DEVELOPMENTALLY REGULATED TRANSCRIPTION FACTORS IN *DROSOPHILA MELANOGASTER*

Thesis by

Xiangdong Fang

In Partial Fulfillment of the Requirements For the

Degree of Doctor of Philosophy

California Institute of Technology

Pasadena, California

May 30, 2001

To my parents

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Carl S. Parker, for giving me the opportunity to conduct the research work in his laboratory. I am especially grateful to his continuous support and encouragement during all these years, which make the completion of this thesis possible. I would also like to thank all current and past members of Parker's research group, especially Tianxin Chen and Kim Tran, for their help and suggestions. I thank Dr. Kulkani Prakash for providing me basic technical trainings. And finally, I would like to extend my sincere thanks to Dr. Yue Lei for all the support she has given to me.

ABSTRACT

During early stages of Drosophila development the heat shock response cannot be induced. It is reasoned that the adverse effects on cell cycle and cell growth brought about by Hsp70 induction must outweigh the beneficial aspects of Hsp70 induction in the early embryo. Although the Drosophila heat shock transcription factor (dHSF) is abundant in the early embryo, it does not enter the nucleus in response to heat shock. In older embryos and in cultured cells the factor is localized within the nucleus in an apparent trimeric structure that binds DNA with high affinity. The domain responsible for nuclear localization upon stress resides between residues 390 and 420 of the dHSF. Using that domain as bait in a yeast two-hybrid system we now report the identification and cloning of a nuclear transport protein Drosophila karyopherin-α3 (dKap-α3). Biochemical methods demonstrate that the dKap-α3 protein binds specifically to the dHSF's nuclear localization sequence (NLS). Furthermore, the dKap-α3 protein does not associate with NLSs that contain point mutations which are not transported in vivo. Nuclear docking studies also demonstrate specific nuclear targeting of the NLS substrate by dKap-α3. Consistant with previous studies demonstrating that early Drosophila embryos are refractory to heat shock as a result of dHSF nuclear exclusion, we demonstrate that the early embryo is deficient in dKap-α3 protein through cycle 12. From cycle 13 onward the transport factor is present and the dHSF is localized within the nucleus thus allowing the embryo to respond to heat shock.

The pair-rule gene *fushi tarazu (ftz)* is a well-studied zygotic segmentation gene that is necessary for the development of the even-numbered parasegments in *Drosophila melanogastor.* During early embryogenesis, *ftz* is expressed in a characteristic pattern of seven stripes, one in each of the even-numbered parasegments. With a view to understand how *ftz* is transcriptionally regulated, cDNAs that encode transcription factors that bind to the zebra element of the *ftz* promoter have been cloned. Chapter III reports the cloning and characterization of the cDNA encoding zeb-1 (zebra element binding protein), a novel steroid receptor-like molecule that specifically binds to a key regulatory element of the *ftz* promoter. In transient transfection assays employing *Drosophila* tissue culture cells, it has been shown that zeb-1 as well as a truncated zeb-1 polypeptide (zeb480) that lacks the putative ligand binding domain function as sequence-specific trans-activators of the *ftz* gene.

The Oct factors are members of the POU family of transcription factors that are shown to play important roles during development in mammals. Chapter IV reports the cDNA cloning and expression of a *Drosophila* Oct transcription factor. Whole mount *in-situ* hybridization experiments revealed that the spatial expression patterns of this gene during embryonic development have not yet been observed for any other gene. In early embryogenesis, its transcripts are transiently expressed as a wide uniform band from 20-40% of the egg length, very similar to that of gap genes. This pattern progressively resolves into a series of narrower stripes followed by expression in fourteen stripes. Subsequently, transcripts from this gene are expressed in the central nervous

system and the brain. When expressed in the yeast *Saccharomyces cerevisiae,* this *Drosophila* factor functions as a strong, octamer-dependent activator of transcription. The data strongly suggest possible functions for the Oct factor in pattern formation in *Drosophila* that might transcend the boundaries of genetically defined segmentation genes.

TABLE OF CONTENTS

page

CHAPTER ONE

INTRODUCTION

Transcriptional regulations

The fundamental of molecular biology is that DNA produces RNA which in turn produces protein. The process of transcription, whereby an RNA product is produced from the DNA, is therefore an essential element in gene expression. The failure of this process to occur will obviously render redundant all the other steps which follow the production of the initial RNA transcript in eukaryotes, such as RNA splicing, transport to the cytoplasm or translation into protein (Nevins, 1983; Latchman, 1998). The central role of transcription in the process of gene expression also renders it an attractive control point for regulating the expression of genes in particular cell types or in response to a particular signal. Indeed, it is now clear that, in the vast majority of cases, where a particular protein is produced only in a particular tissue or in response to a particular signal, this is achieved by control processes which ensure that its corresponding gene is transcribed only in that tissue or in response to such a signal (Darnell, 1982; Latchman, 1998). For example, the genes encoding the immunoglobulin heavy and light chains of the antibody molecule are transcribed at high level only in the antibody-producing B cells, whilst the increase in somatostatin production in response to treatment of cells with cyclic AMP is mediated by increased transcription of the corresponding gene. Therefore, while post-transcriptional regulation affecting, for example, RNA splicing or stability, plays some role in the regulation of gene expression (Ross, 1996; Wang and Manley, 1997), the major control point lies at the level of transcription.

The central role of transcription both in the basic process of gene expression and its regulation in particular tissues has led to considerable study of this process. Initially such studies focused on the nature of the DNA sequences within individual genes which were

essential for either basal or regulated gene expression. In prokaryotes, such sequences are found immediately upstream of the start site of transcription and form part of the promoter directing expression of the genes. Sequences found at this position include both elements found in all genes which are involved in the basic process of transcription itself and those found in a more limited number of genes which mediate their response to a particular signal (Travers, 1993; Muller-Hill, 1996).

Early studies of cloned eukaryotic genes, therefore, concentrated on the region immediately upstream of the transcribed region where, by analogy, sequences involved in transcription and its regulation should be located. Putative regulatory sequences were identified by comparison between different genes and the conclusions reached in this way confirmed either by destroying these sequences by deletion or mutation, or by transferring them to another gene in an attempt to alter its pattern of regulation. This work, carried out on a number of different genes encoding specific proteins, identified many short sequence elements involved in transcriptional control (Davidson et al., 1983; Jones et al., 1988). The elements of this type present in two typical examples, the human gene encoding HSP70 protein (Williams *et* al., 1989) and the human metallothionein IIA gene (Lee *et* al., 1987) are illustrated in Figure 1.1.

Comparisons of these and many other genes revealed that, as in bacteria, their upstream regions contain two types of elements: firstly, sequences found in very many genes exhibiting distinct patterns of regulation which are likely to be involved in the basic process of transcription itself, and, secondly, those found only in genes transcribed in a particular tissue or in response to a specific signal which are likely to produce this specific pattern of expression.

Although they are regulated very differently, both the *hsp70* and metallothionein genes contain a TATA box. This is an AT-rich sequence (consensus TATAA/TAA/T) which is located about 30 base pairs upstream of the transcriptional start site in most of genes. Mutagenesis or relocation of this sequence has shown that it plays an essential role in accurately positioning the start site of transcription (Breathnach and Chambon, 1981). The region of the gene bracketed by the TATA box and the site of transcriptional initiation (the Cap site) has been operationally defined as the gene promoter or core promoter (Goodwin et al., 1990). It is likely that this region binds several proteins essential for transcription, as well as RNA polymerase II itself (Sentenac, 1985). Although the TATA box is found in most eukaryotic genes, it is absent in some genes, notably housekeeping genes expressed in all tissues and in some tissue specific genes (Weis and Reinberg, 1992). In these promoters, a sequence known as the initiator element, which is located over the start site of transcription appears to play a critical role in determining the initiation point and acts as a minimal promoter capable of producing basal levels of transcription.

In promoters with or without TATA box, the very low activity of the promoter itself is dramatically increased by other elements located upstream of the promoter. These elements are found in a very wide variety of genes with different patterns of expression indicating that they play a role in stimulating the constitutive activity of promoters. Thus inspection of the *hsp70* and metallothionein IIA genes reveals that both contain one or more copies of a GC-rich sequence known as the Sp1 box which is found upstream of the promoter in many genes both with and without TATA boxes (Dynan and Tjian, 1985).

In addition, the *hsp70* promoter but not the metallothionein promoter contains another sequence, the CCAAT box, which is also found in very many genes with *disparate* patterns of regulation. Both the CCAAT box and the Sp1 box are typically found upstream of the TATA box. Some genes, as in the case of *hsp70,* may have both of these elements, whereas others such as the metallothionein gene have single or multiple copies of one or the other (McKnight and Tjian, 1986). In every case, however, these elements are essential for transcription of the genes, and their elimination by deletion or mutation abolishes transcription. Hence these sequences play an essential role in efficient transcription of the gene and have been termed upstream promoter elements (UPE) (Goodwin et *al.,* 1990).

Inspection of the *hsp70* promoter (Figure 1.1) reveals several other sequence elements which are only shared with a much more limited number of other genes and which are interdigitated with the UPE. Indeed, one of these, which is located approximately 90 bases upstream of the transcriptional start site, is shared only with other heat-shock genes whose transcription is increased in response to elevated temperature. This suggests that this heat-shock element may be essential for the regulated transcription of the *hsp70* gene in response to heat.

To prove this directly, however, it is necessary to transfer this sequence to a non-heat-inducible gene and show that this transfer renders the recipient gene heat inducible. Pelham (1982) successfully achieved this by linking the heat-shock element to the non-heat inducible thymidine kinase gene of the eukaryotic virus herpes simplex. This hybrid gene could be activated following its introduction into mammalian cells by raising the temperature. Hence the heat-shock element can confer heat inducibility on another

gene, directly proving that its presence in the *hsp* gene promoters is responsible for their heat inducibility. Moreover, although these experiments used a heat-shock element taken from the *hsp70* gene of the fruit fly *Drosphila melanogaster,* the hybrid gene was introduced into mammalian cells. Not only does the successful functioning of the fly element in mammalian cells indicate that this process is evolutionarily conserved, it also permits a further conclusion about the way in which the effect operates. Thus, in the cold blooded *Drosophila,* 37^0C represents a thermally stressful temperature and the heat-shock response would normally be active at this temperature. The hybrid gene was inactive at 37^0C in the mammalian cells, however, and was only induced at 42^0C, the heat-shock temperature characteristic of the cell into which it was introduced. Hence this sequence does not act as a thermostat set to go off at a particular temperature, since this would occur at the *Drosophila* heat-shock temperature. Rather, this sequence must act by being recognized by a cellular protein which is activated only at an elevated temperature characteristic of the mammalian cell heat-shock response. This experiment, therefore, not only directly proves the importance of the heat-shock element in producing the heat inducibility of the *hsp70* gene but also shows that this sequence acts by binding a cellular protein which is activated in response to elevated temperature. The binding of this transcription factor then activates transcription of the *hsp70* gene.

The presence of specific DNA sequences, which can bind particular proteins, will therefore confer on a specific gene the ability to respond to particular stimuli. Thus the lack of a heat-shock element in the metallothionein IIA gene means that this gene is not heat inducible. In contrast, however, this gene, unlike the *hsp70* gene, contains a glucocorticoid response element (GRE). Hence it can bind the complex of the

glucocorticoid receptor and the hormone itself which forms following treatment of cells with glucocorticoid. Its transcription is therefore activated in response to glucocorticoid, whereas that of the *hsp70* gene is not. Similarly, only the metallothionein gene contains metal response elements (MRE), allowing it to be activated in response to treatment with heavy metals such as zinc and cadmium (Thiele, 1992). In contrast, both genes contain binding sites for the transcription factor AP2 which mediates gene activation in *sp70* response to cyclic AMP and phorbol esters.

Similar DNA sequence elements in the promoters of tissue-specific genes play a critical role in producing their tissue specific pattern of expression by binding transcription factors which are present in an active form only in a particular tissue where the gene will be activated. For example, the promoters of the immunoglobulin heavy- and light-chain genes contain a sequence known as the octamer motif (ATGCAAAT), which can confer B cell specific expression on an unrelated promoter (Wirth et al., 1987). Similarly, the related sequence ATGAATAA/T is found in genes expressed specifically in the anterior pituitary gland such as the prolactin gene and the growth hormone gene, and binds a transcription factor known as Pit-1 which is expressed only in the anterior pituitary (Andersen and Rosenfeld, 1994). If this short sequence is inserted upstream of a promoter, the gene is expressed only in pituitary cells. In contrast the octamer motif which differs by only two bases will direct expression only in B cells when inserted upstream of the same promoter (Elsholtz et al., 1990). Hence small differences in control element sequences can produce radically different patterns of gene expression.

One of the characteristic features of eukaryotic gene expression is the existence of sequence elements located at great distances from the start site of transcription which

can influence the level of gene expression. These elements can be located upstream, downstream or within a transcription unit, and function in either orientation relative to the start site of transcription (Figure 2). They act by increasing the activity of a promoter, although they lack promoter activity themselves and are hence referred to as enhancers (Hatzopoulos et al., 1988; Muller *et* al., 1988). Some enhancers are active in all tissues and increase the activity of a promoter in all cell types whilst others function as tissue-specific enhancers which activate a particular promoter only in a specific cell type. Thus the enhancer located in the intervening region of the immunoglobulin genes is active only in B cells and the B-cell-specific expression of the immunoglobulin genes is produced by the interaction of this enhancer and the immunoglobulin promoter, which, as we have previously seen, is also B-cell specific (Garcia et al., 1986).

As with promoter elements, enhancers contain multiple binding sites for transcription factors which interact together (Carey, 1998). In many cases these elements are identical to those contained immediately upstream of gene promoters. Thus the immunoglobulin heavy-chain enhancer contains a copy of the octamer sequence (Sen and Baltimore, 1986) which is also found in the immunoglobulin promoters. Similarly, multiple copies of the heat-shock consensus element are located far upstream of the start site in the *Xenopus hsp70* gene and function as a heat-inducible enhancer when transferred to another gene (Bienz and Pelham, 1986). Enhancers, therefore, consist of sequence elements which are also present in similarly regulated promoters and may be found within the enhancer associated with other control elements or in multiple copies.

The typical eukaryotic gene will therefore consist of up to four distinct transcriptional control elements (Figure 3). These are as follows: firstly, the promoter

itself; secondly, upstream promoter elements located close to it which are required for efficient transcription in any cell type; thirdly, other elements adjacent to the promoter which are interdigitated with the UPEs and which activate the gene in particular tissues or in response to particular stimuli; and, lastly, more distant enhancer elements which increase gene activity either in all tissues or in a regulated manner.

Such sequences often act by binding positively acting transcription factors which then stimulate transcription. Interestingly, however, although most sequences act in such a positive way, some sequences do appear to act in a negative manner to inhibit transcription. Such silencer elements have been defined in a number of genes including the cellular oneogene *c-myc* and those encoding proteins such as growth hormone or collagen type II. As with activating sequences, some silencer elements are constitutively active whilst others display cell-type specific activity. Thus, for example, the silencer in the gene encoding the T-lymphocyte marker CD4 represses its expression in most T cells where CD4 is not expressed but is inactive in a subset of T cells allowing these cells to express the CD4 protein actively (Sawada et al., 1994). In many cases silencer elements have been shown to act by binding transcription factors which then act to reduce the rate of transcription.

Obviously the balance between positively and negatively acting transcription factors which bind to the regulatory regions of a particular gene will determine the rate of gene transciption in any particular situation. In some cases, binding of the RNA polymerase and associated factors to the promoter and of other positive factors to the UPEs will be sufficient for transcription to occur, and the gene will be expressed constitutively. In other cases, however, such interactions will be insufficient and transcription of the gene will

occur only in response to the binding, to another DNA sequence, of a factor which is activated in response to a particular stimulus or is present only in a particular tissue. These regulatory factors will then interact with the constitutive factors allowing transcription to occur. Hence their binding will result in the observed tissue-specific or inducible pattern of gene expression.

Such interaction is well illustrated by the metallothionein IIA gene. As illustrated in Figure 1, this gene contains a binding site for the transcription factor AP1 which produces induction of gene expression in response to phorbol ester treatment. The action of AP1 on the expression of the metallothionein gene is abolished, however, both by mutations in its binding site and by mutations in the adjacent Sp1 motif which prevent this motif binding its corresponding transcription factor Spl (Lee *et al.*, 1987). Although these mutations in the Sp1 motif do not abolish AP1 binding, they do prevent its action indicating that the inducible AP1 factor interacts with the constitutive Sp1 factor to activate transcription. Clearly such interactions between bound transcription factors need not be confined to factors bound to regions adjacent to the promoter, but can also involve the similar factors bound to more distant enhancers. It is likely that this is achieved by a looping out of the intervening DNA allowing contact between factors bound at the promoter and those bound at the enhancer (Latchman, 1998).

This need for transcription factors to interact with one another to stimulate transcription means that transcription can also be stimulated by a class of factors which act indirectly by binding to the DNA and bending it so that other DNA-bound factors can interact with one another. Thus, the LEF-1 factor, which is specifically expressed in T lymphocytes, binds to the enhancer of the T-cell receptor α gene and bends the DNA so

that other constitutively expressed transcription factors can interact with one another thereby allowing them to activate transcription. This results in the T-cell-specific *expression* of the gene even though the directly activating factors are not expressed in a T-cell specific manner (Werner and Burley, 1997).

Inducible gene expression and heat shock response

All cells from bacteria to mammals respond to various treatments by activating or repressing the expression of particular genes. Genes that are activated in response to a specific treatment share a short DNA sequence in their promoters or enhancers whose transfer to another gene renders that gene inducible by the specific treatment. In turn, such sequences act by binding a specific transcription factor which becomes activated in response to the stimulus once activated, this factor interacts with the constitutive transcription factors resulting in increased transcription of the gene. A selection of DNA sequences which enable a gene to respond to a particular stimulus and the transcription factors which bind to them is given in Table 1.1 (Latchman, 1998).

When living cells are under stressful condition, such as exposure to elevated temperature, toxin, heavy metals, and bacterial and viral infection, they must cope with a rapid change in the patterns of gene expression, resulting in the elevated expression of a family of heat shock proteins (HSPs) and molecular chaperones (Craig, 1985; Lindquist Craig, 1988; Morimoto, 1994). HSPs have essential roles in the synthesis, transport and translocation of cellular proteins and in the regulation of protein confirmation. They constitute 5 to 10 percent of the total protein mass in the cells growing under ambient condition. These heat-inducible genes share a common DNA sequence when

Table 1.2 Sequences that confer response to a particular stimulus

Consensus sequences	Response to	Protein factor	Gene containing sequences
CTNGAATNTT CTAGA	Heat	Heat-shock transcription factor	*hsp7O, hsp83, hsp27,* etc.,
T/G T/A CGTCA	Cyclic AMP	CREB/ATF	Sornatostatin fibronectin, a-gonadotrophin *c-fos, hsp70*
TGAGTCAG	Phorbol esters	API	Metallothionein IIA, (xl-antitrypsin, collagenase
CC(A/T),GG	Growth factor in serum	Serum response factor	*c-fos, Xenopus* y-actin
RGRACNNN TGTYCY	Glucocorticoid	Glucocorticoid receptors	Metallothionein IIA, tryptophan oxygenase, uteroglobin, lysozyme
RGGTCANNN TGACCY	Oestrogen	Oestrogen receptor	Ovalbumin, conalbumin, vitellogenin
RGGTCAT GACCY	Thyroid hormone retinoic acid	Thyroid hormone receptors	Growth hormone, myosin heavy chain
TGCGCCCGCC	Heavy metals	Mep-1	Metallothionein genes
AGTTTCNN TTTCNC/T	Interferon-a	Stat- 1 Stat-2	Oligo A synthetase guanylate-binding protein
TTNCNNNAA	Interferon-y	Stat- 1	Guanylate-binding protein, Fe

N indicates that any base can be present at that position, R indicates a purine, i.e., A or G, Y indicates a pyrimidine, i.e., C or T.

transferred to another gene, can render the second gene inducible. This sequence is known as the heat-shock element (HSE). The manner in which a *Drosophila* HSE when introduced into mammal cells functioned at the mammalian rather than the *Drosophila* heat shock temperature suggested that this sequence acted by binding a protein rather than by acting directly as a thermosensor.

Direct evidence that this was the case was provided by studying the proteins bound to the promoters of the *hsp* genes before and after heat shock. Thus, prior to heat shock, the TFIID complex is bound to the TATA box and another transcription factor known as GAGA is bound upstream (Figure 4A) (Tsukiyama *et* al., 1994). Following heat shock, however, an additional factor is observed which is bound to the HSE (Figure 4B). The amount of this factor bound to the HSE increased with the time of exposure to elevated temperature and with the extent of temperature elevation. Moreover, increased protein binding to the HSE was also observed following exposure to other agents that also induce the transcription of the heat-shock genes such as 2, 4-dinitrophenol (Figure 5). Thus activation of the heat shock genes, mediated by the HSE, is accompanied by the binding of a specific transcription factor to this DNA sequence, and is now generally known as the heat-shock factor (HSF). The genes that encode HSF have been cloned from a variety of species, such as yeast, *Drosophila*, mouse, chicken, human and tomatoes (Wiederrecht et al., 1988; Sorger and Pelham, 1988; Scharf et al., 1990; Schuetz et al., 1991). Despite of functional similarities, HSFs from different species share little primary sequence homology besides their DNA binding and oligomerization domains (Sorger and Nelson, 1989).

Prior to heat shock, the heat-shock genes are poised for transcription. The bulk of cellular DNA is associated with histone proteins to form a tightly packed chromatin structure, the binding of the GAGA factor to the heat-shock gene promoters has resulted in the displacement of the histone-containing nucleosomes from the promoter region (Wilkins and Lis, 1997). This opens up the chromatin and renders the promoter region exquisitely sensitive to digestion with the enzyme DNaseI. Although such a DNaseI hypersensitive site marks a gene as poised for transcription (Latchman, 1998), it is not in itself sufficient for transcription. This role for the GAGA factor in chromatin remodeling is not confined to the heat-shock genes. Mutations in the gene encoding GAGA result in the *Drosophila* mutant trithorax in which a number of homeobox genes (which control the formation of the correct body plan) are not converted from an inactive to an active chromatin state and are hence not transcribed (Schumacher and Magnuson, 1997). This mutation thus produces a fly with an abnormal body pattern and has a similar effect to the *brahma* mutation in the SWI2 component of the SWI/SNF chromatin remodeling complex. Indeed, the GAGA factor has been shown to be associated with a multi-protein complex known as nucleosome remodeling factor (NURF), which, like SWI/SNF, can hydrolyse ATP and alter chromatin structure (Tsukiyama and Wu, 1997).

Hence, following binding of GAGA, the gene is in a state poised for the binding of an activating transcription factor which in turn will result in transcription of the gene. In the case of the heat-shock genes, this is achieved following heat shock by the binding of the HSF to the HSE (Figure 6). This factor then interacts with TFIID and other components of the basal transcription complex resulting in the activation of transcription. The critical role

of the HSF in this process obviously begs the question of how this factor is activated in response to heat.

Activation of HSF can be observed following heat treatment of cell extracts in *vitro* when new protein synthesis would not be possible (Larson et al., 1988), suggesting that this factor must pre-exit in non-shock cells in an inactive form. Indeed, HSFs are ubiquitously expressed in cells. In all eukaryotic species, with the exception of budding yeast, HSFs are kept in a latent form at normal condition and are activated by obtaining high-affinity DNA-binding ability upon stress (Kingston et al., 1987). Analysis of the activation process using in *vitro* systems from human cells (Larson et al., 1988) has indicated that it is a two-stage process. In the first stage, the HSF is activated to a form which can bind to DNA by an ATP-independent mechanism which is directly dependent on elevated temperature. Subsequently, HSF is further modified allowing it to activate transcription. Interestingly, the second of these two stages appears to be disrupted in murine erythro-leukaemia (MEL) cells in which heat-shock results in increased binding of HSF to DNA but transcriptional activation of the heat-shock genes is not observed (Hensold et al., 1990). The activation of HSF into a form capable of binding DNA involves its conversion from a monomeric to a trimeric form which can bind to the HSE (Morimoto, 1993). The maintenance of the monomeric form of HSF prior to heat shock is dependent on a region at the C-terminus of the molecule since when this region is deleted, HSF spontaneously trimerizes and can bind to DNA even in the absence of heat shock (Rabindran et al., 1993). The C-terminal region contains a motif known as the leucine zipper which contains a leucine residue every seven amino acids. As leucine zippers are known to be able to interact with one another, it is thought that this region acts by

interaction with another leucine zipper located adjacent to the N-terminal DNA-binding domain promoting intra-molecular folding which masks the DNA-binding domain. Following heat shock, HSF unfolds, unmasking the DNA-binding domain and allowing a DNA-binding trimer to form.

The two-stage process described above represents a common mechanism for the activation of HSF in higher eukaryotes such as *Drosophila* and mammals. However, in yeast *Saccharomyces cerevisiae*, HSF binds to the DNA all the time in the form of trimer (Wiederrecht et al., 1988; Nieto-Sotelo et al., 1990; Chen et al., 1993). It is only under the stress conditions that the transcriptional activity of HSF gets stimulated. The mechanisms of how *S. cerevisiae* HSF (ScHSF) is activated are still unclear. Extensive deletion and mutagenesis studies have been carried out, which revealed a map of functional domains of ScHSF and suggest a possible de-repression mechanism for the temperature-regulated transcriptional activation. Interestingly, in *Schizosaccharomyces pombe* (fission yeast) HSF regulation follows the *Drosophila* and mammalian system with HSF becoming bound to DNA only following heat-shock (Gallo *et* al., 1991).

Hence in mammals, *Drosophila* and fission yeast, activation of HSF is more complex than in budding yeast, involving an initial stage activating the DNA-binding ability of HSF in response to heat as well as the stage. Initially it was suggested that HSF obtains its DNA-binding ability through phosphorylational modification. This is supported by the evidences that phosphatase treatment of HSF increased its mobility on SDS-PAGE gels (Sorge et al., 1987) and follow-up experiments showing a possible hyper-phosphorylation of yeast HSF under heat shock. But it is still not clear whether phosphorylation is prerequisite for heat shock response or the consequence of the heat

stress, and there is hardly any direct evidence which links the phosphorylation to the activation of HSF. Indeed, some of newer reports seem to support the proposal that phosphorylation may serve the role of deactivation of HSF after heat treatment (Larson et al., 1988). The most appealing evidences so far are the finding of that phosphorylation of serine residues adjacent to heptapeptide CK2 region in K. Lactis results the deactivation of HSF (Hoj and Jackobsen, 1994), and the report of that sequential phosphorylation by MAP kinase and GSK3 (Glycoge Sythase Kinase 3) can repress the activity of human HSF-1 (Chu et al., 1996).

In summary, therefore, heat-shock induces the increased transcription of a small number of cellular genes via the post-translational modification of a pre-existing transcription factor, HSF. This factor binds to a sequence known as the HSE which is located in a region of the heat-shock gene promoters that is free of nucleosomes and contains a DNaseI hypersensitive site even prior to heat shock. The activated form of HSF is capable, following binding, of interacting with components of the constitutive transcriptional apparatus which are bound at other sites in the promoter region and thereby stimulating transcription.

Developmentally regulated transcription and homeobox transcription factors

Homeobox proteins

A very large number of mutations which affect the development of *Drosophila melanogaster* have been isolated and their corresponding genes named on the basis of the observed phenotype of the mutant fly (Ingham, 1988; Lawrence and Morata, 1994).

Thus mutations in the so-called 'gap' genes result in the total absence of particular segments, and mutation in the 'homeotic' genes result in the transformation of one particular segment of the body into another.

The products of genes of these types therefore play critical roles in *Drosophila* development. Given that these processes are likely to require the activation of genes whose protein products are required in the particular segment, it is not surprising that many of these genes have been shown to encode transcription factors. For example, products of two gap genes, *Knirps* and *Kruppel,* contain multiple zinc finger motifs characteristic of DNA-binding transcription factors. Similarly, the *tailless* gene, whose product plays a key role in defining the anterior and posterior regions of the *Drosophila* embryo, has been shown to be a member of the nuclear receptor super gene family. It is clear therefore that the genes identified by mutation as playing a role in *Drosophila* development can encode several different types of transcription factors. However, of the first 25 such genes identified, well over half (15) contain a motif known as the homeobox (Gehring et al., 1994) which was originally identified in the homeotic genes of *Drosophila*. Subsequently, other transcription factors have been identified which contain the homeobox as part of a more complex structure. Two classes of such factors, the POU proteins and the PAX proteins, play a key role in developmentally regulated gene expression.

When the first homeotic genes were cloned, it was found that they shared a region of homology approximately 180 base pairs long, named the homeobox (Gehring et al., 1994). Subsequently, the homeobox was shown to be present in many other *Drosophila* regulatory genes. These include the *Fushi-tarazu* gene (*Ftz*), which is a member of the

pair-rule class of regulatory loci whose mutation causes alternate segments to be absent, and the *engrailed* gene (*eng*), which is a member of the class of genes whose products regulate segment polarity. The close similarity of the homeoboxes encoded by *Ftz* and by homeotic genes *Antennapedia* (*Antp*) *and Ultrabithorax (Ubx)* is shown in Figure 7.

The first indication that the homeobox proteins were indeed transcription factors came from the finding that the homeobox was also present in the yeast mating type a and α gene products. These proteins are known to be transcription factors which regulate the activity of a and α-specific genes. Direct evidence that this is the case is now available from a number of different approaches. It has been shown that many of these proteins bind to DNA in a sequence-specific manner as expected for transcription factors (Hoey and Levine, 1988). Moreover, binding of a specific homeobox protein to the promoter of a particular gene correlates with the genetic evidence that the protein regulates expression of that particular gene. For example, the Ubx protein has been shown to bind to specific DNA sequences within its own promoter and in the promoter of the *Antp* gene, in agreement with the genetic evidence that Ubx represses Ant*p* expression.

The ability of the homeobox-coniaining proteins to bind to DNA is directly mediated by the homeobox itself. If the homeobox of the Antp protein is synthesized in either bacteria or by chemical synthesis, it is capable of binding to DNA in the same sequence-specific manner characteristic of the intact protein. Although DNA binding is a prerequisite for the modulation of transcription, it is necessary to demonstrate that the homeobox proteins do actually affect transcription following such binding. In the case of the Ubx protein, this was achieved by showing that co-transfection of a plasmid expressing Ubx with plasmid in which a reporter gene was controlled by the *Antp*

promoter resulted in the repression of expression of such reporter gene. Most interestingly, the Ubx expression plasmid was able to up regulate activity of its own promoter in co-transfection experiments, this ability being dependent on the previously defined binding sites for Ubx within its own promoter. Similarly, although Ubx normally has no effect on expression of the alcohol dehydrogenase (Adh) gene, it can stimulate the Adh promotor following linkage of the promoter to a DNA sequence containing multiple binding sites for Ubx. Hence a homeobox protein can produce distinct effects following binding, Ubx activating its own promoter and a hybrid promoter containing Ubx binding sites but repressing the activity of the Antp promoter.

A similar transcriptional activation effect *of* DNA binding has been demonstrated for the *Fushi-tarazu* (Ftz) protein. This protein binds specifically to the sequence TCAATTAAATGA. As with Ubx, linkage of this sequence to a reporter gene confers responsibility to activation by Ftz, such activation being dependent upon binding of Ftz to its target sequence, a one-base-pair change which abolishes binding, also abolishing the induction of transcription.

At first sight it is remarkably difficult to understand how the relatively simple process of transcriptional regulation by homeobox proteins could in turn control development. Although this process is of course not yet fully understood, a number of mechanisms have been defined which indicate ways in which the complex regulatory networks needed to regulate development might be built up. It has already been shown in multiple cases that a single factor can repress some target genes whilst activating others, thereby increasing the range of effects mediated by one factor. For example, the Ubx protein can activate transcription from its own promoter whilst repressing that of the *Antp* gene. At

least in neuronal cells, this effect appears to operate by the Ubx protein binding to the Antp promoter and preventing the binding to the same site of the Antp factor itself which would normally activate the promoter. Hence this represents an example of gene repression by interfering with the binding of another activating factor, rather than by direct repression, such as in the case of the eve protein.

Whatever its mechanism the activation and repression of different promoters by the Ubx protein has important consequences in terms of the control of development. Thus the ability of Ubx to induce its own transcription provides a mechanism for the long-term maintenance of *Ubx* gene expression during development since once expression has been switched on and some Ubx protein made, it will induce further transcription of the gene via a simple positive feedback loop even if the factors which originally stimulated its expression are no longer present. This long-term maintenance of *Ubx* expression is essential since, if the *Ubx* gene is mutated within the larval imaginal disc cells which eventually produce the adult fly, the cells which would normally produce the haltere (balancer) will produce a wing instead. Although these cells are known to be committed already to form the adult haltere at the larval stage, the continued expression of the Ubx gene is essential to maintain this commitment and allow eventual overt differentiation (Hadorn, 1968).

Similarly, the inhibition of the *Antp* gene by *Ubx* indicates that the homeobox factors do not act simply by activating the transcription of genes for structural proteins or enzymes required in particular cell types. Although this must be one of their functions, they can also regulate the transcription of each other, creating the potential for regulatory networks. Since *Antp* has been shown to stimulate both its own promoter and that of

Ubx, this creates the possibility of complex interactions in which the synthesis of one particular factor at a particular time will create changes in the levels of numerous other factors and ultimately result in the activation or repression of numerous target genes. Indeed, the homeobox proteins evidently also regulate the expression of genes whose protein products are required in a particular cell type such as the cell-surface adhesion molecules (Edelman and Jones, 1993; Gruba et al., 1997). Hence by both regulating each other's expression and that of non-homeobox target genes both positively and negatively, the homeobox transcription factors can create regulatory networks of the type which are necessary for the control of development.

