INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

MOLECULAR MECHANISMS OF *tinman* INDUCTION BY Dpp IN THE DORSAL MESODERM OF *DROSOPHILA*

by

Xiaolei Xu

A dissertation submitted to the Graduate Faculty in Biomedical Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1999

UMI Number: 9917715

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

This manuscript has been read and accepted for the Graduate Faculty in Biomedical Science in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Leslie Pick

Date Dr. Leslie Pick
Chair of Examining Committee

1/26/99

Date Dr. Terry A. Knilwich
Executive Officer

Dr. Manfred Frasch

Dr. Robert Krauss

Dr. Thomas Lufkin

Dr. Richard Padgett

Dr. David Sassoon

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

MOLECULAR MECHANISMS OF *tinman* INDUCTION BY Dpp IN THE DORSAL MESODERM OF *DROSOPHILA*

by

Xiaolei Xu

Advisor: Dr. Manfred Frasch

decapentaplegic (dpp) is a *transforming growth factor-β (TGF-β)* homologue in *Drosophila*, which plays essential roles as an inductive molecule in diverse steps of cell fate determination. During gastrulation, Dpp is secreted from the dorsal ectoderm and induces the formation of dorsal mesoderm, which later develops into visceral mesoderm, dorsal muscles and heart. The restriction of the expression of *tinman*, a NK homeobox gene, from the whole mesoderm to the dorsal mesoderm reflects this inductive event. In this study, *tin*-D, a 349bp Dpp-responsive element from the *tinman* enhancer, was identified and dissected to address the molecular mechanisms underlying this inductive process. A 32bp sub-element D3 was defined as a minimal Dpp-responsive element that is able to drive gene expression in dorsal areas of the embryo. The tissue-specific induction in the dorsal mesoderm requires the presence of another sub-element, D1, which specifically binds Tinman protein. D1 mediates a mesoderm-specific activity that consists of *tinman* autoregulation and acts in concert with the inductive signals. Additional observations suggested that a repressor can bind to D1 and prevent *tin*-D

activity in the Dpp signaling tissue, i.e., the dorsal ectoderm. The minimal Dpp-responsive element D3 was used to isolate binding factors through yeast one hybrid screening, which resulted in the cloning of *Medea*, the common-mediator *Smad* of *Drosophila*. A variety of *in vitro* and *in vivo* experiments showed that MEDEA mediates the dorsal inductive signal of Dpp by directly binding to several GC-rich motifs in the *tin*-D element, including the D3 sub-element. The transduction of the Dpp signal to the *tinman* promoter may additionally require a CAATGT binding transcription factor. The potential functions of other factors in this pathway that were identified in the yeast one hybrid screens, including *HMG-D*, *ccf-β*, *DEAF-1* and *Adf-1*, are also discussed. A comprehensive model combining the synergistic and antagonistic interactions is proposed to explain the molecular mechanisms of both the establishment of the dorsal-ventral polarity in *Drosophila* mesoderm and the signal transduction of TGF-β at the transcriptional level.

To my Dad, Routing Xu

and my Mom, Min Chen

ACKNOWLEDGMENTS

First of all, I would like to thank my preceptor, Dr. Manfred Frasch, for his guidance over the last five years. I am lucky to be trained in his lab who is dedicated to the science of the best quality. This training will benefit my academic career for a lifetime. I extend my thanks to his wife, Dr. Hanh Nguyen, as both a mentor and a collaborator.

I would like to thank the members of my advisory committee: Dr. Thomas Lufkin, Dr. Leslie Pick, Dr. Miki Rifkin and Dr. David Sassoon. From their suggestions and evaluations, I learned to be logical and critical. I am also grateful to current and past members of Manfred's lab who provided me precious friendships and daily help throughout this study. Deep thanks are due to Stefan Knirr, Hsiu-Hsiang Lee, Patrick Lo, Zhizhang Yin and Stephane Zaffran.

Special thanks go to Xueying Lin, Huayu Qi and Yong Xue. my best friends in Mount Sinai. Through stimulating discussions, we confirmed each other to pursue the common aim as both a successful scientist and a decent human being. In addition, I am grateful to all of my friends and colleagues of the Mount Sinai Chinese Student and Scholar Association (MSCSSA) and the Association of Chinese Students and Scholars in New York (ACSSNY). They have provided me so many heart-warming moments.

Finally, I would like to thank my parents, Routing Xu and Min Chen, who always encourage me to pursue what I love. As typical Chinese scholars, they keep me aware of the splendid happiness of an intelligent life. As a marker of the starting point toward my academic career, this thesis is thus dedicated to them.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

The mesoderm autonomous gene hierarchy during *Drosophila* embryogenesis

A fundamental question in developmental biology is how cell fates are determined during embryogenesis. This study focused on the molecular mechanisms underlying the formation of the *Drosophila* dorsal mesoderm. During the blastoderm stage, the cells in the ventral quadrant of the embryo are determined to form the mesoderm. These presumptive mesoderm cells invaginate into the embryo to form a ventral hollow tube during gastrulation, and then spread dorsally along the neighboring ectoderm to form a monolayer of mesoderm cells. Until this stage, these mesoderm cells are not yet committed to particular developmental fates (Beer et al., 1987). The establishment of dorsal-ventral polarity represents one of the first subdivisions of cell fates in the mesoderm. Later, the dorsal mesoderm cells differentiates into the visceral musculature, the dorsal somatic musculature and heart progenitor cells.

A cell-autonomous gene hierarchy plays an essential role in the formation and differentiation of the *Drosophila* mesoderm (Frasch and Nguyen, 1998). The establishment of dorsal-ventral polarity of the *Drosophila* body plan is determined by a nuclear gradient of the maternal morphogen *dorsal* (Roth et al., 1989). *dorsal* transcriptionally activates two zygotic genes, *twist* and *snail*, in the ventral most cells

(Rusch and Levine, 1996). Both of the genes are essential for the mesoderm formation (Grau et al., 1984; Simpson, 1983). *snail* belongs to the zinc finger transcription factor family (Thisse et al., 1988), whose major role is to repress the expression of non-mesodermal genes (Leptin, 1991; Rao et al., 1991). *twist* belongs to the basic helix-loop-helix (bHLH) transcription factor family which recognize so called E-boxes (CANNTG) (Boulay et al., 1987). *twist* trascriptionally activates a number of downstream target genes which have important functions during mesoderm cell fate determination (Kosman et al., 1991; Leptin, 1991; Rao et al., 1991). For example, *tinman*, a homeobox containing gene, is activated by *twist* and is required for the formation of the dorsal mesoderm (Bodmer et al., 1990; Yin et al., 1997, see below). *mef2*, a MADS-box containing gene, is activated by *twist* and later determines the differentiation of the three musculatures in *Drosophila* (Lilly et al., 1994; Nguyen et al., 1994; Nguyen and Xu, 1998). Other candidate genes include *Zfh-1* (Lai et al., 1991), *Heartless* (Shishido et al., 1993), *How* (Baehrecke, 1997; Zaffran et al., 1997; Lo and Frasch, 1997) and genes with undefined functions (Casal and Leptin, 1996).

tinman, a homeobox gene which is essential for mesoderm patterning

tinman belongs to the NK homeo box gene family which was first identified by screening a *Drosophila* cDNA library with degenerate homeodomain oligonucleotides as PCR primers (Kim and Nirenberg, 1989). *tinman* was mapped to the 93E region, together with several of the other members of this gene family, such as *bagpipe (bap)* (Azpiazu and

Frasch, 1993), *S59* (Dohrmann et al., 1990) and *ladybird (lbl)* (Jagla, et al., 1994). Like the homeotic genes, they form a gene cluster.

tinman plays an important role in the autonomous gene hierarchy to determine the cell fates in the mesoderm. In *tinman* mutant embryos, all of the dorsal mesoderm derived tissues are missing, including the visceral mesoderm, heart progenitor cells (Azpiazu and Frasch, 1993; Bodmer, 1993) and dorsal body wall muscles (Yin and Frasch, 1998). The potential downstream target genes of *tinman* include the NK homeobox gene *bap* in the visceral mesoderm (Azpiazu and Frasch, 1993), the homeo box containing gene *msh* in the dorsal body wall muscles (Yin and Frasch, 1998) and the *mef2* in the heart musculature (Gajewski et al., 1997). Tinman could activate its downstream genes by directly binding to enhancer elements. Its DNA binding specificity is different from the ATTA motifs for homeotic genes. Random oligonucleotides selection has been used to determine the optimal binding sites for two NK family proteins: *Drosophila vnd* (Tsao et al., 1994) and mouse *Nkx2.5* (Chen et al., 1995). Both studies revealed a 5'-TNNAGTG-3' consensus sequence, with TCAAGTG as the motif with the highest binding affinity. This sequence has been found in the *mef2* promoter region and *in vitro* data confirm that it can be bound by Tinman protein (Gajewski et al., 1997).

tinman has a dynamic expression pattern in three phases during mesoderm formation and differentiation. The phase I expression is detected in the whole presumptive mesoderm

corresponding to the cells with a high level of Twist protein in late cellular blastoderm stage (Bodmer et al., 1990). During the invagination and spreading of the mesoderm into the interior of the embryo, the phase II expression of *tinman* is restricted to the dorsal mesoderm only. Later on, the phase III expression is further restricted to the heart progenitor cells which are located at the dorsal tips of the mesoderm (Azpiazu and Frasch, 1993). To identify the *cis*-regulatory elements that are responsible for these expression patterns, a series of constructs containing *tinman* flanking regions fused to the *lacZ* reporter gene were tested in the transgenic fly system. Three modular *cis*-regulatory elements were identified which reflected the three phases of *tinman* expression (Figure 1, for original data see Yin et al., 1997). Surprisingly, none of them is located in the 5'-upstream region of the *tinman* gene. *tin*-B, a 379bp element which is located in the first intron, activates *lacZ* expression in the whole mesoderm. *tin*-C, a 300bp element which is located downstream of the *tinman* gene, activates *lacZ* expression in the heart progenitor cells. I contributed to the project by further defining *tin*-D, a 349bp downstream element, which specifically drives reporter gene expression in the dorsal mesoderm. Because of the essential function of *tinman* in the dorsal mesoderm, the identification of the *tin*-D element provides an excellent entry point to uncover the molecular mechanisms of dorsal-ventral polarity establishment in the *Drosophila* mesoderm.

Figure 1. Location of modular *cis*-regulatory elements from the *tinman* gene of *Drosophila melanogaster*.

Exons are shown as black and introns as white boxes. Enhancer elements are shown as dotted boxes (for original data, see Yin et al., 1997). *tin*-B, a 180bp element located in the first intron, drives the reporter gene in the trunk mesoderm at stage 5, thus mimicking the phase I expression of *tinman*. *tin*-C, a 300bp element located in the downstream of the *tinman* encoding region, drives the reporter gene expression in the heart progenitor cells, thus mimicking the phase III expression of *tinman*. *tin*-D, a 349bp element which was identified in this study, drives the reporter gene expression in the dorsal trunk mesoderm, thus mimicking the phase II expression of *tinman*.

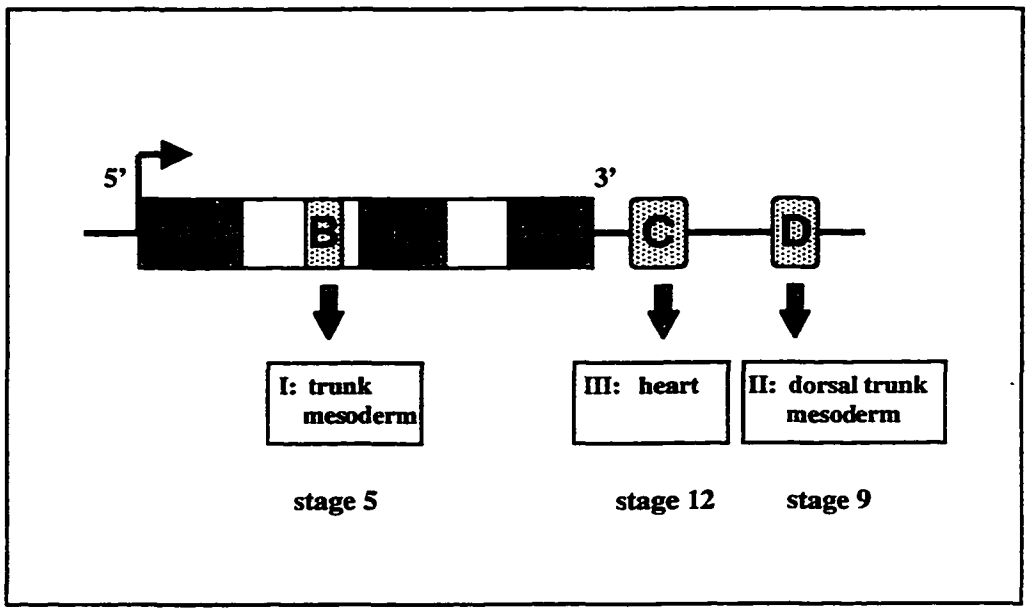


Figure 1

Mesoderm induction in *Drosophila*

In addition to the cell-autonomous gene hierarchy within the mesoderm, inputs from exogenous signals are also important for mesoderm cell fate determination. A cell ablation experiment performed in lacewing flies in the 1930s indicated that there are inductive influences from ectoderm to define mesoderm cell fates (Seidel et al., 1940). Recent experiments have shown that the signal is Decapentaplegic (Dpp), a member of Transforming growth factor–β (TGF–β) family (Staehling-Hampton et al., 1994, Frasch, 1995). When *tinman* is still in its phase I expression in the presumptive mesoderm, the expression domain of Dpp is already restricted to the dorsal ectoderm (Figure 2A, from Frasch, 1995). During the invagination process, only those mesoderm cells that contact the dorsal ectoderm and receive the Dpp signal will maintain *tinman* expression (Figure 2B&C). Mesoderm cells that are located ventrally fail to receive the Dpp signal and *tinman* expression gradually disappears (Figure 2D). In *dpp* mutant embryos, the phase II expression of *tinman* was not detected. When Dpp was ectopically expressed in the ventral ectoderm, *tinman* expression was extended to the ventral mesoderm (Frasch, 1995). Moreover, in a genetic background in which fewer mesoderm cells were produced, the mesoderm was unable to spread dorsally as far as in the wild type embryos. Only those very few cells that could contact the dorsal ectoderm maintained the expression of *tinman* (Maggert et al., 1995). All of the above data pointed to the conclusion that Dpp is secreted from the dorsal ectoderm and induces across the germlayer the maintained expression of *tinman* in the dorsal mesoderm.

