a structural and a functional homology between a yeast transcription factor and a vertebrate oncogene.

The consensus binding site of GCN4 and Jun, ATGA(C/G)TCAT, is closely related to that of the previously characterized mammalian transcription factor, AP-1. AP-1 (activator protein-1) was described initially as a DNA-binding activity in HeLa cell extracts that specifically recognizes the enhancer elements of simian virus 40 (SV40) (Lee *et al.* 1987). AP-1 binding sites also occur in the control regions of viral and cellular genes that are stimulated by treatment of cells with phorbol ester. Are AP-1 and Jun related? Antisera generated against peptides of the Jun sequence react with a $40\times10^3 M_r$ AP-1 protein and the tryptic peptides from this protein that react with Jun antibodies have amino acid sequences that match perfectly that of Jun (Bohmann *et al.* 1987). This suggests that Jun and AP-1 are identical. However, it turns out that AP-1 preparations contain several proteins and Jun is only one of the proteins that contribute to AP-1 activity. In fact, at present it is unclear how many cellular proteins regulate transcription *via* AP-1 sequence elements.

One such protein is Fos. The Fos oncogene (v-*fos*) causes osteogenic sarcomas *via* the FBJ murine sarcoma virus (Curran and Teich, 1982). Its normal cellular homologue, c-*fos*, encodes a nuclear protein (Fos), which participates in protein complexes with a $39\times10^3 M_r$ protein (p39). In fact Jun-specific antibodies enable workers to identify p39 as Jun (Rauscher *et al.* 1988). The expression of Fos is induced by a variety of extracellular stimuli, and so it was proposed that Fos is involved in gene regulation. A combination of studies revealed that Fos binds to the AP-1 consensus recognition sequence (Franza *et al.* 1988). However, it was not clear whether Fos and Fos-related proteins bind to the

AP-1 site directly or indirectly through binding to p39 (or Jun) (now called p39^{jun}). It turns out that they do both.

The AP-1 consensus sequence TGACTCA has dyad symmetry, suggesting that proteins interacting with this sequence bind as dimers. The *fos*-p39^{jun} complex is a heterodimer (Rauscher *et al.* 1988). Jun, like GCN4, can form homodimers that bind to DNA. However, heterodimers between Fos and Jun show increased affinity for the AP-1 binding site. The dimerization occurs by a novel protein–protein interaction called a 'leucine zipper' (Landshulz *et al.* 1988). Leucine zippers are amphipathic α-helices that contain four or five leucines at seven-residue intervals. Adjacent to the amino termini of the zipper domain are regions of basic residues that appear to constitute the DNA-contact points. Potential leucine zippers have also been found in other DNA-binding proteins, e.g. GCN4, Myc, C/EBP protein. Because there are several different Jun- and Fos-related proteins (JunB, JunD and Fra proteins) that associate as heterodimers, there is a possibility for achieving diverse combinational specificities in their interaction on promoters. (For anyone particularly interested in the transcription factors that are associated with oncogenic transformation, there are two excellent reviews: Curran and Franza, 1988; Vogt and Bos, 1989).

Is MyoDI a sequence-specific DNA binding protein?

Many proteins involved in transcriptional regulation have been discovered as a result of experiments in which DNA sequences responsible for tissue-specific expression have been identified by transfection of mutant or chimeric genes into cultured cells followed by the purification of proteins that bind to these sequences. While this biochemical approach remains unparalleled as a means of identifying DNA-binding factors that directly modulate transcription, obviously it will not be useful in finding regulators that act indirectly to control the expression of tissue-specific genes. An alternative approach, used previously to isolate oncogenes, has recently been successfully applied to the identification of genes that regulate mammalian development. This method can potentially help isolate regulatory genes irrespective of whether or not they act directly on the DNA. In this approach, DNA from a donor cell is transfected into cultured cells; the recipient cells are then assayed for the heritable expression of novel gene products.