In addition, however, it is also possible for the same target gene to be regulated by multiple homeobox factors, with the effects observed with one factor being different depending on whether or not another factor is also present. In one such mechanism the DNA-binding specificity of one factor is altered in the presence of another factor. Thus several homeobox proteins such as *Ubx* and *Antp* bind to the same DNA sequences when tested *in vitro* (Hoey and Levine, 1988) yet, paradoxically, the effects of mutations which inactivate the genes encoding each of these proteins are different, indicating that they cannot substitute for one another. Similarly, *in vivo, Ubx* binds to a site in the promoter of the *decapentaplegic (dpp)* gene and activate its expression, whereas *Antp* does not.

The paradox is explained by the presence in the *dpp* promoter of a binding site for another homeobox protein extradenticle (exd), which lies adjacent to the site where Ubx binds. The exd protein interacts with the Ubx protein and modifies its DNA-binding specificity so it can bind to the dpp gen promoter and activate transcription (Mann and

Chan, 1996). This interaction with exd is dependent on the homeobox and C-terminal region of Ubx which have several amino differences from the corresponding region of Antp. Because of these differences, Antp does not interact with exd and hence cannot bind to the site in the *dpp* promoter. This suggests that the modification of the DNA-binding specificity of a homeobox factor by interaction with other homeobox factors represents an evolutionarily conserved strategy for targeting these factors to different genes in different situations. Interestingly, a similar interaction promoting DNA binding has recently been described between the homeobox protein fushi-tarazu (Ftz) and a member of the nuclear receptor transcription factor family, Ftz-F1, indicating that such interactions of homeobox proteins can also occur with members of other transcription factor families (Guichet et al., 1997; Yu et al., 1997).

As well as interactions which alter DNA binding, other interactions operate at a functional level, with combinations of factors either synergizing with each other or interfering with one another so that functional effects are observed with combinations of factors which are not observable with either factor alone. For example, it has been shown that the homeobox proteins engrailed (eng), Ftz, paired (prd) and zerknult (zen) can all bind to the sequence TCAATTAAAT (Hoey and Levine, 1988). When plasmids expressing each of these genes are co-transfected with a target promoter carrying multiple copies of this binding site, the Ftz, prd and zen proteins can activate transcription of the target promoter (Jaynes and O'Farrell, 1988; Han et al., 1989). In contrast, the eng protein has no effect on the transcription of such a promoter. It does, however, interfere with the ability of the activating proteins to induce transcription presumably by blocking the binding of the activating factor. Thus, for example, whilst Ftz can stimulate the target

promoter when co-transfected with it, it cannot do so in the presence of eng (Jaynes and O'Farrell, 1988). Hence the expression of Ftz, alone in a cell, would activate particular genes, whereas its expression in cell also expressing engrailed would not have any effect (Figure 8).

Similar types of interaction can also take place between different positively acting factors binding to the same site. Although the Ftz, prd and zen products can all activate transcription of a target promoter when transfected alone, this effect is relatively small, producing only approximately two-fold activation. In contrast, much larger effects can be obtained by activating the target promoter with two of these proteins in combination, producing 10-20-fold activation, or by all three activators together producing 400-fold activation of the target (Han et al., 1989).

Hence the complex effects of single homeobox factors on the expression of other genes can be rendered still more complex by means of synergistic or inhibitory effects of combinations of factors creating effects which would not be obtained with a single factor alone. Indeed, such interactions of different factors can be used to generate models which predict complex spatial distributions of responder gene activity in response to relatively simple expression patterns of homeobox protein distribution. One such model (Jaynes and O'Farrell, 1988) is based on the interaction of activator and repressor molecules which bind to the same binding site in the manner of the Ftz and eng products. By assuming that target genes vary in the affinity of their binding sites for an activator and two repressor molecules whose areas of expression are overlapping but not identical, it is possible to generate different patterns of responder gene activity in each cell type depending on which particular factors are present.

Hence the activation and repression of target genes by different homeobox factors both alone and in combination can generate complex overlapping patterns of target gene expression of the sort which must occur in development. In fact, however, a further level of complexity exists since many homeobox factors are not expressed in a simple on-off manner but rather show a concentration gradient ranging from high levels in one part of the embryo via intermediate levels to low levels in another part. For example, the bicoid (bcd) protein, whose absence leads to the development of a fly without head and thoracic structures, is found at high levels in the anterior part of the embryo and declines progressively toward posterior part, being absent in the posterior one-third of the embryo.

Most interestingly, genes which are activated in response to *bcd* contain binding sites in their promoters which have either high affinity or low affinity for the bcd protein. If these sites are linked to a reporter gene, it can be demonstrated that genes with low-affinity binding sites are only activated at high concentrations of bcd protein and are therefore expressed only at the extreme anterior end of the embryo. In contrast, genes which have higher affinity binding sites are active at much lower protein concentrations and will be active both end, and expressed more posteriorly. Moreover, the greater the number of higher affinity sites, the greater the level of gene expression which will occur at any particular point in the gradient.

The gradient in bicoid expression can be translated therefore into the differential expression of various bicoid-dependent genes along the anterior part of the embryo. Each cell in the anterior region will be able to 'sense' its position within the embryo and respond by activating specific genes. One of the genes activated by bicoid is the homeobox-containing segmentation gene *hunchback.* In turn, this protein regulates the

expression of the gap genes *Kruppel* and *giant* (Struhl et al., 1992). All four of these proteins then act on the *eve* gene, with *bicoid* and *hunchback* activating its expression while *kruppel* and *giant* repress it. Those concentration gradients result in the spatial localization of *eve* gene expression in a defined region of the embryo where it exerts its inhibitory effects on gene expression (Small *et* al., 1991) (Figure 9).

The *bicoid* protein therefore has all the properties of a morphogen whose concentration gradient determines position in the anterior part of the embryo. This idea is strongly supported by the results of genetic experiments in which the *bicoid* gradient was artificially manipulated, cells containing artificially increased levels of *bicoid* assuming a phenotype characteristic of more anterior cells and vice versa.

The anterior to posterior gradient in *bicoid* levels is required to produce the opposite posterior to anterior gradient in the level of another protein, caudal. However, the caudal mRNA is equally distributed throughout the embryo, indicating that the *bicoid* gradient does not regulate transcription of the *caudal* gene. Rather, the *bicoid* protein binds to the *caudal* mRNA and represses its translation into protein so that *caudal* protein is not produced when *bicoid* levels are high (Carr, 1996; Chan and Struhl, 1997). As well as providing further evidence for the key role of the *bicoid* protein, this finding also shows that homeodomain proteins can bind to RNA as well as to DNA and that they may therefore act at the post-transcriptional level as well as at transcription.

POU proteins

Another set of transcription factors which possess a homeobox as part of a much larger motif and which were first identified in mammalian cells, the octamer-binding transcription factors, play an important role in regulating the expression of specific genes such as those encoding histone H2B, the SnRNA molecules and the immunoglobulins. Similarly, the transcription factor, Pit-1, which binds to a sequence two bases different from the octamer sequence, plays a critical role in pituitary-specific gene expression.

When the genes encoding these factors were cloned, they were found to share a 150-160-amino-acid sequence which was also found in the protein encoded by the nematode gene, unc-86, whose mutation affects sensory neuron development. This common POU (Pit-Oct-Unc) domain contains both a homeobox sequence and a second conserved domain, the POU-specific domain (Prakash et al., 1992; Verrijzer and Van der Vliet, 1993; Ryan and Rosenfeld, 1997).

Interestingly, while the homeoboxes of the different POU proteins are closely related to one another (53 out of 60 homeobox residues are the same in Oct-1 and Oct-2, and 34 out of 60 in Oct-1 and Pit-1), they show less similarity to the homeoboxes of other mammalian genes lacking the POU-specific domain, sharing at best only 21 out of 60 homeobox residues. Hence they represent a distinct class of homeobox proteins containing both a POU-specific domain and a diverged homeodomain. As with the *Drosophila* homeobox proteins, however, the isolated homeodomains of the Pit-1 and Oct-1 proteins are capable of mediating sequence-specific DNA binding in the absence of the POU-specific domain. The affinity and specificity of binding by such an isolated homeodomain is much lower, however, than that exhibited by the intact POU domain, indicating that the POU-specific domain plays a critical role in producing high-affinity

binding to specific DNA sequences. Hence the POU homeodomain and the POU-specific domain form two parts of a DNA-binding element which are held together by a flexible linker sequence.

In addition to its role in DNA binding, the POU domain also plays a critical role in several other features of the POU proteins which are not found in the simple homeobox-containing proteins. For example, the ability of both Oct-1 and Oct-2 to stimulate DNA replication as well as transcription is also a property of the isolated POU domains of these factors. Similarly, the ability of Oct-1 and not Oct-2 to interact with the herpes simplex virus trans-activator protein VP16 is controlled by a single difference in the homeodomain region of the POU domains in the two proteins. Thus the replacement of a single amino-acid residue at position 22 in the homeodomain of Oct-2 with the equivalent amino acid of Oct-1 allows Oct-2 to interact with V-PI6, which is normally a property only of Oct-1.

Interestingly, the key role of position 22 in the homeodomain is not confined to the interaction of Oct-l/Oct-2 with VP16. Thus, the closely related mammalian POU factors Brn-3a and Brn-3b differ in that Brn-3a activates the promoter of several genes expressed in neuronal cells, whereas Brn-3b represses them. Alteration of the isoleucine residue found at position 22 in Brn-3b to the valine found in Brn-3a converts Brn-3b from a repressor into an activator, whereas the reciprocal mutation in Brn-3a converts it into a repressor (Dawson et al., 1996). This effect suggests that the activating/repressing effects of Brn-3a/Brn-3b are mediated by their binding of cellular co-activator or co-repressor molecules whose binding to Brn-3a/Brn-3b is affected by the nature of the amino acid at position 22. More generally, this finding provides the first example of a

single aminoacid change which can reverse the functional activity of a transcription factor, from activator to repressor and vice versa.

As in the case of the homeobox-containing proteins, the POU proteins appear to play a critical role in the regulation of developmental gene expression and in the development of specific cell types. The unc-86 mutation in the nematode results, for example, in the lack of touch receptor neurons or male-specific cephalic companion neurons, indicating that this POU protein is required for the development of these specific neuronal cell types. Similarly, inactivation of the gene encoding Pit-1 leads to a failure of pituitary gland development resulting in dwarfism in both mice and humans (Andersen and Rosenfeld, 1994). Interestingly, however, one type of dwarfism in mice (the Ames dwarf) is produced not by a mutation in Pit-1 but by a mutation in a gene encoding a homeobox-containing factor which was named Prophet of Pit-1. This factor appears to control the activation of the Pit-1 gene in pituitary cells so that Pit-1 is not expressed when this factor is inactivated. This example illustrates how hierarchies of regulatory transcription factors are required in order to control the highly complex process of development.

Following the initial identification of the original four POU factors, a number of other members of this family have been described both in mammals and other organisms such as *Drosophila, Xenopus* and zebra fish. Like the original factors, these novel POU proteins also play a critical role in the regulation of developmental gene expression. Thus, for example, the *Drosophila* POU protein drifter (CFla) has been shown to be of vital importance in the development of the nervous system (Anderson et al., 1995), whilst mutations in the gene encoding the Brn-4 factor appear to be the cause of the most

common form of deafness in humans (de Kok et al., 1995). Moreover, all the novel POU domain-containing genes isolated by He *et* al. (1989) from the rat, on the basis of their containing a POU domain, are expressed in the embryonic and adult brain suggesting a similar role for these proteins in the regulation of neuronal-specific gene expression. Such a close connection of POU proteins and the central nervous system is also supported by studies using the original POU domain genes, which revealed expression in the embryonic brain even in the case of Oct-2 which had previously been thought to be expressed only in B cells (He et al., 1989).

It is clear therefore that, like the homeobox proteins, POU proteins occur in a wide variety of organisms and play an important role in development. Moreover, these proteins may be of particular importance in the development of the central nervous system.

References

Andersen, B. and Rosenfeld, M.G. (1994). Pit-1 determines cell types during development of the interior pituitary gland. *Journal of Biological Chemistry* **269,** 335-338.

Anderson, M.G., Perkins, G.L., Chittick, R., Shrigley, R.J. and Johnson, W.A. (1995). Drifter, a *Drosophila* POU-domain transcription factor, is required for correct differentiation and migration of tracheal cells and midline glia. *Genes and Development* **9,** 123-127.

Bienz, M. and Pelham, H.R.B. (1986). Heat-shock regulatory elements function as an inducible enhancer when linked to a heterologous promoter. Cell **45,** 753-760.

Breathnach, R. and Chambon, P. (1981). Organization and expression of eukaryotic split genes coding for proteins. Annual *Review of Biochemistry* **50,** 349-383.

Carey, M. (1998). Enhanceosome and transcriptional synergy. Cell **92,** 5-8.

Carr, K. (1996). RNA bound to silence. *Nature* **379,** 676.

Chan, S.K. and Struhl, G. (1997). Sequence specific RNA binding by bicoid. *Nature* **388.** 634.

Chen, Y.Q., Barlev, N.A., Westergarrd O. and Jakobsen, B.K. (1993). Identification of the C-terminal activator domain in yeast heat-shock factor- independent control of transient and sustained transcriptional activity. *EMBO J.* **12,** 5007-5018.

Chu, B.Y, Soncin, F., Price, B.D., Stevenson, M.A. and Calderwood, S.K. (1996). Sequential phosphorylation by mitogen-activated protein kinase and glycogen synthase kinase 3 represses transcriptional activation by heat shock factor-1. *J. Bio. Chem.* **271,** 30847-30857.

Craig E.A. (1985). Change in eukaryotic gene-expression in response to environmental-stress. *Science* **230**, 800-801,

Davidson, E.H., Jacobs, H.T. and Britten, R.J. (1983). Very short repeats and coordinate induction of genes. *Nature* **301,** 468-470.

Darnell, J.E. (1982). Variety in the level of gene control in eukaryotic *cells. Nature* **297,** 365-371.

Dawson, S.J., Morris, P.J. and Latchman, D.S. (1996). A single amino acid change converts a repressor into an activator. *Journal of Biological Chemistry* **271,** 11631-11633.

de Kok, Y.J.M., Van der Maarel, S.M., Bitner-Glindzicz, M., Huber, I., Monaco, A.P., Malcolm, S., Pembrey, M.E., Ropers, H.H. and Cremers, F.P.M. (1995). Association between X-linked mixed deafness and mutations in the POU domain gene POU3F4. *Science* **267,** 685-688.

Dynan, WS. and Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature* **316,** 774-778.

Edelman, G.M. and Jones, F.S. (1993). Outside and downstream of the bomeobox. *Journal of Biological Chemistry* **268,** 20683-20686.

Elshotz, H.P., Albert, VR., Treacy, M.N. and Rosenfeld, M.G. (1990). A two-base change in a POU factor binding site switches pituitary-specific to lymphoid-specific gene expression. *Genes and Development* **4,** 43-51.

Gallo, G.J., Schuetz, T.J. and Kingston, R.E. (1991). Regulation of heat shock factor in *Schizosaccharomyces pombe* more closely resembles regulation in mammals than in Saccharomyces cerevisiae. Molecular *and Cellular Biology* 11, 281-288.

Garcia, J.V., Bich-Thuy, L., Stafford, J. and Queen, C. (1986). Synergism between immunoglobulin enhancers and promoters. *Nature* **322,** 383-385.

Gehring, W.J., Affolter, M. and Burglin, T. (1994). Homeodomain proteins. *Annual Review of Biochemistry* **63,** 487-526.

Goodwin, G.H., Partington, G.A. and Perkins, N.D. (1990). Sequence specific DNA binding proteins involved in gene transcription. In: *Chromosomes: Eukaryotic,*

Prokaryotic and Viral (Adolph, K.W, ed.), Vol. 1, pp. 31-85. Boca Raton, Florida: CRC Press.

Gruba, Y., Aragnol, D. and Pradel, J. (1997). *Drosophila* Hox complex targets and the function of homeotic genes. *BioEssays* **19,** 379-388.

Guichet, A., Copeland, J.W.R., Erdelyi, M., Hlousek, D., Zavaroszky, P., Ho, J., Brown, S., Percival-Smith, A., Krause, H.M. and Ephrussi, A. (1997). The nuclear receptor homologue Ftz-F1 and the homeodomain protein Ftz are mutually dependent co-factors. *Nature* **385,** 348-552.

Hadorn, E. (1968). Tansdetermination in cells. *Scientific American* **219**, 110-120.

Han, K., Levine, M.S. and Manley, J.L. (1989). Synergistic activation and repression of transcription by *Drosophila* homeobox proteins. Cell ***56,*** 573-583.

Hatzopoulous, A.K., Schlokat, U. and Gruss, P. (1988). Enhancers and other cis-acting sequences. In: *Conscription and Splicing* (Hames, B.D. and Glover, D.M., eds), pp. 43-96. Oxford: IRL Press.

He, X., Treacy, M.N., Simmons, D.M., Ingraham, H.A., Swanson, L.S. and Rosenfeld, M.G. (1989). Expression of a large family of POU-domain regulatory genes in mammalian brain development. *Nature* **340**, 35-42.

Hensold, J.0., Hunt, C.R., Calderwood, S.K., Housman, D.E. and Kingston, R.E. (1990). DNA-binding of heat shock factor to the heat shock element is insufficient for transcriptional activation in murine erythroleukemia cells. *Molecular and Cellular Biology 10,* 1600-1608.

Hoey, T. and Levine, M. (1988). Divergent homeobox proteins recognize similar DNA sequences in *Drosophila. Nature* **332,** 858-861.

Hoj, A. and Jackobsen, B.K. (1994). A short element required for turning off heat-shock transcription factor – evidence that phosphorylation enhances deactivation. *EMBO J.* **13**, 2617-2624.

Ingham, P.W. (1988). The molecular genetics of embryonic pattern formation in *Drosophila. Nalure* **335**, 25-34.

Jaynes, J.B. and O'Farrell, P.H. (1988). Activation and repression of transcription by homeodomain-containing proteins that bind a common site. *Nature* **336**, 744-749.

Larson, J.S., Schuetz, T.J. and Kingston, R.E. (1988). Activation in vitro of sequence specific DNA binding by a human regulatory factor. *Nature* **335**, 272-275.

Latchman, D.S. (1998). *Gene Regulation: A Eukaryotic Perspective, 3rd* ed. London: Chapman and Hall.

Lawrence, P.A. and Morata, G. (1994). Homeobox genes: their function in *Drosophila* segmentation and pattern formation. *Cell* **78,** 181-189.

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* **325,** 369-372.

Lindquist S. and Craig, E.A. (1988) The heat shock proteins. *Annu. Rev. Gene.* **22**, 631-677.

Mann, R.S. and Chan, S.K. (1996). Extra specificity from extradenticle: the partnership between Hox and PBX/EXD homeodomain proteins. *Rends in Genetics* **12,** 258-262.

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

Morimoto, R. (1993) Cells in stress: transcriptional activation of heat shock genes. *Science* **259,** 1409-1410.

Morimoto, R.I. (1994) The Bio. Of. Heat Shock Gene Trans. By a Family of Heat Shock Factors, Cold Spring Harbor Press.

Muller, M.M., Gerster, T. and Schaffner, W. (1988) Enhancer sequences and the regulation of gene transcription. *European Journal of Biochemistry* **176,** 485-495.

Muller-Hill, B.W. (ed.) (1996). The lac operon: a short history of a genetic paradigm. Berlin: de Grayler Co.

Nevins, J.R. (1983). The pathway of eukaryotic mRNA transcription. *Annual Review of Biochemistry* **52,** 441-446.

Nieto-Sotole, J., Wiederrecht, G., Okuda, A. and Parker C.S. (1990). The yeast heat shock transcriptional factor contains a transcriptional activation domain whose activity is repressed under nonshock conditions. *Cell* **62**, 807-817.

Pelham, H.R.B. (1982). A regulatory upstream promoter element in the Drosophila hsp70 heat-shock gene. *Cell* **30,** 517-528.

Prakash, K., Fang, X., Engelberg, D., Behal, A. and Parker, C.S. (1992). dOct2, a *Drosophila* Oct transcription factor that functions in yeast. *Proc. Natl. Acad. Sci.* **89**, 7080-7084.

Rabindrin, S.K., Haroun, R.I., Clog, J., Wisnieski, J. and Wu, C. (1993) Regulation of heat shock factor trimer formation: role of a conserved Leucine zipper. *Science* **259,** 230-234.

Ryan, A.K. and Rosenfeld, M.G. (1997). POU domain family values: flexibility, partnerships and developmental codes. *Genes and Development* **11**, 1207-1225.

Ross, J. (1996). Control of messenger RNA stability in high eukaryotes. *Cell* **74,** 413-421.

Schumacher, A. and Magnuson, T. (1997). Murine polycomb and trithorax group genes regulate homeotic pathways and beyond. *Trends in Genetics* **13,** 167-170.

Sen, R. and Baltimore, D. (1986). Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell **46,** 705-716.

Scharf, K.D., Rose, S., Zott, W., Schoff, F. and Nover, L. (1990). Three tomato genes code for heat stress transcription factors with a region of remarkable homology in the DNA-binding of the yeast HSF. *EMBO J.* **9**, 4495-4501.

Schuetz, T.J. and Larin, Z. (1987). Heat inducible human factor that binds to a human hsp70 promoter. *Mol. Cell. Biol.* **7**, 1540-1534.

Schuetz, T.J., Gallo, G.J., Scheldon, L., Tempst, P. and Kingston, R.E. (1991). Isolation of a cDNA factor HSF2: evidence for the heat shock factor genes in humans. *Proc. Natl. Acad. Sci.* **88**, 6911-6915.

Sentenac, A. (1985) Eukaryotic RNA polymerases. *CRC Critical Reviews in Biochemistry 1,* 31-90.

Small, S., Krant, R., Hoey, T., Warrior, R. and Levine, M. (1991). Transcriptional regulation of a pair-rule stripe in *Drosophila. Genes and Development* **5**. 827-839.

Sorger, P.K., Lewis M.J. and Pelham, H.R.B. (1987). Heat-shock factor is regulated differently in yeast and Hela-cells. *Nature* **329**, 81-84, 1987;

Sorger, P.K. and Nelson, H.C.M. (1989). Trimerization of a yeast transcriptional activator via a coiled-coil motif. *Cell* **59**, 807-813.

Sorger, P.K. and Pelham H.R.B (1988). Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phorsphorylation. *Cell* **54**, 855-864.

Struhl, G., Johnston, P. and Lawrence, P.A. (1992). Control of Drosophila body pattern by the hunchback morphogen gradient. *Cell* **69**, 237-249.

Thiele, D.J. (1992). Metal regulated transcription in eukaryotes. Nucleic *Acids Research* **20,** 1183-1191.

Travers, A. (1993) *DNA-Protein Interactions.* London: Chapman and Hall.

Tsukiyama, T., Becker, R.B. and Wu, C. (1994). ATP-dependant nucleosome disruption at a heat shock promotor mediated by DNA-binding of GAGA transcription factor. *Nature* **367,** 525-532.

Tsukiyama, T. and Wu, C. (1997). Chromatin remodelling and transcription. *Current Opinion in Genetics and Development* **7,** 182-191.

Turner, B.M. (1993). Decoding the nucleosome. Cell **75,** 5-8.

Verrijzer, C.R. and Van der Vliet, P.C. (1993). POU domain transcription factors. *Biochimica et Biophysica Acta* **1173,** 1-21.

Wang, J. and Manley, J.L. (1997). Regulation of pre-mRNA splicing in metazoa. *Current Opinion in Genetics and Development* **7,** 205-211.

Weis, L. and Reinberg, D. (1992) Transcription by RNA polymerase 11 initiator directed formation of transcription competent complexes. *FASEB Journal* **6,** 3300-3309.

Werner, M.H. and Burley, S.K. (1997). Architectural transcription factors: proteins that remodel DNA. *Cell* **88,** 733-736.

Wiederrecht, G., Seto, D. and Parker, C.S. (1988). Isolation of the gene encoding the *S. Cerevisiae* heat shock transcription factor. *Cell* **54**, 842-853.

Wilkins, R.C. and Lis, J.T. (1997). Dynamics of potentiation and activation. GAGA factor and its role in heat shock gene regulation. *Nucleic Acids Research* **25,** 3963-3968.

Williams, G.T., McClanahan, T.K. and Morimoto, R.I. (1989). Ela transactivation of the human hsp70 promoter is mediated through the basal transcriptional complex. Molecular *and Cellular Biology* **9,** 2574-2587.

Wirth, T., Staudt, L. and Baltimore, D. (1987). An octamer oligonucleotide upstream of a TATA motif is sufficient for lymphoid specific promoter activity. Nature **329,** 174-178.

Wolffe, A. (1995). *Chromatin: Structure and* Function, 2nd edn. London, San Diego. Academic Press.

Yu, Y., Li, W., Su, K., Yussa, M., Han, W., Perriman, N. and Pick, L. (1997). The nuclear hormone receptor Ftz-F1 is a co-factor for the *Drosophila* homeodomain protein Ftz. *Nature* **385,** 552-555.

Figures and figure legends

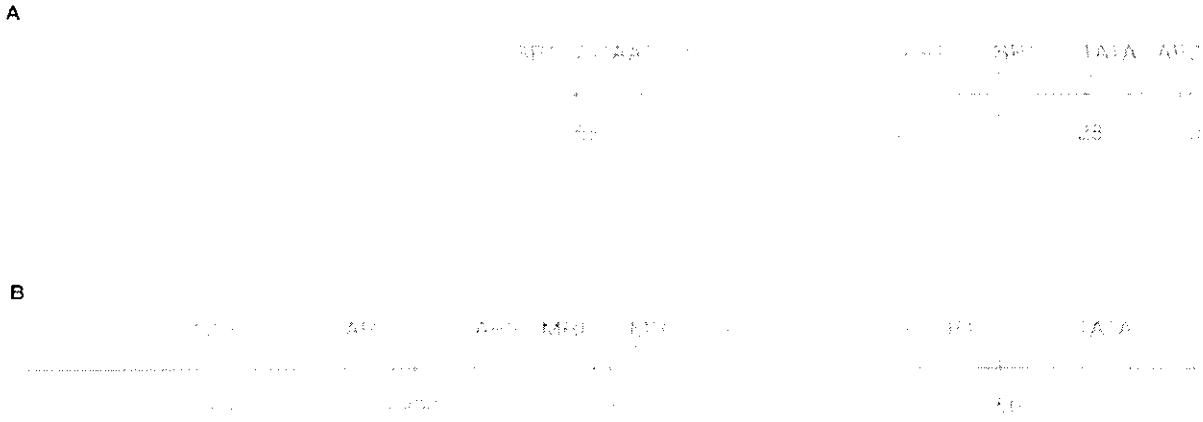

Figure 1. Transcriptional control elements upstream of the transcriptional start site.

A) Human *hsp70* gene

B) Human methallothionein IIA gene

The TATA,Sp1 and CCAAT boxes bind factors involved in constitutive transcription, while GRE, HSE, MRE, AP1 And AP2 sites bind factors involved in the inducible transcription in response to specific stimuli.

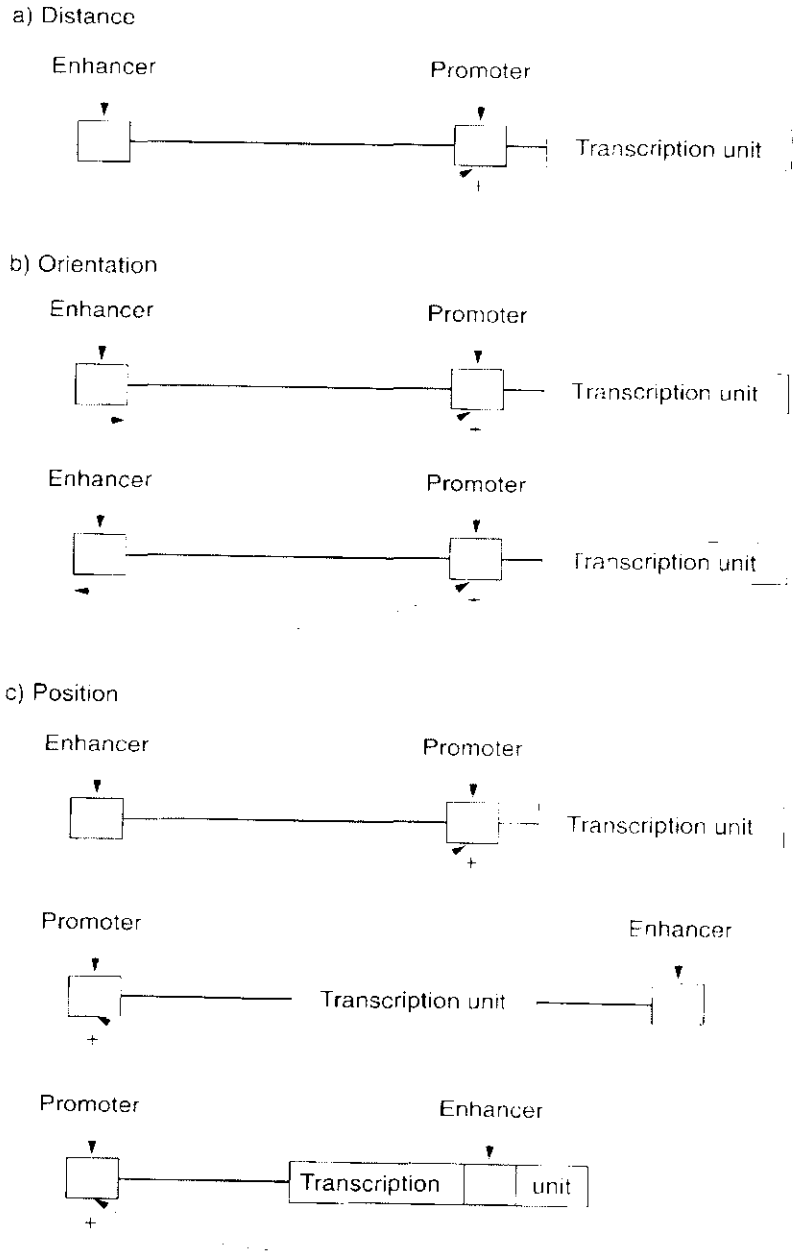

Figure 2. Characteristics of an enhancer element which activates a promoter at a distance (a); in either orientation (b); and at different position (c).

Figure 3. Structure of a typical transcriptional regulatory region. It contains a TATA-box promoter, upstream promoter elements such as CCAAT and Sp1 boxes, regulatory elements such as CRE and GRE, and other elements within more distant enhancer.

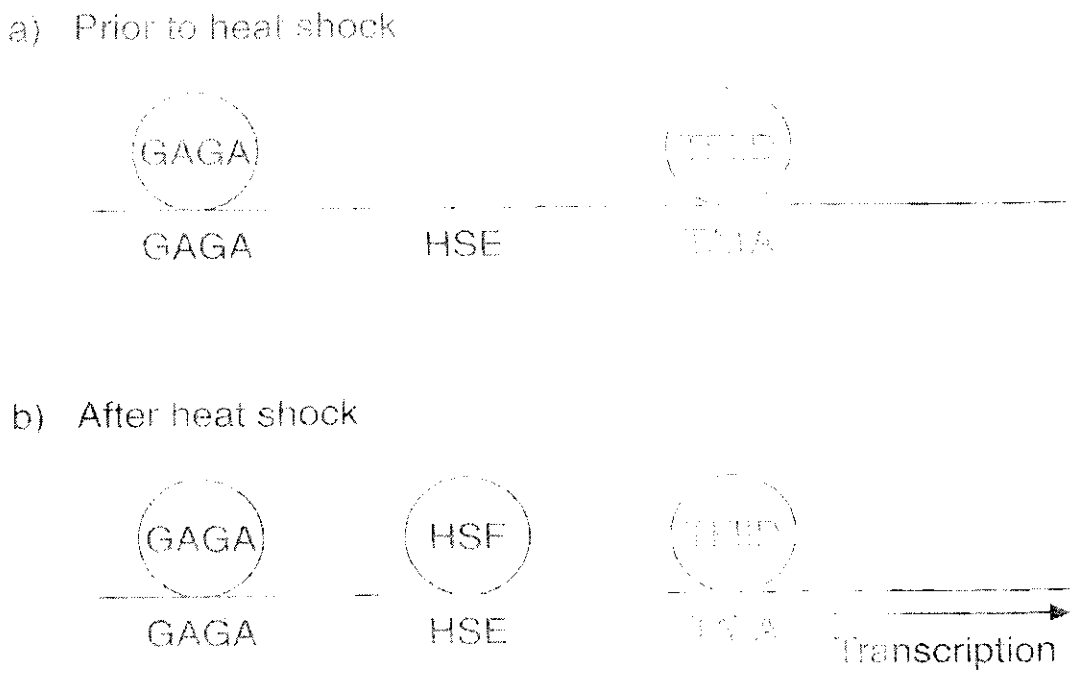

Figure 4. Proteins that bind to *hsp70* gene promoter.

a) before and b) after heat shock.

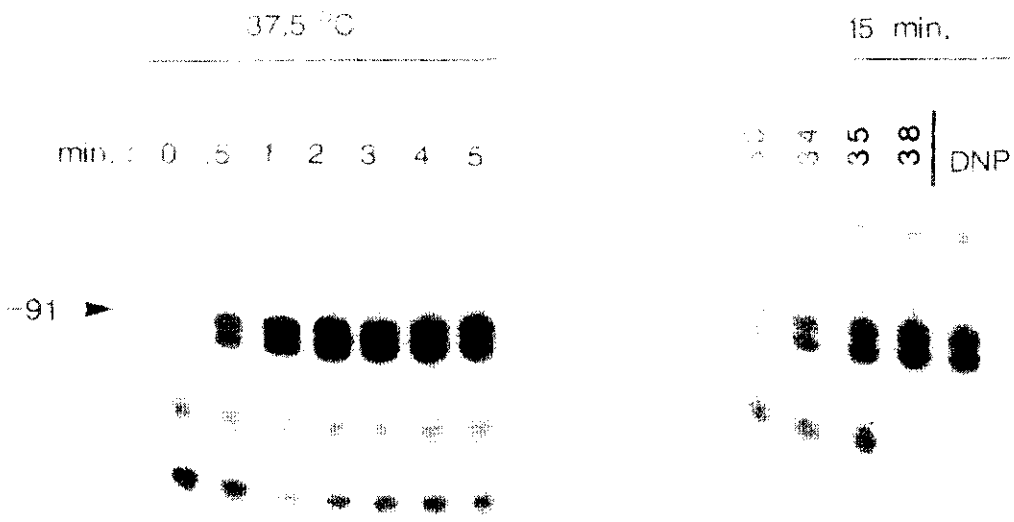

Figure 5. Binding of HSF to the HSE of *Drosophila hsp82* gene. HSF binds at the site 91 bases upstream of transcription starting point and protects this region from exonuclease III digestion. HSF binding affinity increases with increasing time of heat shock (left panel) or increasing severity of heat shock (right panel). HSF binding is also induced by exposure to 2,4-dinitrophenol (DNP).