Figure 2. Comparison of *tinman* and *dpp* expression during early mesoderm differentiation.

(for original data, see Frasch, 1995). Shown are cross-sections stained for *tinman* mRNA expression (inner layer) and *dpp* expression (out layer). *dpp* expression is monitored with a *lacZ* reporter gene driven by *dpp* regulatory sequences. (A) The phase I expression pattern of *tinman* in the whole trunk mesoderm at stage 8. At the same stage, *dpp* expression is already restricted to the dorsal ectoderm. (B-C) During the process of invagination, mesoderm cells migrate toward the dorsal side to form a single layer of cells. The expression of *tinman* is gradually decreased in the ventral mesoderm and increased in the dorsal portions. (D) The phase II expression pattern of *tinman* in the dorsal mesoderm at stage 10. Notably, the border of *tinman* expression corresponds to that of *dpp* expression in the dorsal ectoderm. (ms.) mesoderm. (ec.) ectoderm. (V) ventral. (D) dorsal.

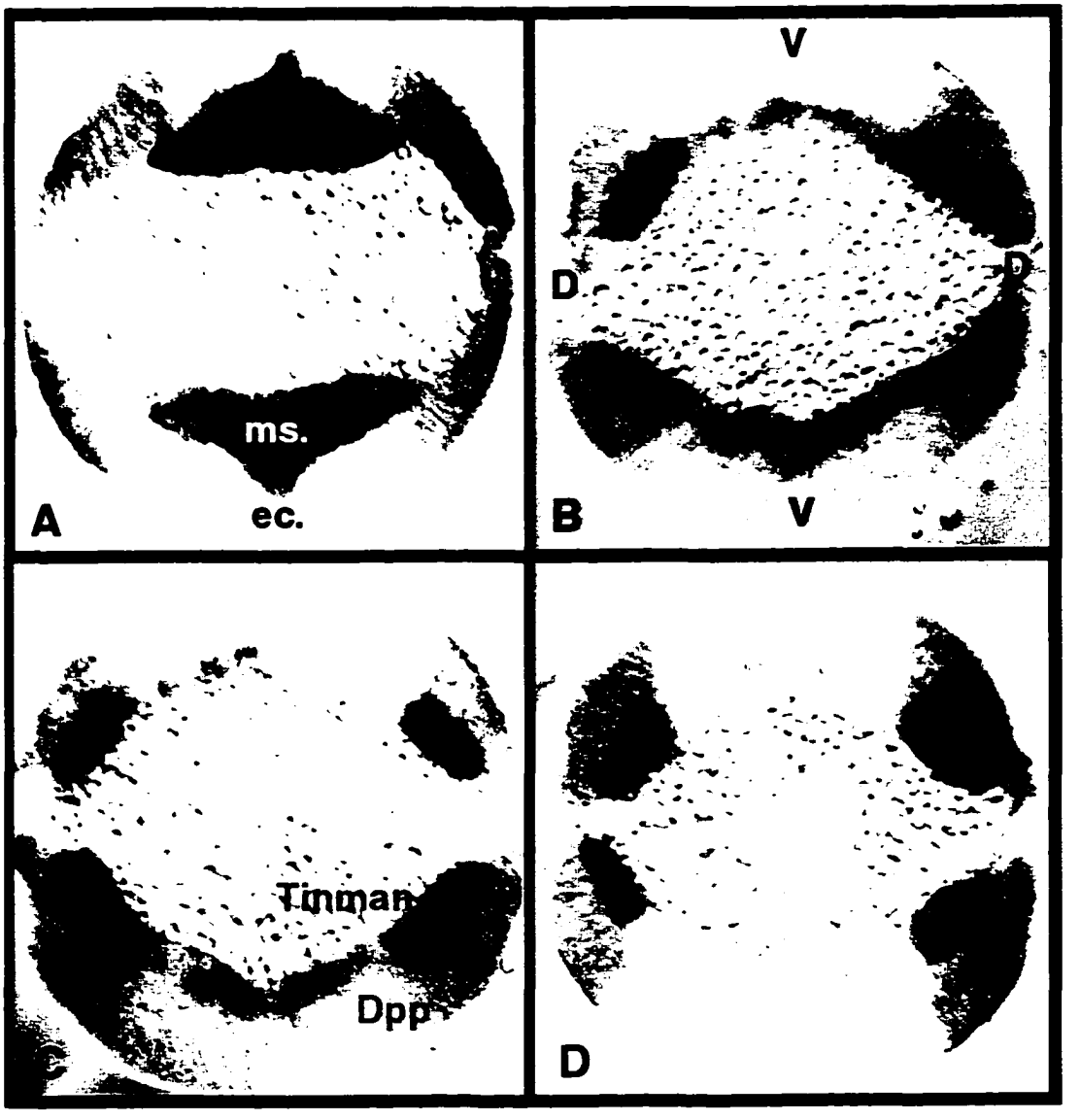


Figure 2

Other inductive interactions have been identified during *Drosophila* mesoderm cell fate determination. For example, *S59* and *nautilus* expression are induced in the ventral lateral mesoderm by Wingless signals (Baylies et al., 1995; Ranganayakulu et al., 1996), while the dorsal median (DM) cell fate is induced by CNS midline progenitors (Luer et al. 1997). Inductive interactions between cell layers is a common theme during cell fate determination processes. However, how the autonomous gene hierarchy and the inductive signals cooperate to perform their specific functions has remained unclear. The dissection of the *tin*-D element was expected to permit further insights into the molecular mechanisms underlying this interplay of regulatory mechanisms.

Dpp-mediated induction

Dpp (Padgett et al., 1987) is a well studied signaling molecule and plays an essential function in several developmental processes. Dpp is first expressed in the dorsal-most cells of blastoderm stage embryos, i.e., in areas with low levels of nuclear Dorsal protein. During the establishment of dorsal-ventral polarity in the *Drosophila* embryo, Dpp functions as a morphogen. Higher levels of Dpp determine the formation of the amnioserosa dorsally, whereas lower activities prevent neurogenesis in the adjacent areas and allow the formation of the dorsal ectoderm (Ferguson and Anderson, 1992; Wharton et al., 1993). Dpp expression is then confined to the dorsal ectoderm only (St. Johnston and Gelbart, 1987), from where it is secreted across germ layers to induce the establishment of dorsal-ventral polarity in the neighboring mesoderm. Because of the first

phase of *dpp* expression, the analysis of the gene functions during mesoderm induction in *dpp* mutant embryos could be compromised. In these ventralized mutant embryos, the expression of *dpp* in the dorsal ectoderm itself is strongly reduced, which could prevent the mesoderm from receiving Dpp signals. This problem can be overcome by restoring the *dpp* expression in the ectoderm using ectopic gene expression systems (see results on *thick veins* germline mutants).

After germ band retraction, *dpp* is expressed in a portion of the visceral mesoderm, where it regulates midgut morphogenesis by maintaining the homeotic gene *Ultrabithorax (Ubx)* expression within the visceral mesoderm parasegment 7 and inducing another homeotic gene, *labial*, in the neighboring endoderm (Bienz, 1994). During dorsal closure, *dpp* is expressed in the leading edge of the epithelia cells. The Dpp pathway interacts with the *Drosophila* Jun-amino-terminal kinase (DJNK) pathway to determine the cell shape change and migration of the epithelia cells during the closing process (Glis and Noselli, 1997; Hou et al., 1997; Sluss and Davis, 1997; Riesgo and Hafen, 1997). *dpp* also controls the migration of the tracheal cells along the dorsal-ventral axis (Vincent et al., 1997). During metamorphosis, *dpp* is expressed in a narrow domain of cells at the anterior-posterior compartment boundary in the wing imaginal disc. It is thought to function as a morphogen to activate downstream target genes such as *spalt (sal)*, *optomotor-blind (omb)* and *vestigial (vg)* and thus organize growth and anterior-posterior patterning of the wing (Nellen et al., 1996; Lecuit et al., 1996; Goto and Hayashi, 1997).

In addition, Dpp plays important roles in the patterning of the eye and leg imaginal discs (Jiang and Struhl, 1996; Lecuit and Cohen, 1997; Chanut and Heberlein, 1997; Pignoni and Zipursky, 1997). The pleiotropic nature of Dpp functions raises questions of temporal-spatial specificity: how does the dpp signaling pathway interpret the Dpp signal differently in different tissues?

The *dpp* receptors and Dpp-signaling pathway

Dpp is a *Drosophila* member of the Transforming growth factor–β (TGF–β) superfamily (Padgett et al., 1987). With TGF–β as the prototypic member, this superfamily can be divided into three sub-families, including the TGF–βs, the activins and the bone morphogenetic proteins (BMPs) (Massague, 1990). Dpp belongs to the BMP family, which is the largest and most heterogeneous group. Like Dpp in *Drosophila*, the vertebrate members of the BMP sub-family are involved in many inductive interactions during early development (Hogan, 1996). *screw(scw)* and *60A* are another two *Drosophila* members of the BMP sub-family (Arora et al., 1994; Khalsa et al., 1998). No TGF-β or activin homologues in *Drosophila* has been reported yet.

The receptors for the TGF–β superfamily are Serine/Threonine kinase receptors which consist of two types (Massague, 1992; Attisano et al, 1994). Upon ligand stimulation, the constitutively active type II receptor recruits the type I receptor into the ligand-

receptor complex to form a heterodimer. The type I receptor is then phosphorylated by the type II receptors at its GS domain (a glycine/serine rich domain which is located at the cytoplasmic side of the receptor) and then propagates the signal to downstream targets (Wrana et al., 1994). Replacement of the threonine 204 in the GS domain of the TGF−β type I (TβR-I) receptor with a negatively charged aspartate (D) yields a constitutively active version of type I receptor (Wieser et al., 1995).

In *Drosophila*, the type II receptor for Dpp is *punt* (Letsou et al., 1995; Ruberte et al., 1995). Two type I receptors, *thick veins (tkv)* and *saxophone (sax)*, were cloned through either a degenerate PCR approach or a low stringency hybridization approach (Penton et al., 1994, Brummel 1994, Nellen et al, 1994, Xie et al., 1994). The constitutively active version of *tkv*, $tkv^{Q\text{-}D}$, has also been generated by the replacement of the Glutamine 253 (Q), whose position corresponds to the threonine 204 in TβR-I, to an aspartate (D) (Nellen et al., 1996). Recent studies indicated that Sax mediates the Scw signal, while the Tkv is required for both Dpp and Scw (Neul and Ferguson, 1998; Nguyen et al., 1998).

Smads, the major signal transducers for the TGF-β superfamily

The identification of the *Smad* gene family defines a simple signaling pathway to confer TGF−β signals directly from the receptors to the nucleus (Padgett et al., 1998a; Attisano and Wrana, 1998). The first member of this family, *Mothers against dpp* (*Mad*), was

identified during a genetic screen for maternal effect mutations that enhance the phenotype of weak alleles of *dpp* (Raftery et al., 1995). The gene was cloned through a germline transformation rescue approach (Sekelsky et al., 1995; Newfeld et al., 1996). In *Caenorhabditis elegans*, three genes (*sma-2*, *sma-3* and *sma-4*) which were implicated in a pathway regulated by *daf-4*, a TGF−β receptor homologue, were cloned subsequently (Savage et al., 1996). Comparison of their sequences with *Mad* defined a novel gene family, which was thus renamed as *Smad* to distinguish from unrelated proteins that were also abbreviated as Mad (Derynck et al., 1996).