This method has recently been used to identify two genes that regulate muscle cell development. All these experiments are based on the fact that a short exposure to 5-azacytidine converts a mouse embryonic fibroblast cell line, 10T½ cells, into myoblasts (Taylor and Jones, 1979). Hypomethylation of specific sequences was found to be involved in the myogenic conversion (Jones and Taylor, 1980; Konieczny *et al.* 1986). These studies suggested a simple model of myogenic conversion in which one or a few myogenic determination genes became activated by hypomethylation. That a single gene could convert 10T½ cells into myoblasts was demonstrated first by transfection of a short fragment of genomic DNA from mouse myoblast into these non-muscle cells (Lassar *et al.* 1986) and later by transfection of a mouse cDNA designated MyoDI (Davis *et al.* 1987). This cDNA was isolated using a technique called subtraction hybridization. MyoDI has subsequently been shown to be a nuclear phosphoprotein that is expressed in skeletal muscle *in vivo* and in certain muscle cell lines *in vitro*.

A second regulatory locus in the same pathway has now been reported by Pinney *et al.* (1988). They found that transfection into 10T½ cells of a human genomic fragment (called Myd) resulted in the accumulation of muscle proteins and of MyoDI mRNA, suggesting that the expression of Myd precedes that of MyoDI. Thus, DNA transfection has led to the identification of two regulatory loci that are sequentially expressed during myogenesis.

While *myoDI* and *myd* clearly play important roles in establishing the myogenic lineage, there is considerable evidence for the involvement of additional regulatory genes in myogenic determination and differentiation. Exposure of 10T½ cells to 5-azacytidine gives rise to myoblasts at high frequency, whereas other fibroblast lines are converted at low frequency, suggesting that 10T½ cells may already express one or more genes that are involved in conversion to the myogenic lineage (Taylor and Jones, 1979). Similarly, cells of mesodermal origin are more efficiently converted to myoblasts by MyoDI than are nonmesodermally derived cells (Davis *et al.* 1987). These observations suggest that MyoDI may normally cooperate with additional genes to confer myogenic phenotype and it is reasonable to predict that expression of a differentiated muscle phenotype may involve interactions between multiple regulatory genes. In fact a third gene, myogenin, with homology to both *myoDI* and c-*myc* has been found recently that is expressed during myogenesis and is sufficient to activate the muscle differentiation programme (Edmondsen and Olson, 1989). However, it is still not clear what the function is of the 22 amino acid domain that is conserved among myogenin, MyoDI and c-Myc. The exact mechanism of action of these proteins remains unclear although there is some evidence that MyoDI acts as a sequence-specific DNA-binding protein (Tapscott *et al.* 1988). It will indeed be ironic if these proteins should turn out to be conventional DNA-binding proteins, when the assay used to find them was specifically chosen to permit the discovery of proteins acting indirectly.

Autoregulation

The results from both the work on myogenic differentiation and that on homeoboxes in development suggest that a complex network of regulatory circuitry is involved in the development of higher organisms and it seems likely that this circuitry exists in the form of a hierarchy in which the protein from one gene binds to and activates or represses the promoter of other genes and the proteins from these genes bind to and regulate the expression of yet more genes. An interesting recent discovery in the study of eukaryotic transcription factors is that of auto-

regulation. Numerous homeobox-containing genes in *Drosophila* control their own expression by positive autogenous regulation. One example is the fushi tarazu (*ftz*) gene where the Ftz product recognizes the sequence TCAATTAAAT located within its own enhancer and autoregulates its own expression (for review, see Serfling, 1989). Similarly the mammalian factor Oct-2, which controls the expression of immunoglobulin (Ig) genes, positively regulates its own expression (Thali *et al.* 1988). The proto-oncogene c-*jun*, is also positively autoregulated by its own gene product (Angel *et al.* 1988). *fos* on the other hand can act as both an activator and a repressor of gene activity and generally represses its own transcription (Sassone-Corsi *et al.* 1988). Why autoregulate? In response to external stimuli such as growth factors or other 'morphogens', cells initiate biochemical cascades that end with the reprogramming of gene expression. This involves signal transduction from the outside of the cell *via* receptors to the nucleus. The existence of regulatory networks between genes coding for factors that control the transcription of structural genes associated with terminal differentiation and autoregulation of such genes enables short-term external signals to become permanent decisions in cellular commitment.

So, work over the last few years has shown us some ways of approaching the problem of discovering how a particular group of genes is turned on, in a particular tissue at a particular time of development. One can use the biochemical approach, by studying the DNA sequences in and around your gene of interest, looking for sequences involved in transcriptional regulation and then looking for proteins that will bind these sequences *in vitro*. Alternatively, one can use the 'transfection' approach, by transfecting undifferentiated cells with DNA fragments from a differentiated cell type and then assaying for those cells that have acquired specific DNA sequence that switches on the gene of interest. In this article I have picked just a few examples where these methods have been used to great effect. I am sure that the particular instances that I have discussed will turn out to be the tip of a very large iceberg.