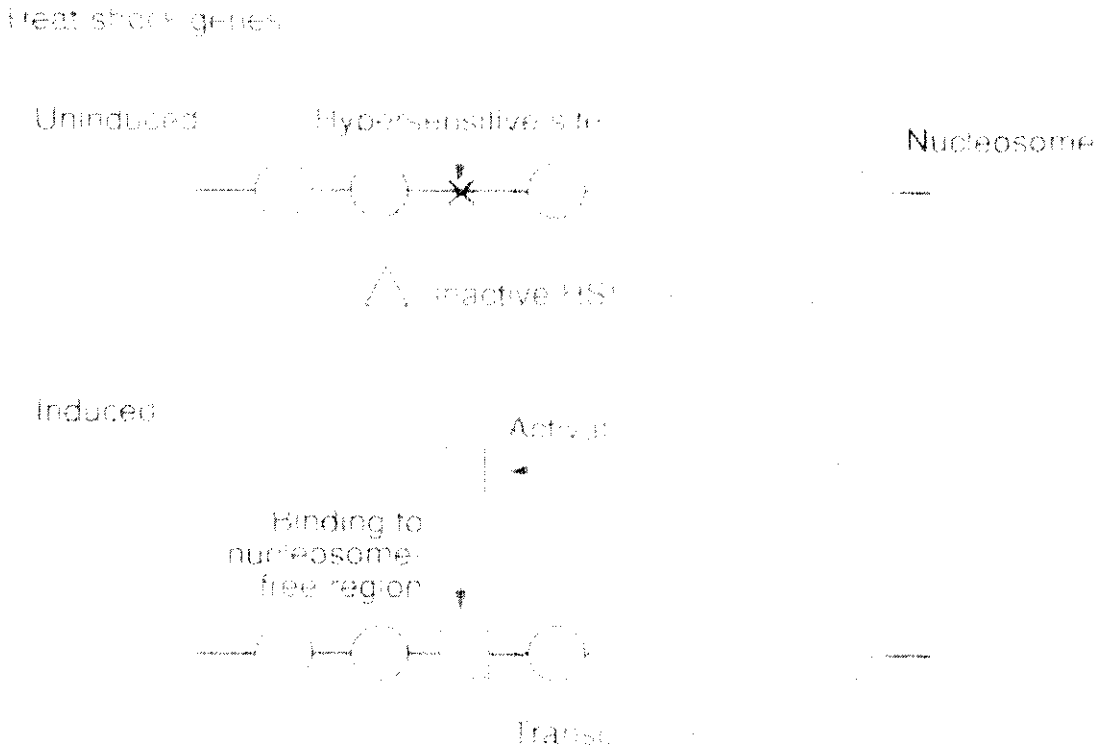

Figure 6. Activation of HSF by heat is followed by its binding to a pre-existing nucleosome free region in the heat-shock gene promoters. This site is a DnaseI hypersensitive site and is produced by the prior binding of the GAGA factor.

Figure 7. Conserved helical motifs of several *Drosophila* homeodomain proteins. Differences between the amino-acid sequences of Ubx and Ftz homeodomains from that of Antp are indicated. Blanks denote the identities in the sequences. The helix-turn-helix region also indicated.

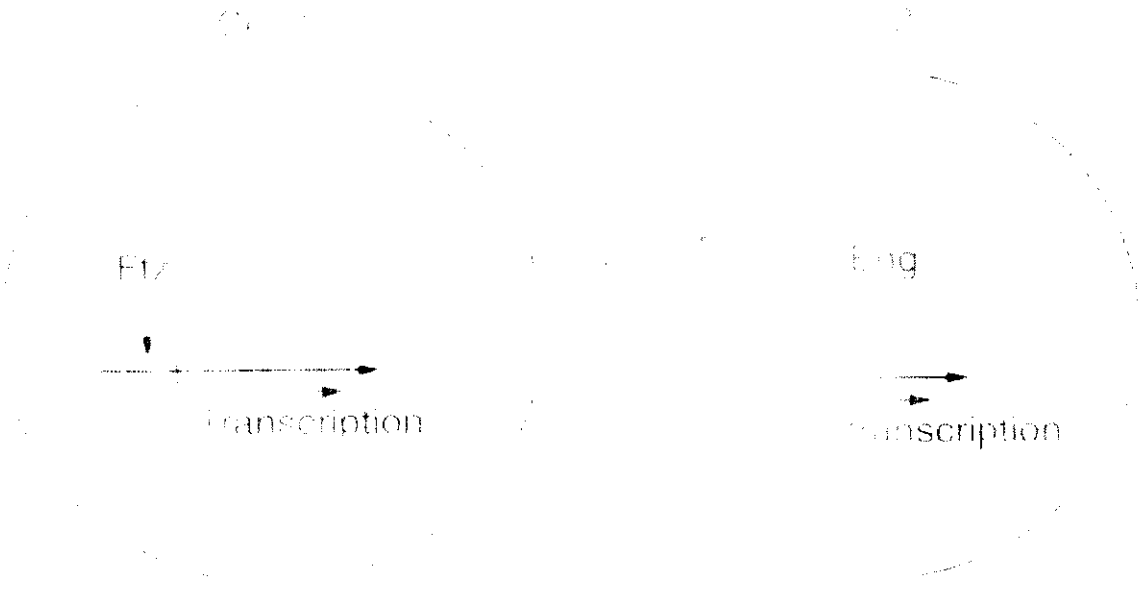

Figure 8. Trancriptional induction by Ftz is blocked in the cells expressing Eng protein, which binds to the same sequence as Ftz.

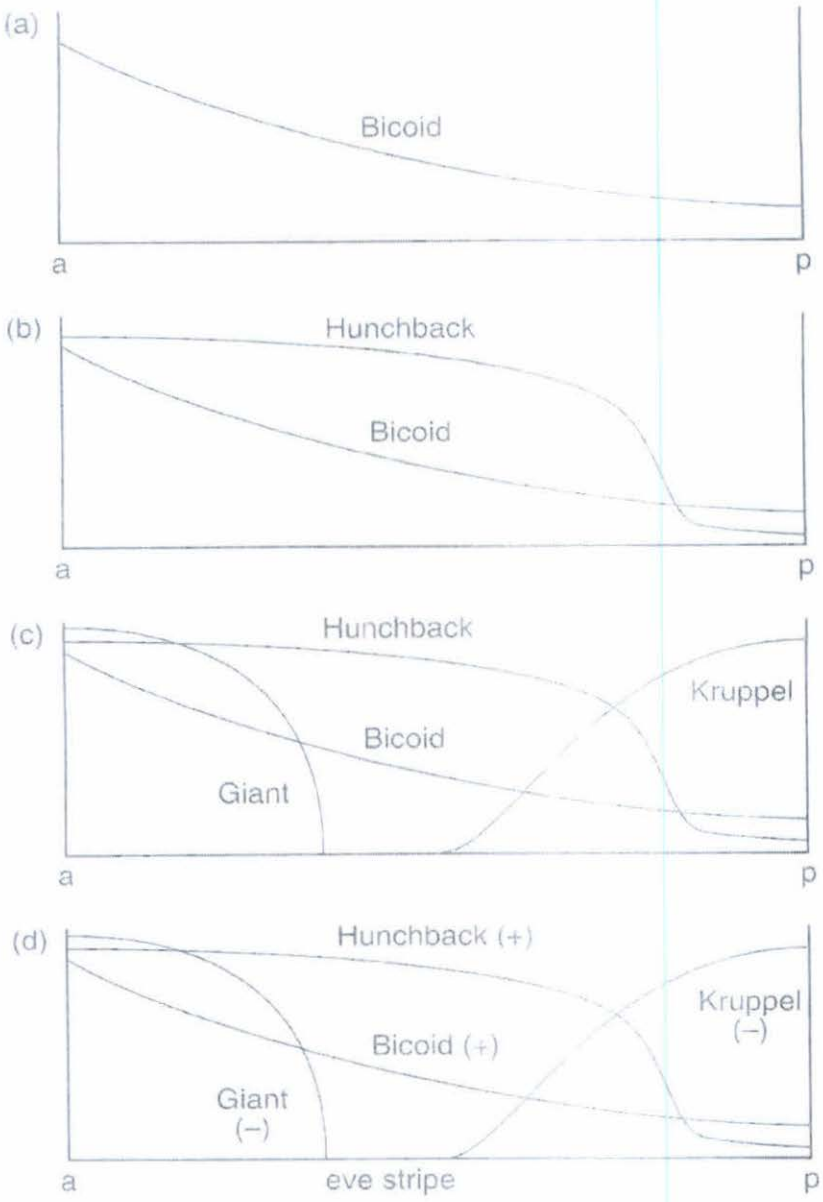

Figure 9. Model illustrating how the concentration gradients of activators bicoid and hunchback, and repressors kruppel and giant, produce a stripe of eve gene expression. The bicoid gradient affects hunchback expression (a and b), which in turn affects giant and kruppel expression (c). Eve gene expression is activated (+) by hunchback and bicoid, and repressed (-) by giant and kruppel (d).

CHAPTER TWO

DEVELOPMENTAL REGULATION OF THE HEAT SHOCK RESPONSE BY NUCLEAR TRANSPORT FACTOR KARYOPHERIN–α3

The work in this chapter was done in collaboration with Tianxin Chen and Kim Tran

Introduction

Temporal and spatial regulation of specific transcription factor activity is a key process in embryonic development. Activation of specific transcription factor activity in response to extracellular stimuli is also an important process both during and post-embryonic development. Some of the regulatory mechanisms include restricted transcription factor expression by promoter regulatory elements during development (Arnone and Davidson, 1997), activation of a dormant pre-existing molecule into an active form by protein modification (Karin et al., 1997; Schindler and Darnell, 1995), and restricted nuclear entry (Kaffman and O'Shea, 1999). In mammals the well-characterized NF-κB transcription factor is released from an inhibitory molecule IκB bound to the NLS of NF-κB by specific phosphorylation and degradation of the inhibitor (Woronicz et al., 1997; Zandi et al., 1998; Zandi et al., 1997a). The released NF-κB molecule can now enter the nucleus and activate specific gene transcription. Conditional nuclear import is also used by the NF-AT transcription factors. NF-ATs are cytoplasmic in unstimulated T cells but upon engagement of the T cell receptor and CD28 coreceptor they rapidly translocate into the nucleus and activate transcription of an array of cytokine genes (Zhu et al., 1998). NF-AT nuclear import is induced by the calcium-dependent phosphatase calcineurin (Jain et al., 1992). In Drosophila regulated nuclear entry of the NF-κB/Rel-related protein Dorsal plays a key role in the establishment of the dorsal-ventral axis during early embryogenesis (Morisato and Anderson, 1995; Steward, 1987). Dorsal is

retained in the cytosol by Cactus, which shares significant homology with IκB family members (Geisler et al., 1992). Cactus is phosphorylated in response to a signaling cascade that leads to its degradation allowing Dorsal to enter the nucleus (Kidd, 1992; Rushlow et al., 1989).

Nuclear restriction also plays a role in the regulation of the heat shock response during early development in Drosophila (Wang and Lindquist, 1998). Although many heat shock proteins (HSPs) of Drosophila are maternally supplied including Hsp83 and the small HSPs, Hsp70 is not (Arrigo and Tanguay, 1991; Zimmerman and Cohill, 1991). In fact Hsp70 is not inducible by heat shock in the early Drosophila embryo, nor is any other HSP gene (Graziosi et al., 1980). Indeed Hsp70 is not inducible in embryos from a wide variety of organisms in addition to flies including mice, frogs, and sea urchins (Heikkila et al., 1985; Morange et al., 1984; Roccheri et al., 1982). Despite the fact that in the Drosophila embryo the heat shock transcription factor (dHSF; which is responsible for the heat-induced transcription activation of HSP genes) is maternally supplied and abundant, the embryo remains refractory to heat shock until cycle 13. Wang and Lindquist have shown that the dHSF does not enter the nucleus until cycle 13 at which point Hsp70 induction can occur (Wang and Lindquist, 1998). Developmental regulation of the nuclear localization of dHSF, therefore, plays a key role in the establishment of the heat shock response in the early embryo.

The nuclear localization sequence of the dHSF has been identified and characterized (Zandi et al., 1997b). To identify protein(s) which bind to the NLS

and may be involved in the nuclear transport of dHSF, a yeast two-hybrid screen was conducted using the NLS as 'bait'. Several positive Drosophila cDNAs were identified of which one belongs to the karyopherin family of nuclear transport proteins, and has been designated Drosophila karyopherin-α3 (dKap-α3) (Gorlich et al., 1994; Kohler et al., 1997). Biochemical analysis of dKap-α3 demonstrates specific and functional interactions of the nuclear transporter with dHSF in vitro. Examination of the temporal and spatial expression pattern of dKap-α3 revealed that it is not expressed until cycle 13 of embryogenesis. These observations strongly support the notion that dKap-α3 is the nuclear transporter of the dHSF in vivo and that developmental regulation of dKap-α3 synthesis determines the time at which the heat shock response can be activated in the early embryo.

Results

The Drosophila HSF contains a 33-residue bi-partite NLS that is required for nuclear localization of the HSF during heat stress (Zandi et al., 1997b). Deletion of the NLS prevents nuclear localization yet allows the spontaneous oligomerization of the HSF in the cytosol, thus generating an active DNA binding form of the factor. These observations suggest that cytosolic factor(s) may interact, at least in part, with the NLS to prevent oligomerization as well as nuclear localization. The cytosolic factor(s) might function as a direct repressor of nuclear entry, similar to Iκ-B. Alternatively, a novel nuclear localization

process may be involved in which the monomer HSF is associated with a specific nuclear localization transport molecule.

To identify components of this regulatory system, a 43 amino acid segment of the Drosophila HSF which includes the NLS (Figure 1A) was used as bait in a yeast two-hybrid system screen (Bartel et al., 1993; Chien et al., 1991; Fields and Song, 1989). This screen allowed a search of a Drosophila cDNA embryonic library for proteins capable of specifically binding to the NLS. The screen revealed one primary class of NLS-binding proteins with sequence similarity to nuclear transport proteins.

Cloning Drosophila karyopherin-α3

Three strong positive clones were isolated and sequenced from the yeast two-hybrid screen of $5x10^6$ cDNAs from a Drosophila embryonic library (see Experimental Procedures). All three clones contained overlapping amino acid sequences (Figure 1B). One of the three contained an apparent full length-cDNA sequence encoding a 514-residue protein. Gene bank sequence search identified a matching sequence with a previously unpublished Drosophila karyopherin-α3 protein (dKap-α3). Based on sequence homology dKap-α3 is a member of the nuclear import α–proteins (importins) (Gorlich et al., 1994; Kohler et al., 1997). Also based on homology the dKap-α3 contains an internal NLS and Drosophila karyopherin-β (dKap-β) binding domain in the N-terminus as other previously characterized karyopherins (Moroianu et al., 1996), and eight "arm-repeats" (armadillo repeats) in the carboxy portion of the protein (Figure 1A and

1C) (Conti et al., 1998). At the amino acid level dKap-α3 shows 45% identity to Drosophila-α1, 44% identity to Drosophila-α2 and 67% identity to human-α3.

Karyopherin-α3 binds specifically to the Drosophila HSF NLS

In vitro protein cross-linking was used to determine the capability and specificity of dKap-α3 binding to the Drosophila HSF NLS. A 191 amino acid protein segment derived from the dHSF containing a centrally located NLS was cloned into pET11a and expressed in E. coli (Figure 1A). The purified protein was labeled in vitro with MAP kinase and γ-^{32}P ATP at fortuitous MAP kinase sites present within the protein segment; this probe was termed the 'mini-NLS'. An otherwise identical protein segment containing a deletion of the NLS was similarly labeled with MAP kinase and termed 'mini-ΔNLS' (Figure 1A). Mini-NLS was incubated with recombinant dKap-α3, molecularly cross-linked using disuccinimidyl suberate (DSS) and the product of this reaction analyzed by SDS-PAG electrophoresis (Figure 2 lane 1). A complex of approximately 90kD was observed by autoradiography of the gel. This complex is only observed when dKap-α3 was present in the reactions and immunoprecipitation with anti-Kap-α3 antibodies demonstrates that dKap-α3 is present in the complex (data not shown). Similar analysis with the mini-ΔNLS revealed no complex formation (Figure 2 lane 9). These results demonstrate that the NLS is required for dKap-α3 binding to the mini probes.

To determine the specificity of dKap-α3 interaction with the NLS, a series of point mutants within the NLS that are known to affect NLS function in vivo

(Zandi et al., 1997b, also summarized in the upper portion of Figure 2) were tested for their in vitro binding properties. Biochemical cross-linking studies with labeled mutant mini-NLS probes revealed a close correlation between the transport properties of these mutants and their ability to bind to the mini probes. As shown in Figure 2, those mutants, which are not transported in vivo, do not bind dKap-α3 in vitro: (K405M, lane 3; L404P, lane 4; N408S/R411L, lane 5 and K400E/R401L, lane 6). Two mutants, which are transported in vivo, do bind dKap-α3 in vitro although one more weakly than wild type: Q399I/K400R/L404R, lane 2; E406P, lane 8. A constitutively nuclear localized mutant Q403L binds efficiently to dKap-α3 (lane 7). Overall these results support the view that the cloned dKap-α3 protein can specifically bind to the dHSF NLS in vitro.

Karyopherin-β enhances dKap-α3 binding to the NLS of Drosophila HSF

Active nuclear transport complexes in vivo include an α/β-karyopherin heterodimer with the α subunit bound to the NLS of the cargo (Enenkel et al., 1995; Rexach and Blobel, 1995). It has been shown that in vitro binding of recombinant yeast kap-α (kap 60) to NLS domains is cooperatively enhanced by kap-β (kap 95) (Rexach and Blobel, 1995). The cross-linking experiments described above show that in the absence of dKap-β the NLS and dKap-α3 bind specifically to each other. To determine what quantitative role dKap-β may have in the dNLS-dKap-α3 interaction we cloned the Drosophila Kap-β cDNA using primers derived from the known genomic sequence (see Experimental Procedures). dKap-β was expressed in E. coli and purified using the GST-tag

system and the GST-tag removed prior to use (Smith and Johnson, 1988). α and β proteins were combined and incubated with ^{32}P-labeled mini probes cross-linked with DSS and examined by SDS-PAGE. The addition of dKap-β to the reactions modestly stimulated dKap-α3 binding to the NLS (Figure 3A; compare lanes 1, 2 and 3). Interestingly, the molecular weight of the complex was not altered by the presence of dKap-β suggesting that either the dKap-β association is transient in vitro or the interaction is such that it cannot be cross-linked with DSS.

The karyopherin-β binding domain of dKap-α3 is required for NLS binding in vitro

A comparison of the human and mouse dKap-α3 protein sequences to the Drosophila α3 sequence reveals several regions of homology. One homologous segment corresponds to the internal NLS of the human and mouse proteins, and is located between residues 22 to 41 of Drosophila dKap-α3 (Moroianu et al., 1996). A second homologous region present in dKap-α3 is the karyopherin-β binding domain (Moroianu et al., 1996). This domain is located in the first 115 amino acids of the N-terminus of karyopherin-α3 (see Figure 1A).

To investigate the role of these domains in dKap-α3 binding to the Drosophila HSF NLS, N-terminal deletions of dKap-α3 were examined. One deletion removed the internal NLS (deleting residues 1-47 of the N-terminus) and the other had the entire N-terminal 99 amino acid β-binding domain deleted (Figure 1A). These proteins were cross-linked to ^{32}P-labeled mini-NLS and mini-ΔNLS probes and the results are shown in Figure 3B, lanes 3 to 6. Drosophila

dKap-α3 deleted of its internal NLS bound the mini-NLS probe very well but not the mini-ΔNLS probe (lanes 3 and 4, respectively). Deletion of the entire β-binding domain, however, eliminates mini-NLS probe binding (lane 5). This observation suggests that the β-binding domain of Drosophila Kap-α3 is necessary for dHSF-NLS binding. This is unexpected because previous biochemical and structural studies with yeast and human proteins have demonstrated that karyopherin-α3 binds to its target NLS-peptide cargo within the arm repeats. Indeed, an N-terminal β-binding domain-deleted form of the yeast α3 protein was sufficient for crystallization with an SV-40 NLS peptide (Conti et al., 1998). It is possible that the use of a significantly larger cargo in our experiments requires a more significant portion of dKap-α3 for docking.

Karyopherin-α3 is required for nuclear docking in vitro

Nuclear docking experiments were performed to determine if dKap-α3 and dKap-β can target the dHSF-NLS to the nuclear pore complex. These experiments employed Schneider (SL2) cells, which were depleted of nuclear transport factors along with other cytosolic proteins by digitonin permeabilization (Smith and Johnson, 1988). Nuclear transport factors are added to the cells as purified recombinant proteins along with NLS-EGFP fusion proteins to serve as the cargo protein. As shown in Figure 4 wild type NLS fused to EGFP was effectively docked on the nuclear periphery in the presence of both α and β proteins (Figure 4 panel B). The NLS-EGFP protein alone was not able to dock on the nuclear membrane (Figure 4, panel A), indicating that dKap-α3 and dKap-β are required

for docking. Consistent with the biochemical crosslinking experiments, and in vivo localization studies, the K405M NLS mutant did not show nuclear docking in the presence of dKap-α3 and dKap-β (Figure 4 panel C). The constitutive nuclear-localized mutant, Q403L NLS, was able to dock on the nuclear pore complexes in the presence of dKap-α3 and β, as expected (Figure 4, panel D).

Drosophila karyopherin–α3 is excluded from the nucleus after heat shock

Biochemical and immuno-cytochemical studies of HeLa and Drosophila SL2 cells have shown that vertebrate HSF1 and dHSF are distributed in a diffused pattern over the cytoplasm and nucleus under normal growth conditions. Upon heat shock, HSF is localized into the nucleus rapidly and forms large nuclear granules (Cotto et al., 1997; Zandi et al., 1997b). Because dKap-α3 appears to be the nuclear import factor involved in this process in Drosophila, we examined the cellular distribution of dKap-α3 protein in response to heat shock. In normally growing SL2 cells the immunofluorescent staining pattern of dKap-α3 is uniformly distributed throughout the cell. After five minutes of heat shock the majority of dKap-α3 is localized in the cytoplasm (Figure 5). However, after 15 minutes of heat shock dKap-α3 is localized exclusively in the cytoplasm around the nuclear periphery. This peripheral nuclear staining increased strikingly with increasing duration of heat shock.

Developmental regulation of karyopherin–α3 protein synthesis correlates with nuclear entry of dHSF

The developmental profile of dKap-α3 expression was examined by Western blotting. Extracts were prepared from early Drosophila embryos at 0-2, 0-4 and 0-6 hours after egg deposition. The results of the Western analysis are shown in Figure 6 using anti-dKap-α3 monoclonal antibodies (the epitopes for these antibodies are present in the α-3 specific N-terminus). It is evident from the blot that dKap-α3 is expressed at very low levels if at all at 0-2 hours (Figure 6; lanes 1 and 2). At 0-4 and 0-6 hours dKap-α3 is abundantly expressed (Figure 6; lanes 3, 4, and 5). The total amount of protein present in each lane was normalized by Coomassie staining; and demonstrated that equivalent amounts of total protein were loaded for each developmental stage (Figure 6B).

In situ immunostaining of Drosophila embryos was used to assess the temporal and spatial expression of both dHSF and dKap-α3 proteins. In Figure 7A non-heat shocked embryos are shown on the left (NS) and heat shocked embryos on the right (HS). The embryonic nuclei are stained with DAPI to facilitate nuclear positioning and determination of the division cycle of the embryo. Panel A shows a representative cycle 11 embryo and panel B shows a cycle 13 embryo. Panels C and D show cycle 12 and 13 heat shocked embryos, respectively. The distribution of dHSF is revealed by immunostaining with an anti-dHSF monoclonal antibody. At cycle 11 dHSF is clearly seen in the cytosol of the embryo and excluded from the nucleus (panel E). At cycle 13 dHSF is localized within the nucleus in the absence of heat shock (panel F). Similarly staged embryos were also stained with anti-dKap-α3 antibodies. Non-shocked embryos demonstrated essentially no staining in early cycles (cycle 11, panel I),

but significant staining is seen in cycle 13 embryos predominantly in the cytosol (panel J).

Analysis of the distribution of dHSF in heat shocked embryos shows that dHSF is present only in the cytosol of heat shocked cycle 12 embryos (panel G). Despite the fact that these embryos have been heat shocked, the transcription factor cannot enter the nucleus. Heat shocked cycle 13 embryos, however, demonstrate a striking puntate/granular pattern to the nuclear staining (panel H). A similar pattern of staining is seen in heat shocked Drosophila SL2 cells (Zandi et al., 1997b). Note the absence of detectable dKap-α3 protein in the heat shocked cycle 12 embryo (panel K), correlating well with the absence of dHSF nuclear entry. At cycle 13 there is significant dKap-α3 protein observed in the cytosol, and the dHSF can now enter the nucleus.

Similar analysis of the mRNA distribution of dHSF and dKap-α3 are in agreement with the protein distribution patterns both temporally and spatially. The absence of appreciable dKap-α3 mRNA in the early embryo is evident in Figure 7B (panel 1), where a cycle 10-11 embryo is compared against a cycle 13-14 embryo that shows considerable mRNA accumulation. It is interesting to note, however, the presence of a small amount of dKap-α3 mRNA in the posterior of the cycle 10-11 embryo where the pole cells will arise (indicated by the arrow, panel 1). These observations demonstrate that dKap-α3 expression during the first 12 cycles of embryogenesis is restricted to the very posterior of the embryo. Further analysis of the posterior region of cycle 12 embryos with anti-dHSF antibodies shows clear localization of dHSF protein within the nuclei of

the poll cells (Figure 7B panel 2). These data further strengthen the correlation between the presence of dKap-α3 and the nuclear localization of dHSF.

Discussion

Early embryos as diverse as flies, frogs, sea urchins and mice are unable to induce Hsp70 transcription (Heikkila et al., 1985; Morange et al., 1984; Roccheri et al., 1982). The presence of large quantities of Hsp70, or the modification of the transcription apparatus that heat shock might cause, could be deleterious to the early embryo. It has been shown that elevated levels of Hsp70 can impede cell growth and division at normal growth temperatures (Feder et al., 1992; Krebs and Feder, 1997). This may also be the case at stress temperatures in the early embryo. The interruption of the normal developmental transcription program by the activation of heat shock protein synthesis may simply be too damaging to the early embryo and provide little advantage to the organism. Wang and Lindquist previously demonstrated that nuclear exclusion of the dHSF in the early embryo is correlated with the absence of Hsp70 induction (Wang and Lindquist, 1998). In this report we show that the mechanism which restricts the nuclear entry of the dHSF is the absence of a key nuclear transporter, Drosophila karyopherin-α3.

Biochemical analysis of dKap-α3

dKap-α3 is likely to be the bona fide nuclear transporter for dHSF for a number of reasons. First, the two-hybrid system used to screen for NLS binding proteins

selected dKap-α3 as the primary interacting protein. Second, dKap-α3 was the only gene isolated from this family; indeed, neither dKap-α1 or dKap-α2 were identified in this screen although both cDNAs are present in the embryonic library that was used (unpublished observations; Kussel and Frasch, 1995; Torok et al., 1995). Third, point mutations in the NLS of dHSF, which block nuclear entry in vivo, also prevent dKap-α3 binding in vitro. Finally, dKap-β promotes binding of the NLS to dKap-α3 in vitro and together the α3 and β proteins allow NLS-EGFP fusion proteins to dock to nuclei in digitonin-treated cells. Successful docking to the nuclei is also sensitive to mutations in the NLS that prevent nuclear entry in vivo.

Localization of dKap-α3

In normally growing SL-2 cells, dKap-α3 protein is uniformly localized throughout the cells. After heat shock, the transporter relocates to the nuclear membrane and within 15 minutes is entirely excluded from the nucleus. The significance of this localization may be to prevent interactions of dHSF with the transporter while it is involved in the activation of transcription. Alternatively, the dKap-α3 may become associated with the nuclear membrane and not be able to undergo the normal nuclear transporter recycling events during heat stress.

The spatial and temporal aspects of dKap-α3 expression in the early embryo demonstrate that the dKap-α3 protein is not expressed until 13 cycle, at which point significant RNA and protein accumulation is observed. Remarkably, this correlates precisely with that of dHSF nuclear entry and Hsp70 heat-

inducibility. The absence of dKap-α3 expression is coincident with the refractory period of Hsp70 induction and the nuclear exclusion of dHSF. Taken together with the biochemical analysis, these data present a compelling case that dKap-α3 is the nuclear transporter of the dHSF in vivo. Further genetic analysis will be necessary to demonstrate that mutations in dKap-α3 alter dHSF subcellular localization and function.

The role of dKap-α3

Western blotting and immunoprecipitation of dKap-α3 in cultured Drosophila Kc cells have shown that dKap-α3 is present in significantly greater quantities than dHSF (data not shown). It is therefore reasonable to suppose that dKap-α3 transports a number of other proteins into the nucleus. Indeed, a recent report using similar methods to identify proteins interacting with the Drosophila germ cell-less protein identified the same dKap-α3 described in this report (Dockendorff et al., 1999). Interestingly, the Gcl protein also contains a bi-partite NLS of approximately 30 amino acids. Comparison of the two NLSs revealed essentially no sequence similarity other than the presence of basic residues. Computer projections of the putative structure of the dHSF NLS suggests that it is α-helical (Zandi et al., 1997b), whereas the Gcl NLS contains two proline residues that would interrupt an α-helical structure.

Previous northern analysis and whole mount in situ hybridization results have suggested that dKap-α3 is ubiquitously expressed throughout early development (Dockendorff et al., 1999). These observations do not coincide with

our analysis of protein and RNA expression. Although we do not observe any appreciable accumulation of either dKap-α3 protein or mRNA until cycle 13, there is a small amount mRNA in the posterior of cycle 10-11 embryos as described in the Results section. This RNA may provide dKap-α3 to the developing pole cells and hence transport Gcl; it is clear that in the cycle 12 pole cells dHSF is nuclear. Thus, early expression of dKap-α3 in the posterior of the embryo may facilitate nuclear entry of critical proteins like Gcl into the developing pole cells.

Domains of dKap-α3

Remarkably, deletion of the dKap-α3 β-binding domain eliminates binding of the dHSF NLS to dKap-α3 in vitro. Previous structural studies have shown that for a fragment of the yeast Kap-α3 protein, which lacks the β-binding domain, two binding sites exist for an SV-40 NLS peptide within the arm-repeat domain (Conti et al., 1998). A recent structural study of the mouse importin α using full-length protein shows that the N-terminal β-binding domain is capable of interacting intramolecularly with the arm repeats to form a self-inhibitory structure (Kobe, 1999). In this case no exogenous NLS was present in the crystals. It is likely that the significant size difference between the dHSF mini-NLS cargo used in this report and the SV-40 peptide may explain why other domains of the α-3 protein are needed for binding.

Early embryonic transcription and nuclear transport

Early development in Drosophila is characterized by series of rapid zygotic nuclear divisions without appreciable transcription until cycles 8 and 9 (Erickson and Cline, 1993). It has been demonstrated that components of the basal transcription machinery are not transported into the nuclei at different division cycles. The RNA polymerase IIC subunit is found within the nucleus at cycle 7 whereas TFIID's TATA-binding protein (TBP) is localized within the nucleus between cycles 8 and 9 (Wang and Lindquist, 1998). The timing of dHSF entry into the nucleus is independent of these two general factors and this is likely due to the requirement of other nuclear transport molecules for the nuclear localization of these basal factors.

Developmental regulation of the heat shock response by a nuclear transporter represents a novel form of transcription regulation for a specific group of genes. It is possible that the absence of transcription during early embryonic stages may, in general, be due to the absence of specific nuclear transporters, at least for those transcription factors that are maternally provided. Indeed, this mechanism could represent a general explanation for the lack of transcription of early acting genes in embryonic nuclei. It will be very interesting for future studies to determine whether the nuclear entry of specific transcription factors as well as members of the basal transcription machinery correlates with the presence of specific nuclear transporters.

Experimental procedures

Two-hybrid screening and protein expression

A 129bp dHSF DNA fragment that encodes the bait protein containing NLS was amplified by PCR and cloned into the pAS2 GAL4-DB (DNA binding domain) vector using BamHI and NdeI restriction sites. The fusion plasmid DNA was then amplified and purified from E.Coli. A Drosophila embryonic c-DNA library was carried in pAC2 GLA4-AD (activation domain) vector (Clonetech). Both plasmids were co-transformed into yeast y190 cells and plated on SD/-His/-Trp/-Leu to select for His^+ transformants. All positives were then tested for expression of the second reporter gene by colony lift β-galactosidase assays. Finally, individual positive clones were co-transformed with a pAS2 vector that carries the NLS-deleted bait to eliminate the false positives. Three positive pAC2 plasmids were isolated from yeast and sequenced to assemble the full-length coding region of dKap-α3. PCR primers were then designed according to the c-DNA sequence, and a full-length dKap-α3 DNA was cloned by PRC using a λgt11 cDNA library. dKap-α3 was expressed in E.Coli cells by subcloning the full-length cDNA into a GST fusion protein vector pGEX-2t. The GST-tagged dKap-α3 was then bound to a GST-affinity resin (Stratagene) and eluted with 10mm reduced glutathione in 50mM Tris (pH8). The GST tag was cleaved with thrombin for 60 minutes at a ratio of 3 units protease per mg of recombinant protein in 50mm Tris (pH8), 150mm NaCl and 2.5mm Ca_2Cl .