Subsequently, at least nine vertebrate *Smad* homologues were cloned (Zhang et al., 1996; Chen, Y. et al., 1996; Riggins et al., 1996; Yingling et al., 1996; Padgett et al., 1998b). According to their sequence similarity and function, they were divided into three sub-families: pathway-restricted *Smads*, common-mediator *Smads* and inhibitory *Smads* (Heldin et al., 1997). The C-terminal tail of the pathway-restricted *Smads*, such as *Smad*1/2/3/5 in vertebrates and *Mad* in *Drosophila*, contains a characteristic SSXS motif. Upon ligand stimulation, a specific pathway-restricted *Smad* receives the signal from the specific type I receptor by being phosphorylated at two serines in this SSXS motif (Abdollah et al., 1997; Liu, X. et al., 1997; Souchelnytski et al., 1997). For example, *Smad1/5* are phosphorylated by the *BMP* receptors (Yamamoto et al., 1997; Kretzschmar et al., 1997), while *Smad2/3* are phosphorylated by the *activin* or *TGF-β* receptors (Macias-Silva et al., 1996). The members of the common-mediator *Smads* include *Smad4*

in vertebrates and *sma-4* in *C. elegans*. *Smad4*, which was originally named as *DPC4 (homozygously deleted in pancreatic carcinoma, locus4)*, was cloned as a tumor suppressor gene through positional cloning (Hahn et al., 1996). Smad4 is a central mediator for the functions of *Smads*, possibly participating in signaling pathways of all *TGF-β*, *activins* and *BMPs* (Lagna et al., 1996; Zhang et al, 1997). Upon ligand stimulation, Smad4 proteins associate with pathway-restricted Smads, which have been phosphorylated by specific receptors (Abdollah et al., 1997; Wu et al., 1997; Souchelnytskyi et al., 1997). The heteromeric Smads complexes will then translocate from the cytoplasm to the nucleus and activate downstream target genes (Zhang et al., 1997; Nakao et al., 1997a). The members of the inhibitory Smads include Smad6/7/8 in vertebrates and *Dad* in *Drosophila* (Tsuneizumi et al., 1997). In comparison to the other two Smad sub-families, they contain less conserved sequences at their N-terminal domain. When over-expressed, they inhibit signaling of the TGF−β superfamily, probably by binding to the type I receptor and interfering with the activation of the pathway-restricted Smads (Imamura et al., 1997; Nakao et al., 1997b; Hayashi et al., 1997; Nakayama et al., 1998).

Functional mechanisms of Smad signaling

The studies of the functional domains of Smad proteins provide insights into the mechanisms of this family of TGF-β signal transducer. The members of both the

pathway-restricted Smads and the common-mediator Smads share a highly conserved N-terminal MH1 domain and a highly conserved C-terminal MH2 domain, with a less conserved linker region in between. The conserved MH2 domain mediates the homomeric and heteromeric interactions between Smads proteins (Wu et al., 1997). Transition of the Smad proteins from monomers to oligomers is a critical event in the signal transduction of the *TGF–β* superfamily (Kawabata, et al., 1998). The crystal structure analysis of the Smad4 MH2 domain and cross-linking experiments suggest that the functional Smad proteins could be a trimer or even a hexamer (Shi et al., 1997; Kawabata, et al., 1998).

It was proposed that the MH2 domain is the effector domain of the Smad proteins (Liu, F. et al., 1997), based on two reasons. First, when the MH2 domain of either Smad1 or Smad4 were fused to a yeast GAL4 DNA–binding domain, a BMP4 inducible transcriptional activation activity was detected (Liu et al., 1996). In agreement with this notion, recent evidence indicated that Smad proteins can interact with the transcriptional adaptor CBP/p300 through their MH2 domain to contact the basal transcriptional machinery (Feng et al., 1998; Janknecht et al., 1998). Second, the MH2 domain of Smad2 alone can induce a full range of *activin* responses, while the MH1 domain has an inhibitory effect (Baker and Harland, 1996). This inhibitory effect may result from an intramolecular interaction between the MH1 domain and the MH2 domain. Mutations that increase this internal interaction enhance the auto-inhibition effect, thus inactivating tumor suppressors Smad2 and Smad4 (Hata et al., 1997). Upon ligand stimulation,

hetero-oligomization through the MH2 domains could open this internal interaction, thus releasing the inhibitory effect.

Much progress has been made recently on how Smad proteins are activated by the type I receptor and how they form hetero-oligomers and translocate from cytoplasm to nucleus. However, it has remained unclear how Smad proteins activate downstream target genes. One breakthrough on how Smad proteins function at the transcription level came from the cloning of *FAST-1 (forkhead activin signal transducer-1)* from *Xenopus* (Chen, X. et al., 1996). FAST-1 is a novel winged-helix protein which was cloned by yeast one-hybrid screening for transcription factors that could bind to an activin responsive element (ARE) from the *Mix.2* promoter. FAST-1 can interact with both Smad2 and Smad4 to form an activin-responsive factor in a ligand regulated fashion (Chen et al., 1997). Therefore, one possible mechanism for Smad proteins to activate target gene transcription is through recruiting a transcription factor like FAST-1.

Signaling specificity of the TGF-β superfamily

There are two challenging problems when studying the signaling pathway which interprets the complex functions of TGF-β. The first is to explain how TGF-βs function as morphogens, which are defined by their ability to specify multiple cell types in a concentration-dependent manner. This capacity has been shown for several members of

the TGF-β superfamily including Dpp in *Drosophila* (Ferguson and Anderson, 1992; Wharton et al., 1993). In *Xenopus*, BMP4 can evoke distinct responses in ectoderm cells at high vs. low concentrations. Interestingly, different amounts of Smad1 protein accurately reproduce the responses to different activities of the BMP4 protein (Wilson et al., 1997). It was thus proposed that the extracellular BMP4 gradient functions through establishing an intracellular and then nuclear gradient of Smad activity within neighboring cells.

The second challenging problem is to explain how the members of TGF-β superfamily evoke different responses with spatial-temporal specificity. At the receptor level, it has been shown that the combination of different type I/II receptors may determine the affinity and specificity of ligand binding (Derynck and Feng, 1997; Massague 1998). The type I receptors appear to be the primary determinants of the signaling specificity (Carcamo et al., 1994). At the signal transducer level, the specificity seems to be determined by the pathway-restricted *Smads*. The L3 loop of the pathway-restricted Smad MH2 domain and the L45 loop of the type I receptor kinase domain determine the specificity of receptor-Smad interactions (Chen et al., 1998). Swapping two amino acids inside the L3 loop causes a switch in the phosphorylation of Smad1 and Smad2 by the BMP and TGF–β receptors, respectively (Lo et al., 1998). In summary, a pathway-restricted Smad protein in the cytoplasm awaits instructions from a specific TGF-β type

I receptor and then confers the information to the nucleus. For example, in *Xenopus*, Smad2 receives the signal from *activin* receptors and induces the formation of the dorsal mesoderm, while Smad1 transduces the signal from BMP2/BMP4 and induces the formation of the ventral mesoderm (Graff et al., 1996). However, pathway-restricted *Smads* can not be the only explanation for the functional specificity of the *Smads* as signal transducers. As have been summarized above, there are multiple functions for *dpp* during *Drosophila* embryogenesis. Nevertheless, a single pathway-restricted Smad protein (Mad) plays a pivotal role in many signaling events induced by Dpp (Raftery et al., 1995; Sekelsky et al., 1995; Newfeld et al., 1997). It remains to be investigated how Mad regulates a specific target gene at the level of transcription during these diverse processes.

The identification of the Dpp-responsive *tin*-D element provides an excellent opportunity to address questions on the molecular mechanisms of Dpp signaling in the context of a specific target gene, *tinman*. In the course of this study, the common-mediator *Smad* in *Drosophila*, *Medea*, was cloned as a *tin*-D binding protein. Additional evidence confirms that Smad proteins can bind to DNA. Based on the detailed analysis of this Dpp-responsive element, a model was proposed to explain the tissue specificity of Dpp signaling during *Drosophila* mesoderm induction.

MATERIALS AND METHODS

Construction of P-transformation plasmids

1. Series of constructs to define the dorsal element *tin*-D

The series of fragments located downstream of the *tinman* gene shown in Figure 5 were generated by PCR. The sequences of the primer pairs are listed below. The fragments in *tin*-D-#1~#4 were amplified by PCR using KS-*tin*-D(1-225) as a template, which is the original dorsal element identified by Zhizhang Yin in this lab. The PCR products were digested to generate EcoRI site at their 5' end and BamHI site at their 3' end. The fragments were then cloned into EcoRI/BamHI sites of the pCasperhs43 vector and the pBluescriptKS+ vector. The inserts in *tin*-D-#5~#13 were amplified by PCR using phage 31/3 which contains the larger flanking region around the original *tin*-D(1-225) element. For constructs #5/#6, #12/#13, the PCR products were treated with Klenow to generate blunt ends and then cloned into CIP treated SmaI site of the pBluescriptKS vector. For *tin*-D-#8/#9, #10/#11, the PCR products were first cut with BamHI and then treated with Klenow polymerase to generate blunt ends. The resulting mixture were loaded into an agarose gel and the bands corresponding to 240bp and 180bp fragments were recovered and cloned into the SmaI site of pBluescriptKS vector. The directions of the inserts in the KS vector were determined by restriction enzyme digestion. The inserts were then cut out with NotI/XhoI to be cloned into the NotI/XhoI sites of the pCasperhs43 vector.

For the construct *tin*-D-#7, the PCR product was digested with EcoRI/XhoI and then cloned into the EcoRI/XhoI sites of the pCasperhs43 vector and pBluescriptKS, respectively. These two constructs were later renamed as pCasper-*tin*-D and KS-*tin*-D. The PCR primer pairs used to generate these constructs are:

#1:	200RIA	(GAGAATTCATGTCAAGTGGCACTA)
	& 200BA	(ACGGATCCGTAATTGTTTATGGGT)
#2:	RB306R	(from Yin Z.)
	& 200BB	(CGGGATCCATGTTTAGTGCCACTTG)
#3:	200RIB	(GCGAAGGCGGAATATGAATATGC)
	& M13F	
#4:	200RIA & M13F	
#5/6:	200RIA & BuXho2	(CACTCGAGCAATTGTAGCGTTTTCATAC)
#7:	200RIA & BuXho1	(CACCTCGAGGTGGGAGGCTCGCAGCT)
#8/9:	200RIB & BuXho2	
#10/11:	RIB & BuXho1	
#12/13:	BuXho2 & Bu300	(TCGTGGGCATTTCCTCAATG)

2. Constructs to define insulators

The stripe 2 and stripe 3 enhancer from *even-skipped* promoter were gifts from Dr. M. Levine, UCSD. The stripe 2 enhancer, a 480bp element (-1.55~-1.1kb), was in the SK-Str.2 construct, which had been generated by cloning the EcoRI/BssHII(blunted) fragment from the *even-skipped* promoter into EcoRI/SmaI sites of SK+ vector. The stripe 3 enhancer, a 500bp element (-3.8~-3.3kb), was in the SK-Str.3 construct, which had been generated by cloning the BamHI(blunted)/SacI(blunted) fragment into the SmaI site of the SK+ vector. To make the constructs listed in Figure 7, an intermediate construct pCasper-St.2 was first generated by cloning the EcoRI/BamHI fragment from the SK-St.2 into the EcoRI/BamHI sites of the pCasperhs40 vector. The control construct was then

generated by cloning the XhoI/BamHI insert from SK-St.3 into the BamHI/XhoI sites of the intermediate construct. The constructs 8.1 and 8.2 were generated by a two fragment ligation: a ~500bp XhoI/BamHI fragment from SK-St.3 and a ~350bp BamHI/XhoI fragment from the KS-*tin*-D-#8 were ligated together into the CIP treated BamHI site of the pCasper-St.2 intermediate construct. The transformants that contained two fragments were identified by BamHI/XhoI double digestion. The directions of the inserts were determined by XhoI digestion. Similarly, the constructs 9.1 and 9.2 were generated by two fragment ligation with a ~500bp XhoI/BamHI fragment from SK-St.3 and a ~350bp BamHI/XhoI fragment from the KS-*tin*-D-#9. The insert from either KS-*tin*-D-#8 or KS-*tin*-D-#9 is fragment B+C (see Figure 5), although in opposite direction.