References

ANGEL, P., HATTORI, K., SMEAL, T. AND KARIN, M. (1988). The *jun* proto-oncogene is positively autoregulated by its product, Jun/AP-1. *Cell* **55**, 875–885.

BALL, A. R., JR, BOS, T. J., LÖLIGER, C., NAGATA, L. P., NISHIMURA, T., SU, H., TSUCHIE, H. AND VOGT, P. K. (1989). *Cold Spring Harbor Symp. quant. Biol.* **53** (in press).

BOHMANN, D., BOS, T. J., ADMON, A., NISHIMURA, T., VOGT, P. K. AND TJIAN, R. (1987). Human proto-oncogene c-*jun* encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. *Science* **238**, 1386–1392.

CURRAN, T. AND FRANZA, B. R., JR (1988). Fos and Jun: the AP-1 connection. *Cell* **55**, 395–397.

CURRAN, T. AND TEICH, N. M. (1982). Candidate product of the FBJ murine osteosarcoma virus oncogene: characterization of a 55,000-dalton phosphoprotein. *J. Virol.* **42**, 114–122.

DAVIS, R. L., WEINTRAUB, H. AND LASSAR, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000.

DESPLAN, C., THEIS, J. AND O'FARRELL, H. (1988). The sequence specificity of homeodomain-DNA interaction. *Cell* **54**, 1081–1090.

DRESSLER, G. R. (1989). An update on the vertebrate homeobox. *Trends Genet.* **5**, 129–131.

EDMONDSEN, D. G. AND OLSON, E. N. (1989). A gene with homology to the *myc* similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. *Genes Dev.* **3**, 628–640.

FRANZA, B. R., JR, RAUSCHER, F. J. III, JOSEPHS, S. F. AND CURRAN, T. (1988). The fos complex and fos-related antigens recognise sequence elements that contain AP-1 binding sites. *Science* **239**, 1150–1153.

HERR, W., STURM, R. A., CLERC, R. G., CORCORAN, L. M., BALTIMORE, D., SHARP, P. A., INGRAHAM, H. A., ROSENFELD, M. G., FINNEY, M., RUVKUN, G. AND HORVITZ, H. R. (1988). The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2 and unc-86 gene products. *Genes Dev.* **2**, 1513–1516.

HOEY, T. AND LEVINE, M. (1988). Divergent homeo box proteins recognise similar DNA sequences in *Drosophila*. *Nature, Lond.* **332**, 858–861.

INGRAHAM, H. A., CHEN, R., MARGALAM, H. J., ELSHOLTZ, H. P., FLYNN, S. E., LIN, C. R., SIMMONS, D. M., SWANSON, L. AND ROSENFELD, M. G. (1988). A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. *Cell* **55**, 519–529.

JONES, P. A. AND TAYLOR, S. M. (1980). Cellular differentiation, cytidine analogs and DNA methylation. *Cell* **20**, 85–93.

KONIECZNY, S. F., BALDWIN, A. S. AND EMERSON, C. P., JR (1986). Myogenic determination and differentiation of 10T½ lineages: evidence for a single genetic regulatory system. In *Molecular and Cellular Biology, New Series* (ed. C. Emerson, D. A. Fischman, B. Nadal-Ginard and M. A. W. Siddiqui), vol. 29, pp. 21–34. A. R. Liss, New York.

LANDSCHULZ, W. H., JOHNSON, P. F. AND MCKNIGHT, S. L. (1988). The leucine zipper: a hypomethylated structure common to a new class of DNA binding proteins. *Science* **240**, 1759–1764.

LASSAR, A. B., PATERSON, B. M. AND WEINTRAUB, H. (1986). Transfection of a DNA locus that mediates the conversion of 10T½ fibroblasts to myoblasts. *Cell* **47**, 649–656.

LAUGHON, A. AND SCOTT, M. P. (1984). Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. *Nature, Lond.* **310**, 25–31.

LEE, W., HASLINGER, A., KARIN, M. AND TJIAN, R. (1987). Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. *Nature, Lond.* **325**, 368–372.

MAKI, Y., BOS, T. J., DAVIS, C., STARBUCK, M. AND VOGT, P. K. (1987). Avian sarcoma virus 17 carries the jun oncogene. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2848–2852.