Colony lift β-galactosidase filter assay

The primary His^+ transformants were grown in 5ml selective SD medium at 30^0C to O.D. 0.5. Cells were then re-streaked onto a 150mm SD/-His/-Trp/-Leu agar plate and incubated for 72 hr at 30^0C. A sterile Whatman #5 filter was placed over the surface of the agar plate. As soon as the filter was completely wet it was carefully lifted off the plate and quick frozen in liquid nitrogen for 1 min. A second sterile Whatman #5 filter was pre-soaked in 3.5 ml of Z buffer/X-gal solution. The frozen filter was thawed at room temperature and carefully placed on the 2^{nd} filter, colony side up. Both filters were incubated at room temperature for up to 8 hr until the appearance of blue colonies.

In-vitro cross-linking assay

A 191aa 6HIS-tagged polypeptide from dHSF (mini-probe) containing the NLS was over-expressed in DE3 cells and purified by Ni-NTA affinity column (Qiagen). In addition, a similar 154aa polypeptide with the NLS deleted (mini-ΔNLS probe) was also over-expressed in DE3 as were and all polypeptides with point mutant NLS (Mini-mNLS probes). All of the protein probes were labeled with γ^{32}P-ATP by MAPK phosphorylation in vitro.

dKap-α3 (1 μg/μl) was incubated in D buffer (25mm Hepes(pH7.9), 100mm KCl, 1mm EDTA and 0.2% Triton-x 100) with 2 μl of γ^{32}P-Mini probes (~80000cpm/μl) for 20 min at 25^0C in a 20 μl reaction volume. Cross-linking was subsequently carried out by addition of 2μl of 20mM DSS and incubation for 15 min at 25^0C. The reactions were quenched by the addition of 2μl of 200mM

Lysine for 10 min. Protein-protein adducts were analyzed by 6% SDS-PAGE and autoradiography.

Cloning, expression and purification of Drosophila karyopherin (importin) β

Four PCR primers were designed according to the partial genomic sequence published in GeneBank (accession number g92598391):

5'GCGCGCGAATTCCATATAGAGAGGAAAAGAG3'

5'GCGCGCCTCGAGCATAGTGCTTGGACAC3'

5'GCGCGCCTCGAGGTGCTCTGCAGTTCCTG3'

5'GCGCGCTCTAGACTACTGTGCGATGGACCTGGGT3'

Two amplified fragments corresponding to –26 to 962 and 963 to 2655 of the cDNA sequence of karyopherin β were obtained using the above primers and a Drosophila embryonic cDNA library (λgt11) as PCR template. The two fragments were ligated into pBluescript (Stratagene) and sequenced. The full-length karyopherin β was fused to a GST tag using vector pGEX-2t (Smith and Johnson, 1988), expressed in E coli, and purified with GST affinity resin (Stratagene). The GST tag was removed with thrombin.

In-vitro nuclear docking assays

In vitro nuclear docking assays were performed according to the methods developed by (Stochaj and Silver, 1992). To study the binding of Drosophila heat shock factor NLS, Schneider cells were allowed to attach to polylysine-coated slides for 20 min on ice. Then the cells were permeabilized with 45μg/ml

digitonin in Buffer A (20mM Hepes (pH7.3), 110mM potassium acetate, 5mM sodium acetate, 2mM magnesium acetate, 1mM EGTA, 2mM DTT, 1mM PMSF, and "Complete Protease Inhibitors" from Boehringer for 20 min on ice. NLS-EGFP fusion proteins, dKap-α3 and dKap-β were pre-incubated on ice for 20 min, then were incubated with permeabilized cells at room temperature for 20 min. Cells were washed with Buffer A and fixed with Histochoice Tissue Fixative MB (Amresco). Slides were mounted in A Buffer/90% glycerol containing 1mg/ml o-phenylenediamine.

Immuno-fluorescent staining

Mouse monoclonal antibodies were raised against recombinant dKap-α3 using standard immunization procedures. Three monoclonal lines were characterized: 5E3, 5F6, 6G7 and all three reacted with an epitope present in the N-terminal 100 amino acids of dKap-α3. This domain is unique among the dKap-α family members and therefore should eliminate any cross-reactivity. Western analysis with these antibodies reveal only a single strongly reactive protein species of the correct molecular weight.

18x18 mm #1 cover slips were coated in 1 mg/ml poly-L-lysine for 15 min and air dried. 0.5ml of Drosophila SL2 cells with a density of $4x10^5$ cells/ml were placed onto each cover slide and incubated for 15 min at room temperature. Cells were then heat shocked at 37°C for various times, washed with PBST, and fixed with Histochioce Tissue Fixative MB (Amresco) for 15 min on ice. After several washes to remove the fixative, the fixed cells were incubated with 1:1000

dilution of monoclonal anti-dKap-α3 antibody (5E3) in PBS buffer containing 0.5% bovine serum albumin (BSA) for 2 hr at room temperature. Cells were then washed with PBST 4 times for 10 minutes each to remove unbound first antibody. Fluorescein-coupled goat anti-mouse IgG (Pierce) secondary antibody was then added at 1:100 dilution in 0.5% BSA in PBS and incubated at 4°C overnight. Cell nuclei were visualized by co-staining with DAPI (4', 6-Diamidine-2'-phenylindole dihydrochloride) for 10 min. Finally, the cover slips were washed with PBST 4 times and mounted onto microscope slides in 90% glycerol/PBS containing 2.5% DABCO (1,4-diazabicyclo [2,2,2] octane (Sigma)). The fluorscent images were viewed and photographed with Zeiss Axioplan microscopy with UV irradiation and appropriate filters.

Fixed Drosophila embryos were re-hydrated by a series methanol/PBST mixtures: 15 each min in 75%, 50%, 25% methanol/PBST and 30 min in PBST. Immuno-fluorescent staining was then carried out as described by (Patel, 1994). All antibodies were pre-incubated with 0-12 hr embryos overnight at 4^0C. Anti-dKap-α3 antibodies were diluted 1:500 and anti-dHSF antibodies were diluted at 1:250. Fluorescent dye-labeled secondary antibodies were diluted 1:100. All primary antibody incubations were at 4^0C overnight and secondary antibody incubations were at room temperature for 90 min. Images were taken by confocal microscopy (Zeiss, LSM310) or Axioplan microscopy. The embryonic stages were determined by co-staining with DAPI (4', 6-Diamindine-2'-phenylinedole dihydrochloride).

Developmental Western

0-2 hr, 0-4 hr and 0-6 hr embryos were collected and washed with 0.03% Triton X-100/0.9% NaCl. Non-shocked or heat shocked (37^{0}C for 15 min) embryos were rinsed twice with homogenization buffer (50mM Tris (pH7.5), 140mM NaCl, 5mM $MgCl_2$, 0.05% NP-40, 1mM PMSF, 1μg/ml pepstain A, 1-2μg/ml aprotinin, 1μg/ml leupeptin, and then homogenized in 2V of homogenization buffer. Extracts were then centrifuged to remove cell debris and the supernatants were mixed with SDS-PAGE gel loading buffer and electrophoresed in 8% SDS-PAGE. The separated proteins were then transferred to nitrocellulose and blocked with 5% NFDM overnight. The blot was then probed with anti-dKap-α3 monoclonal antibody 5E3 and developed with anti-mouse alkaline phosphatase conjugated antibodies.

Embryo preparation

Drosophila embryos were collected in the population cages with different time span. No-shock embryos were processed immediately, while heat-shock samples were incubated in a 37^{0}C water bath for 15 min. All embryos were washed with NaCl-Triton (0.9%NaCl, 0.03%Triton-x 100), dechorionated in 50% bleach for 3 min and fixed in n-heptane/formaldehyde/PBS (5:1:5) for 30 min. After removing vitelline membranes in n-heptane/methanol (1:1) by vigorous shaking, embryos were washed three times with methanol and stored in ethanol at –20^{0}C. For western blotting, embryonic extracts were made without the fixation and stored at –70^{0}C.

Acknowledgements

We thank Christine Molodovich and LiJuan Wang for technical assistance. We thank Susan Ou and the Caltech Monoclonal Antibody Facility for generating the monoclonal antibodies used in this study. We thank Dr. Al Courey for generously allowing us to collect staged embryos from his laboratory. We thank Dr. Eric Davidson, Dr. Bruce Hay, Duncan Odom, Phoebe Tzou and Dr. Ebrahim Zandi for helpful comments on the manuscript. This work was supported by NIH Grant GM47381 to CSP.

References

Arnone, M. I. and Davidson, E. H. (1997). The hardwiring of development: Organization and function of genomic regulatory systems, *Development* **124**, 1851-1864.

Arrigo, A. P. and Tanguay, R. M. (1991). Expression of Heat Shock proteins during development in Drosophila, Results and Problems in Cell Differentiation *17*.

Bartel, P., Chien, C. T., Sternglantz, R. and Fields, S. (1993). Elimination of False Positives That Arise in Using the 2-Hybrid System, *Biotechniques* **14**, 920-&.

Chien, C. T., Bartel, P. L., Sternglanz, R. and Fields, S. (1991). The 2-Hybrid System - a Method to Identify and Clone Genes for Proteins That Interact with a Protein of Interest, *Proc. Natl. Acad. Sci. (USA)* **88**, 9578-9582.

Conti, E., Uy, M., Leighton, L., Blobel, G. and Kuriyan, J. (1998). Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha, *Cell* **94**, 193-204.

Cotto, J. J., Fox, S. G. and Morimoto, R. I. (1997). HSF1 granules: a novel stress-induced nuclear compartment of human cells, *Journal of Cell Science* **110**, 2925-2934.

Dockendorff, T. C., Tang, Z. Y. and Jongens, T. A. (1999). Cloning of karyopherin-alpha 3 from Drosophila through its interaction with the nuclear localization sequence of germ cell-less protein, *Biological Chemistry* **380**, 1263-1272.

Enenkel, C., Blobel, G. and Rexach, M. (1995). Identification of a Yeast Karyopherin Heterodimer That Targets Import Substrate to Mammalian Nuclear-Pore Complexes, *Journal of Biological Chemistry* **270**, 16499-16502.

Erickson, J. W. and Cline, T. W. (1993). A b-zip protein sisterliss-a collaborates with bHLH transcription factors early in Drosophila development to determine sex, *Genes & Development* **7**, 1688-1702.

Feder, J. H., Rossi, J. M., Solomon, J., Solomon, N. and Lindquist, S. (1992). The Consequences of Expressing Hsp70 in Drosophila Cells at Normal Temperatures, *Genes & Development* **6**, 1402-1413.

Fields, S. and Song, O. K. (1989). A Novel Genetic System to Detect Protein Protein Interactions, *Nature* **340**, 245-246.

Geisler, R., Bergmann, A., Hiromi, Y. and Nusslein-Volhard, C. (1992). Cactus, a gene involved in dorsalventral pattern formation of Drosophila, is related to the IkB gene family of vertebrates, *Cell* **71**, 613-621.

Gorlich, D., Prehn, S., Laskey, R. A. and Hartmann, E. (1994). Isolation of a Protein That Is Essential for the First Step of Nuclear-Protein Import, *Cell* **79**, 767-778.

Graziosi, G., Christini, F., Marzari, R., Micale, F. and Savoini, A. (1980). Varibility of response of early drosophila embryos to heat shock, *Journal of Embryology and Experimental Zoology* **214**, 141-145.

Heikkila, J. J., Kloc, M., Bury, J., Schultz, G. A. and Browder, L. W. (1985). Acquisition of the Heat-Shock Response and Thermotolerance During Early Development of Xenopus-Laevis, *Developmental Biology* **107**, 483-489.

Jain, J., McCaffrey, P. G., Valgearcher, V. E. and Rao, A. (1992). Nuclear Factor of Activated T-Cells Contains Fos and Jun, *Nature* **356**, 801-804.

Kaffman, A. and O'Shea, E. K. (1999). Regulation of Nuclear Localization: a key to a door. In Annual Review of Cell and Developmental Biology (Palo Alto, Annual Reviews, Inc.), pp. 291-339.

Karin, M., Liu, Z. G. and Zandi, E. (1997). AP-1, Function and Regulation, *Current Opinions in Cell Biology* **9**, 240-246.

Kidd, S. (1992). Characterization of the *Drosophila cactus* locus and analysis of interactions between cactus and dorsal proteins, *Cell* **71**, 623-635.

Kobe, B. (1999). Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin alpha, *Nature Structural Biology* **6**, 388-397.

Kohler, M., Ansieau, S., Prehn, S., Leutz, A., Haller, H. and Hartmann, E. (1997). Cloning of two novel human importin-alpha subunits and analysis of the expression pattern of the importin-alpha protein family, *Febs Letters* **417**, 104-108.

Krebs, R. A. and Feder, M. E. (1997). Deleterious consequences of Hsp7O overexpression in Drosophila melanogaster larvae, *Cell Stress & Chaperones* **2**, 60-71.

Kussel, P. and Frasch, M. (1995). Pendulin, a Drosophila protein with cell cycle-dependent nuclear localization, is required for normal cell proliferation, *Journal of Cell Science* **129**, 1491-1507.

Morange, M., Diu, A., Bensaude, O. and Babinet, C. (1984). Altered Expression of Heat-Shock Proteins in Embryonal Carcinoma and Mouse Early Embryonic-Cells, *Molecular and Cellular Biology* **4**, 730-735.

Morisato, D. and Anderson, K. V., eds. (1995). Signaling pathways that establish the dorsal-ventral pattern of the Drosophila embryo (Palo Alto, Annual Reviews, Inc.).

Moroianu, J., Blobel, G. and Radu, A. (1996). The binding site of karyopherin alpha for karyopherin beta overlaps with a nuclear localization sequence, *Proc. Natl. Acad. Sci. (USA)* **93**, 6572-6576.

Patel, N., ed. (1994). Imaging neuronal subsets and other cell types in whole-mount Drosophila embryos and larvae using antibody probes (San Diego, Academic Press, Inc.).

Rexach, M. and Blobel, G. (1995). Protein Import into Nuclei - Association and Dissociation Reactions Involving Transport Substrate, Transport Factors, and Nucleoporins, Cell **83**, 683-692.

Roccheri, M. C., Sconzo, D., Di, C. M., Di, D. M. G., Pirrone, A., Gambino, R. and Giudice, G. (1982). Heat-shock proteins in sea urchin embryos: transcriptional and postranscriptional regulation, *Differentiation* **22**, 175-178.

Rushlow, C. A., Han, K., Manley, J. L. and Levine, M. (1989). The graded distribution of the dorsal morphogen is initiated by selective nuclear transport in Drosophila, *Cell* **59**.

Schindler, C. and Darnell, J. E. J. (1995). Transcriptional Responses to Polypeptide Ligands: The JAK-STAT Pathway. In Annual Review of Biochemistry, C. Richardson, J. Abelson, A. Meister, and C. Walsh, eds. (Palo Alto, Annual Reviews Inc.), pp. 621-651.

Smith, D. B. and Johnson, K. S. (1988). Single step purification of polypeptides expressed in Escherichia-Coli as fusions with Glutathione S-Transferase, *Gene* **67**, 31-40.

Steward, R. (1987). *Dorsal*, an embryonic polarity gene in *Drosophila*, is homologous to the vertebrate proto-oncogene *c-rel*., **Science 238**, 692-694.

Stochaj, U. and Silver, P. A. (1992). A conserved phosphopeptide that specifically binds nuclear-localization sequences is involved in nuclear import, *Journal of Cell Biology* **117**, 473-482.

Torok, I., Strand, D., Schmitt, R., Tick, G., Torok, T., Kiss, I. and Mechler, B. M. (1995). The Overgrown Hematopoietic Organs-31 Tumor-Suppressor Gene of Drosophila Encodes an Importin-Like Protein Accumulating in the Nucleus at the Onset of Mitosis, *Journal of Cell Biology* **129**, 1473-1489.

Wang, Z. and Lindquist, S. (1998). Developmentally regulated nuclear transport of transcription factors in Drosophila embryos enable the heat shock response, *Development* **125**, 4841-4850.

Woronicz, J. D., Gao, X., Cao, Z., Rothe, M. and Goeddel, D. V. (1997). I kappa B kinase-beta: NF-kappa B activation and complex formation with I kappa B kinase-alpha and NIK, *Science* **278**, 866-869.

Zandi, E., Chen, Y. and Karin, M. (1998). Direct phosphorylation of I kappa B by IKK alpha and IKK beta: Discrimination between free and NF-kappa B-bound substrate, *Science* **281**, 1360-1363.

Zandi, E., Rothwarf, D. M., Delhase, M., Hayakawa, M. and Karin, M. (1997a). The I kappa B kinase complex (IKK) contains two kinase subunits, IKK alpha and IKK beta, necessary for I kappa B phosphorylation and NF-kappa B activation, *Cell* **91**, 243-252.

Zandi, E., Tran, T. N. T., Chamberlain, W. and Parker, C. S. (1997b). Nuclear entry, oligomerization, and DNA binding of the Drosophila heat shock

transcription factor are regulated by a unique nuclear localization sequence, *Genes & Development* **11**, 1299-1314.

Zhu, J. Y., Shibasaki, F., Price, R., Guillemot, J. C., Yano, T., Dotsch, V., Wagner, G., Ferrara, P. and McKeon, F. (1998). Intramolecular masking of nuclear import signal on NF-AT4 by casein kinase I and MEKK1, *Cell* **93**, 851-861.

Zimmerman, J. L. and Cohill, P. R. (1991). Heat shock and termal tolerance in plant and animal embryogenesis, *New Biology* **3**, 641-650.

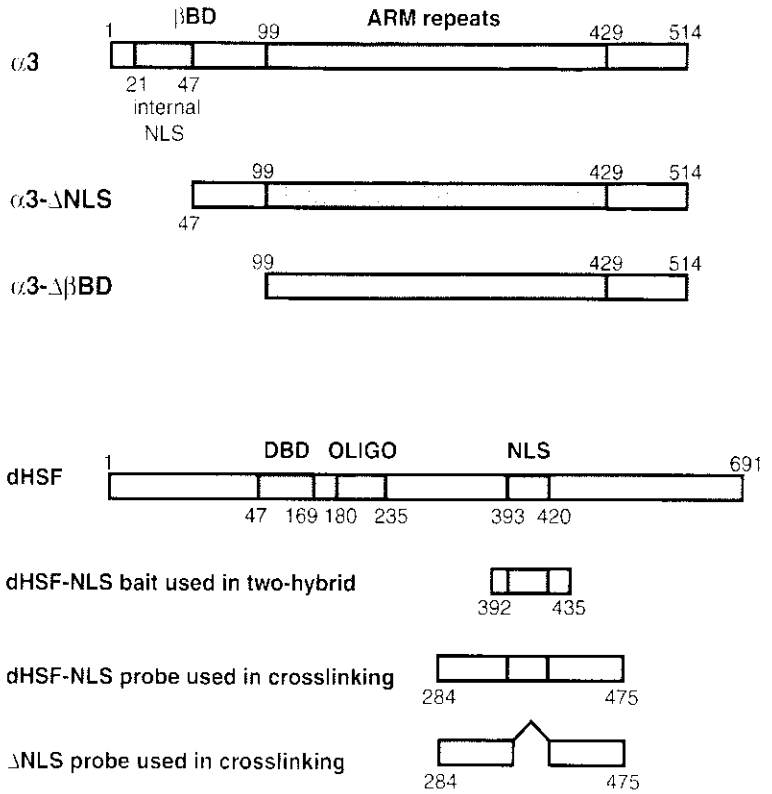
βBD
ARM repeats
α3
1
99
429
514
21
47
internal
NLS
α3-ΔNLS
99
429
514
47
α3-ΔβBD
99
429
514
dHSF
DBD
OLIGO
NLS
1
691
47
169
180
235
393
420
dHSF-NLS bait used in two-hybrid
392
435
dHSF-NLS probe used in crosslinking
284
475
ΔNLS probe used in crosslinking
284
475

Figure 1A

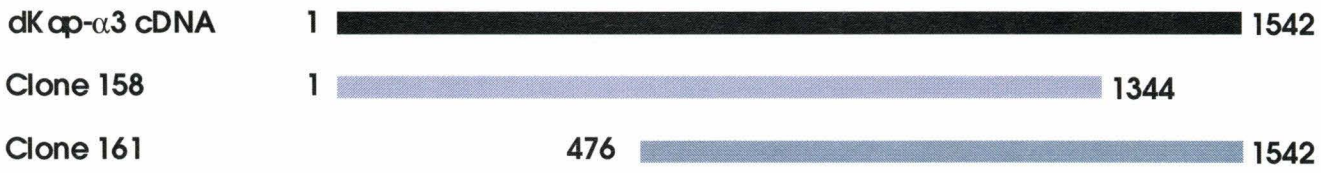

Figure 1B


Internal NLS
1 MTSMEQNRLQNFKNKGKDQD EMRRRRNEVTVELRKNKREETILKRR N VPNLDSNTDEEEQLSSSIDLKKL
71 AKAAADATKPEQQLAAVQAARKLLSSDKNP PINDLIQSDILPILVECLKQHNHTMLQFEAAWALTNIASG
141 TSAQTNEVVAAGAVPLFLQLLNSPAPNVCEQAVWALGNIIGDGPLLRDFVIKHGVVQPLLSFIKPDIPIT
211 FLRNVTWVIVNLCRNKDPAPPTATIHEILPALNVLIHHTDTNILVDTVWAISYLTDGGNDQIQMVIESGV
281 VPKLIPLLGNSEVKVQTAALRAVGNIVTGSDEQTQVVLNYDALSYFPGLLSHPKEKIRKEAVWFLSNITA
351 GNQSQVQAVINVGLLPKIIENLSKGEFQTQKEAAWAISNLTISGNREQVFTLIKEGVIPPFCDLLSCQDT
421 QVINVVLDG LNNMLKVADSHVEAVANCIEECEGLAKIERLQSHENVEIYKLAYEIIDQYFTDEGEQTNMA
491 PSSDGAQYNFDPHADRLTMNSFNF
ARM repeats

Figure 1C

Figures and Figure Legends

Figure 1A. Schematic diagram of the domain organization of Karyopherin-α3 and Drosophila HSF. Amino acid endpoints for each region, as well as their proposed functions, are indicated. Two Kap-α3 deletion mutants used in the crosslinking experiments, α3-ΔNLS and α3-ΔβBD, as well as three dHSF-NLS probes used in two hybrid screening or crosslinking experiments, are also shown. DBD: DNA-binding domain of dHSF, OLIGO: oligomerization domain of dHSF.

Figure 1B. Two-hybrid dKap–α3 clones. Schematic diagram of the dKap-α3 cDNAs isolated by two-hybrid screening. Two of the three distinctive positive clones, #158 and #161, were found to cover the entire coding region of dKap-α3.

Figure 1C. Amino acid sequence of Drosophila Karyopherin-α3. Amino acid sequence of full length Karyopherin-α3 is shown. The functional domains of dKap-α3 are illustrated as follows: boxed sequences include the internal NLS of dKap-α3 and ARM repeats, respectively; dKap-α3 binding domain is underlined and in italics.

Lane	1	2	3	4	5	6	7	8	9
NLS probe	wtNLS	Q399L K400R L404R	K405M	L404P	N408S R411L	K400E R401L	Q403L	E406P	ΔNLS
Localization (25°C)	Cyt	Cyt	Cyt	Cyt	Cyt	Cyt	Nuclear	Nuclear	Cyt
Localization (37°C)	Nuclear	Nuclear	Cyt	Cyt	Cyt	Cyt	Nuclear	Nuclear	Cyt

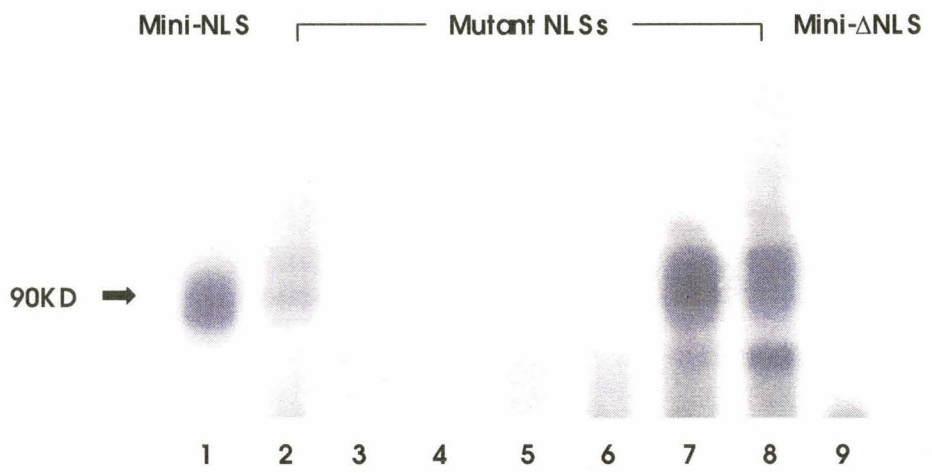

Figure 2

Figure 2. Cross-linking of dKap-α3 to mutant forms of dHSF-NLS polypeptides. The upper table summarizes the in vivo subcellular localization of the wild type and mutant NLSs determined by immuno-staining (Zandi, et al. 1997b). Recombinant wild type dKap-α3 protein (1μg) was cross-linked to the following γP^{32}-labeled forms of mini-NLS: mini-NLS (lane 1); Q399L, K400R, L404R mini-NLS (lane2); K405M mini-NLS (lane 3); L404P mini-NLS (lane 4); N408S, R411L mini-NLS (lane 5); K400E, R401L mini-NLS (lane 6); Q403L mini-NLS (lane 7); E406P mini-NLS (lane8) and mini-ΔNLS (lane 9). All probes were labeled with γ^{32}P-ATP and MAPK. 2ng of probe (approximately 150,000cpm) was used in each reaction. The NLS mutation Q399L/K400R/L404R (lane 2) and two constitutive nuclear entry mutations (Q403L, lane 7 and E406P, lane 8) formed complexes with dKap-α3 of approximately the same size (~90KD) as the wild type mini-NLS probe (lane 1). The remaining mutant forms, which are not transported in vivo (lanes 3-6), did not form specific complexes in vitro. The lower molecular weight complexes seen are dimers of probe and are mini-NLS probe specific.

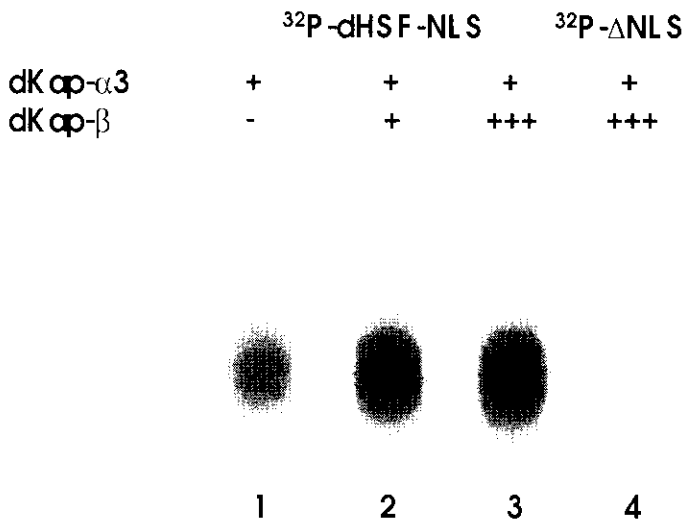

Figure 3A

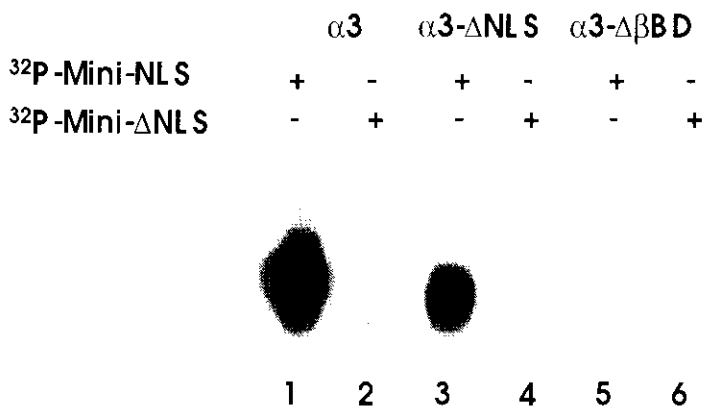

Figure 3B

Figure 3A. The effect of dKap-β on dKap-α3 binding to Drosophila HSF's NLS.

200ng of recombinant dKap-α3 protein was cross-linked to either mini-NLS or mini-ΔNLS probe. Increasing amounts of recombinant dKap-β was added to the indicated reactions with mini-NLS probe (lanes 1 –3). Lane 1, no dKap-β; lane 2, 200ng dKap-β; lane 3, 1μg dKap-β. No enhancement of dKap-α3 binding to the mini-ΔNLS reaction was observed even with the addition of 1μg of dKap-β (lane 4).

Figure 3B. Domains of dKap-α3 required for Drosophila HSF NLS binding.

Recombinant wild type or mutant dKap-α3 protein (1μg) was cross-linked to either mini-NLS or mini-ΔNLS probes, which were labeled with γ^{32}P-ATP using MAP kinase. 2ng of probe (150,000cpm) was used in each reaction. Both wild type dKap-α3 and dKap-α3-ΔNLS cross-linked to mini-NLS probe very efficiently (lanes 1 and 3). Deletion of the dKap-β binding domain, dKap-α3-ΔβBD, eliminated binding to the mini-NLS probe (lane 5). The mini-ΔNLS probe did not bind to any of the dKap-α3 proteins, and no complex is observed (lanes 2,4 and 6).

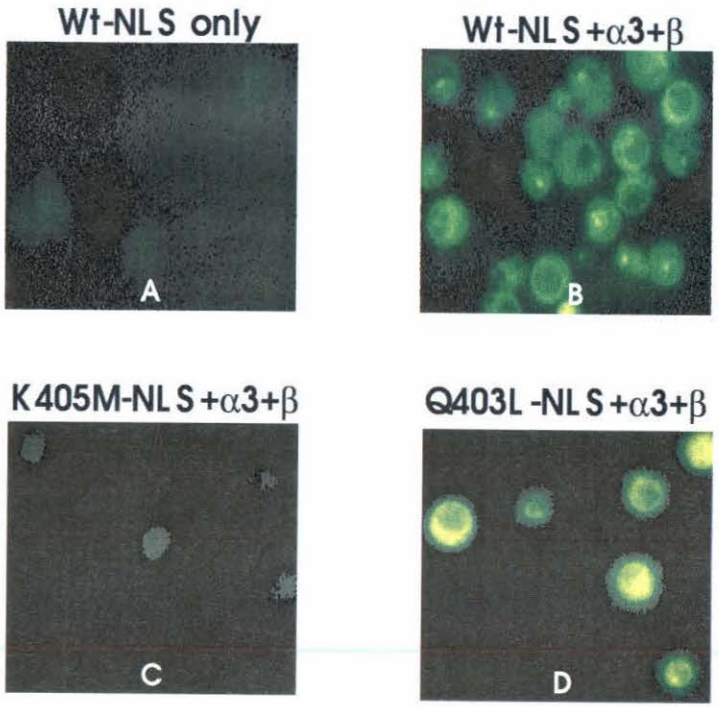

Figure 4

Figure 4. Karyopherin-α3 is required for nuclear docking in vitro.

Digitonin permeabilized SL2 cells were incubated with NLS-EGFP fusion proteins, recombinant dKap-α3 and dKap-β. The nuclear docking of EGFP fusion proteins was examined by fluorescent microscopy. Panel A: wild type NLS-EGFP fusion protein only; Panel B: wild type NLS-EGFP fusion protein, dKap-α3 and kap-β; Panel C: K405M mutant NLS-EGFP fusion protein, dKap-α3 and dKap-β; Panel D: Q403L mutant NLS-EGFP fusion protein, dKap-α3 and Kap β.

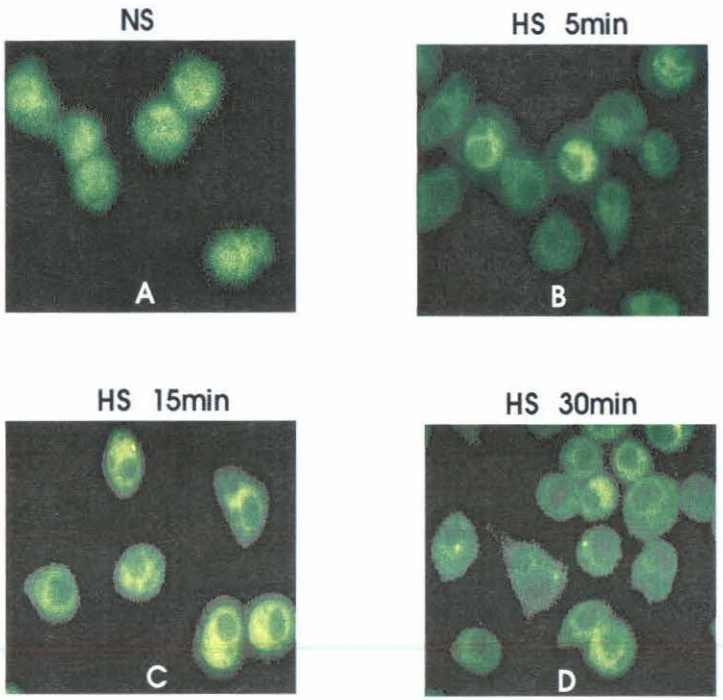

Figure 5

Figure 5. Effect of heat shock on the sub-cellular distribution of dKap-α3 in SL2 cells. Schneider cells stained with DAPI are displayed on the left side and immunofluorescent staining with anti-dKap-α3 monoclonal antibodies at 1:1000 dilution is shown on the right side. (NS) Non shocked cells; (HS) Heat shocked cells at 37°C.