3. Internal deletion derivatives of *tin*-D

Internal deletion derivatives in Figure 10 were generated using PCR-based *in vitro* mutagenesis. The sequences of the primer pairs are listed below. *tin*-D-ΔD6 was generated from the construct *tin*-D-#5 in pBluescript KS+ vector. A fragment was amplified by PCR using the primer pair 200RIA and 6A. The PCR product was then digested with EcoRI/XhoI and cloned into the EcoRI/XhoI sites of the KS vector to generate KS-*tin*-D-ΔD6. To generate the construct *tin*-D-ΔD1, D1a was deleted first by digesting the PCR product amplified by primer pair 1A and BuXho1 with EcoRI/XhoI, which was then cloned into the EcoRI/XhoI sites of the KS vector. The D1b sequence was deleted subsequently with a PCR based *in vitro* mutagenesis kit (ExSite™ PCR-

Directed Mutagenesis Kit, Stratagene) using primers that flank the deleted sequences. Constructs *tin*-D-ΔD2~ΔD5 were generated from KS-*tin*-D using the same *in vitro* mutagenesis kit as above. Upon sequence confirmation of the deletion constructs in the KS vector, the mutated fragments were cloned into the NotI/XhoI sites of the pCasperhs43 vector.

The primer pairs used to generate these constructs are:

D1a: 1A (CAGAATTCACTAAACATGACCTAATG) & BuXhoI
D1b: 1B (AGAGCTGCGCCGGTCTAG) & 1C (AGAATTTGTCATTCCCAT)
D2: 2A (GTGCCACTTGACATGCAT) & 2B (CACGCCCGTCGCGAAGG)
D3: 3A (AAAGACGACACCGTAATTG) & 3B (ACGTTTGCCCCTCCGTCG)
D4: 4A (CTCCACTTGAGAGCTGCG) & 4B (CGAAGGATCGGGGAGTTG)
D5: 5A (CAACTCCCCGATCCTTCG) & 5B (GTTCAAGAGCCGCTGTCG)
D6: 6A (CCCTCGAGGTTTGATCCTTTTCGGTA) & 200RIA

4. Constructs with multiple-copy sub-elements for P-transformation

To generate the tandemly aligned multiple-copies of sub-element D1, D3, and D6, oligonucleotide pairs were designed to have asymmetric AvaI sites at both ends (the sequences are shown below). The annealed oligonucleotides were ligated with a AvaI/XbaI digested 489S vector (provided by Dr. T. Lufkin). The asymmetric AvaI site will guarantee the addition of the sub-elements in one direction. The resulting linear DNA mixture were treated with Klenow polymerase, blunt-end ligated, and transformed into XL-1 bacteria competent cells. The copy number of positive clones was determined by sequencing. Four copies of D1, D6, and five copies of D3 were generated in this way, namely 489S-$(D1)_4$, 489S-$(D3)_5$, 489S-$(D6)_4$. The inserts were then excised with

EcoRI/BamHI and cloned into pCasperhs43 vector, generating constructs pCasper-$(D1)_4$, pCasper-$(D3)_5$ and pCasper-$(D6)_4$ for P-transformation.

The combinatory P-transformation construct $(D1)_4$+$(D3)_5$ was generated by cloning the BamHI/SalI insert from 489S-$(D3)_5$ into the BamHI/XhoI sites of the pCasper-$(D1)_4$. The construct $(D3)_5$+$(D6)_4$ was generated analogously by cloning the BamHI/SalI insert from 489S-$(D6)_4$ into the pCasper-$(D3)_5$. The orientations were $(>>D1>>)_4$ - $(>>D3>>)_5$ /lacZ and $(>>D6>>)_4$ _ $(>>D3>>)_5$ /lacZ (where > denotes 5' to 3').

The oligonucleotide pairs used to generate the above constructs are the following. The sequences in *tin*-D that were covered by these oligonucleotides were indicated as black boxes in Figure 11. Sub-element D1 includes two separate Tinman binding sites: D1a and D1b.

D1: TCGGGTGTCAAGTGGCATCTCAAGTGGAG
& CACAGTTCACCGTAGAGTTCACCTCAGCC
D3: TCGGGTTTCAATGTCGGCGGCAATGTTGCGGCGACG
& CAAAGTTACAGCCGCCGTTACAACGCCGCTGCAGCC
D6: TCGGGAGCCGCTGTCGCAGCTGCGAGCCTCCCAC
& CTCGGCGACAGCGTCGACGCTCGGAGGGTGAGCC

5. *tin*-D* and its derivatives with internal mutations

The strategy to make the series of constructs with elements carrying base pair changes in D3 and D6 sub-elements shown in Figure 22 was to add two sets of oligonucleotides (see below for sequences) to pCasper-*tin*-D- #1. The pCasper-*tin*-D-#1, which is shown

schematically in Figure 5, covers nucleotides 1-100 of the *tin*-D sequence. The first set of oligonucleotides, E6 and e6, include the second Tinman binding site (nucleotide 195-204) and the essential D6 sub-element, which contains the two 3' most Mad/Medea binding sites (nucleotide 316-349 of *tin*-D) (see below and Figure 22). E6 corresponds to the wild type sequence while e6 contains mutations in two GC-rich motifs of the Mad/Medea binding sites, M6 and M7 (see below). The complementary strains were annealed and digested with BamHI/XhoI. The resulting fragments were cloned into the BamHI/XhoI sites of pCasper-*tin*-D-#1 to generate two intermediate constructs, pCasper-*tin*-D-#1-D6 and pCasper-*tin*-D-#1-d6. The second sets of oligonucleotides, E3(#94-#143, wild type), e3g..c (mutated in the GC-rich motifs) and e3c..t (mutated in two CAATGT motifs), covered the essential sub-element D3 (see below and Figure 22). The complementary strains were annealed and digested with BamHI. The resulting fragments were then cloned into the CIP treated BamHI site of the pCasper-*tin*-D-#1-D6 and pCasper-*tin*-D-#1-d6 constructs, respectively. The native orientations were verified by PCR reaction, using the sense strand of the second sets of oligonucleotides as 5'-primer, together with the "HS-upstream" oligo located in the basal promoter region of the pCasperhs40 vector as 3'-primer. Six constructs with different combination were thus generated. The construct generated from *tin*-D-#1 with E3 and E6 was named as *tin*-D*. The construct generated with e3g..c and E6 was named as *tin*-D*-d3g..c. Analogously, *tin*-D*-d6, *tin*-D*-d3,d6 and *tin*-D*-dc..t were named. The sequence of these constructs were shown in Figure 22. The structure of these constructs is summarized schematically in Figure 24.

The sequences for the sense strains of these two sets of oligonucleotides are as follows. The sequences in *tin*-D that are covered by these oligonucleotides are indicated as shadowed sequences in Figure 22. The mutated residues are underlined.

E6:
GGATCCTCAAGTGGACAAGAGCCGCTGTCGCAGCTGCGAGCCTCCCACCTC
GAG
e6:
GGATCCTCAAGTGGACAAGAGTATCTATCGCAGCTGCGATATTCACACCT
CGAG

E3:
GGATCCATTACGGTCTCGTCTTTCAATGTCGGCGGCAATGTTGCGGCGACG
TTTGCAGATCT
e3c..t:
GGATCCATTACGGTCTCGTCTTTTACTACCGGCGGTACTACTGCGGCGACG
TTTGCAGATCT
e3g..c:
GGATCCATTACGGTCTCGTCTTTCAATGTATGCATCAATGTTATGGATACG
TTTGCAGATCT

One-hybrid screening

The MATCHMAKER One-Hybrid System (Clontech) was used. The protocols are based on the Clontech protocol (PT1031-1) with some modifications.

1. Construction of the reporter plasmids

In this study, four copies of D1 and five copies of D3 were used to increase the sensitivity of the screening. There were two *HIS3* reporter vectors available: pHISi-1 and pHISi. The former has a lower background expression of *HIS3* and thus was further

used in this study. To help to eliminate false positives, another reporter vector, pLacZi, was included in the same yeast reporter cells.

The EcoRI/BamHI inserts from the 489S-$(D1)_4$ and 489S-$(D3)_5$ were cloned into the pBluescript KS+ vector. The fragments were then recloned as EcoRI /XbaI inserts into the pHisi-1 vectors, generating reporter constructs $(D1)_4$/His and $(D3)_5$/His, respectively. The EcoRI/BamHI inserts from the pCasper-$(D1)_4$ and pCasper-$(D3)_5$ were cloned into the EcoRI/BamHI sites of the pLacZi vector, generating reporter constructs $(D1)_4$/LacZ and $(D3)_5$/LacZ respectively.

2. Integrating the reporter constructs into the yeast strain YM4271

The reporter strains for the screening were generated by integrating *HIS3* reporter constructs first into the genome of the yeast strain, YM4271, using selection medium SD/-Histidine. pLacZi reporter plasmids were then integrated using Uracil as a selection marker.

The competent cells were prepared as the following: A single colony of YM4271 from a fresh plate was inoculated into 50ml YPD medium. The flask was incubated at 30°C for 16-18 hr. with shaking at 250 rpm until reaching the stationary phase (OD600>1.5). The yeast cells were inoculate into fresh 300ml YPD medium to produce an initial OD_{600}=0.2-0.3 and incubated at 230 rpm for 3hr. at 30°C. The yeast culture was centrifuged in 50ml

tubes at 1000Xg for 5 minutes and the pellets were washed with sterile TE. The pellets were suspended in 1ml TE/LiAc buffer (10mM Tris pH7.5, 1mM EDTA, 100mM Lithium acetate). The competent cells were then ready to be used.

To be integrated into the yeast genome, the reporter plasmids should be digested with an unique restriction enzyme before transformation. pHisi-1 constructs were digested with XhoI, while pLacZ constructs were digested with NcoI. 1ug of digested reporter plasmid in 20ul volume, together with 100ug salmon sperm DNA, was mixed together with 100ul yeast competent cells. 600ul PEG/LiAc buffer (40% PEG 4000, 10mM Tris pH7.5, 1mM EDTA, 100mM Lithium acetate) was added and mixed well by vortexing. The tube was incubated with shaking at 30°C for 30min. 70ul DMSO was added and mixed well. The cells were heat shocked at 42 °C for 15 min. and then chilled on ice for 2min. After having been spinned down at top speed for 10sec., the pellets were resuspended in 150ul YPD medium. The cells were then plated on SD/-His selective medium for pHisi-1 constructs or SD/-Ura selective medium for pLacZi reporter constructs. The plates were incubated at 30 °C for 4-6 days.

3. Determination of 3-AT concentration.

3-aminotriazole (3-AT), a competitive inhibitor of the yeast HIS3 protein, was used to inhibit the background level activity of *HIS3*. An appropriate concentration of 3-AT should be determined for each pHisi-1 reporter strain before the screening. Upon library

transformation, positive clones use the GAL4 activation domain to activate the *HIS3* gene transcription and overcome the inhibition by 3-AT, and thus survive the selection medium SD/-His/+3-AT. The appropriate concentration of the 3-AT should be high enough to inhibit the growth of the background colonies, while still be sensitive enough to allow positive clones to survive the selection.

To determine the appropriate concentration of 3-AT, colonies of the yeast reporter strains were inoculated using toothpicks to the SD/-His selection medium plates supplemented with 0mM, 15mM, 30mM 45mM and 60mM 3-AT, respectively. The plates were grown for 4-6 days to observe the inhibition of basal growth. 15mM 3-AT was just sufficient to inhibit the growth of the yeast reporter strain $(D3)_5$/His. 45mM 3-AT was required to inhibit the yeast reporter strain $(D1)_4$/His. These concentrations of 3-AT were then used for the screening of the libraries.

4. Amplification of the libraries

Three *Drosophila* Gal4 activation domain-fused cDNA libraries were used in this study. Library #1 is a phage library in λACT vector with 10^7 complexity (gift from Dr. Leslie Pick, Mount Sinai Medical School). The inserts came from mRNA extracts of 0 - 6 hrs. staged embryos. The phage library was amplified and converted to a plasmid library according to a protocol from Dr. Yan Yu's dissertation (Yu, 1996). During the amplification, the library was plated on five 150-mm plates to culture the bacteria. At

this density, the bacteria are in confluent state, which could potentially result in biased amplification.

Library #2 is a 0 - 16 hrs. embryonic library in pGAD10 vector (Clontech). It is a plasmid library with a complexity of $2X10^6$. The library was in a form of being transformed into bacteria, whose titer was determined to be $1.2X10^8$ cfu/ml. To avoid the biased amplification, around $2X10^6$ isolated colonies were generated by distributing 20ul bacteria to one hundred fifty 150mm LB-Amp plates. After overnight growth at 37°C , the bacteria were scraped into 2L LB with Amp and then incubated at 37°C for 3hr with shaking. The bacteria were collected by centrifugation at 5000rpm for 15 min. and then divided into 6 bottles for plasmid extraction. Standard large-scale alkaline lysis method was used (Maniatis et al., 1982). The plasmids were purified using CsCl/EtBr gradient ultracentrifugation method.