MÜLLER, M., AFFOLTER, M., LEUPIN, W., OTTING, G., WÜTHRICH, K. AND GEHRING, W. J. (1988). Isolation and sequence-specific DNA binding of the *Antennapedia* homeodomain. *EMBO J.* **7**, 4299–4304.

NELSON, C., ALBERT, V. R., ELSHOLTZ, H. P., LU, L. AND ROSENFELD, M. G. (1988). Activation of cell-specific expression of rat growth hormone and prolactin genes by a common transcription factor. *Science* **239**, 1400–1405.

OTTING, G., QIAN, Y., MÜLLER, M., AFFOLTER, M., GEHRING, W. AND WÜTHRICH, K. (1988). Secondary structure determination for the *Antennapedia* homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif. *EMBO J.* **7**, 4305–4309.

PINNEY, D. F., PEARSON-WHITE, S. H., KONIECZNY, S. F., LATHAM, K. E. AND EMERSON, C. P., JR (1988). Myogenic lineage determination and differentiation: evidence for a regulatory gene pathway. *Cell* **53**, 781–793.

PTASHNE, M., JEFFREY, A., JOHNSON, A. D., MAURER, R., MEYER, B. J., PABO, C. O., ROBERTS, T. M. AND SAUER, R. T. (1980). How the λ repressor and cro work. *Cell* **19**, 1–11.

RAUSCHER, F. J. III, COHEN, D. R., CURRAN, T., BOS, T. J., VOGT, P. K., BOHMANN, D., TJIAN, R. AND FRANZA, B. R., JR (1988). Fos-associated protein p39 is the product of the jun proto-oncogene. *Science* **240**, 1010–1016.

SASSONE-CORSI, P., SISSON, J. C. AND VERMA, I. M. (1988). Transcriptional autoregulation of the proto-oncogene *fos*. *Nature, Lond.* **334**, 314–319.

SCHEIDEREIT, C., HEGUY, A. AND ROEDER, R. G. (1987). Identification and purification of a human lymphoid-specific octamer-binding protein (OTF-2) that activates transcription of an immunoglobulin promoter *in vitro*. *Cell* **51**, 783–793.

SCOTT, M. P. AND CARROLL, S. B. (1987). The segmentation and homeotic gene network in early *Drosophila* development. *Cell* **51**, 689–698.

SERFLING, E. (1989). Autoregulation – a common property of eukaryotic transcription factors? *Trends Genet.* **5**, 131–133.

SHEPHARD, J. C. W., MCGINNIS, W., CARRASCO, A. E., DE ROBERTIS, E. M. AND GEHRING, W. J. (1984). Fly and frog homoeodomains show homologies with yeast mating type regulatory proteins. *Nature, Lond.* **310**, 70–71.

STRUHL, K. (1988). The Jun oncoprotein, a vertebrate transcription factor, activates transcription in yeast. *Nature, Lond.* **322**, 649–650.

TAPSCOTT, S. J., DAVIS, R. L., THAYER, M. J., CHENG, P.-F., WEINTRAUB, H. AND LASSAR, A. B. (1988). MyoD1: a nuclear phosphoprotein requiring a myc homology region to convert fibroblasts to myoblasts. *Science* **242**, 405–411.

TAYLOR, S. M. AND JONES, P. A. (1979). Multiple new phenotypes induced in 10T½ and 3T3 cells treated with 5-azacytidine. *Cell* **17**, 771–779.

THALI, M., MÜLLER, M. M., DELORENZI, M., MATTHIAS, P. AND BIENZ, M. (1988). Drosophila homeotic genes encode transcriptional activators similar to mammalian OTF-2. *Nature, Lond.* **336**, 598–601.

VOGT, P. K. AND BOS, T. J. (1989). The oncogene *jun* and nuclear signalling. *Trends Biothem.* **14**, 172–175.

VOGT, P. K., BOS, T. J. AND DOOLITTLE, R. F. (1987). Homology between the DNA-binding domain of the GCN4 regulatory protein of yeast and the carboxyl-terminal region of a protein coded for by the oncogene *jun*. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3316–3319.

WRIGHT, C. V. E., CHO, K. W. Y., OLIVER, G. AND DE ROBERTIS, E. M. (1989). Vertebrate homeodomain proteins: families of region-specific transcription factors. *Trends Biochem.* **14**, 52–56.