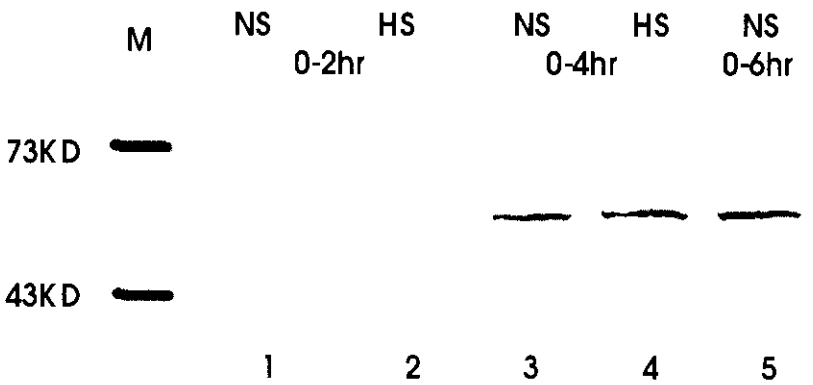

Figure 6A

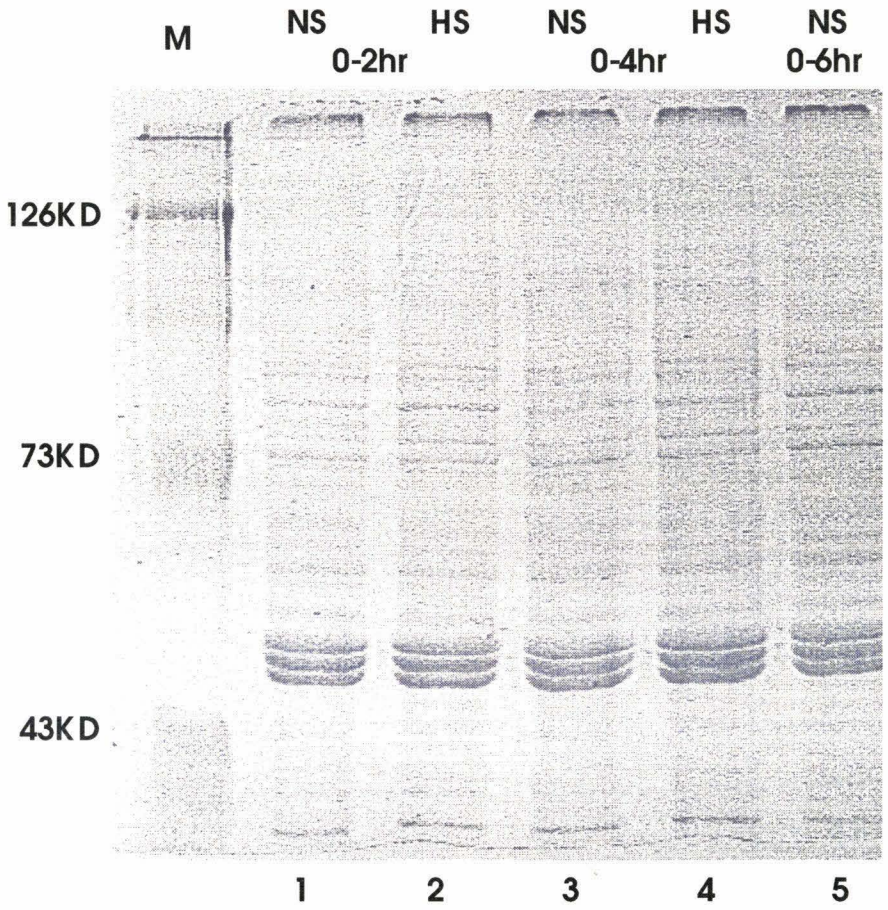

Figure 6B

Figure 6. Developmental western analysis of dKpa-α3 expression in Drosophila embryos.

0-2 hr, 0-4 hr and 0-6 hr (after egg laying) embryos were collected and homogenized. Whole embryo extracts obtained from homogenization were then analyzed by SDS-PAGE, probed with anti- α3 antibody for western blotting analysis (Figure 4A) or stained with coomassie blue (4B).

M: molecular weight standards; NS: non-shock embryos; HS: heat-shocked embryos. Lane 1: NS 0-2hr embryos, lane 2: HS 0-2 hr embryos, lane 3: NS 0-4 hr embryos, lane 4: HS 0-4 hr embryos, lane 5: NS 0-6 hr embryos.

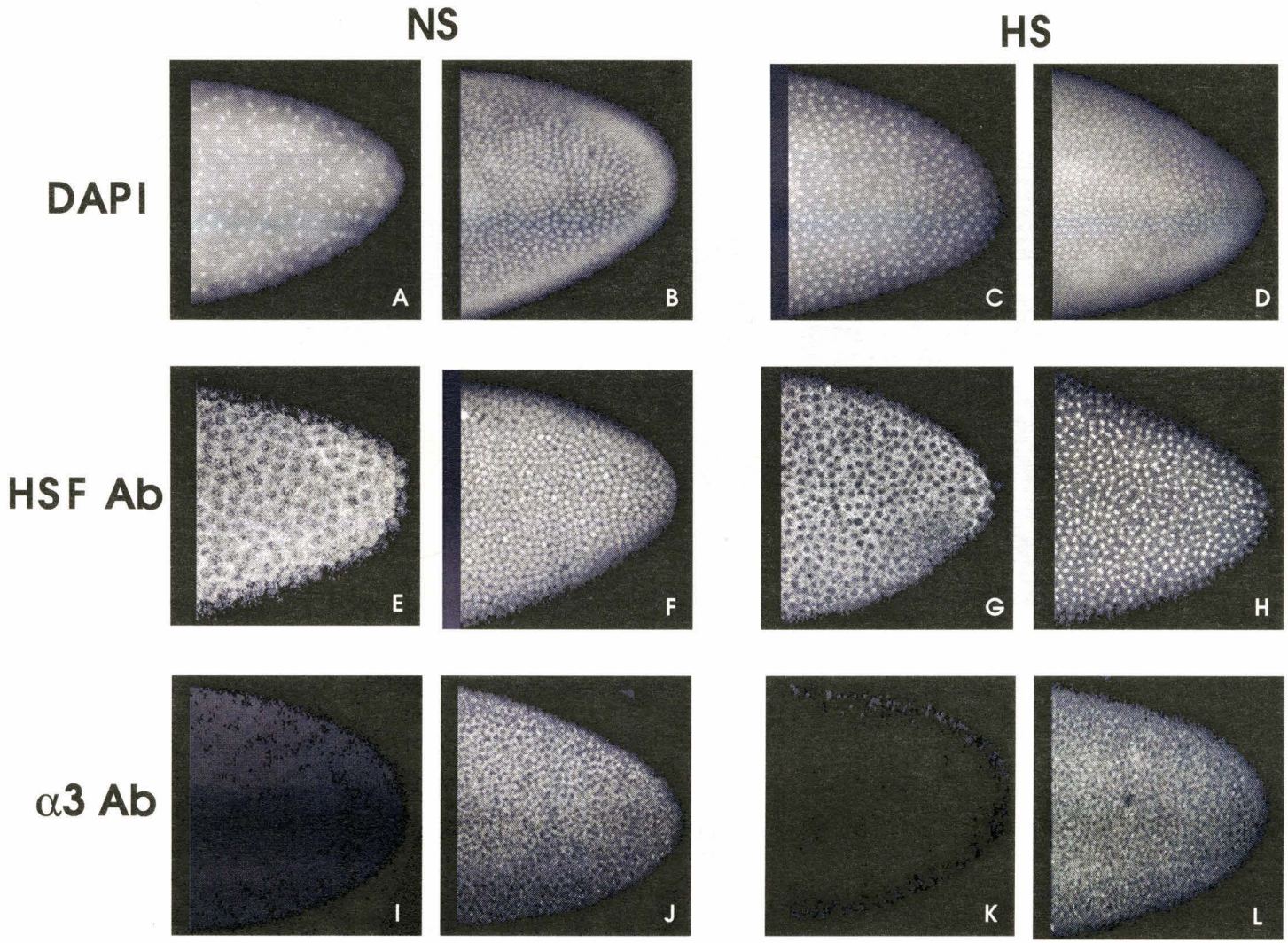

Figure 7A

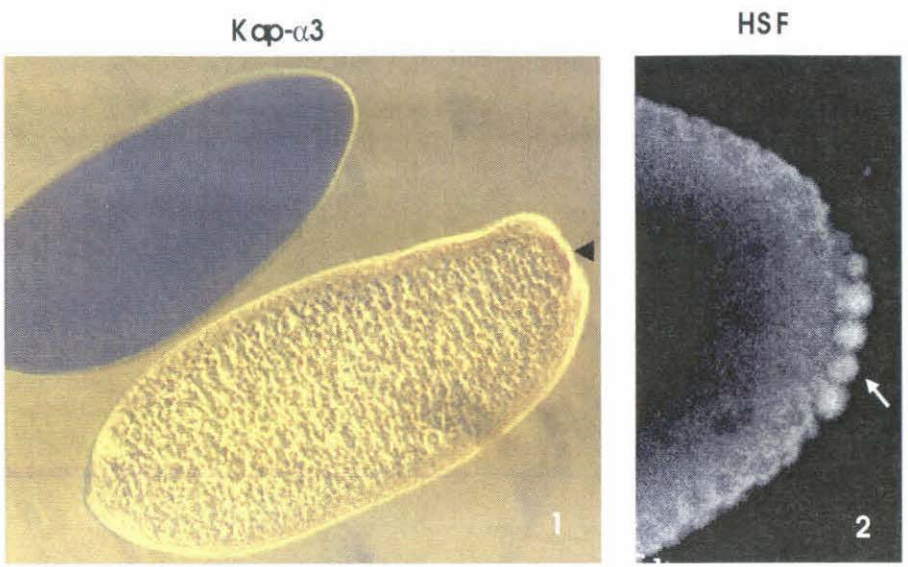

Figure 7B

Figure 7. Developmental timing of dKap–α3 expression correlates with nuclear entry of dHSF.

A. Cycle 11 to cycle 13 embryos were stained with either DAPI, anti-dHSF monoclonal antibodies (HSF panels) or anti-dKap-α3 monoclonal antibodies (Kap-α3 panels). NS; non-shocked Drosophila embryos and HS; heat-shocked embryos. Cycle 11 embryos are shown on the left under NS and cycle 12 embryos shown on the left under HS, respectively. Cycle 13 embryos are shown under both the NS and HS on the right side. Clearly present but excluded from the nucleus is dHSF in both non-shocked cycle 11 embryos (panel E) and heat shocked cycle 12 embryos (panel G). Similarly staged embryos display essentially no dKap-α3 staining (panels I and K). At cycle 13 in non-shocked embryos the dHSF is constitutively nuclear and dKap-α3 is expressed yet largely excluded from the nucleus (panel J). In heat shocked cycle 12 embryos the dHSF remains excluded from the nucleus (panel G) and no dKap-α3 is observed as in non-shocked embryos (panel K). At cycle 13 in heat shocked embryos the characteristic puntate pattern of dHSF nuclear staining is observed (panel H) and dKap-α3 remains excluded from the nucleus (panel L).

B. Analysis of the dKap-α3 mRNA distribution in the cycle 9-10 embryo demonstrates essentially no RNA except in the posterior region (indicated by the arrow, panel 1). Adjacent to this embryo is a cycle 13-14 embryo that shows significant levels of dKap-α3 mRNA (panel 1). Analysis of dHSF protein distribution in the cycle 12 embryo shows that dHSF is clearly present within the

nucleus of pole cells at this early stage (panel 2), corresponding to the presence of dKap-α3 mRNA in this region of the pre-cycle 13 embryo (panel 1).

CHAPTER THREE

A STEROID RECEPTOR-LIKE PROTEIN THAT *TRANS*-ACTIVATES THE *FUSHI TARAZU* SEGMENTATION GENE

The work in this chapter was done in collaboration with Dr. Kulkarni Prakash and Dr. Amita Behal.

Introduction

The process of segmentation in the early *Drosophila* embryo requires the coordinated function of three classes of zygotically active segmentation genes (Akam, 1987). These different classes of genes are called gap, pair-rule and segment polarity genes based on morphological defects that ensue in the lack of wild-type activity of the corresponding segmentation gene. During oogenesis, the primary determinants of the body plan are deposited in the maturing oocyte by maternally active genes. Immediately following fertilization, these determinants trigger the activity of the zygotically expressed segmentation genes to determine the number and organization of the body segments. The gap genes are the first set of zygotic genes that respond to the maternal information and are expressed in a series of overlapping domains. Subsequently, and in response to the gap gene products, the pair-rule class of genes are activated in a repeating 'zebra stripe' pattern suggesting that the overlapping distribution of the products encoded by the gap genes provide the spatial cues responsible for pair-rule gene expression. Thus, it appears that a major function of the pair-rule genes is to decode the spatial information into a periodic expression pattern (reviewed Nusslein-Volhard, 1991).

The pair-rule gene *fushi tarazu (ftz)* is a well-studied zygotic segmentation gene that is necessary for the development of the even-numbered parasegments (Wakimoto et al., 1984) and for the correct expression of certain homeotic and segmentation genes (Duncan, 1986; Howard & Ingham, 1986; Ingham & Martinez-Arias, 1988). In wild-type embryos, *ftz* is expressed in a 'zebra' stripe fashion in the

even-numbered parasegmental primordia of the embryonic blastoderm (Hafen et al, 1984; Weir & Komberg, 1985). We are interested in elucidating the molecular mechanisms underlying the regulation of *ftz* expression at the transcriptional level. Earlier work from this laboratory has identified several cis-sequences in the *ftz* zebra promoter that are critical for its striped expression (Dearolf et al., 1989; Topol et al., 1991). In particular, it was shown that in transgenic *Drosophila* embryos one of these *cis*-elements, fDE1, was able to function both as an activator as well as a repressor of *ftz* transcription (Topol et al. 1991). In an attempt to identify the transcription factors that bind to this site and mediate the spatially restricted expression of the *ftz* gene, we decided to clone the genes encoding these factors. In this report we present the cDNA cloning and characterization of zeb- 1, a novel steroid receptor-like protein that interacts with the fDE1 (zeb-1) site in the zebra promoter. Further, we show that in tissue culture cells zeb-1 functions as a sequence-specific *trans*-activator of *ftz* gene by binding to the zeb-1 site.

Results

Molecular Cloning and Characterization of the zeb-1 CDNA

To clone the zeb- 1 CDNA we screened a 0-20 hr embryonic λgt11 cDNA library with a concatenated zeb-1 binding site as described (Vinson et al., 1988). One positive zeb-1 clone was identified from an initial screen of 1.2×10^6 recombinant clones that specifically interacted with the wild-type zeb-1 recognition sequence (Fig. 1a) but not with an unrelated site (Fig. 1b). This cDNA was used to re-screen a 9-12 hr

embryonic cDNA library. Of the several overlapping clones that were obtained in this screen, the one that had the largest insert of 3.1 kb was selected for further analysis. To further authenticate the identity of the zeb-1 CDNA, it was expressed in *E. coli* and bacterially produced zeb-1 was assayed for its ability to interact with its target sites by DNAse footprinting. As shown in Fig. 1c, recombinant zeb-1 specifically recognized its binding sites in the zebra element suggesting that this cDNA indeed encoded a sequence-specific DNA binding protein.

The zeb-1 cDNA Encodes a Steroid Receptor-Like Molecule

Nucleotide sequence determination of the zeb-1 cDNA revealed that it contained a single open reading frame that encodes a polypeptide of 808 amino acids with a predicted molecular size of about 87 kDa (Fig. 2). A search of the protein sequence database revealed two regions of sequence similarity between zeb-1 and members of the steroid receptor superfamily: The punitive DNA binding domain that contains the two Cys2-Cys2 zinc finger motifs, also called the C-region, and second, the more C-terminal putative ligand binding domain of the hormone receptor superfamily, also called the E-region (Krust et al., 1986). A comparison of the predicted zeb-1 protein sequence with representative members of this superfamily showed that within the C-region, it shares a greater homology (76%) with a closely related *Drosophila* protein FTZ-F1 than it does with most others (Fig. 3a). In fact, FTZ-F1 binds to the same binding site in the zebra element, and is also implicated as an activator of *ftz* (Ueda et al., 1990; Lavorgna et al., 1991). Recently, Ohno and Petkovich (1993) have reported the cDNA sequence of FIZ-F1β that shows high homology to FTZ-F1 (referred to as

FTZ-F1α) in the C-region. FTZ-F1β and zeb-1 appear to be the same clones that map at position 39C and show identical patterns of RNA expression in developmental Northern blots (Ohno & Petkovich, 1993) (Fig. 4). However, a significant portion of the 5' ends of the CDNA sequences is completely divergent (see legend to Fig. 2). This divergence in the 5' sequences is most likely due to an alternative splicing event which results in the two *transcripts,* the 'early' and the 'late' transcripts (Fig. 4), that differ only at their 5' non-coding ends but otherwise code for the same protein.

The C-region is well conserved across the nuclear receptor superfamily with 20 invariant amino acid residues. However, based on certain key amino acid residues in the C-region that confer DNA recognition specificity, this superfamily of receptors can be divided into two major subfamilies: The estradio/thyroid subfamily and the glucocorticoid subfamily (Mader et al., 1989; Umesono & Evans, 1989; Danielson et al., 1989). In the former, the recognition amino acids flanked by the second pair of cysteines are CEGC while in the latter CGSC are encountered. Interestingly, both zeb-1 and FTZ-F1 share the same amino acid residues in these key namely positions, CGSC that constitute a composite of the estradiol and glucocorticoid receptor subfamilies and, hence, can be grouped into additional subfamilies (Fig. 3b). Similarly, sequence alignments in the E-region indicate that the three core domains are also fairly well conserved between zeb-1 and most other *Drosophila* receptor-like proteins, and the human estrogen and glucocorticoid receptors (Fig. 3a). Again, while regions I and II in the zeb-1 molecule showed as much as 50% homology with the corresponding regions of FTZ-F1, slightly lower levels were observed between the other receptor molecules. However, a relatively higher degree of conservation is seen

in region III, the smallest of the three core regions, in most cases except in the case *of tailless (tll)*. Considered together, the sequence similarity between zeb-1 and other steroid receptors suggested that zeb-1 might be a ligand-responsive transcription factor.

zeb-1 is a Sequence-Specific Trans-Activator in Tissue Culture Cells

The transcription activity of zeb-1 was determined by transient transfection assays in *Drosophila* tissue culture cells. For this purpose, we constructed a reporter plasmid pBLCATZ4, by cloning four copies of the zeb-1 binding site upstream of the tk promoter in the pBLCAT2 vector. The expression plasmid pPACzeb-1 was constructed by cloning the zeb-1 cDNA in the pPAC vector that contained the *Drosophila* actin 5C promoter and polyadenylation signals. As shown in Fig. 5 no CAT activity was seen when Schneider cells were transfected with pBLCATZ4 alone (lane 1). Further, even in the presence of pPACzeb-1 that contained the full-length zeb-1 cDNA, no CAT activity was observed (lane 2), suggesting that if indeed zeb-1 is a transcription factor, it may be functional only in the presence of its cognate ligand. Since in the present study the transfected cells were not exposed to any exogenously added ligand, it seemed reasonable to infer that zeb-1 may indeed be a ligand-dependent transcription factor. In contrast, the truncated form of the protein that lacked the ligand binding domain (pPACzeb480) was able to *trans*-activate the reporter gene by more than 100-fold (lane 3), suggesting that in the absence of the ligand binding domain, zeb-1 may function as a constitute *trans*-activator. These data

are consistent with similar observations made for other steroid receptors (Godowski et al., 1987).

In order to ascertain that the trans-activation that we observed was due to a specific interaction between the zeb-1 polypeptide and its consensus binding sites, a series of control experiments were conducted. As can be seen from the data in Fig. 5, the parent reporter plasmid alone (pBLCAT2), that lacked the binding sites (lane 7) or in the presence of the zeb-1 polypeptide (lane 5), showed no detectable CAT activity. Similarly, the expression plasmid alone (pPAC), without the zeb-1 cDNA, in the presence of pBLCATZ4, also showed no activity (lane 6). Finally, zeb-1 showed no trans-activation in the presence of unrelated (octamer) binding sites (lane 9). Taken together, these data demonstrated that zeb-1 is a sequence-specific *trans*-activator.

We next wanted to investigate the possibility that zeb480 could *trans*-activate the reporter gene by interacting with the native *ftz* promoter. For this purpose two plasmids were constructed. In the first construct called pBasicCATZE, the entire zebra element of the *ftz* gene was cloned upstream of the CAT gene in the pBasicCAT vector that lacks a functional promoter element. In the second, a truncated fragment of the zebra element that contained only 300 bp from the *ftz* cap site was cloned in the same vector. In this construct called pBasicCATZEΔPst, both the distal and the proximal zeb-1 binding sites were still present. These reporter plasmids were then transfected individually or together with the zeb-1 or zeb480 expression plasmids.

As expected, zeb480 showed good *trans*-activation of the reporter gene both from the pBasicCATZEΔPst construct as well as from the pBasicCATZE by 50 and 100 fold respectively (Fig. 6, lane 3 & Fig. 7, lane 4). In control experiments,

pBasicCAT alone showed no activity (Fig. 6 & 7, lane 1). Similarly, the expression plasmid alone without the zeb-1 cDNA showed little or no *trans*-activation when co-transfected together with pBasicCAT (Fig. 6, lane 4 & Fig. 7, lane 2). We were surprised, however, when zeb-1 showed significant transactivation (25-fold) from the pBasicCATZEΔPst construct (Fig. 6, lane 2) and even higher activation from the pBasicCATZE construct (50-fold, see Fig. 7, lane 3).

A major difference in the two reporter plasmids employed in the present study, aside from the basal promoters themselves, is the presences of zebra promoter sequences and the spacing between the adjacent zeb-1 binding sites. As regards the full-length zeb-1 molecule, the spacing of the cognate binding sites appears to be critical; closely spaced sites may cause steric hinderance between adjacently bound proteins. In contrast, spacing does not appear to affect the activity of the truncated polypeptide. Indeed, a P-element construct that contained four copies of the zeb-1 site in tandem and immediately upstream to the *ftz* basal promoter was also functional, albeit weakly, in transformed embryos (Topol et al., 1991). It is possible that zeb-1 interacts with a component of the basic transcriptional machinery of the *ftz* promoter or with adjacent *ftz* zebra element sequences, and in the case of PBLCAT2 construct that utilizes the tk promoter, such a component might be missing.

Discussion

In this paper we have described the cDNA cloning and characterization of zeb-1 a novel steroid receptor-like molecule from *Drosophila*. Further, in transient transfection

experiments we have been able to demonstrate that zeb-1 is a sequence-specific *trans*-activator of the *ftz* gene. Although we cannot exclude the possibility that a receptor-specific ligand-like molecule exists for zeb-1, the data presented in this paper argue that during early embryonic development when the *ftz* gene is actively transcribed, both full-length zeb-1 and/or the truncated polypeptide zeb480 may positively regulate *ftz* transcription even in the absence of any exogenously added ligand. Indeed, it is noteworthy that such a truncated form of zeb-1 that lacks the ligand-binding domain is also made in the fly embryo (Ayer et al., 1993).

A rather intriguing aspect that emerges from the present study is that if zeb-1 and zeb480 are uniformly distributed in the developing embryo, then how are they functional only in the even-numbered parasegments? That they may be involved in selectively transcribing *ftz* in the seven stripes is gathered from two independent lines of evidence. First, the present data support the fact that they positively regulate *ftz* transcription in tissue culture cells and second, in P-element transformed embryos, a reporter gene that contained four copies of the zeb-1 binding site was selectively transcribed in the even-numbered parasegments. One explanation is that a strong repressor of *ftz* transcription that also binds to the zeb-1 site is present only in the odd-numbered parasegments. Indeed, in an earlier work (Topol et al., 1991) it was shown that the zeb-1 site can also function as a repressor of transcription of a reporter gene in transformed embryos. A good candidate for such a stripe-specific repression is the pair-rule segmentation gene *hairy* (Ingham et al., 1985; Howard & Ingham, 1986; Carroll & Scott, 1986; Isn-Horowicz & Pinchin, 1987; Hiromi & Gehring, 1987; Ingham & Gergen, 1988; Carroll et al., 1988; Carroll & Vavra, 1989). Unfortunately, however,

it is not clear at present how *hairy* acts to mediate repression of the *ftz* gene. Secondly, it is possible that a tissue-specific covalent modification of the zeb-1 polypeptide is necessary for its function. There is precedence for such modifications in the fly embryo. For instance, the *ftz* polypeptide itself is selectively phosphorylated in the embryo (Krause et al., 1987). Thirdly, it is conceivable that an additional component such as a cofactor with which zeb-1 interacts is essential for *trans*-activation and that the distribution of such a putative factor is restricted to the even-numbered parasegments. Finally, it is likely that zeb-1 can be activated by a novel ligand-independent mechanism similar to the one that has been observed in activating certain other receptors such as the chicken COUP transcription factor, and the human estrogen, vitamin D and thyroid hormone receptors (Power et al., 1991). Such mechanisms, however, do not preclude the possibility that zeb-1 becomes ligand-responsive during later stages of development and thereby regulates the transcription of additional target genes.

It is well known that steroid hormones are involved in the transcriptional regulation of development and homeostasis in phylogenetically diverse organisms. Although there is no direct evidence at present that implicates a steroid-like molecule in early embryonic development, the identification of a number of nuclear hormone receptor-like molecules implicated in regulating various aspects of embryonic development in *Drosophila* such as *seven-up (svp*), a gene required in photoreceptor cell precursors during eye development (Modzik et al., 1990); *tll,* a gap segmentation gene (Pignoni et al., 1990); dHNF-4, a gene implicated to have a determinative role in gut formation (Zhong et al., 1993) and *ultraspiracle (usp),* a locus involved in pattern

formation (Oro et al., 1990), suggests that similar clues may also underlie morphogenetic signaling in the fruit fly. Thus, the tissue culture system that we have established to assay for the transcription activity of zeb-1 should provide us with an opportunity to elucidate the mechanism by which zeb-1 functions as a transcription factor. Such studies should provide additional insights into the molecular mechanisms of transcriptional control in *Drosophila* segmentation.

Experimental procedures

cloning and characterization of the zeb-1 cDNA

A 0-20 hr embryonic cDNA library constructed in λgt11 (gift from Dr. T. Sheih, Duke University Medical Center) was screened as described (Vinson et al., 1988) using the oligomerized zeb-1 binding site as probe. The sequence of the top strand for this site is 5'GATCTCTCAAGGTCGCCGAGTAGG3'. Rescreening of the 9-12 hr embryonic cDNA library (gift from Dr. Kai Zinn, Caltech) with the cDNA, isolation and sequencing of cDNAs were done following standard protocols (Sambrook et al., 1989). For footprinting experiments, the zeb-1, protein was over-produced in bacteria using the system of Studier & Moffatt (1986). Crude extracts from the supernatant of lysed bacteria and DNaseI digestion of the 32P-labeled zebra element were done as described earlier (Topol et al., 1991).

Northern blotting

Various developmentally staged embryos were collected at four-hour intervals and total RNA from these embryos as well as from pupae, larvae and adult flies was extracted following the procedure of (Chomczynsid & Sacchi, 1987). Poly(A+) RNA was isolated from each sample by two rounds of oligodt cellulose chromatography. Ten micrograms of poly(A+) RNA was fractionated in an agarose gel containing formaldehyde and transferred to a nitrocellulose membrane from Amersharn. The filter was probed with the nick-translated zeb-1 cDNA, washed and developed as described by the supplier. As a control, the same blot was stripped off the zeb-1 probe and reprobed with the *Drosophila* TFIID probe.

Transient transaction assays in Drosophila *(Schneider S2) tissue culture cells*

The zeb-1 expression vector pPACzeb-1 contains the entire cDNA inserted into the PPAC vector while the reporter gene contains four copies of the zeb-1 binding site cloned at the BGIH site in the pBLCAT2 vector. The pPACzeb480 construct containing the truncated zeb-1 (from amino acid residues 1 to 480) that lacks the putative ligand binding domain was constructed as follows: First, the Bluscript plasmid containing the zeb-1 cDNA was digested with KpnI and SacI and the CDNA insert was gel purified. The purified insert was then digested with PvII and the KpnI/PvuII fragment was recloned into Bluescript at the KpnI and SmaI sites, and excised as a KpnI/SacI fragment that was eventually subcloned into the pPAC vector to generate the pPACzeb480 construct. Transient transfections in Schneider S2 cells were done as described by Krasnow et al. (1989). After transfection, cells were grown for 48 hr at 24^{0}C following which they were harvested. Total cell extract preparation and CAT

assays were done as described (Sambrook et al., 1989). In a control set of experiments, cells were cotransfected with a pPAClacZ plasmid which contained the P-galactosidase coding region cloned in the PPAC vector, as an internal control.

Acknowledgements

We thank Drs. Tao Hseih and Kai Zinn for their generous gifts of the cDNA libraries. KP thanks Drs. Girish Nallur, Pat Sankhavaran, Ajay Bhargava and N. Baskaran for their comments on the manuscript. This work was supported by a grant from the NIH GM42671 to CSP.

References

Akam, A. 1987. The molecular basis for metameric pattern in the *Drosophila* embryo. *Development* **101**: 1-22.

Ayer, S., N. Walker, M. Mosammaparast, L. Nelson, B. Shflo, and C. Benyajati. 1993. Activation and repression of *Drosophila* alcohol dehydrogenase distal transcription by steroid hormone receptor superfamily members binding to a common response element. *NuclAcids Res.* **21**:1619-1627.

Carroll, S.B. and M.P. Scott. 1986. Zygotically active genes that affect the spatial expression of *the fushi tarazu* segmentation gene during early *Drosophila* embryogenesis. *Cell* **45**:113-126.

Carroll, S.B., A. Laughon, and B.S. ThaRey. 1988. Expression, function, and regulation of the hairy segmentation protein in the *Drosophila embryo. Genes Dev.* **2**:883-890.

Carroll, S.B. and S.H. Vavra. 1989. The zygotic control of *Drosophila* pair-rule gene expression. H. Spatial repression by gap and pair-rule products. *Development* **107**:673-683.

Chomczynski, P. and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. *Anal.Biochem.* **162:** 156-159.

Danielson, M., L. Hinck, and G. Ringold. 1989. Two amino acids within the knuckle of the first zinc finger. *Cell* **57**:1131-1138.

Dearolf, C.R., 1. Topol, and C.S. Parker. 1989. Transcriptional control of *Drosophila fushi tarazu* zebra stripe expression. *Genes Dev.* **3**: 384-398.

Duncan, I. 1986. Control of *Bithorax* complex functions by the segmentation gene *fushi tarazu of Drosophila melanogaster. Cell* **47**:297-309.

Godowsid, P.J., S. Rusconi, R. Nfiesfeld, and KR. Yamamoto. 1987. Glucocorticoid receptor mutants that are constitutively activators of transcriptional enhancement. *Nature* **325**:365-368.

Greene, G.L., P. Gilna, M. Waterfield, A. Baker, Y. Host, and J. Shine 1986. Sequence and expression of human estrogen receptor complementary DNA. *Science* **231**:1150-1154.

Hafen, E.A., Kuroiwa, A. and Gehring, W.J. Spatial distribution of transcripts from the segmentation gene *fushi tarazu* during *Drosophila* embryonic development. *Cell* **37**:833-841.

Hiromi, Y. and W.J. Gehring. 1987. Regulation and function of the *Drosophila* segmentation gene *fushi tarazu*. *Cell* **50**:963-974.

Howard, K. and P. Ingham. 1986. Regulatory interactions between the segmentation genes *fushi tarazu, hairy, and engrailed* in the *Drosophila* blastoderm. *Cell* **44**:949-957.

Hollenberg, S.M., C. Weinberger, E.S. Ong, G. CereRi, A. Oro, R. Lebo, E.B. T'hompson, M.G. Rosenfeld, and R.M. Evans. 1985. Primary structure and expression of a functional glucocorticoid receptor cDNA. *Nature* **318**:635-*64* 1.

Ingham, P.W. and A. Martinez-Arias. 1988. The correct activation of *Antennapedia* and *Bithorax* complex genes requires *the fushi tarazu gene*. *Nature* **335**:25-31.

Ingham, P., KR. Howard, and D. Ish-Horowicz. 1985. Transcription pattern of *the Drosophila gene hairy*. *Nature* **318**:439-445.

Ingham P. and P. Gergen. 1988. Interaction between the pair-rule genes runt, *hairy, even-skipped and fushi tarazu* and the establishment of periodic pattern in the *Drosophila embryo*. *Development* (suppl.) **104**:51-60.

Ish-Horowicz, D. and S.M. Pinchin. 1987. Pattern abnormalities induced by ectopic expression of the *Drosophila gene* hairy are associated with repression of *fushi tarazu* transcription. *Cell* **51**:405-415.

Koelle, NLR., W.S.Talbot, W.A. Segraves, M.T. Bender, P. Cherbast and D.S. Hogness. 1991. The *Drosophila EcR* gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. *Cell* **67**:59-77.

Krasnowe M.A, E.E. Saffman, I.C. Komfeld, and D.S. Hogness. 1989. Transcriptional activation and repression by *ubx* proteins in cultured *Drosophila* cells. *Cell* **57**:1031-1043.

Krust, A., S. Green, P. Argos, V. Kumar, P. Walter, J.M. Bomet, and P. Chambon. 1986. The chicken oestrogen receptor sequence: homology with *v-erbA* and the human oestrogen and gluocorticoid receptors. *EMBO J.* **5**:891-897.