Library #3 is a 3 - 12 hrs. embryonic library in pGADNot vector (gift from Dr. Richard Mann, Columbia University). It is a plasmid library with a complexity of $2X10^5$, which was ready to be transformed into the yeast reporter strains.

5. Library screening

The *HIS3* reporter gene was used as the first selection marker during the screening in the presence of appropriate concentrations of 3-AT. 50ug library DNA and 2mg salmon

sperm DNA were mixed with 1ml competent reporter yeast cells in a 50ml falcon tube. 6ml LiAc/PEG buffer was added and mixed well by vortexing for 10sec. The tube was incubated at 30 °C for 30min. with shaking. 700ul DMSO was added and mixed. The cells were heat shocked for 15min. in 42 °C and then chilled on ice for 2min. The cells were collected by centrifugation at 1000Xg for 5min and then resuspended in 2X25ml of SD/-His liquid medium in two 50ml tubes. The tubes were incubated at 30 °C for 1hr. at 200rpm. The pellets from the two tubes were collected by centrifugation and pooled together in 7ml TE buffer. 500ul cells were plated per 150-mm plates containing SD/-Leu/-His with appropriate concentrations of 3-AT. The plates were incubated at 30 °C for 5-10 days. Typically around 120 large colonies appeared during this time.

The pLacZi reporter gene was used for the second step of the screening. This step will typically eliminate ~30% false positive colonies. The large colonies from the screening were inoculated using toothpicks to a new 150mm plate for the β-galactosidase assay. A 125-mm Whatman filter was presoaked in a 150-mm plate containing 3.5ml Z buffer/X-gal solution (16.1g/L $Na_2HPO_4 \cdot 7H_2O$, 5.5g/L $NaH_2PO_4 \cdot H_2O$, 0.75g/L KCl, 0.246g/L $MgSO_4 \cdot 7H_2O$, 0.27% β-mercaptoethanol, 0.335mg/ml X-GAL). Another dry filter was placed over the surface of the agar plates containing the transformants. The relative position of the filter to the plate was oriented by poking unsymmetrical holes. The filter was frozen in liquid nitrogen for 10 sec. and then thawed at room temperature. This filter

was placed carefully on the presoaked filter with the colony side up. The filters were incubated at room temperature for 0.5-8hr. The positive colonies will turn blue.

6. Plasmid dependency tests

The survival of some colonies does not depend on the existence of the transformed plasmids from the library. Theoretically, a drop-out selection procedure was performed first to eliminate these false positive clones. The yeast colonies were grown in a rich medium to eliminate the transformed plasmids. The yeast clones without the plasmids should loss their resistance to the appropriate concentration of 3-AT which has been used for screening. They also should be negative for lacZ activity. However, it was tedious to get pure clones in which the transformed plasmid was eliminated completely. Therefore, I directly used the re-transformation test to test the plasmid dependency of the positive clones. The plasmids from the positive yeast clones were isolated and transformed into competent bacteria cells. The bacterial plasmids were re-transformed into the corresponding fresh reporter yeast strains. The transformants were then tested for their ability to grow at plates with appropriate concentration of 3-AT and for their β-galactosidase activity using the same transformation procedure as above. In this case, the plasmids were not integrated into the genome and were not digested with the restriction enzymes before transformation. This selection procedure usually eliminated ~60% false positive plasmids.

The yeast plasmid DNA was extracted as follows. The yeast cells were grown in 3ml SD/-His/-Leu liquid medium overnight until saturation. 1.5 ml cultures was spinned down in an eppendorf tube. The pellet was resuspended in 250ul yeast digestion mix (0.1M EDTA, 14mM β-mercaptoethanol). 12.5 ul zymolyase (15mg/ml in water) was added and the mixture were incubated at 37 °C for 30min. 50ul yeast miniprep mix (0.25M EDTA, 0.5 M Tris pH 8.0, 2.5% SDS) was added and the mixture were incubated at 65°C for 30min. 63ul 5M potassium acetate was added. The tube was inverted to mix the solutions and then incubated on ice for 30min. The pellets were spinned down at top speed for 15min. The supernatant was transferred to a new tube. 720ul ethanol was added to precipitate the DNA. The pellets was resuspended in 120ul TE. 5ul yeast extracted DNA was used to transform bacteria XL1 competent cells. The standard miniprep method was used to amplify the plasmids (Maniatis et al., 1982).

7. Sequencing and further tests

The plasmids from the clones which had passed the re-transformation test were sequenced from both 5' end and 3' end. BLAST search was then performed to identify homologous known sequences in GenBank. Based upon their sequences, ~30% plasmids turned out to be false positive, since their cDNAs encoded mitochodrial, ribosomal or membrane proteins, etc. The positive plasmids were labeled as RNA or DNA probes for *in situ* hybridization. Only the plasmids which were expressed in areas including the

dorsal mesoderm region were pursued further. They were cloned to generate GST-fused proteins to confirm their *in vitro* DNA binding activity to the *tin*-D element.

When $(D1)_4$ was used as a bait, the library #1 was screened. 55 colonies were positive in β-galactosidase assays. After the re-transformation test, only two plasmids remained positive. Sequencing results indicated that the inserts from both plasmids were identical *tinman* cDNAs containing the entire homeo box region.

When $(D3)_5$ was used as a bait, all three libraries were screened. In β-galactosidase assays, 70 colonies from Library #1, 118 colonies from Library #2 and 60 colonies from Library #3 were positive. 21 plasmids from library #1, 31 plasmids from Library #2 and 24 plasmids from Library #3 passed the re-transformation tests. When these plasmids were sequenced, 55 potentially interesting cDNAs were identified and the results are summarized in Figure 13.

Yeast assays for lacZ activities

The procedure is based on a protocol from Steven Hanes (SUNY-Albany). Fresh yeast cells were inoculated into 3ml selective culture medium and grown at 30 °C to saturation. These cultures were used to start new cultures of 3ml at an OD_{600} of 0.1, which were grown at 30 °C to a final OD_{600} of ~0.4. After chilling on ice the OD_{600} was measured,

1ml of the cells were pelleted, and the resuspend pellets were vortexed in 200 µl 0.1M Tris pH7.5/0.05%Triton X-100. After freezing on dry ice, cells were thawed on ice. 1ml ONPG solution (16.1g/L $Na_2HPO_4 \cdot 7H_2O$, 5.5g/L $NaH_2PO_4 \cdot H_2O$, 0.75 g/L KCl, 0.246g/L $MgSO_4 \cdot 7H_2O$, 0.8g/L ONPG, 1.25mM DTT, 0.00625%SDS, 0.27% β-mercaptoethanol) was added and the suspension was incubated at 30 °C. When the color reaction appeared to have reached maximum levels, reaction times (t) were recorded and 500µl 1M Na_2CO_3 were added. After centrifugation, the OD_{420} were taken from the supernatants. The lacZ activities of five independent clones were measured for each construct and calculated with the formula:

β-Gal units = $(OD_{420} \times 1000) / [OD_{600} \times t \text{ (min.)}]$

Construction of plasmids to generate derivatives of *Medea*

The cDNA insert from the isolated GAL4 fusion clone 3.15 (pGAD-*Medea*A_{1-681}) was used to isolate full-length cDNA clones from a 4 - 8 hrs. embryonic cDNA library (Brown and Kafatos, 1988). Two classes of positive clones were obtained by Zhizhang Yin and were sequenced in their entirety. These two classes represent two splicing products of *Medea*. The clone #14 (pNB40-*Medea* A_{1-692}) is a truncated form of *Medea A*, while the clone #29 (pNB40-*Medea* B) corresponds to the *Medea* B product, which contains an internal deletion compared with *Medea* A (see Figure 15). The XmnI/BglII insert of the clone 3.15 (pGAD-MedeaA_{1-681}) was cloned into BamHI(blunted)/BglII sites

of the pGAD424 vector to generate pGAD424-3.15. These clones were used to generate the following derivatives in pGAD vector.

pGAD-*Medea* A_{1-692}: A 1.9 kb fragment cut out from NotI partial/PstI digestion of cDNA clone #14 (pNB40-*Medea* A_{1-692}) was cloned to substitute PstI/NotI fragment in pGAD424-3.15.

pGAD-*Medea* A_{1-505}: The PstI/SmaI insert from clone #14 (pNB40-*Medea* A_{1-692}) was cloned to substitute PstI/NotI(blunted) fragment in pGAD424-3.15.

pGAD-*Medea* A_{1-265}: An intermediate construct, KS-*Medea*-3.15, was first generated by cloning the XmnI/NotI insert from pGAD-*Medea* A_{1-681} into SmaI/NotI sites of the KS+ vector. The BamHI/KpnI (blunted) insert from KS-*Medea*-3.15 was then cloned into BamHI/EcoRI(blunted) sites of pGAD10 vector.

pGAD-*Medea* A_{1-303}, pGAD-*Medea* A_{1-65}: Nested 3'-deletions were generated by DNaseI treatment of pGAD424-3.15, followed by NotI digestion, Klenow treatment and ligation. The plasmids that did not contain truncations were eliminated by SalI digestion. The mixed products were then transformed into bacteria and the individual clones were sequenced. Two of the clones contained truncations at residues 303 and 65, respectively.

pGAD-*Medea* $A_{167-505}$: Nested 5'-deletions were generated by DNaseI treatment of KS-*Medea*-3.15 with the similar procedures as above. One of the clone, KS-*Medea* $A_{167-681}$, was cut with EcoRI(blunted)/XhoI, and the insert was cloned into the

BglII(blunted)/SalI sites of a modified pGAD424 vector. The reading frame of this vector had been adjusted by cutting with BamHI, filling the end with Klenow polymerase and re-ligation.

pGAD-*Medea* $A_{506-692}$: The SmaI/BglII insert from pGAD-*Medea* A_{1-692} was cloned into the SmaI/BglII sites of a modified pGAD424 vector. The reading frame of this vector had been adjusted by cutting with EcoRI, filling the end with Klenow polymerase and re-ligation.

pGAD-*Medea* B_{1-432}: The PstI/SmaI insert from clone #29 (pNB40-*Medea* B) was cloned to substitute the PstI/NotI (blunted) fragment in pGAD424-3.15.

pGAD-*Medea* B: pGAD-*Medea* A was first generated by cloning the NotI insert from clone #29(pNB40-*Medea* B) into the NotI site of pGAD424-3.15. The direction of the insert was confirmed by PstI digestion. The PstI insert from clone#29(pNB40-*Medea* B) was then cloned to substitute the PstI fragment in pGAD-*Medea* A to generate pGAD-*Medea* B. The direction of the insert was confirmed by NotI digestion.

Construction of plasmids to generate bacterially expressed proteins

For the purpose of *in vitro* DNA binding experiments, the following constructs were made to generate bacterially expressed proteins.

GST-*Medea*$_{MH1+L}$: The plasmid encoding the GST-fused MH1 domain and linker region of Medea protein was generated by cloning the XmnI/EcoRI fragment from pGAD-*Medea* A_{1-681} into the SmaI/EcoRI sites of pGEX-3X vector.

GST-*Medea*$_{MH1}$: The plasmid encoding the GST-fused MH1 domain of Medea was generated by cloning the BamHI/Kpn(blunted) fragment from KS-*Medea*-3.15 into BamHI/SmaI sites of pGEX-2T vector.

GST-*Mad*$_{MH1+L}$: To generate GST-fused Mad proteins, a 200bp fragment of the *Mad* gene was amplified by PCR from *Drosophila* genomic DNA using the primer pair: CAGGATCCACACCGACAGCAGCGCGATG (which deletes 5 codons at the N-terminus) and CCAGACTGTCGACGGCCTTC. The fragment was cloned into BamHI/SalI sites of KS+ vector and used to isolate the full-length *Mad* cDNA from a *Drosophila* cDNA library (Brown and Kafatos, 1988). The BamHI/SalI digested 200bp PCR product and the SalI/NotI fragment of *Mad* cDNA(in pNB40 vector) were cloned together into BamHI/NotI sites of KS+ vector to generate an intermediate construct KS-*Mad*. The plasmid encoding GST-fused MH1 domain and linker region of Mad protein was then generated by cloning the BamHI/PvuII(blunted) fragment from KS-*Mad* into the BamHI/SmaI sites of pGEX-3X vector.

GST-*Mad*$_{MH1}$: The plasmid encoding GST-fused MH1 domain of Mad protein was generated by cloning the BamHI/EcoRI(blunted) fragment from KS-Mad into the BamHI/SmaI sites of pGEX-3X vector.

GST-*tinman*: The plasmid encoding GST fused Tinman protein was generated by cloning the BamHI/EcoRV fragment from pQE-*tinman*(NK-4-6) and the EcoRV/EcoRI fragment from *tin* cDNA(NK4-8) into the BamHI/EcoRI sites of pGEX-3X vector.

GST-*tinman*-CT: Generated by Dr. S. Zaffran in this lab. Contains amino acids 301~416

GST-*tinman*-NT: Generated by Dr. S. Zaffran in this lab. Contains amino acids 1~124

GST-*tinman*-ΔTN: Generated by Dr. S. Zaffran in this lab. Deletes amino acids 29~42

GST-*twist*: Gift from Dr. Ip, Y. T., UCSD.

GST-*snail*: Gift from Dr. Ip, Y. T., UCSD.

His-*schnurri*: The plasmid encoding His-tagged Schnurri was provided by Dr. Rahul Warrior, USC (Arora et al., 1995). The insert is a 908bp NcoI fragment covering the second set of the zinc-fingers which was cloned into the NcoI site of pRSET B vector.

GST-*FAST-1*: The plasmid encoding GST-fused FAST-1 was provided by Dr. Malcolm Whitman, Harvard Medical School (Chen, X. et al., 1996). The insert is a MscI fragment (amino acid 40-364) covering the FAST-1 folkhead domain which was cloned into the SalI (filled-in by Klenow) site of pGEX-4T-1 vector.

GST-*dSmad2*$_{MH1+L}$: The plasmid encoding the GST-fused MH1 domain and linker region of the dSmad2 protein was generate by Zhizhang Yin in this lab. The insert covers amino acids 1-378.

GST-*dSmad2*$_{MH1+L}$(mut): To substitute amino acids 69-77 of dSmad2, an intermediate construct, KS- *dSmad2* A (contains amino acids 43-378) was generated by cloning the EcoRI/NotI insert from GST-*dSmad2* $_{MH1+L}$ into the EcoRI/NotI sites of the KS+ vector. *In vitro* mutagenesis was then performed using a PCR approach. PCR primer pair NM-L2 and M13 reverse primer were used to amplify a cDNA fragment encoding aa. 43-73. NM-L2 was designed to generate mutations in aa. 69-73 and to have a PstI site at the end of the PCR product. PCR primer pair NM-R2 and M13 forward primer were used to amplify another cDNA fragment encoding aa. 73-378. NM-R2 was designed to generate mutations in aa. 73-77 and also to contain a PstI site at the end of the PCR product. The PstI sites were designed to guarantee in frame ligation. The two PCR products were digested with EcoRI/PstI and PstI/NotI, respectively. The two inserts were then cloned together into the EcoRI/NotI sites of the KS+ vector to generate KS-*dSmad2* A(mut). Upon sequence confirmation of the full-length insert, the EcoRI/NotI insert from KS-*dSmad2* A(mut) was cloned back to substitute the EcoRI/NotI fragment in GST-*dSmad2*$_{MH1+L}$, generating GST-*dSmad2*$_{MH1+L}$(mut). The sequences of the two primers are:

NM-L2:: GCTGCAGTCGTCCGTCCAGGCTGCGCGGCACCGTCACGCAC

NM-R2:: GCCTGCAGGTCTCCCATCGCAAGGGTCTGCCGCACGTC

GST-*HMG-D*: The pGAD-*HMG-D* plasmid identified from the one-hybrid contained stop codons before the translation start site. The protein-coding region was thus

re-amplified by PCR reaction. The primers were designed to have a BamHI site at the 5' end and a SmaI site at the 3'-end. The PCR product was digested with BamHI/SmaI and then cloned into the BamHI/SmaI sites of the pGEX-2T vector. The insert was confirmed by sequencing. The sequences of the primers are:

HMG-D5': GCGGATCCATGTCTGATAAGCCAAAAC

HMG-D3': GGCCCGGGATGAGCCAGTCTACTC.

GST-*DEAF-1*: The full-length coding sequence came from the pET-DEAF-1 clone, which was provided by Dr. William McGinnis, UCSD (Gross and McGinnis, 1996). The plasmid encoding GST-fused DEAF-1 protein was generated by cloning the NdeI(Blunted)/NotI insert from pET-DEAF-1 into the SmaI/NotI sites of the pGEX-5X-1 vector.

GST-*Adf-1*: The full-length coding sequence came from the pET-3a-Adf-1 clone, which was provided by Dr. Robert Tjian, UC Berkly. The plasmid encoding GST-fused Adf-1 protein was generated by cloning the NdeI(Blunted)/NotI insert from pET-3a-Adf-1 into the SmaI/NotI sites of the pGEX-5X-1 vector.

GST-*ccf-β*: The clone 3.19 isolated from yeast one-hybrid screening using D3 as a bait was later renamed as pGAD-*ccf-β*. The BamHI/NotI insert was cloned into the BamHI/NotI sites of the pGEX-5X-1 vector to generate the plasmid encoding GST-fused ccf-β protein.

Extraction of GST-fusion protein

The plasmids encoding the GST-fused proteins were transformed into BL21(DE3) lysS competent cells. A single colony was inoculated in 2ml 2XYT medium with 150ug/ml Ampicillin and 32 ug/ml Chloroamphenicol. After being cultured in a 37°C shaker for 3-5hr., the bacteria were transferred to 200ml 2XYT liquid medium containing 150ug/ml Ampicillin and 20mM glucose, which were further grown in 37°C shaker until OD_{600} reached 0.4-0.6. The bacteria were spin down at 5000rpm for 15 minutes. The pellet was resuspended in 200ml of fresh 2XYT medium containing 150ug/ml Ampicillin and 0.8mM IPTG. The bacteria were then incubated at 37°C for 2hr to induce the expression of the GST-fused protein. The bacteria were collect by centrifugation at 5000rpm for 15 minutes. The pellets were stored at -80°C. The pellets were resuspended in 4ml lysis buffer (25mM HEPES pH7.5, 20mM KCl, 2.5mM EDTA, 1%Triton X-100, 1mM DTT, 0.5mM PMSF 1ug/ml antipain, lug/ml leupeptin, lug/ml aprotinin) and sonicated for 30sec. to mix the pellets. Three freeze and thaw cycles and sonication for another 2X30 seconds were performed to disrupt the cell membranes. 0.4ml 5M NaCl was added and the mixture was placed on ice for 5 min. The pellets were eliminated by centrifugation at 12,000rpm for 15 min. at 4 °C. The supernatant was transferred to a 15ml. Falcon tube. 0.8ml 50% pre-treated glutathion-agarose beads were added and incubated for 1hr. at 4 °C . The beads were pelleted at 4000 rpm for 5min. and washed three times with 8ml lysis buffer supplemented with 0.8ml 5M NaCl. The beads were again washed once with 8ml 50mM Tris (pH8.0). The bound GST-fused proteins were

eluted two times with 400ul elution buffer (50mM Tris pH8.0, 10mM Glutathion, 1mM DTT). The eluates were dialyzed overnight against 250 volumes of dialysis buffer (25mM HEPES pH7.5, 50mM KCl, 1mM DTT, 1mM EDTA, 6mM $MgCl_2$, 10% glycerol). The final products were aliquoted and stored at -80 °C.

***In vitro* DNA binding assays**

DNaseI footprinting assays were used to identify specific protection sites by GST-fused proteins. Gel shift assays were used to confirm the binding specificity, to define the protein recognition sites by competitions with mutated oligos, to compare binding affinities through cross competition experiments, and to define protein-protein interactions.

1. Probes for the gel shift assays

5pmol of annealed oligonucleotides were labeled by filling in with Klenow enzyme in the presence of 5ul ^{32}P-dCTP. The free isotopes were eliminated on a G50 column. The flow-through was then loaded on a 8% native PAGE gel. The "crush and soak" method was used to purify the DNA from the PAGE gel (Maniatis et al., 1982). The sense strands of the oligonucleotides are listed below. Mutated nucleotides in the derivatives that were used as specific competitors are underlined.

For the analysis of *tin*-D:
D1: TCGGTGTCAAGTGGCATCTCAAGTGGAG
D3 up: GGCATTACGGTCTCGTCTTT

D3: TCGGGTTTCAATGTCGGCGGCAATGTTGCGGCGACGTCGG
d3g..c: TCGGGTTTCAATGTATGCATCAATGTTATGGATACGTCGG
d3c..t: TCGGGTTTCGGTTTCGGCGGCGGTTTTGCGGCGACGTCGG
E6:GGATCCTCAAGTGGACAAGAGCCGCTGTCGCAGCTGCGAGCCTCCCACC
TCGAG
e6:GGATCCTCAAGTGGACAAGAGTATCTATCGCAGCTGCGATATTCACACC
TCGAG

For the analysis of I-D_L element from *mef2*:
GC-1:
TTATGGCCCTCGCCTCTCGGCGGCGCCACAATGTTGCATTTTGA
GC-1mt:
TTATGGACCACGACTATCAGCAGCACCACAATGTTGCATTTTGA
GC-1c..t:
TTATGGCCCTCGCCTCTCGGCGGCGCCATGGCACTGCATTTTGA
GC-2:
TCGACAGTCTTAATTCAATAAACGCCGCCCGGCGATTTGCGCAT
GC-2mt:
TCGACAGTCTTAATTCAATAAACACCACCAGGAGATTTGCGCAT
GC-3:
ATGTACCCCGATGCTGTGCGCCGTACGGTTGATGCTGCATGTTG
GC-3mt:
ATGTACCCCGATGATGAGCACCATACGGTTGATGCTGCATGTTG

For the analysis of dSmad2:
CAGA Box: TCGAGAGCCAGACAAGGAGCCAGACAC
Smad optimal binding element (SBE): TCGAGTAAGTCTAGACGGCAGTCTAGAC

2. Gel shift assay

Binding reactions were performed in a 10ul volume for 30 minutes on ice with 10,000cpm of the probe, appropriate amounts of GST-fused proteins, 0.5ug of non-specific competitor poly-[(dA)(dT)] and 10-1000-fold molar excess of specific competitor DNA in a final buffer composition of 4% Ficoll, 20mM HEPES pH7.6, 50mM KCl, 1mM EDTA, 1mM DTT and 0.25mg/ml BSA. The mixtures were then loaded directly without dyes on a pre-run (1hr.) 4% polyacrylamide gel (30:1 crosslink ratio). The

electrophoresis was performed at 12V/cm in 0.5X TBE buffer for 3hr. at 4°C. Gels were dried and exposed to X-ray films overnight.

3. Probes for DNaseI footprinting assays

The probes for footprinting assays were single-end labeled by digesting with the first restriction enzyme, labeling as above with ^{32}P-dCTP, followed by digestion with the second restriction enzyme. The free isotopes were eliminated on a G50 column. The flow-through was loaded on a 8% native PAGE gel to separate the probe from the vector band. After being recovered from the gel, the probes were purified by QIAGEN spin columns.

The *tin*-B-A probe was generated by first digesting KS-Intron400 with BamHI, followed by Klenow labeling, heat denaturation and a second digestion with NotI. The *tin*-B-B probe was generated by first digesting KS-Intron400 with XhoI, followed by Klenow labeling, heat denaturation and a second digestion with BamHI. The *tin*-D probe for Schnurri binding was generated by first digesting KS-*tin*-D with NotI, followed by Klenow labeling, heat denaturation and a second digestion with XhoI. For the *tin*-D-A probe, KS-*tin*-D-#4 was first digested with NotI, Klenow labeled, heat denatured and then digested with EcoRV. For the *tin*-D-B probe, the KS-*tin*-D-#10 was first digested with SalI, Klenow labeled, heat denatured and then cut with BamHI. The footprinting probe for the *mef2* dorsal element was generated from the plasmid pCR-MEF-2, which

was provided by Dr. Hanh Nguyen, Albert Einstein College of Medicine. The plasmid was first digested with NotI, Klenow labeled, heat denatured and then digested with XhoI.

4. Chemical sequencing for the G+A reaction

10^5 cpm of the footprinting probe were co-precipitated with 3.3 ug salmon sperm DNA. The pellet was re-dissolved in 10 ul of 1mM EDTA buffer. 1ul 1M formate (pH2) was added and incubated at 37 °C for 20 minutes. 150ul 10% piperidine was added and incubated at 90°C for 30 minutes. The eppendorf tube was filled with butanol to precipitate the DNA. The solution was mixed well and centrifuged for 5 minutes at top speed. The supernatant was removed and the pellet was dried by a speed vacuum which was re-dissolved in 150ul 1%SDS. The butanol precipitation was repeated once. The pellets were then disolved in 8ul gel loading buffer (95% formamide, 20mMEDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF).