Lavorgna, G., I.L. Ueda, J. Clos, and C. Wu. 1991. FTZ-F1, a steroid hormone receptor-like protein implicated in the activation *of fushi tarazu. Science* **252**:848-851.

Mader, S., V. Kumar, I-L de Vemeuil, and P. Chambon.1989. Three amino acids of the oestrogen receptor are essential to its ability to distinguish an oestrogen from a glucocorticoid responsive element. *Nature* **338**:271-275.

Mlodzik., M., Y. Hiromi, U. Weber, C.S. Goodman, and G.M. Rubin. 1990. The *Drosophila seven-up* gene, a member of the steroid receptor gene superfamily, controls photoreceptor cell fate. *Cell* **60**:211-224.

Nusslein-Volhard, C. 1991. Determination of the embryonic axes of *Drosophila. Development* (suppl.) **1**:1-10.

Ohno, C.K. and M. Petkovich. 1993. FTZ-F1β, a novel member of the *Drosophila* nuclear receptor family. *Mech.Dev.* 40:13-24.

Oro, A.E., M. Mckeown, and R. Evans. 1990. Relationship between the product of the *Drosophila ultraspiracle* locus and the vertebrate retinoid X receptor. *Nature* **347**:298-301.

Pignoni, F., R.M. BaldaveRi, E. Steingrimsson, R.J. Diaz, A. Patapoutian, J.R. Merriam, and J.A. Lengyel. 1990. The *Drosophila gene tailless* is expressed at the embryonic and is a member of the steroid receptor superfamily. *Cell* **62**:151-163.

Power, R., S.K. Mani, J. Codina, O.M. Conneely, and B.M. O'Malley. 1991. Dopaminergic and ligand-independent activation of steroid hormone receptors. *Science* 254:1636-1639.

Sambrook, J., E.F. Fritsch and T. Maniatis. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989).

Segaves, W.A. and D.S. Hogness 1990. The E75 ecdysone-inducible gene responsible for the 75B early puff in *Drosophila* encodes two new members of the steroid receptor superfamily. *Genes Dev.* **4**:204-219.

Studier, F.W. and Moffatt, B.A. *J.Mol.Biol.***189**,*113-130* (1986).

Topol, J., C.R. Dearolf, K. Prakash, and C.S. Parker. 1991. Synthetic oligonucleotides recreate *Drosophila fushi tarazu* zebra-stripe expression. *Genes Dev.* **5**:855-867.

Ueda, H., S. Sonoda, J.L. Brown, M.P. Scott, and C. Wu. 1990. A sequence specific DNA-binding protein that activates *fushi tarazu* segmentation gene expression. *Genes Dev.* **4**:624-635.

Umesono, K. and R. Evans. 1989. Determinants of target gene specificity for steroid/thyroid hormone receptors. *Cell* **57**:1139-1146.

Vinson, C.R., LaMarco, K.C., Johnson, P.F., Landshulz, W.I.L. and McKnight, SJ,. 1988. In situ detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. *Genes Dev.* **2**:801-807.

Wakimoto, B.T., F.B.Tumer, and T.C. Kaufman. 1984. Defects in embryogenesis in mutants associated with the *Antennapedia* gene complex of *Drosophila metanogaster. Dev.Biol.* **102**:147-172.

Weir, M.P. and T. Komberg. 1985. Pattems of *engrailed and fushi tarazu* transcripts reveal novel intermediate stages in *Drosophila segmentation. Nature* **318**:433-439.

Zhong, W., F.M. Sladek, and J.E. Damefl. 1993. The expression pattern of *a Drosophila* homolog to the mouse transcription factor BNF-4 suggests a determinative role in gut formation. *EMBO.J.* **12**:537-544.

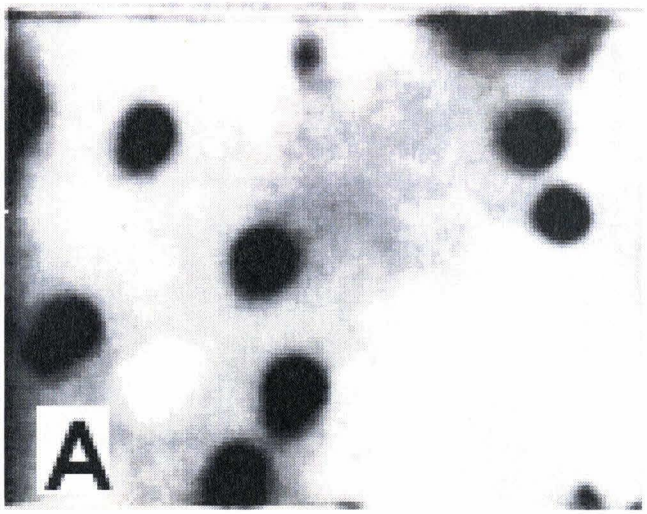

Figure. 1A

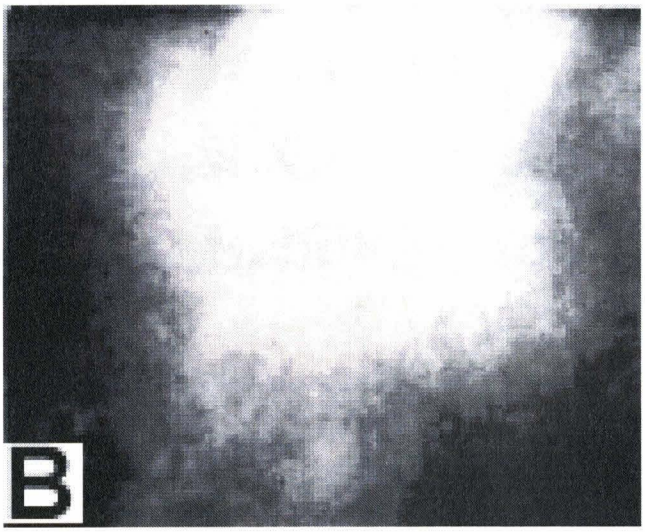

Figure. 1B

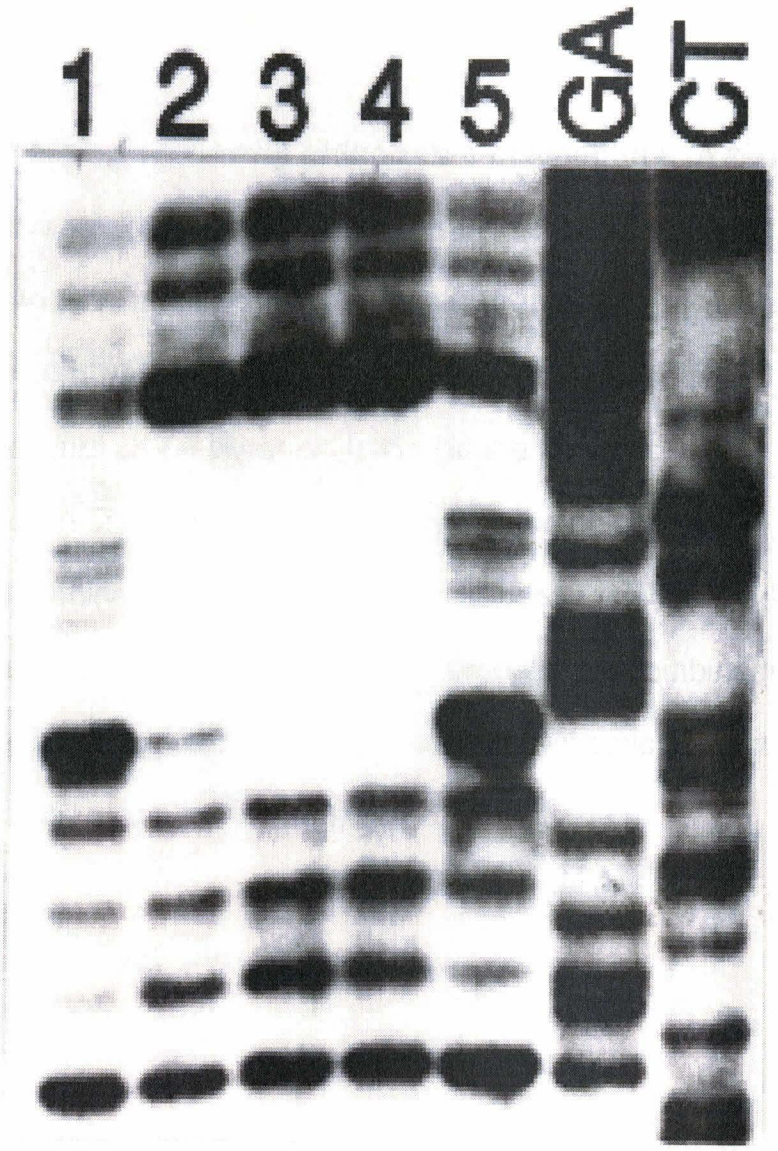

Figure. 1C

Figures and Figure Legends

Figure 1. DNA binding specificity of the zeb-1 polypeptide. A zeb-1 positive clone that was taken through four rounds of plaque purification was plated at low density. The fusion protein was induced with IPTG and transferred to a nitrocellulose filter. The filter was cut into two halves and each half was probed separately with the target probe (A) and a control probe (B) as described in the Methods.

C. DNaseI protection of the *ftz* zebra element by recombinant zeb-1. Lanes 1 & 5 are control samples to which no protein extract was added. Lanes 2-5 received increasing amounts of the extract prepared from bacteria producing recombinant zeb-1. GA and CT indicate sequencing products obtained by chemical sequencing of the corresponding region of the zebra element.

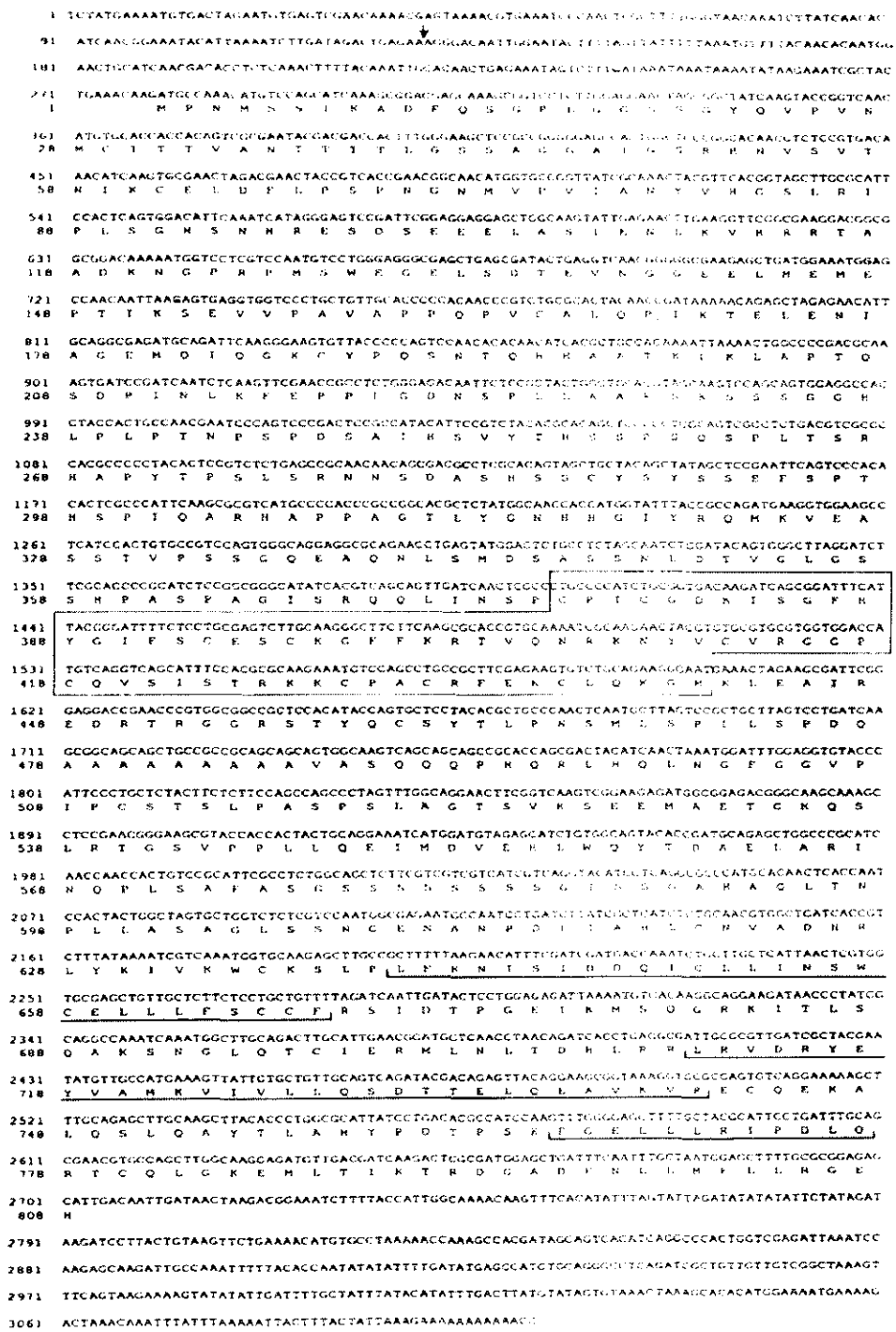

Figure. 2

Figure 2. Nucleotide sequence of the zeb-1 cDNA and the predicted amino acid sequence. The arrow at the 5' end of the nucleotide sequence indicates the end point of sequence divergence between zeb-1 and FTZ-F1β. The C-region is boxed and the three regions that constitute the E-region are underlined.

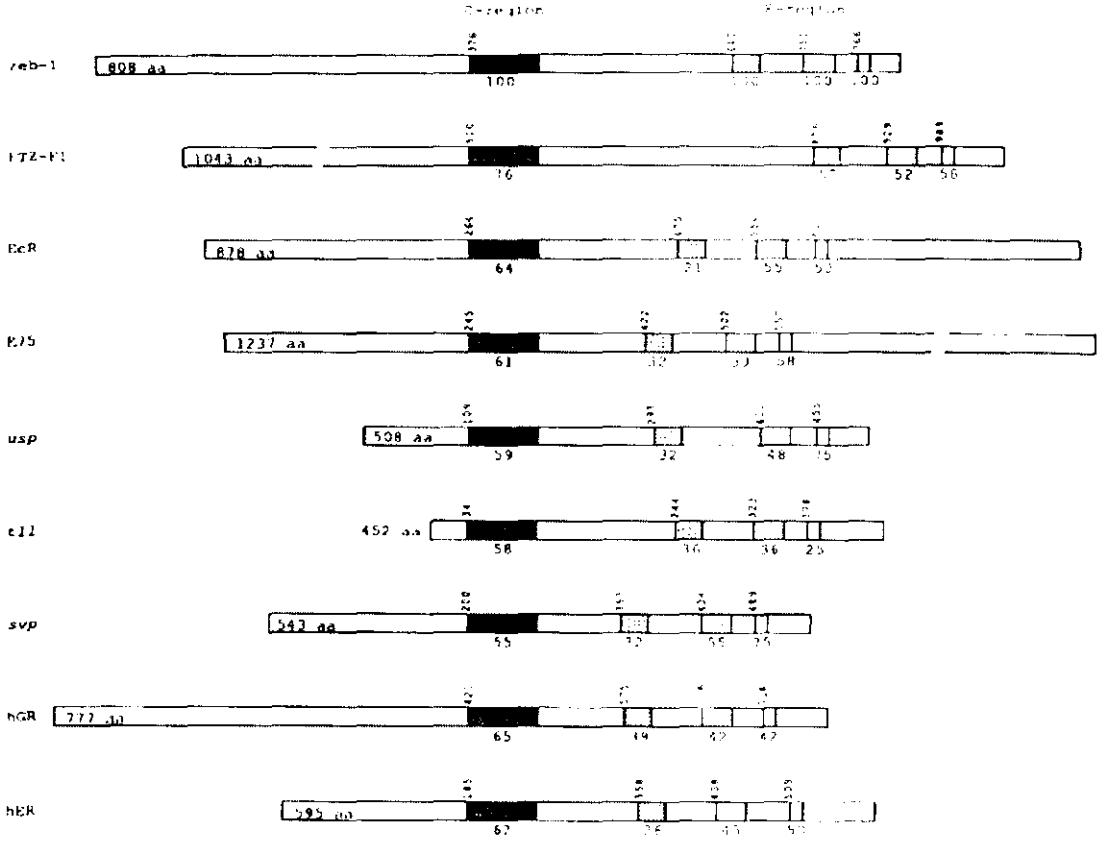
zeb-1
808 aa
100
FTZ-F1
1043 aa
76
52
56
EcR
878 aa
64
55
E75
1237 aa
61
usp
508 aa
59
32
48
35
ecR
452 aa
58
36
36
25
svp
543 aa
55
55
25
hGR
777 aa
65
49
42
hER
595 aa
62

Figure. 3A

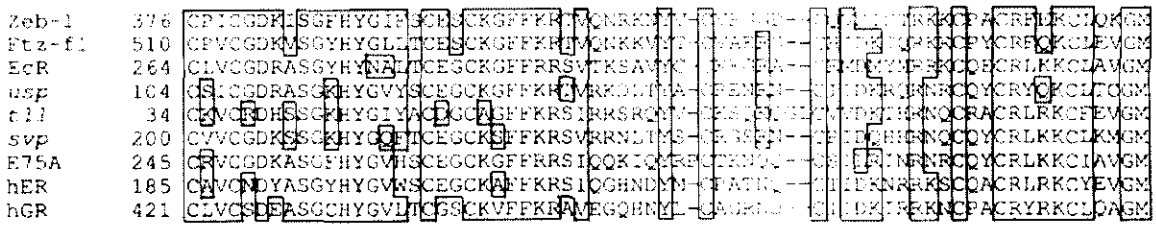

Fig. 3B

Figure 3. A. Sequence comparison of the conserved C- and E-regions of zeb-1 with those of other *Drosophila* receptors and the human estradiol and glucocorticoid receptors. The C-region (dark hatched box) and the E-region comprising the three core segments of the putative ligand-binding domain (lightly hatched boxes) are indicated. The sizes of the individual proteins are indicated as number of amino acids at the respective amino termini. Numbers on the top of each diagram shown vertically represent the amino acid residue at the start of each domain while the numbers below each domain represents the percent homology in each of them. Numbers indicating amino acids are as detailed from the following: FTZ-F1 (Lavorgna et al., 1991), EcR (Koelle et al., 1991), E75 (Segraves & Hogness, 1990), *usp* (Oro et al., 1990), tll (Pignoni et al., 1990), *svp* (Ngodzik et al., 1990), hGR (Hollenberg et al., 1985) and hER(Greene et al., 1986). The nucleotide sequence presented in this paper has been submitted to the GenBank. The accession number is XYZ.

B. Comparison of the amino acid sequences of zeb-1 and other receptors in the C-region. Identical and functionally similar residues are boxed.

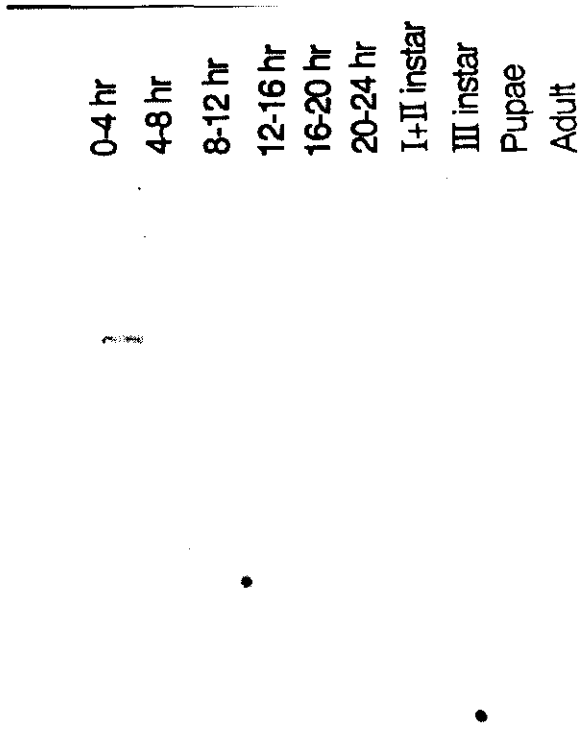

Figure 4. Developmetal Northern blotting experiment. The source of the RNA in each lane is indicated on top of the respective lane.

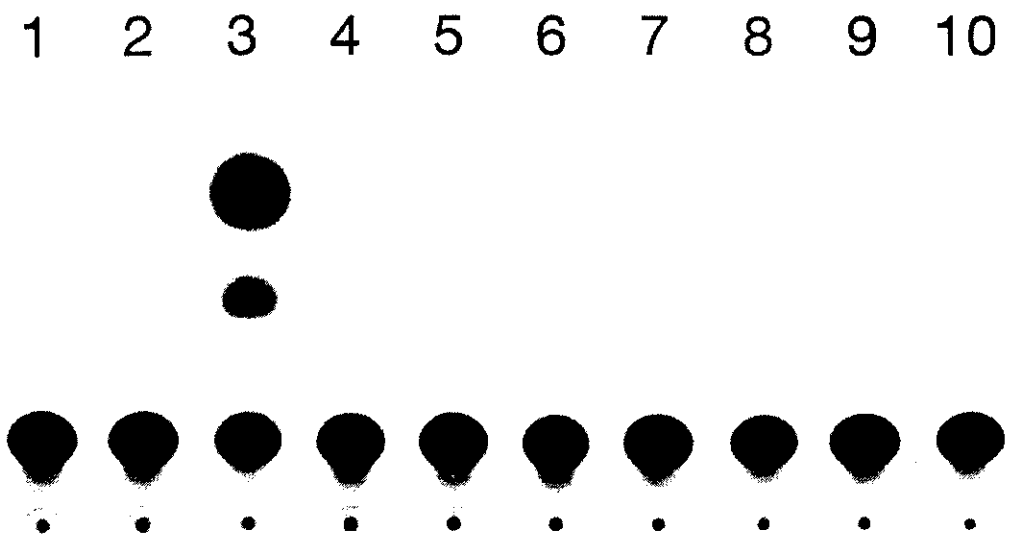

Figure. 5

Figure 5. zeb480 is a sequence-specific trans-activator from a heterologous promoter in *Drosophila* tissue culture cells. Lane 1 shows CAT activity by the reporter plasmid (pBLCATZ4) alone while lane 2 shows CAT activity when Schneider cells were transfected with pBLCATZ4 and pPACzeb-1. As described in the Results, no *trans*activation is observed. In contrast, zeb480 showed about 100-fold trans-activation from the same promoter (lane 3) in the presence of the Z4 sites. In control experiments, zeb480 showed no activation from the parent reporter plasmid (lane 4). Similarly, no activity was observed when the expression plasmid PPAC alone was cotransfected with pBLCATZ4 (lane 5) or together with just pBLCAT2 (lane 6). Likewise, neither the empty reporter plasmid (lane 7) nor the empty expression plasmid together with pBLCAT2 showed no detectable activity (lane 8). Finally, no trans-activation of the reporter gene occurred when the Z4 sites were replaced by unrelated ones (lanes 9 & 10).

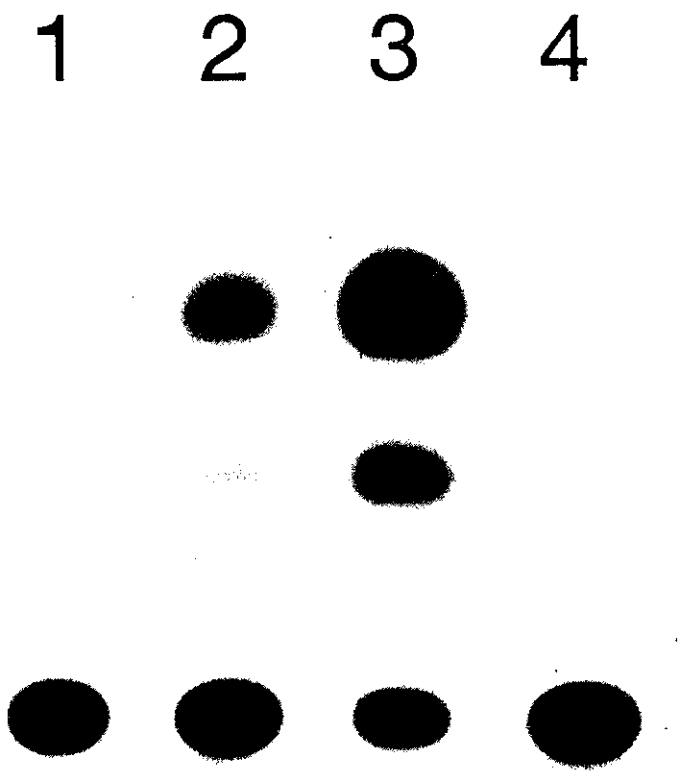

Figure 6. Trans-activation by zeb-1 and zeb480 from the truncated zebra promoter. Lane 1 shows CAT activity with the reporter plasmid pBLCATZEΔPst alone, while lanes 2 and 3 show activation in the presence of zeb-1 and zeb480, respectively. Lane 4 shows the control experiment in which the reporter plasmid is cotransfected with just the pPAC expression plasmid.

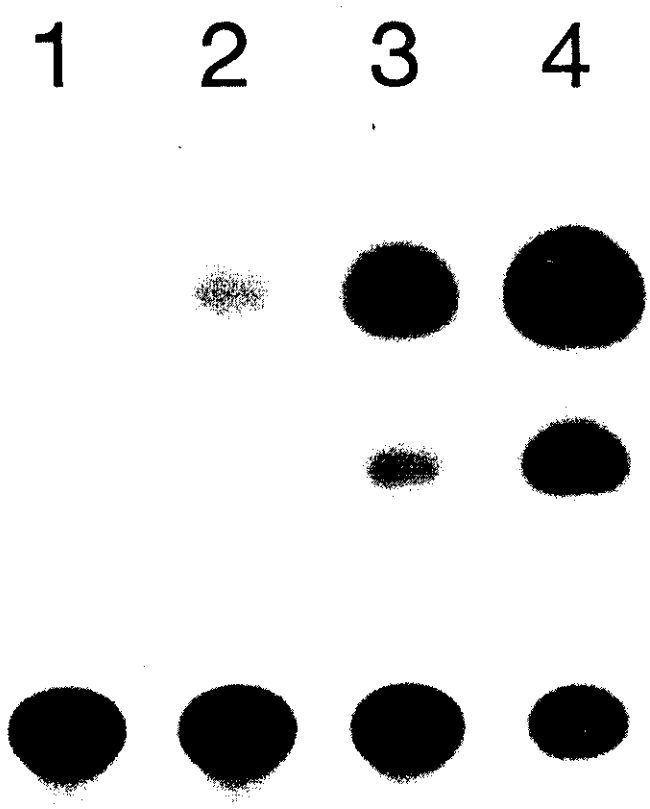

Figure 7. Trans-activation by zeb-1 and zeb480 from the entire zebra promoter. In these experiments, lanes 1 and 2 are the control lanes with only the reporter plasmid alone (lane 1) or together with pPAC (lane 2). Lane 3 shows activity in the presence of the reporter plasmid and the zeb-1 expressing plasmid while lane 4 shows the CAT activity with zeb480 in the presence of the reporter plasmid.

CHAPTER FOUR

cDNA Cloning of An Octamer-Binding Factor From Drosophila That Shows Novel Patterns of Expression During Embryonic Development

The work in this chapter was done in collaboration with Kulkarni Prakash, David Engelberg and Amita Behal.

Kulkani Prakash, Xiangdong Fang, David Engelberg, Amita Behal and Carl S. Parker
dOct2, a Drosophila Oct transcription factor that functions in yeast
Proc. Natl. Acad. Sci. USA Vol. 89, 7080-7084, 1992

Introduction

The establishment of the segmental body pattern in Drosophila requires the coordinated function of a large number of maternally and zygotically expressed segmentation genes. Based on patterns of larval cuticular defects observed in mutant embryos, the segmentation genes are divided into three classes: The gap genes controlling the formation of several body segments, the pair-rule genes affecting pattern formation in pairs of segments and, finally, the segment polarity genes affecting development within each segment (reviewed in Scott & O'Farrell, 1986; Akam, 1987; Nusslein-Volhard et al., 1987; Scott & Carroll, 1987; Ingham, 1988). Several of the segmentation genes have been cloned and found to contain a conserved domain called the homeodomain. This domain contains a helix-trn-helix DNA binding motif, first identified in bacterial repressors. Homeodomain proteins have been shown to bind to specific DNA sequences and function as transcription regulators (reviewed in Scott et al., 1989; Prakash et al., 1991). Highly conserved homeodomain-containing genes have since been isolated from several other species including mammals (for reviews see Manley & Levine, 1985; Holland & Hogan, 1988; Goulding & Gruss, 1989; Wright et al., 1989; Akam, 1989). The spatial and temporal patterns of expression of these genes strongly suggest that these genes may also be involved in segmentation, anteroposterior axis determination and in cell type specification of the mammalian embryo (Holland & Hogan, 1988).

Recently, a new family of transcription factors called the POU family has been identified in mammals and in Drosophila that, like the homeodomain *proteins*, regulates various aspects of cell-fate specification during development. Members of this family share two homologous regions in their DNA binding domain: First, a 60 amino acid homeodomain quite divergent from the classical Drosophila homeodomain, called the POU-homeodomain, distinguished by the presence of tryptophan-phenylalanine-cystine (WFC) motif in the C-terminal domain of the DNA recognition helix. And second, a 68-70 amino acid region that is unique to this class of transcription factors referred to as the POU-specific domain. These two regions together with a non-conserved spacer region between them constitute the POU-domain (Herr et al., 1988). Like the homeodomain-containing proteins, the POU proteins bind to specific DNA sequences and function as transcription regulators (reviewed, Rosenfeld, 1991).

The Oct factors are also members of the POU family of proteins and interact with a specific DNA sequence motif called the octamer. As many as ten different Oct factors have been identified in mammals (Scholer et al., 1989) that show highly restricted patterns of temporal and spatial expression. Thus, for example, while Oct-1, Oct-2 and Oct-6, like all other POU proteins, are expressed widely in the developing nervous system (He et al., 1989), Oct-3 is expressed in undifferentiated embryonal carcinoma cells (stem cells) and in primordial germ cells (Okamoto et al., 1990; Rosner et al., 1990). Oct-4, however, is expressed only in the primordial germ cells (Scholer et al., 1990). The co-expression of different POU proteins in the developing

brain suggests that these transcription factors may, at least in part, regulate the phenotypic maturation of certain neuronal cell types.

In contrast to the vertebrate and mammalian Oct factors, little is known about the Oct factors from Drosophila. Here we report the cloning of a Drosophila Oct (doct) cDNA. Sequence comparisons with known POU transcription factors reveal that dOct is a new member of this family with a highly conserved POU domain that is 90% identical to that of the human Oct factors. dOct is expressed very early in embryogenesis and exhibits a novel spatial pattern that has not been observed for any other Drosophila regulatory gene thus far. These findings suggest that the POU proteins may in fact also participate in pattern formation in Drosophila.

Results

Identification of a Drosophila Oct factor

To identify the presence of a nuclear factor that could specifically interact with the octamer sequence motif, we carried out electrophoretic mobility shift experiments with crude embryonic nuclear extracts and the consensus ('wild-type') octamer oligonucleotide. As shown in Fig. 1a, several protein-DNA complexes were observed when nuclear extracts were incubated with the labeled consensus oligonucleotide (lane 2). When excess of unlabeled competitor oligo was included in the binding reactions together with the labeled probe, all the complexes showed a dose-dependent decrease in binding (lanes 3-7). However, when challenged with a mutant competitor oligo that carried point mutations in two residues within the octamer

consensus, only one complex (indicated by an asterisk) disappeared completely even at the lowest concentration of the competitor used (Fig. 1b, lanes 3-5). The other complexes remained relatively unaffected although there seemed to be a general decrease in the band intensities. Similar results were obtained when totally unrelated oligos of similar sizes were used as competitors (Fig. 1c, lanes 3 & 4). We believe that this complex is due to the presence of a non-specific DNA binding protein not related to the Oct factors. Taken together, these data demonstrate the presence of multiple nuclear factors that can specifically bind to the consensus octamer sequence motif. A DNase I footprint analysis also revealed the presence of protein factors from crude embryonic nuclear extracts that can specifically bind to the octamer motif (data not shown).

cDNA cloning of the dOct gene

Having confirmed the presence of nuclear factors that interact with the octamer motif, we screened a Drosophila 0-20 hr embryonic λgt11 CDNA library (a gift from Dr. Tao Hseih, Duke University Medical Center) using the wild-type octamer oligonucleotide as described in the Methods. In the expression screens of $6x10^5$ recombinant plaques, two positive clones λOct A and λOct B were identified that interacted with the wild-type octamer recognition site but not with the non-specific binding sites. The cDNA inserts contained in the two phage clones were identical at their 5' ends but λOct B was longer at the 3' end. Rescreening the same library with the nick-translated λOct A cDNA probe yielded another overlapping clone, λOct C. This clone was identical to λOct A at the 3' end but was longer at the 5' end. Sequence analysis of the 2125

nucleotides covered by the three overlapping cDNA clones revealed a single continuous open reading frame beginning at position 194 and ending at position 1345 (Fig. 2a). Several additional methionine residues are also present within this frame. The dOct protein predicted from this open reading frame has 384 amino acids with a molecular weight of 42.3 kDa containing a POU domain that lies in the C-terminal half of the molecule and terminates 21 residues from the carboxyl terminus. *In vitro* translation of the dOct mRNA obtained by *in vitro* transcription of the λOct C cDNA with T7 polymerase resulted in a major translation product of 46 kDa, in good agreement with the predicted molecular weight (Fig. 2b). In addition, one minor species of higher molecular weight (about 60 kDa) and several minor ones of lower molecular weights are also observed (Fig. 2b). While the faster migrating forms could be due to initiation at internal methionine residues, the slower migrating species is likely to be due to post-translational modification of the dOct protein in the reticulocyte lysate system employed for translation. No translation products were observed in an unprogrammed lysate or when the lysate was programmed with the anti-sense dOct RNA obtained by transcription with T3 polymerase (not shown).