5. DNaseI footprinting assay

The binding reactions were performed on ice for 1hr in a 50ul volume with 10,000cpm probe, appropriate amounts of GST-fused proteins, 1ug non-specific competitor poly-[(dA)(dT)] in a final buffer composition of 110mM KCl, 47.5mM HEPES pH7.9, 13.75mM $MgCl_2$, 1mM DTT, 17% glycerol, 0.05% NP-40. 50ul of buffer containing 10mM $MgCl_2$ and 5mM $CaCl_2$ was then added. 0.02 units of DNaseI (Boehringer

Mannheim) was used to digest the DNA for 2 minutes on ice, and the reaction was stopped by adding 90 ul stop buffer (1% SDS, 20mM EDTA, 200mM NaCl, 25ug/ml yeast tRNA). The mixtures were extracted twice with phenol-chloroform and the DNA was precipitated with ethanol. The pellets were resuspended in the loading buffer. After carefully adjusting the number of counts for each reaction, 5,000cpm were loaded for each lane on a 8% polyacrylamid gel containing 7.5M urea. The electrophoresis were performed in 0.5XTBE running buffer for 3hr. at 64W of power. The gel was then dried and exposed with X-ray films.

***Drosophila* strains and embryo stainings**

tkv^{IO} and *tkv*^{SZ3} were gifts from Dr. J. A. Lengyel, UCLA (Terracol and Lengyel, 1994). The crosses to generate germline mutant *tkv* genetic background are shown in Figure 3.

The P-transformed pCasper-*tin*-D line 7.25 was determined to contain a P-insertion in the X chromosome. This insertion was crossed into a dpp^{H46} (Wharton et al., 1993) genetic background to test the *tin*-D activity in *dpp* mutant embryos through β-galactosidase antibody staining.

To test the possible *in vivo* function of *ccf-β*, a deficiency line W(118); Df(3R) 6-7/TM3, Sb (082D3-08; 082F) was used. This line deleted a region including the *ccf* gene. The

embryos were collected and stained with either Tinman antibody or Even-skipped antibody.

To test the *in vivo* function of DEAF-1 in mesoderm formation, a deficiency line w; Df(3L) 25-21/TM3, Sb, Ubx-lacZ was used (a gift from Dr. William McGinnis, UCSD). This line was obtained in a P-element excision screening, which removes all of the DEAF-1 locus and four or five additional genes.

To test the *in vivo* function of Adf-1 in mesoderm formation, a null mutant line (w;+); nal^{le60}/Cyo(actin-lacZ) was used (a gift from Dr. Jim DiZazzo, Cold Spring Harbor Laboratory).

The antibody stainings, *in situ* hybridizations, whole mountings and sectionings of embryos were performed as described in Azpiazu et al., 1996.

RESULTS

tkv is the receptor that mediates the Dpp signal regulating *tinman* expression

Two genes for type I receptors have been identified in *Drosophila*, *thick veins* (*tkv*) and *saxophone* (*sax*). Compared with *sax*, the expression level of *tkv* is higher in the mesoderm around gastrulation (Penton et al., 1994; Yin and Frasch, 1998). Therefore, the possibility that *tkv* is the type I receptor that mediates the *dpp* signal to regulate *tinman* expression was tested. As *tkv* has strong maternal expression, genetic crosses were designed to reduce or eliminate both the maternal and the zygotic *tkv* contribution (Figure 3). This is possible because it had been shown that some combinations of *tkv* alleles, including tkv^{JO}/tkv^{SZ3}, are viable and the females are fertile (Terracol and Lengyel, 1994). Because of the early embryonic function of *tkv* and *dpp* in dorsal-ventral body axis formation, embryos with this genotype are ventralized and Dpp is therefore only expressed in very restricted dorsal regions of the ectoderm (Figure 4C & 4D, arrowheads outside the embryo indicate the borders of *dpp* expression). In this situation, the mesoderm (Figure 4D, arrowheads inside the embryo indicate the dorsal tip of the mesoderm) can not extend far enough to contact the Dpp-expressed ectodermal cells and has no opportunity to be induced and maintain *tinman* expression (Figure 4D, arrowheads inside the embryo show dorsal extent of the mesoderm). Thus it is unclear whether the observed loss of *tinman* expression is due to the absence of *tkv* activity or just results

from this secondary effect. To overcome this problem, the UAS-GAL4 binary system was used to restore broad Dpp expression in the ectoderm and provide the neighboring mesoderm with the opportunity to receive Dpp signals in the *tkv* mutant background. Dpp was ectopically expressed in a portion of the embryo along the anterior-posterior axis using a combination of the Kruppel promoter and *Zen* ventral repression elements (for details of the system, see Frasch, 1995). In the wild type background, this ectopic expression of Dpp can induce ectopic *tinman* expression in the ventral mesoderm (Frasch 1995). In a heterozygous *tkv* mutant background, the wild type allele of *tkv* still can transduce the inductive signal to maintain *tinman* expression according to the border of ectodermal Dpp (Figure 4B). In contrast, in the homozygous *tkv* mutant background, *tinman* can not be induced even when the mesoderm is in contact with the ectoderm that ectopically expresses *dpp* (Figure 4F). This result provides strong evidence that *tkv* is required to mediate the *dpp* signal to regulate *tinman* expression in the dorsal mesoderm.

tin-D is a 349bp Dpp-responsive element driving *tinman* expression in the dorsal mesoderm

Transgenic flies were used to identify *cis*-regulatory elements in the *tinman* flanking region with *lacZ* as a reporter. The originally identified 225bp *tinman* downstream fragment, *tin*-D(1-225), drives weak expression of the reporter gene in the dorsal mesoderm (Figure 5 and 6C). Further deletion into the sequence reduces the level of expression (*tin*-D-#4, contains fragment A, Figure 5) and larger deletions even ablate the

Figure 3. Genetic crosses to generate embryos that lack both the maternal and the zygotic activities of *tkv* and ectopically express Dpp.

The UAS-GAL4 binary expression system was used. For the Kruppel promoter and the UAS-*dpp*4 system used, see Frasch, 1995.

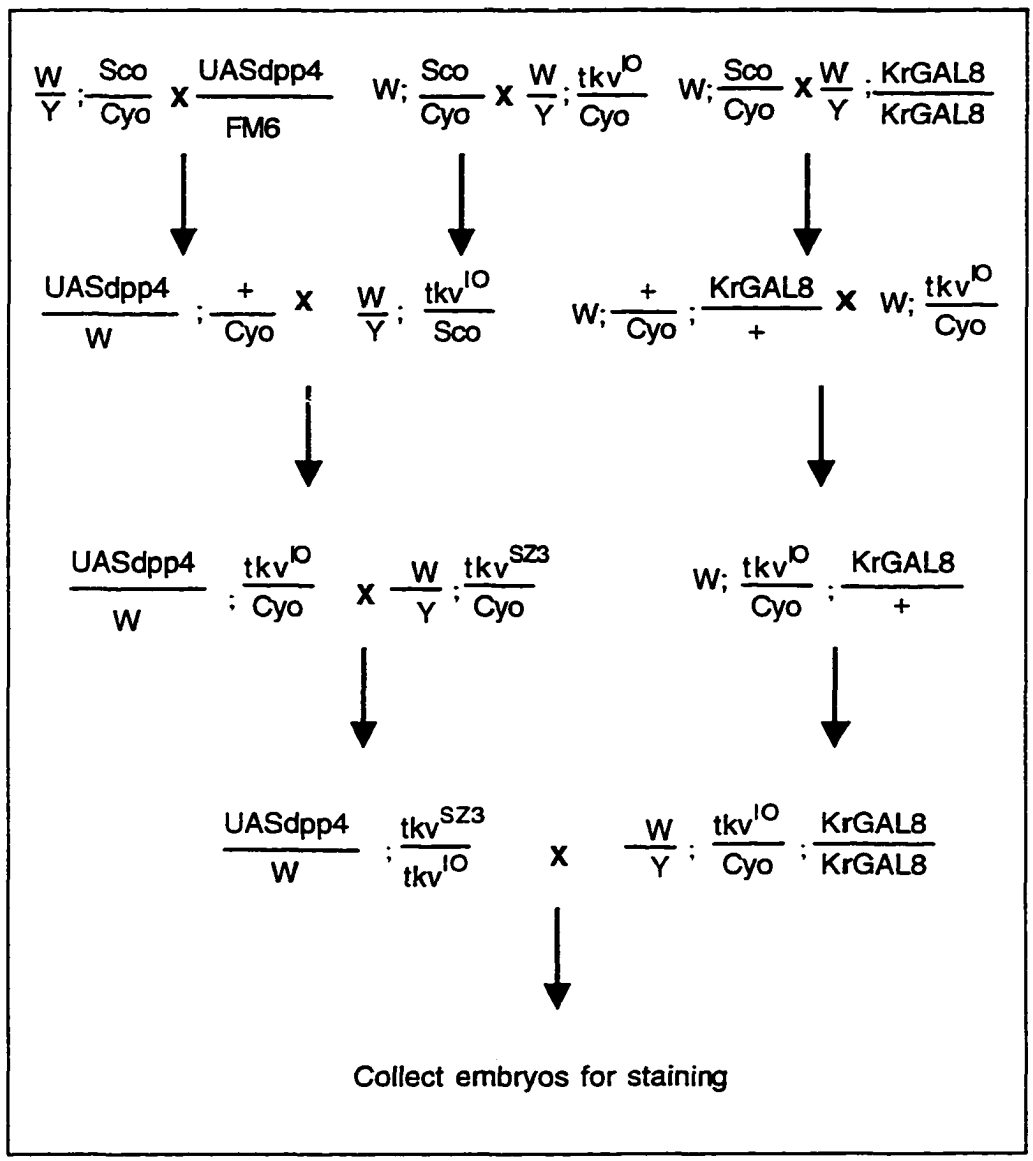


Figure 3

Figure 4. *tinman* activity in *tkv* germ-line mutant embryos.

All panels are double *in situ* hybridization with *dpp* and *tinman* probes. Panels A, C, E are whole mount staining, A & C are dorsal views, E is a lateral view (dorsal is up and ventral is down). Panels B, D, F are cross sections. All embryos contain maternally mutated *tkv*(tkv^{IO} and tkv^{SZ3}). Embryos in (A) and (B) contain a zygotic wild type *tkv* allele ($tkv^{IO\ or\ SZ3}/+$) as indicated by a wide expression pattern of *dpp* in the dorsal ectoderm. The neighboring mesoderm can be induced to express *tinman* corresponding to the Dpp signal (B). Embryos in panels C, D, E, F contain zygotically homozygous mutated alleles of *tkv* (tkv^{IO}/tkv^{IO}) as indicated by narrowed expression of Dpp in the dorsal ectoderm (C and D). In this case, the mesoderm was not able to reach the Dpp expressing ectoderm (D). Embryos in panel (E) contain ectopic Dpp expression in the Kruppel domain, as indicated by brackets. A cross-section in this domain is shown in panel (F). The Dpp expression in the ectoderm is extended ventrally. However, the underlying mesoderm which is in contact with the Dpp signal still fails to maintain *tinman* expression (F). Arrowheads outside the embryos in panels B and D indicate the ventral border of the endogenous Dpp expression. Arrows outside the embryos in panel F indicate the border of the ectopically expressed Dpp. Arrowheads inside the embryos in B, D, F indicate the dorsal border of the mesoderm.