The ability of the dOct protein to bind to the octamer sequence was verified by mobility shift experiments. dOct protein synthesized *in vitro* was allowed to bind to the ^{32}P-labeled wild-type octamer oligo probe and the complexes separated by gel electrophoresis. As shown in Fig. 2c, specific protein-DNA complexes were observed with the wild-type oligo (lane 2). These complexes could be competing with excess of wild-type unlabeled oligo (lanes 3-5). No protein-DNA complexes were observed

when the unprogrammed lysate was incubated with the labeled probe (lane 1). These data provided strong evidence that this cDNA indeed encodes an Oct protein.

Temporally and spatially restricted patterns of doct expression

We used the dOct cDNA clone to analyze the temporal and spatial patterns of expression of the dOct transcripts during embryogenesis. In Northern blotting experiments with mRNA from different embryonic and larval stages of development and from adult flies, the dOct cDNA hybridized strongly to a 3.2 kb transcript (Fig. 3). A low level of this transcript was detected in 0-4 hours of embryogenesis and maximal levels were present in 4-8 hours of development. Relatively high levels of dOct expression persisted until 16 hours of embryogenesis following which the levels fell significantly until they were no longer detectable during the first and the second instar larval stages. Low levels of expression reappeared during the third instar larval and pupal stages but no expression was detectable in adult flies. Weaker reacting species of transcripts that also show interesting patterns of embryonic expression are apparent in the RNA gel blot (Fig. 3). These bands may be due to cross-hybridization between dOct and closely related members of the POU family. It is also possible that they could represent differentially processed transcripts from the same gene. As a control, the blot was stripped and re-probed with the Drosophila TFIID cDNA that has been reported to be uniformly expressed during development (Hoey et al., 1990). A single transcript of 1.6 kb was detected that was almost constant in all the lanes as expected (data not shown).

The spatial distribution of dOct RNA was examined in whole embryos by *in-situ* hybridization. For this purpose, a digoxigenin (DIG)-labeled probe was prepared by PCR amplification of a portion of the 3' untranslated region of the dOct cDNA as described in the Methods. Transcripts from this gene were detectable as early as the cellular blastoderm stage in a strong, uniform, wide band in the posterior part of the embryo between 20 to 40% of the egg length (EL) (Fig. 4a). This expression is transient and within minutes after its appearance as a uniform band, dOct RNA resolves progressively into a series of three narrower stripes that reside between 20 and 40% of the EL (Fig. 4b & c; shown at higher magnification in Fig.5a-e). The three-stripe pattern persists as late as the onset of gastrulation and then disappears (Fig. 4d). At the elongated germ band stage (about 4.5 hr of development), dOct RNA is expressed in fourteen sharp stripes, one in each parasegment (Fig. 4e & f). Still later in embryogenesis the pattern of dOct RNA distribution undergoes significant changes; dOct becomes strongly expressed in a segmentally repeated pattern of cells in the ventral nervous system, and in most cells of the developing brain. By the germ band retracted stage, dOct expression is most prominent in cells of the neural tissue including a cluster of the muscle precursor cells or the peripheral sensory cells (Fig. 4g-j).

At the cellular blastoderm stage, the pair-rule segmentation gene *fushi tarazu (ftz)* is expressed in seven evenly spaced stripes, one in each even-numbered parasegment (Hafen et al., 1984). To determine if the dOct stripes are in register with the parasegments, whole embryos were also hybridized *in-situ* with either the DIG-labeled, single-stranded *ftz* cDNA probe alone (Fig. 6a) or together with the dOct

probe (Fig. 6b-h). The results of these experiments indicated that indeed, at the cellular blastoderm stage, two of the three dOct stripes, stripes 1 and 2, coincided with the *ftz* stripes 4 and 5 (Fig. 6d). Therefore, the dOct stripes 1 and 2 correspond to the even-numbered parasegments 8 and 10. The third dOct stripe, however, was only partially overlapping *ftz* stripe 6. Thus, dOct stripe 3 appears to overlap parasegments 11 and 12 (Fig. 6d). During early gastrulation, no additional stripes of dOct expression are seen, while the seven *ftz* stripes persist (Fig. 6e). However, by the time the germ band is extended, seven additional stripes are apparent in addition to the normal seven stripes of *ftz* which are due to the expression of dOct in fourteen stripes (Fig. 6f). Later in embryogenesis, *ftz* and dOct are transiently expressed in a specific subset of neuronal precursor cells (Fig. 6g & h); in addition, dOct expression is also observed in the peripheral sensory cells (Fig. 6h).

dOct is a novel member of the POU family

A search of the protein database revealed high homology between dOct and the human and vertebrate Oct proteins. High homology was also observed between dOct and other vertebrate and the two Drosophila POU proteins, cf1a and I-POU, described recently (Johnson & Hirsh, 1990; Treacy et al., 1991). Based on the extent of sequence homology within the POU domain, Rosenfeld and co-workers (He et al., 1989; Treacy et al., 1991) have proposed a classification of the POU proteins into five different classes. dOct falls in the Class II type of POU proteins together with other Oct-1 and Oct-2 genes from vertebrate and human species (Fig. 7a & b). While dOct shares between 87-90% identity with other members of this class over the entire POU

domain (Fig. 7c), it shows 91-93% identity in the POU-specific domain and between 81-86% in the POU-homeodomain among these members. Interestingly, dOct shows significantly less sequence identity over the POU domain to the mouse Oct-6 (76%) and even less so with mouse Oct-3 and Oct-4 (only 63%). Among all the POU sequences, within the POU-specific domain, dOct is least homologous to rn-3, I-POU and unc-86 (only 52-54%) which by themselves fall in a distinct class and are thus by far the most divergent of the POU domains. Within the A and B sub-regions of the POU-specific domain, dOct is highly homologous with Oct-1 and Oct-2 (96% and 91% identity, respectively) but showed low homology to Brn-3, I-POU, and unc-86 (only 73-77%).

Certain amino acid residues within the POU domain at which dOct differs from all other members in its class are noteworthy. For instance, within the A sub-region there is only one change in the dOct protein from a G to a C residue at position 24 whereas within the B sub-region there are three changes from a K to a N residue at position 16, from an E to a Q residue at position 26, and from an E to a D residue at position 33 (Fig. 7a). Thus, dOct is a novel member of the Class II type of POU genes and the sequence divergence at these amino acid residues may indeed reflect novel functions of this protein.

As with the POU domain, sequence alignments of the POU-homeodomain revealed a similar trend in the extent of homology between dOct and all other POU proteins with maximal homology with the class H proteins and minimal homology with the class V proteins (Fig. 7b). On the other hand, a significantly higher degree of sequence homology was observed between dOct and all other classes in the WFC

region. The WFC region, lying in the carboxyl terminal third of the homeodomain, comprises helix 3, the DNA recognition helix, and therefore, it is not surprising that there is 100% sequence identity between dOct and members of Classes I, II and III and more than 80% between dOct and members of Classes IV and V in this region. dOct differs from the other members of its class at five residues over the entire POU-homeodomain. Only one change, from an I to a V residue, occurs at position 4 in helix 3. This is a functional replacement implying severe constraints on the evolution of this DNA binding motif. The four other changes are seen in helix 1 and helix 2. Of these, a T residue at position 2 and an A at position 10 of helix 1 are unique to the dOct protein. In contrast, the presence of a G residue at position 5 of helix 1 is common to the Class III members and the occurrence of a Q residue at position 6 of helix 2 is encountered in at least one other POU protein namely, that of the C. elegans protein, ceh-6 (Fig. 7b).

The POU- and homeo-domains in the dOct protein are separated by a spacer of twenty-six amino acids, equal to that found in the human Oct-2 protein. However, no homology is observed between the two proteins in this linker region. Upstream of the POU domain, dOct is particularly rich in Q (17%), S (13%), A (12%) and P (9%) residues. The preponderance of these particular amino acids together with reiterated stretches of A and Q residues is a common feature shared by several Drosophila and other eukaryotic transcription factors (Mitchell & Tjian, 1989; Muhich et al., 1990, and cf therein).

Discussion

In this paper, employing a functional screening technique, we have described the identification and cDNA cloning of an Oct gene from Drosophila for the first time. This gene is a novel member of the class II type of POU genes that comprises all of the known Oct genes. The remarkable evolutionary conservation and extensive sequence homology in the POU domain with previously known genes that are involved in cell-fate specification suggest that the dOct gene is likely to be a component of one of the gene regulatory networks that operates during Drosophila development. Our most important observation is that the spatial patterns of dOct expression transcend the domains of expression of the classical Drosophila segmentation genes. To date we are not aware of any other regulatory gene in Drosophila whose spatial expression patterns cover the entire spectrum of expression patterns that are seen for each class of segmentation genes, i.e., the gap, pair-rule and segment polarity genes. Thus, while the initial domain of dOct RNA expression in a broad uniform band is surprisingly similar (i.e., covering several segments), but not identical, to some of the gap gene expression domains [for example, *Kr* and *knirps* (for reviews see, Gaul & Jackle, 1987, 1990)], the striped expression is comparable to a subset of stripes of the pair-rule class [for example, *hairy, even-skipped* and *ftz* (for a recent review see Carroll, 1990)]. Similarly, whereas the expression of dOct in fourteen narrower stripes is clearly reminiscent of the segment-polarity genes [such as *engrailed* (Komberg et al., 1985)], its expression in the central nervous system parallels those of several of the gap,

segmentation and homeotic genes (Dinardo et al., 1985; Knipple et al., 1985; Bopp et al., 1986; Baker, 1987; Gaul et al., 1987; Frasch et al., 1987; Doe et al., 1988).

What are the developmental functions of the dOct gene? The mere detection of the spatial patterns of transcript localization does not necessarily mean that the resulting protein is functional in these domains since we have not yet defined the dOct protein expression pattern and regulation could also occur at the post-transcriptional level. For instance, in the case of the *Kr* gene, it has been observed that there are regions in the embryo that accumulate high amounts of *Kr* RNA, but little or no *Kr* protein (Gaul et al., 1987). Similarly, in the case of the *caudal* gene, Mlodzik and Gehring (1987) have observed that while the *caudal* RNA is evenly distributed in the cytoplasm up to the 12th nuclear division in syncytial embryos, the *cad* protein is localized mainly in the nuclei to form a gradient along the anteroposterior axis as early as the 9th nuclear division. Nonetheless, our data provide suggestive evidence that the dOct gene is likely to be involved in the spatial programming of gene expression during early embryonic development.

Although the spatial pattern of dOct expression is comparable to some of the gap genes, the timing of its expression suggest that it may function subordinate to these genes in the genetic hierarchy. The subsequent striped expression that still covers only 20-40% EL is intriguing. Given that the dOct stripes overlap those of *ftz* in those parasegments that correspond to dOct expression, it is possible that dOct may function as a *trans*-regulator of *ftz*. Although no octamer-like DNA sequences are obvious in the 'zebra element' that is sufficient to drive the normal seven-stripe expression of *ftz* (Hiromi et al., 1985; Hiromi & Gehring, 1987), it is likely that dOct can

interact with novel cis-sequences (such as the homeodomain binding sites, see below) or through regions of the *ftz* gene other than the zebra element. Two other homeodomain-containing proteins are known to bind to multiple sites in the *ftz* zebra element; the *eve* protein that recognizes the typical TAA rich homeodomain site (J. Topol & C.R. Dearolf, unpublished), and another, the *caudal (cad)* protein, that recognizes a novel DNA sequence (Dearolf et al., 1989a). While *eve* is known to function as a repressor of *ftz* (Carroll & Scott, 1986), *cad* is known to positively regulate the posterior-specific stripes of *ftz* (Dearolf et al., 1989a). The presence of multiple homeodomain binding sites in the zebra element further strengthen the argument that dOct, like *cad*, may indeed positively regulate the more posterior *ftz* stripes 4, 5 and 6. The expression of dOct transcripts in every segment is reminiscent of that observed for the segment-polarity genes and points to another possible function for this gene. dOct could function as a position-specific regulator of these genes by either activating or repressing their expression. The temporal patterns of dOct expression are entirely consistent with such a putative function. Finally, the neuronal expression of dOct is not surprising given that most POU genes (He et al., 1989) including the Drosophila cf Ia and I-POU genes (Treacy et al., 1991), and several homeotic and segmentation genes, are all expressed in the nervous system. However, the precise functions of this gene could be defined by characterizing mutations at this locus and by analyzing the patterns of expression of the dOct protein. The availability of the cDNA clone should facilitate these experiments.

The Oct factors, like the homeoproteins (see Scott et al., 1989), at least *in vitro* (Baumruker et al., 1988) and in tissue culture systems (Thali et al., 1988), have been

shown to bind with varying affinities to several DNA recognition sites that deviate widely from the cannonical octamer sequence. Furthermore, Oct-1 has been shown to function differentially depending on the promoter context by complexing with another trans-activator (Gerster & Roeder, 1988; Stem et al., 1989). Such a remarkable degree of flexibility in DNA recognition and trans-activation may reflect the multitude of functions performed by these transcription factors in their natural cellular contexts. Coordinated interactions between the components of the basic transcription apparatus and the more selective, cell- and tissue-specific ones such as the homeoproteins and POU proteins, may establish critical codes for programming cell fates during development. We suspect that the co-expression of an Oct (POU) gene with several of the classical pattern-forming genes in *Drosophila* may also contribute to a combinatorial code critical for the establishment of the segmental body plan. Future studies on the regulation of expression of dOct and the identification of the genes that interact with it will provide further insights into the molecular mechanisms of pattern formation in *Drosophila*.

Experimental procedures

Electrophoretic mobility shift assays

These assays were done following the procedure described by Calzone et al. (1988). Briefly, 1 ug of synthetic oligonucleotide that contained the wild-type octamer binding site 5' GATCCFTAATAATTTGCATACCCTCA 3' and its complement (Staudt et al., 1988) were labeled separately at the 5' end with [^{32}Pγ]-dATP and T4 polynucleotide kinase. The phosphorylated oligos were then mixed, heated to 55^{0}C and allowed to anneal at RT for 1 hr. The unincorporated label was removed by chromatography over a NACS prepac column (Bethesda Research Laboratories), and the labeled oligo was precipitated with ethanol. After drying the pellet it was dissolved in 100 ul of TE and the radioactivity was measured using a scintillation counter. Each 10 ul binding reaction for the mobility shift assay contained 2 ul of 5X mix (50 mM HEPES pH, 7.9/2.5 mM DTT/330 mM KCV25 mM MgCl2/50 ug/ml of TRNA), 1 ul of dI-dc (100 ug/ml final), 2 ul of gycerol/NP-40 (10% gycerol/0.1% PN-40) and 1 ul (20,000 cpm). The final volume was made up by adding a suitably diluted aliquot of extract or an equal volume of buffer C (20 mM HEPES, pH 7.9/40 mM KCVO. 1 mM EDTA/1 mM DTT/10% glycerol). After incubation on ice for 5 min, 1 ul of sample buffer containing 15% Ficoll, 0.25% bromophenol blue and 0.25% xylene cyanol was added to each reaction and the complexes were separated from unbound probe by electrophoresis in 5% acrylamide in 0.5X TBE at 200 V for 2 hr at RT. Gels were pre-run for 2 hr before use. After electrophoresis, gels were dried and autoradiographed overnight at -70^{0C} to locate the radioactive complexes and the unbound probe. The mutant oligo was prepared by creating two point mutations in the wild-type sequence. The sequence of the mutant site is 5' GATCCTTAATAATCGGCATACCCTCA 3'. Crude embryonic nuclear extracts were prepared from 0-12 hr embryos as described earlier (Dearolf et

al., 1989b). When the recombinant dOct protein was used in the mobility shift assays, reticulocyte lysates were programmed with the dOct RNA (or the anti-sense RNA) and unlabeled D-methionine. One ul (out of 50 ul) of the translation mixture was used in each 10 ul binding reaction.

Library screening and CDNA cloning

A 0-20 hr embryonic λgt11 library was probed with the wild-type octamer oligonucleotide following the procedure of Vinson et al. (1988). The probe for the library screening was constructed by cloning four copies of the oligonucleotide into the BamHI site of the Bluescript SK+ polylinker and ^{32}p-labeling the 200 hp KpnI-SacI fragment by nick-translation. Two positive clones, λOct A and λOct B, were identified that were identical at their 5' ends but different at their 3' ends. A third overlapping clone, λOct C, was obtained by rescreening the same library with the λOct A nick-translated cDNA as a probe following standard procedures (Sambrook et al., 1989). The three cDNAs were subcloned into the Bluescript SK+ plasmid vector at the EcoRI site and the entire sequence of 2125 nucleotides was determined by double-stranded sequencing using a Sequenase Kit (United States Biochemicals) and synthetic oligonucleotides as primers.

In vitro transcription-translation

For cell-free transcription-translation experiments, the Bluescript plasmid containing the λOct C cDNA was linearized at the 3' end with BamHI and transcribed with T7 polymerase using a Stratagene Kit. After transcription, the DNA template was digested with DNase and the RNA was recovered following phenol/chloroforin extraction and ethanol precipitation. The final pellet was dissolved in 25 ul of DEPC-

treated water and 1 ul of the transcribed RNA was translated in reticulocyte lysate system (Amersham) using ^{35}S-methionine in a total of 50 ul as described by the manufacturer. Five ul of the translation Mixture was applied to a 10% SDS-acrylamide gel (Laemmlli, 1970) and electrophoresed at 30 mAmps for 4 hr. After electrophoresis the gel was fixed, treated with Amplify (Amersham), dried and autoradiogahped at -70^{0}C. ^{14}C-methylated molecular weight markers (from Amersham) were used to estimate the sizes of the translated products.

Northern blotting

Total RNA from staged embryos, larvae and adult flies was isolated by the guanidine thiocynate procedure described by (Chomczynski & Sacchi, 1987). Poly(A+) RNA was prepared by oligo(dt) cellulose chromatography. Ten ug of each mRNA sample was fractionated on a 1.5% formaldehyde-agarose gel, blotted onto a GeneScreen membrane (Amersham) in 25 mM sodium phosphate buffer, pH 6.5. The blot was prehybridized in 50% formamide, 5X SSC, 5X Denhradts and 100 ug/ml of salmon sperm DNA at 42^{0}C for 16 hr. Hybridization with 32P-labeled dOct cDNA (1x10^6 cpm/ml) was allowed for 24 hr. After hybridization, the blot was washed four times with 2X SSC containing 0.1% SDS at RT followed by 0.1X SSC containing 0.1% SDS for 3 hr at 67^{0C}. The blot was then dried and autoradiographed at -70^{0C} for three days to visualize the transcripts.

In-situ hybridizations

Embryos were collected and aged at RT, devitelinized, fixed and stored in 100% ethanol at -20^{0C} until used. DIG-labeled single-stranded DNA probes used in the present studies were prepared by the PCR method. Briefly, recombinant plasmid

DNAs were digested with the appropriate restriction enzyme and the linearized DNA fragments were gel purified. 500 ng of the linear DNA fragments were taken in a *final* volume of 25 ul in a 0.5 ml eppendorf tube that contained, in addition, 2.5 ul of the I0X Reaction buffer (Boehringer Mannheim Biochemicals), 5 ul of 5X dNTP mix (Boehringer Mannheim Biochemicals) and 5 ul of the SK/KS primer. 40 ul of mineral oil was added to each tube. The contents were mixed, briefly centrifuged and boiled for 5 min. Two ul of Taq I DNA polymerase (1.25 units) was added to this tube. After mixing the contents, the tubes were centrifuged briefly and placed in a Perkin-Elmer Cetus PCR thermal cycler. The DNA was amplified for 30 cycles under the following conditions: 95^{0C} for 45 seconds, 55^{0C} for 30 seconds and 72^{0C} for I min. After the PCR reaction, 75 ul of water was added to each tube and the phases were separated by centrifugation. The aqueous layer (90-95 ul) was aspirated carefully from beneath the oil and precipitated twice with ethanol. After washing once with 70% ethanol, the pellets were dried in a speed vacuum and were re-suspended in 300 ul of the hybridization buffer. Before use the probe was boiled for 1 hr. Hybridization, washing and detection were carried out following the procedure of Tautz and Pfeifle (1989) with minor modifications. In the case of dOct the recombinant Bluescript plasmid DNA containing the λOct A cDNA with HindIH. The larger fragment that contained 267 bp of the 3' untranslated portion of the cDNA was purified and used in the PCR reaction. For the *ftz* probe the pFI plasmid (gift from Mike Muhich) that contained a 600 bp 5' portion of theftz gene, was linearized with BamHI.

Acknowledgements

We are grateful to Dr. Tao Hsieh for the embryonic lambda gt11 cDNA library, to Dr Mike Muhich for the *ftz* and Drosophila TFIID clones, to Dr. Judith Lengyel for the PCR protocol and to Drs. Judith Lengyel and Walter Gehring for their suggestions and protocols regarding the whole mount *in-situ* hybridizations. We thank Drs. Howard Lipshitz and Kai Zinn for their interest in this work and for many helpful discussions, and for their thoughtful comments on the manuscript.

References

Akam, M. (1987). The molecular basis for metameric pattern in the Drosophila embryo. *Development* **101**, 1-22.

Akam, M. (1989). HOX and HOM: homologous gene clusters in insects and vertebrates. *Cell* **57,** 347-349.

Baker, N.E. (1987). Molecular cloning of sequence from *wingless,* a segment-polarity gene in Drosophila: the spatial localization of a transcript in embryos. *EMBO J.* **6,** 1765-1773.

Baumruker, T., Sturm, R. and Herr, W. (1988). OBP100 binds remarkably degenerate octamer motifs through specific interactions with flanking sequences. *Genes Dev.* **2**, 1400-1413.

Bodner, M., Castrillo, L.T., Theill, T., Deemick, M., Ellisman, M. and Karin, M. (1988). The pituitary transcription factor GHF-I is a homeobox-containing protein. *Cell* **55,** 505-518.

Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. and Noll, M. (1986). Conservation of a large protein domain in the segmentation gene *paired* and in functionally related genes of Drosophila. *Cell* 47, 1033-.1045

Buerglin, T.R., Finney, M., Coulson, A. and Ruvkun, G. (1989). *C. elegans* has scores of homeobox-containing genes. *Nature* **341,** 239-243.

Calzone, F.J., Theze, N., Thieband, P., Hill, R.C., Britten, R.J. and Davidson, E.H. (1988). Developmental appearance of factors that bind specifically to cis-regulatory sequences of a gene expressed in the sea urchin embryo. *Genes Dev.* **2,** 1074-1088.

Campos-Ortega, J.A. and Hartenstein, V. (1985). The embryonic development of Drosophila melanogaster. Springer-Verlag, Berlin.

Carroll, S.B. (1990). Zebra patterns in fly embryos: Activation of stripes or repression of interstripes? *Cell* **60,** 9-16.

Carroll, S.B. and Scott, M.P. (1986). Zygotically active genes that affect the spatial expression of *fushi tarazu* segmentation gene during early Drosophila embryogenesis. *Cell* **45,** 113-126.

Chomczynski, P. and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chlroforrn extraction. *Anal. Biochem.* **162,** 156-159.

Clerc, R.G., Corcoran, L.M., LeBowitz, J.H., Baltimore, D. and Sharp, P.A. (1988). The B-cell-specific Oct-2 protein contains POU box- and homeo box type- domians. *Genes Dev.* **2,** 1570-1581.

Dearolf, C.R., Topol, J. and Parker, C.S. (1989a). The *caudal* gene product is a direct activator of *fushi tarazu* transcription during Drosophila embryogenesis. *Nature* **341**, 340-343.

Dearolf, C.R., Topol, J. and Parker, C.S. (1989b). Transcriptional control of Drosphila *fushi tarazu* zebra stripe expression. *Genes Dev.* **3**, 384-398.

DiNardo, S., Kuner, J.M., Theis, J. and O'Farrell, P. (1985). Development of embryonic pattern in Drosophila melanogastor as revealed by accumulation of nuclear *engrailed* protein. *Cell* **43**, 59-69

Doe, C.Q., Hiromi, Y., Gehring, W.J. and Goodman, C.S. (1988). Expression and function of the segmentation gene *fushi tarazu* during Drosophila neurogenesis. *Science* **239**,170-175.

Finney, M., Ruvkun, G. & Horvitz, H.R. (1988). The C. elegans cell-lineage and differentiation gene unc-86 encodes a protein with a homeodomain and extended sequence similarity to mammalian transcription factors. *Cell* **55,** 787-789.

Frasch, M., Hoey, T., Rushlow, C., Doyle, H. and Levine, M. (i987). Characterization and localization of *even-skipped* protein of Drosophila. *EMBO J.* **6**, 749-759.

Gaul, U. and Jackle, H. (1987). How to fill a gap in the Drosophila embryo? *Trends Genet.* **3**, 127-131.

Gaul, U. and Jackle, H. (1990). Role of gap genes in early Drosophila development *Advances Gent.* **27**, 239-275.

Gaul, U., Seifert, E., Schuh, R. and Jackle, H. (1987). Analysis of *Kruppel* protein distribution during early Drosophila development reveals posttranslational regulation. *Cell* **50**, 639-647.

Gerster, T. and Roeder, R.G. (1988). A herpes virus trans-activating protein interacts with transcription factor OTF-1 and other cellular proteins. *Proc. Natl. Acad. Sci.(USA)* **85**, 6347-6351.

Goulding, M.D. and Gruss, P. (1989). The homeobox in vertebrate development. *Current Opinion Cell Biol.* **1**, 1088-1093.

Hafen, E., Kuroiwa, A. and Gehring, W.J. (1984). Spatial distribution of transcripts from the segmentation gene *fushi tarazu* during Drosphila embryonic development. *Cell* **37**, 833-841.

He, X., Treacy, M.N., Simmons, D.M., Ingraham, H.A., Swanson, L.W. and Rosenfeld, M.G. (1989). Expression of a large family of POU-domain regulatory genes in mammalian brain development. *Nature* **340**, 35-42.

Herr, W., Sturm, R.A., Clerc, R.G., Corcoran, L.M., Baltimore, D., Sharp, P., 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 *Caenorhabditis elegans* unc-86 gene products. *Genes Dev.* **2,** 1513-1516.

Hiromi, Y. and Gehring, W. (1987). Regulation and function of the Drosophila segmentation gene *fushi tarazu. Cell* **50,** 963-974.

Hiromi, Y., Kuroiwa, A. and Gehring, W.J. (1985). Control elements of the Drosophila segmentation gene *fushi tarazu. Cell* **43,** 605-613.

Hoey, T., Dynlacht, B.D., Peterson, M.G., Pugh, B.F. and Tjian, R. (1990). Isolation and characterization of the Drosophila gene encoding the TATA box binding protein, TFIID. *Cell* **61,** 1179-1186.

Holland, P. and Hogan, B. (1988). Expression of homeobox genes during development. *Genes Dev.* **2**,773-782.

Ingham, P.W. (1988). The molecular genetics of embryonic pattern formation in Drosophila. *Nature* **335,** 25-34.

Ingraham, H.A., Chen, R., Mangalam, 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.

Johnson, W.A. and Hirsh, **J.** (1990). Binding of a Drosophila POU-domain protein to a sequence element regulating expression in specific dopaminergic neurons. *Nature* **343,** 467-470.

Knipple, D.C., Seifert, E., Rosenberg, U.B., Preiss, A. and Jackle, H. (1985). Spatial and temporal patterns of *Kruppel* gene expression in early Drosophila embryos. *Nature* **317,** 40-44.

Komberg, T., Siden, I., O'Farrell, P. and Simon, M. (1985). The *engrailed* locus of Drosophila: In situ localization of transcripts reveals compartment-specific expression. *Cell* **40,** 45-53.

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

Manley, J.L. and Levine, M. (1985). The homeobox and mammalian development. Cell **43,** 1-2.

Mitchell, P.J and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245,** 371-378.

Mlodzik, M. and Gehring, W. (1987). Expression of the caudal gene in the germ line of Drosophila: Formation of an RNA and protein gradient during early embryogenesis. *Cell* **48,** 465-478.

Monuki, E.S., Kuhn, R., Weinmaster, G., Trapp, B.D. and Lemke, G. (1991). Expression and activity of the POU transcription factor SCIP. *Science* **249,** 1300-1303.

Muhich, M.L., Iida, C.T., Horikoshi, M., Roeder, R.G. and Parker, C.S. (1990). cDNA clone encoding Drosophila transcription factor TFIID. *Proc. Natl. Acad. Sci.(USA)* **87,** 9148-9152.

Muller, M.M., Ruppert, S., Schaffner, W. and Matthias, P. (1988). A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoter in non-B cells. *Nature* **336,** 544-55 1.

Nusslein-Volhard, C., Frohnhofer, H.G. and Lehman, R. (1987). Determination of anteroposterior polarity in the Drosophila embryo. *Science* **238,** 1675-1681.

Okamoto, K., Okazawa, H., Pkuda, A., Sakai, M., Muramatsu, M. and Hamada, H. (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* **60,** 461-472.

Petryniak, B., Staudt, L.M., Postema, C.E., McCormack, W.T. & Thompson, C.B. (1990). Characterization of chicken octamer-binding proteins demonstrates that POU domain containing homeobox transcription factors have been highly conserved during vertebrate evolution. *Proc. Natl. Acad. Sci. (USA)* **87**,1099-1103.

Prakash, K., Topol, J., Dearolf, C.R. and Parker, C.S. (1991). Transcriptional regulation during early Drosophila development. In *Advances in Developmental Biochemistry,* (Wasserman, P., Ed) Vol. I, pp 85-131, JAI Press, Greenwich.

Rosenfeld, M.G. (1991). POU-domain transcription factors: pou-er-ful developmental regulators. *Genes Dev.* **5**, 894-907.

Rosner, M.H., Vigano, M.A., Ozato, K., Timmons, P.M., Poirier, F., Rigby, P.W. andStaudt, L.M. (1990). A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345,** 686-692.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning a laboratory manual (2nd edition). Cold Spring Harbor Laboratory Press, New York.

Scheidereit, C., Cromlish, J.A., Gerster, T., Kawakami, K., Balmaceda, C.G., Currie, R.A. and Roeder, R.G.(1988). A human lymphoid-specific transcription factor that activates immunoglubulin genes is a homeobox protein. *Nature* **336,** 551-557.

Scholer, H.R., Hatzopoulos, A.K., Balling, R., Suzuki, N. and Gruss, P. (1989). A family of octamer-specific proteins present during mouse embryogenesis: evidence for germlinespecific expression of an Oct factor. (1989). *EMBO J.* **8**, 2543-2550.

Scholer, H.R., Ruppert, S., Suzuki, N., Chowdhury, K. and Gruss, P. (1990). New type of POU domain in germ line-specific protein Oct-4. *Nature* **344,** 435-439.

Scott, M.P. and Carroll, S.B. (1987). The segmentation and homeotic gene network in early Drosophila development. *Cell* **51,** 689-698.

Scott, M.P. and O'Farrell, P.H. (1986). Spatial programming of gene expression in early Drosophila embryogenesis. *Annu.Rev.Cell Biol.* **2**,49-80.

Scott, M.P., Tamkum, J.W. and Hartzell, G.W., III (1989). The structure and function of the homeodomain. *BBA Rev.Cancer* **989,** 25-49.

Smith, D.P. and Old, R.W. (1990). Nucleotide sequence of Xenopus laevus Oct-1 cDNA. *Nucl.Acids Res.* **18**,369.

Staudt, L.M., Clerc, R.G., Singh, H., LeBowitz, J.H., Sharp, P.A. and Baltimore, D. (1988). Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif. *Science* **241,** 577-580.

Stem, S., Tanaka, M. and Herr, W. (1989). The Oct-1 homeodomain directs formation of a multiprotein-DNA complex with the HSV transactivator VP 1 6. *Nature* **341,** 624-630.

Sturm, R.A., Das, G. andHerr, W. (1988). The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeobox subdomain. *Genes Dev.* **2,** 1582-1599. Sturm, R.A. and Herr, W. (1988). The POU domain is a bipartite DNA-binding structure. *Nature* **336,** 601-604.

Suzuki, N., Rohdewohld, H., Neuman, T., Gruss, P. and Scholer, H.R. (1990). Oct-6: a POU transcription factor expressed in embryonal stem cells and in the developing brain. *EMBO J.* **9**, 3723-3732.

Tautz, D. and Pfeifle, C. (1989). A non-radioactive in-situ hybridization method for the localization of specific RNAs in Drosophila embryos reveals translational control of the segmentation gene *hunchback. Chromosoma* **98,** 81-85.

Thali, M., Muller, M.M., DeLorenzi, M., Matthias, P. and Bienz, M. (1988). Drosophila homeotic genes encode transcriptional activators similar to mammalian OTF-2. *Nature* **336,** 598-601

Treacy, M.N., He, Xi and Rosenfeld, M.G. (1991). I-POU: a POU-domain protein that inhibits neuron-specific gene activation. *Nature* **350,** 577-584.

Vinson, C.R., LaMarco, K.C., Johnson, P.F., Landshuiz, W.H. and McKnight, S.L. (1988). *Genes Dev.* **2**, 801-807.

Wright, C.V.E., Cho, K.W.Y., Oliver, G. and DeRobertis, E.M. (1989). Vertebrate homeodomain proteins: families of region-specific transcription factors. *Trends Biochem.Sci.* **14**,52-56.

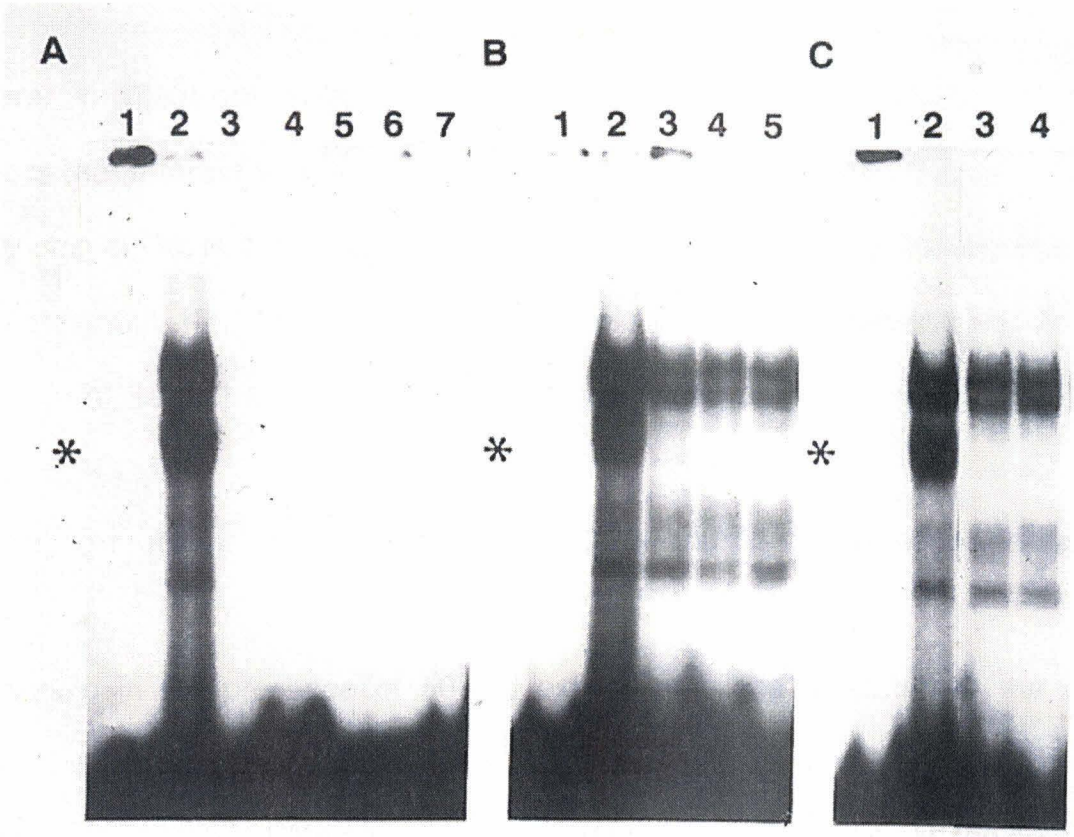

Figure. 1

Figures and Figure Legends

Figure. 1. Electrophoretic mobility shift assay showing binding of nuclear factors that interact with the consensus octamer binding site.

A. Binding of Oct factors to the ^{32}p-labeled wild-type octamer site (lane 2) that can be competed with 10-, 20-, 30-, 50- and 100-fold excess of unlabeled wild-type oligo (lanes 3-7). Lane 1 is the control lane that contained only the probe and no nuclear extract.

B. Binding reactions in the presence of 10-, 20- and 50-fold excess of mutant oligo (lanes 3-5).

C. binding in the presence of 50-fold excess of oligos unrelated to the octamer sequence (Lanes 3 & 4). Lanes 1 & 2 in B and C are same as lanes 1 & 2 in A. The position of the protein-DNA complex indicated by an asterisk indicates the non-specific complex.

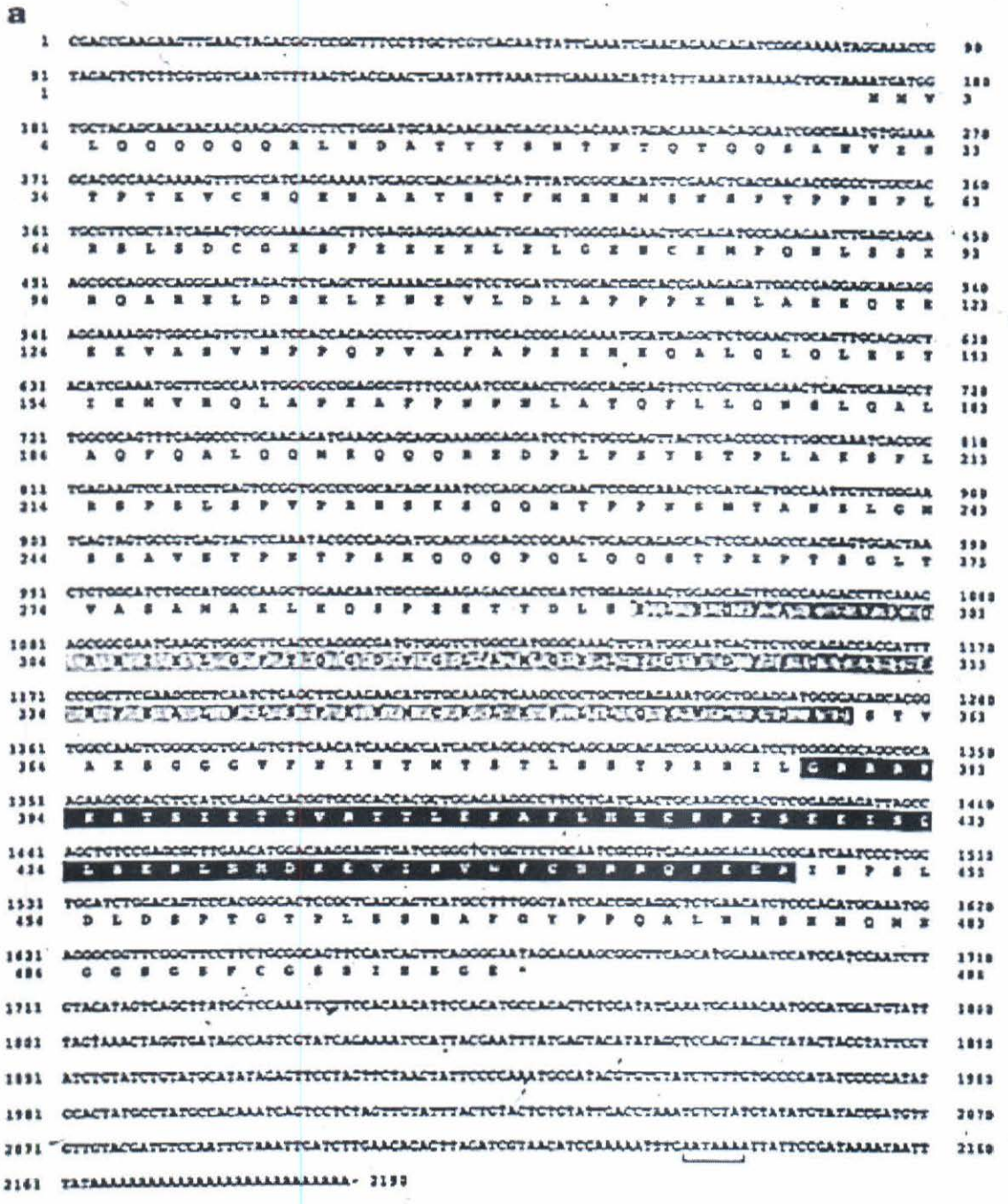

Figure. 2A

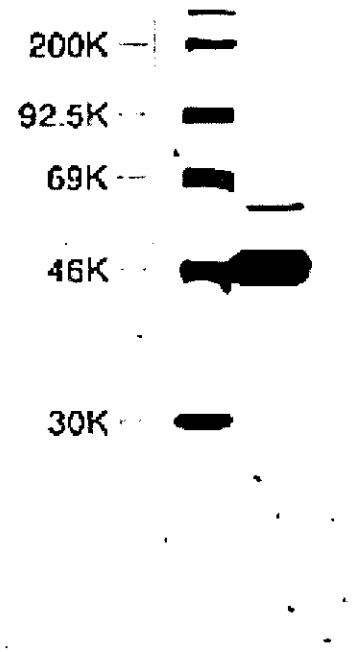

Figure. 2B

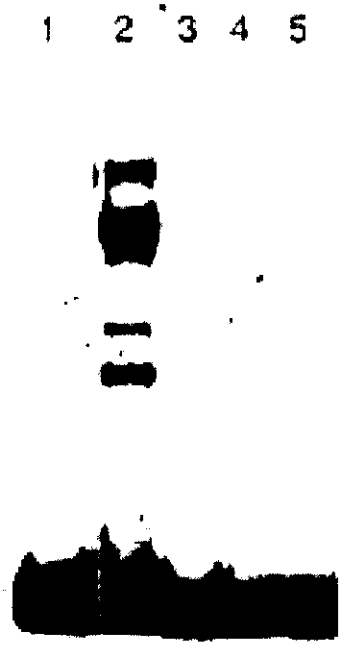

Figure. 2C

Figure. 2 A. Nucleotide sequence of the dOct cDNA and the predicted amino acid sequence indicated in the one-letter code. A putative polyadenylation signal in the nucleotide sequence is underlined. The POU-specific and the POU-homeodomain are indicated by the shaded and dark boxes, respectively.

B. SDS-polyacrylamide gel electrophoresis of the products of the dOct MRNA translation. The molecular weights of the ^{14}C-methylated markers (Amersham) is indicated in kilo daltons.

C. Mobility shift assay of the *in vitro* synthesized recombinant dOct polypeptide. The doct protein binds to the wild-type ^{32}P-labled octamer oligo (lane 2). The protein-DNA complexes are competed by 10-, 20- and 50-fold excess of the unlabeled oligo (lanes 3-5). Lane 1 shows the control experiment in which the unprogrammed lysate was incubated with labeled probe.

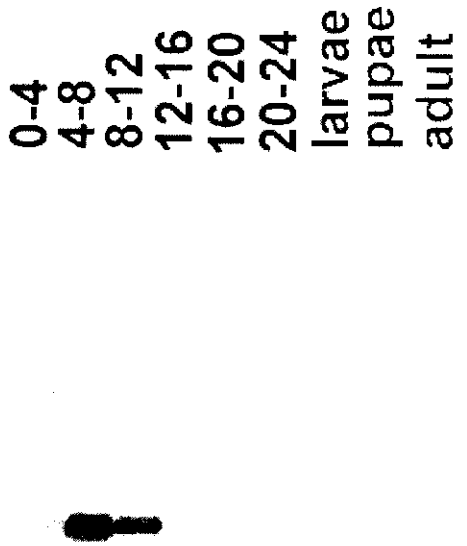

Figure 3. Developmental Northern blot analysis of the dOct transcripts. Poly(A+) RNA from isolated from staged embryos (indicated as hours after egg laying), larvae, pupae and adult flies was hybridized with dOct cDNA as described in the Methods. The size of the transcript was estimated using commercial RNA markers (Bethesda Research Laboratories).

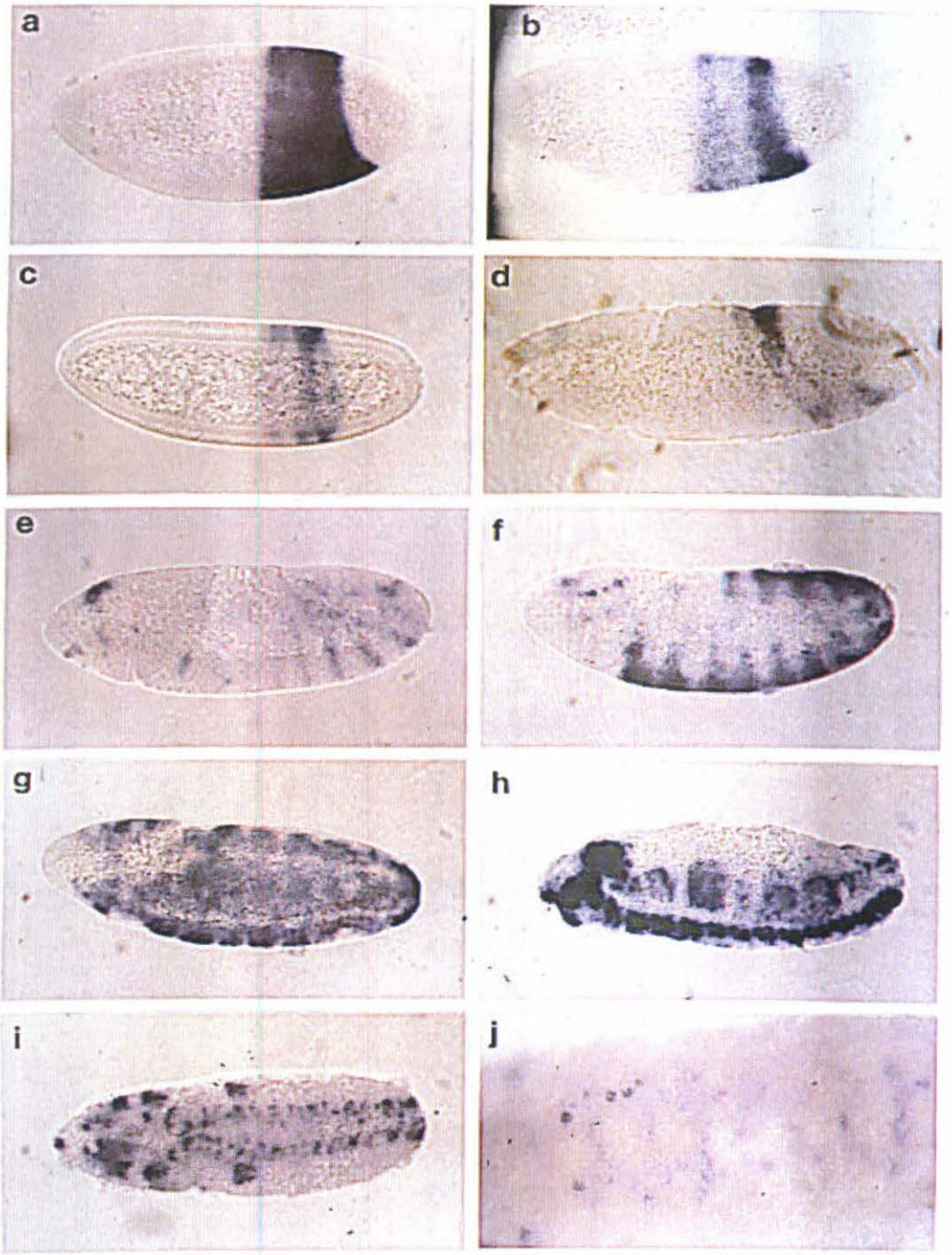

Figure. 4

Figure 4. Spatial distribution of doct RNA during early embryogenesis. In-situ hybridization of doct CDNA to whole, wild-type embryos was as described in the Methods. Staging of the embryos was done as described by Campos-Ortega and Hartenstein (1985). All views are lateral with anterior on the left unless otherwise stated. A-C. Stage 5, cellular blastoderm. D. Stage 6, early gastrulation. E & F. Stages 7 and 10, germband extension. The arrows in F indicate the doct expression in clusters of cells. G. Stage 11, germband elongation completed. H. Stage 15, germband retracted. The arrow indicates expression of doct in the brain. I & J. Ventral views. J. Higher magnifcation showing the @dline precursor cells. The arrows in I indicate the sensory neurons.

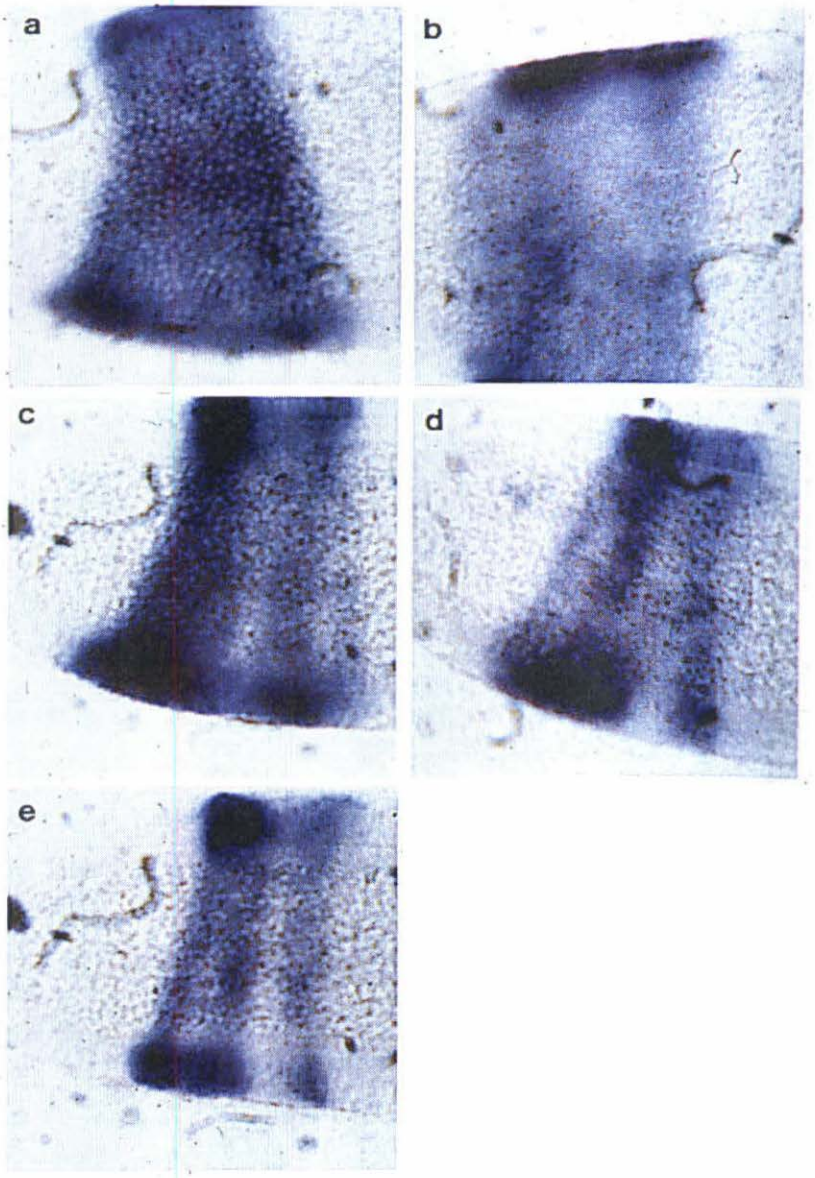

Figure 5. Higher magnification showing the dynamic changes in dOct RNA localization in the cellular blastoderm embryo (Stage 5). A. Uniform wide band that progressively resolves into two (B & C) and finally into three stripes (D & E).

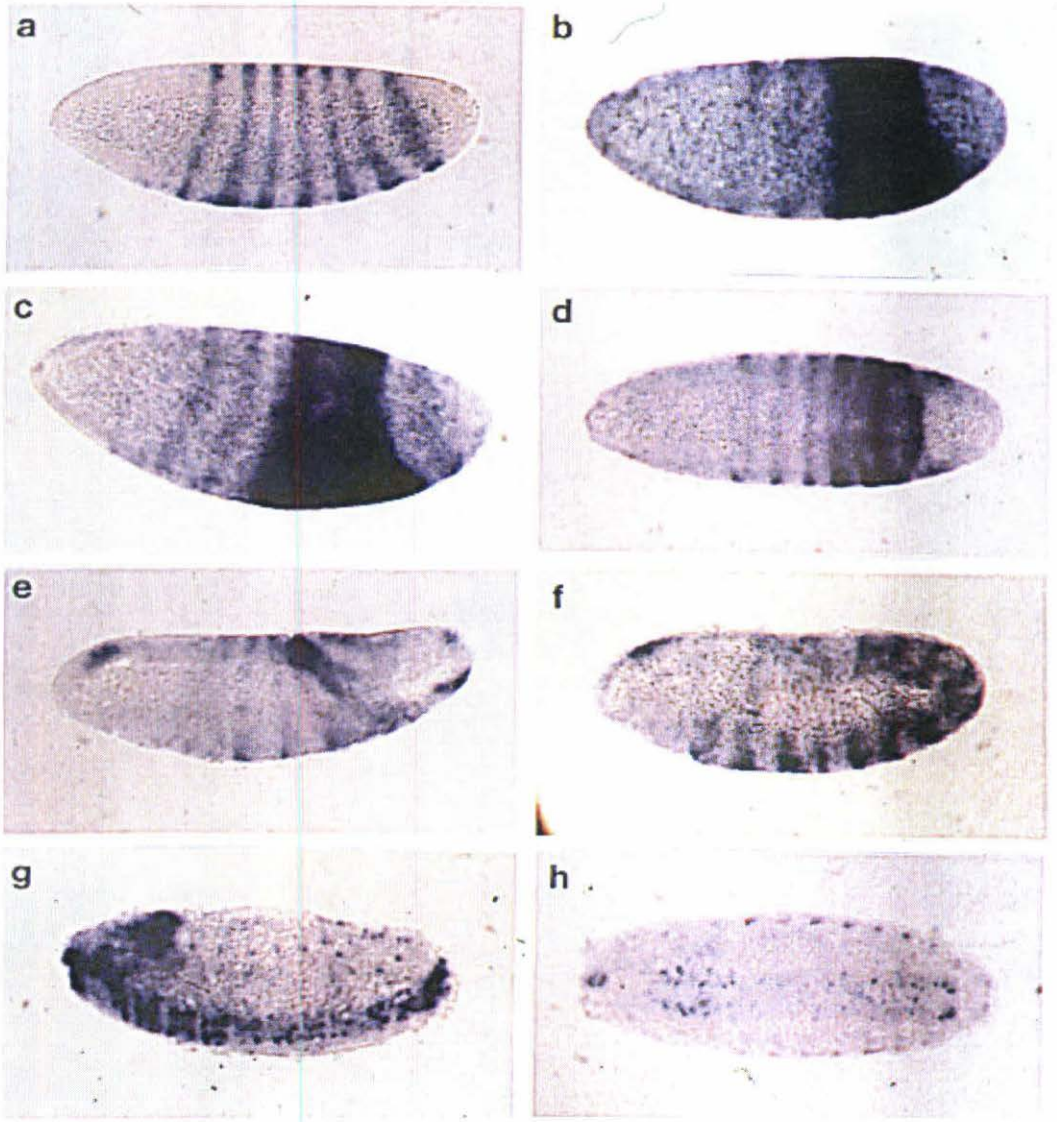

Figure. 6

Figure 6. Co-localization of transcripts from the *ftz* and dOct genes. A. Cellular blastoderm embryo (Stage 5) probed with the *ftz* probe alone showing the characteristic seven stripes. All other panels show embryos of different developmental stages probed with both the *ftz* and the dOct probes (B-H). B. Stage 5 embryo showing strong expression of dOct but weak expression and only a few of the *ftz* stripes. C & D. Embryos that show the striped expression of dOct. The dOct stripes overlap those of *ftz.* E. Stage 6, early gastrulation. F. Stage 8-10, elongated germband embryo showing fourteen stripes. G & R (dorsal view). Stage 11, germband elongation completed. Expression of both the transcripts is seen in the central nervous system. The bigger arrow in G shows expression of dOct in the brain cells and smaller arrows indicate the peripheral sensory/muscle precursor cells. The arrows in H indicate the position of clusters of peripheral sensory cells.

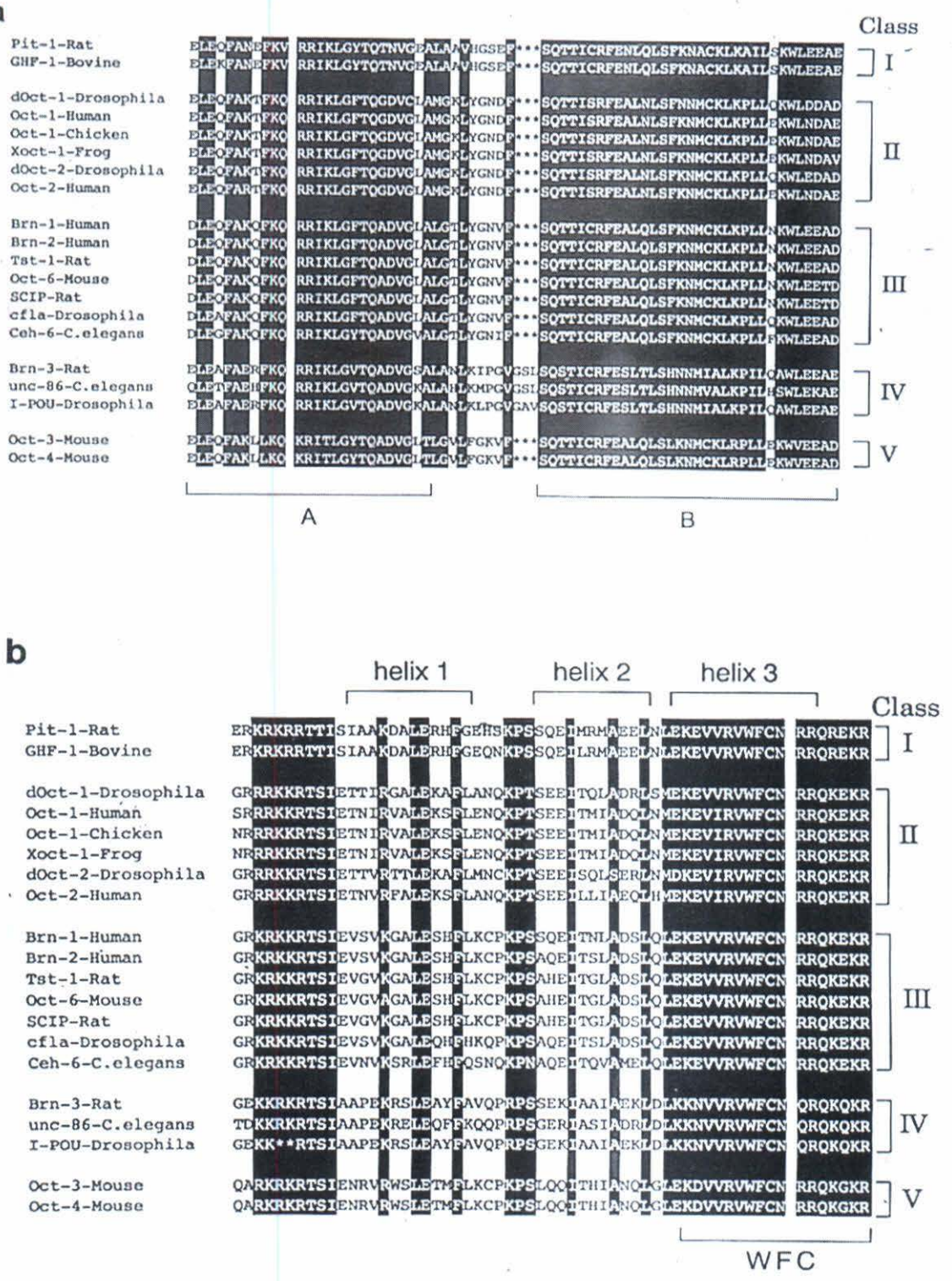
a
Class
Pit-1-Rat
GHF-1-Bovine
dOct-1-Drosophila
Oct-1-Human
Oct-1-Chicken
Xoct-1-Frog
dOct-2-Drosophila
Oct-2-Human
Brn-1-Human
Brn-2-Human
Tst-1-Rat
Oct-6-Mouse
SCIP-Rat
cfla-Drosophila
Ceh-6-C.elegans
Brn-3-Rat
unc-86-C.elegans
I-POU-Drosophila
Oct-3-Mouse
Oct-4-Mouse
I
II
III
IV
V
A
B
b
helix 1
helix 2
helix 3
Class
Pit-1-Rat
GHF-1-Bovine
dOct-1-Drosophila
Oct-1-Human
Oct-1-Chicken
Xoct-1-Frog
dOct-2-Drosophila
Oct-2-Human
Brn-1-Human
Brn-2-Human
Tst-1-Rat
Oct-6-Mouse
SCIP-Rat
cfla-Drosophila
Ceh-6-C.elegans
Brn-3-Rat
unc-86-C.elegans
I-POU-Drosophila
Oct-3-Mouse
Oct-4-Mouse
I
II
III
IV
V
WFC

Figure. 7A&B

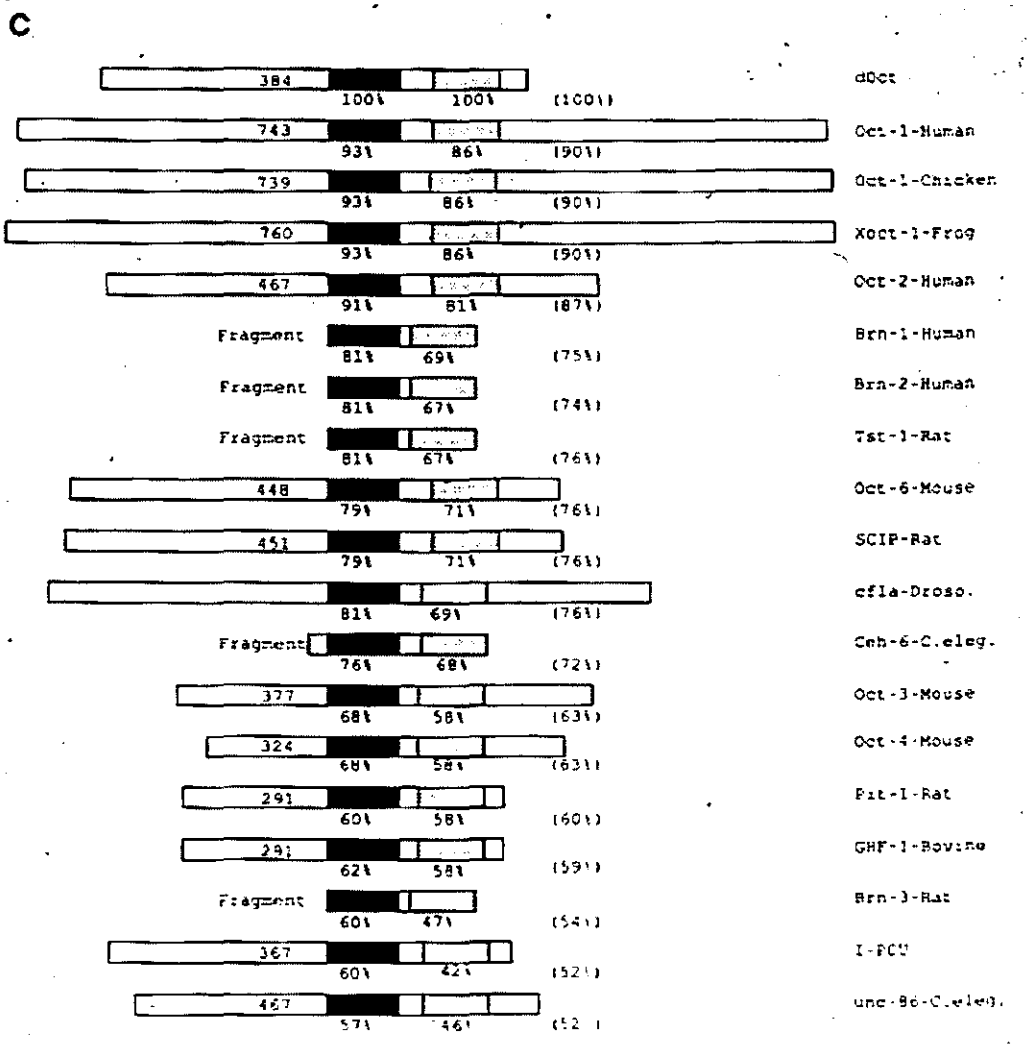
C
384
100%
100%
(100%)
dOct
743
93%
86%
(90%)
Oct-1-Human
739
93%
86%
(90%)
Oct-1-Chicken
760
93%
86%
(90%)
Xoct-1-Frog
467
91%
81%
(87%)
Oct-2-Human
Fragment
81%
69%
(75%)
Brn-1-Human
Fragment
81%
67%
(74%)
Brn-2-Human
Fragment
81%
67%
(76%)
Tst-1-Rat
448
79%
71%
(76%)
Oct-6-Mouse
451
79%
71%
(76%)
SCIP-Rat
81%
69%
(76%)
cfla-Droso.
Fragment
76%
68%
(72%)
Cnh-6-C.eleg.
377
68%
58%
(63%)
Oct-3-Mouse
324
68%
58%
(63%)
Oct-4-Mouse
291
60%
58%
(60%)
Pit-1-Rat
291
62%
58%
(59%)
GHF-1-Bovine
Fragment
60%
47%
(54%)
Brn-3-Rat
367
60%
42%
(52%)
I-POU
467
57%
46%
(52%)
unc-86-C.eleg.

Figure. 7C

Figure 7. Comparison of the dOct POU domain with those of previously characterized POU domains.

A and B. Amino acid sequence comparisons of the POU-specific and POU-homeodomain regions of the POU proteins and dOct, respectively. The darkly shaded regions indicate the invariant or, one or two conservative substitutions of the amino acid residues. The POU sequences shown here were taken from the following references: human Oct-1, Sturin et al. (1988); chicken Oct-1, Petryniak et al. (1990); frog Oct-1, Smith & Old (1990); human Oct-2, Muller et al. (1988); Scheidereit et al. (1988); Clerc et al. (1988); human Bm-1, human Brn-2, rat Bm-3, rat Tst-1, He et al. (1989); mouse Oct-6, Suzuki et al. (1990); rat SCIP, Monuki et al. (1991); Drosophila cfla, Johnson & Hirsh (1990); Treacy et al. (1991); C. elegans ceh-6, Buerglin et al. (1989); mouse Oct-3, Okamoto et al. (1990); Rosner et al. (1990); mouse Oct-4, Scholer et al. (1990); rat Pit-1, Ingraham et al. (1988); bovine GHF-1, Bodner et al. (1988); Drosophila I-POU, Treacy et al. (1991); C. elegans unc-86, Finney et al. (1988).

C. Percent identities between the various POU proteins within the POU-specific (black box) and POU-homeodomain (stippled box), and the entire POU domain (numbers in parentheses) and the dOct protein. The number on the left in each case corresponds to the length of the individual polypeptide. The sources for the individual sequences were the same as in A & B.