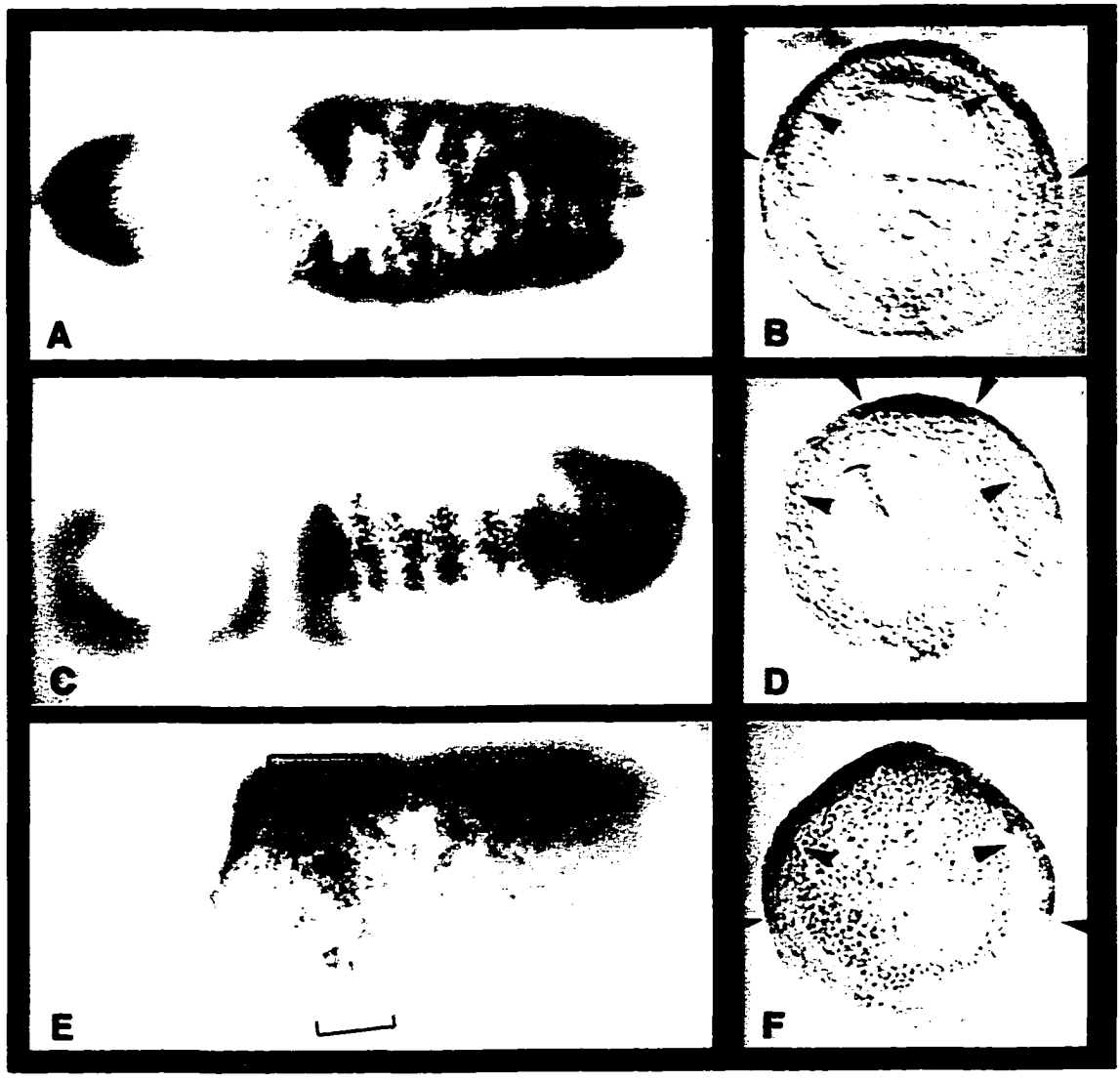


Figure 4

expression (*tin*-D-#1, *tin*-D-#2 and *tin*-D-#3). When the *tinman* downstream region from *Drosophila virilis* was sequenced by Zhizhang Yin in our lab, a highly conserved ~300bp stretch was found just downstream of the 225bp sequence, suggesting that its function is important. Indeed, the constructs that include either all of this stretch (*tin*-D-#6 which consists of fragments A+B+C; for the positions of the three fragments, see Figure 5) or part of it (*tin*-D-#7, contains fragments A+B) drove much stronger expression in the dorsal mesoderm (Figure 6B). The pattern of the *lacZ* expression was comparable to the *tinman* mRNA expression at stage 9 (Figure 6A). The B+C sequence by itself (*tin*-D-#8) only gave segmented expression pattern at late stage 11 (Figure 6D). Further deletions into this stretch (*tin*-D-#10, *tin*-D-#12) resulted in a loss of reporter gene expression. Therefore, the 349bp *tin*-D-#7 was the shortest element which drove strong dorsal mesoderm expression and was subsequently renamed as *tin*-D. The sequence of *tin*-D (Figure 8) is highly conserved between *Drosophila melanogaster* and *Drosophila virilis.*

An orientation effect has been observed when some elements were cloned into the *lacZ* reporter vector in the opposite direction. For example, *tin*-D-#6 drove strong continuous dorsal mesoderm expression, while *tin*-D-#5 drove a segmented expression pattern. *tin*-D-#8 drove a similar segmented pattern, while the *tin*-D-#9 construct gave no expression (Figure 5). One explanation for this observation is that there is an insulator located in the fragment B+C. The insulator is a DNA structure that can block the enhancer function toward the promoter when located between the enhancer and promoter (Kellum and

Schedl, 1991). The insulators may function to isolate neighboring genes by blocking interactions between the two transcriptional machinery. *bagpipe (bap)*, another NK family member, is located just 5kb downstream of *tinman* gene and the B+C fragments are located around 1kb downstream of the *tinman* gene.

The insulator hypothesis has been tested using *even-skipped (eve)* stripe 2 and stripe 3 enhancers (for details of the system, see Cai et al., 1995). The principle of the test is shown in Figure 7A. If the test sequence located between the stripe 2 and stripe 3 enhancer functions as an insulator, it should prevent the *lacZ* reporter gene from being expressed in stripe 2, and prevent the *white* reporter gene from being expressed in stripe 3. In contrast, if the test sequence cannot function as an insulator, then the *lacZ* or the *white* reporter genes will be expressed in both stripe 2 and stripe 3. The inserts from either *tin*-D-#8 or #9 were cloned into this position to test the insulator hypothesis. They contain the fragment B+C in opposite orientations. A series of constructs was generated; their relative position and direction are shown in Figure 7B. For unknown reasons, the control construct in this study did not work for the *lacZ* reporter gene since the stripe 2 expression was not observed as had been expected. Therefore, the lack of *lacZ* expression in stripe 2 for constructs 8.1 and 9.1 may be due to this technical problem and not to the existence of an insulator. In two other constructs, 8.2 and 9.2, the fragment B+C was inserted between both of the stripe enhancers and the *lacZ* promoter, in opposite orientations. If there is an insulator located in the fragment B+C, then the

lacZ expression in these two constructs should be detected in neither stripe 2 nor stripe 3. In contrast, both stripe 2 and stripe 3 expression was observed, which argues against the existence of an insulator in this region (staining not shown; data summarized in Figure 7B). The explanation for the orientation effect of *tin*-D series constructs still needs further investigation.

As the *lacZ* expression pattern driven by the *tin*-D element resembles the *dpp*-dependent pattern of the endogenous *tinman* expression (Figure 6A and 6B; for cross sections, see 6E and 6F), this suggests that *tin*-D is a *dpp*-responsive element. Consistent with this notion, *tin*-D activity was ablated in the *dpp* mutant embryos (Figure 6G). In contrast, when $tkv^{Q\text{-}D}$, the *dpp* receptor in its constitutive active version, was ectopically expressed in the whole mesoderm, *tin*-D activity was also extended toward the ventral mesoderm (see Xu et al, 1998). These observations confirm the identity of *tin*-D as a Dpp-responsive element.

Functional dissection of the *tin*-D element

1. *tin*-D contains Tinman autoregulatory sites and Schnurri binding sites while *tin*-B contains Twist binding sites.

To dissect the *tin*-D element, a candidate approach was first used. Several known transcription factors that are expressed in the mesoderm during stages 6-12 were tested for their ability to bind to *tin*-D *in vitro*. During an effort to establish footprinting

Figure 5. Constructs to define the functional element *tin*-D.

The nucleotide number starts at the 5' end of the original dorsal element *tin*-D(1-225) which was identified by Zhizhang Yin in this lab. The *lacZ* staining level in the dorsal mesoderm of this construct is defined as + (for staining, see Figure 6C). The *lacZ* reporter expression in the dorsal mesoderm driven by other constructs are symbolized by +++(robust expression, see Figure 6B), +/- (even weaker then +), S (segmented expression in the visceral mesoderm, see Figure 6D) and - (no expression detected). All of these elements are located ~1.2Kbp downstream of the *tinman* encoding region (see Yin et al., 1997). There is a BamHI site at nucleotide 225. For details about the generation of these constructs, see Materials and Methods. Three fragments are covered by these constructs: A (ntd. 59-225), B (ntd. 225-414) and C (ntd. 414-564). *tin*-D-#7 (covers fragment A+B) is the shortest element that was able to drive robust expression in the dorsal mesoderm, and was subsequently renamed as *tin*-D. *tin*-D-#4 (covers fragment A) and *tin*-D-#10 (covers fragment B) were renamed as *tin*-D-A and *tin*-D-B, respectively. Notably, the constructs *tin*-D-#8 and *tin*-D-#9 (cover fragment B+C) drive different expression patterns, indicating the existence of an orientation effect. Arrows indicate the direction of the constructs in comparison to the *lacZ* promoter.

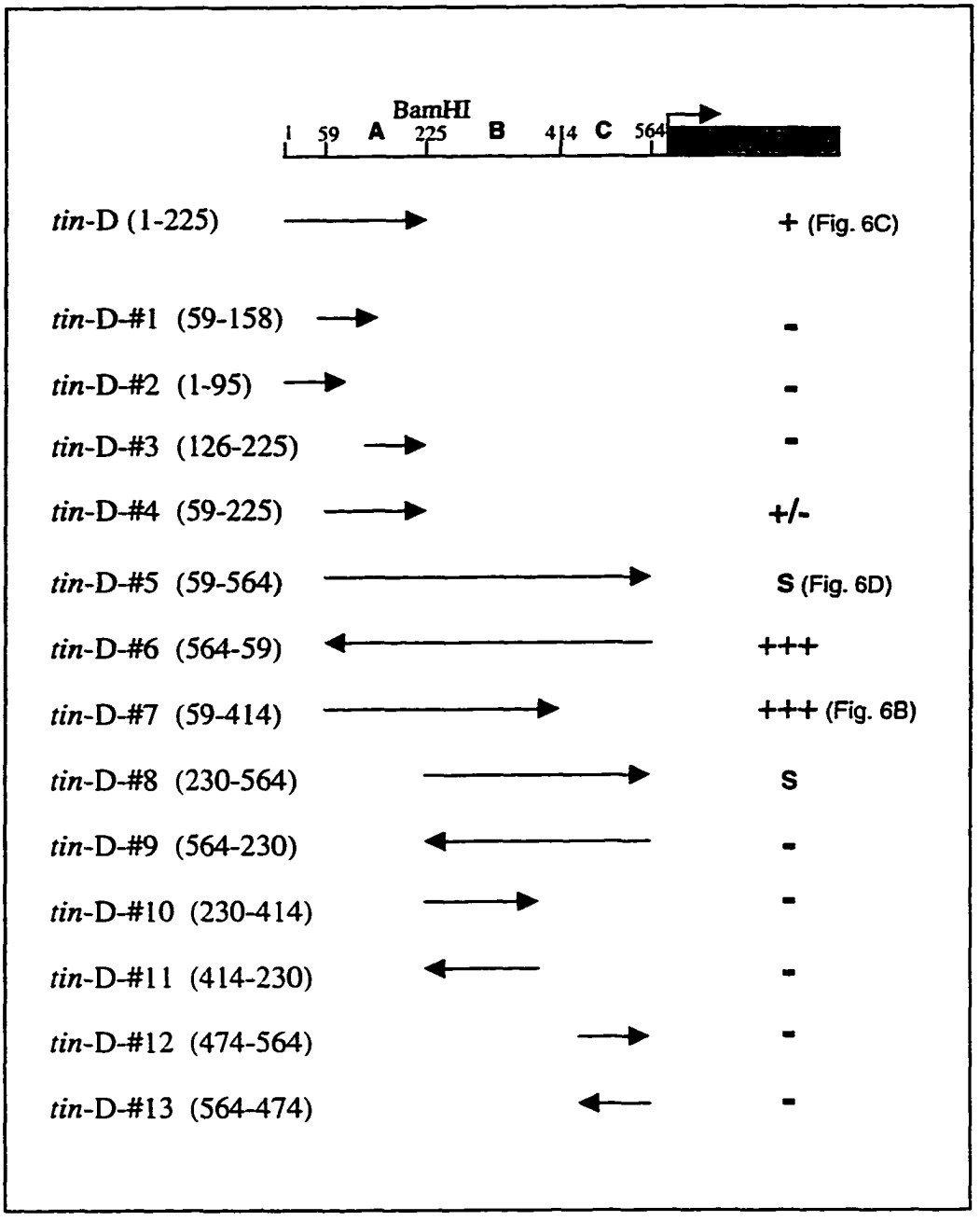


Figure 5

Figure 6. *tin*-D, the minimal Dpp-responsive element with full activity.

Panels A-D show lateral views of whole mount stained embryos at stage 10, with anterior to the left, posterior to the right, dorsal up and ventral down. Arrowheads indicate the position of the dorsal mesoderm. Panels E-G show cross sectioned embryos. (A&E) An embryo stained with a *tinman* mRNA probe, showing the phase II expression in the dorsal mesoderm. (picture provided by Manfred Frasch). (B&F) *lacZ* antibody staining driven by *tin*-D-#7 (59-414) (renamed as *tin*-D). The expression level and pattern is comparable to the *in vivo tinman* mRNA signal (compare B with A, F with E). (C) *lacZ* antibody staining driven by the original dorsal element *tin*-D(1-225) (picture provided by Zhizhang Yin). (D) Segmented expression pattern driven by the construct *tin*-D-#5. (G) Cross section of an embryo with a *lacZ* reporter gene driven by *tin*-D in dpp^{H46} mutant genetic background. No *tin*-D activity was detected. Arrowheads in panel (A-D) indicate the position of the dorsal mesoderm. The arrow in panel (C) indicates ectopic staining generated by the pCasper vector.

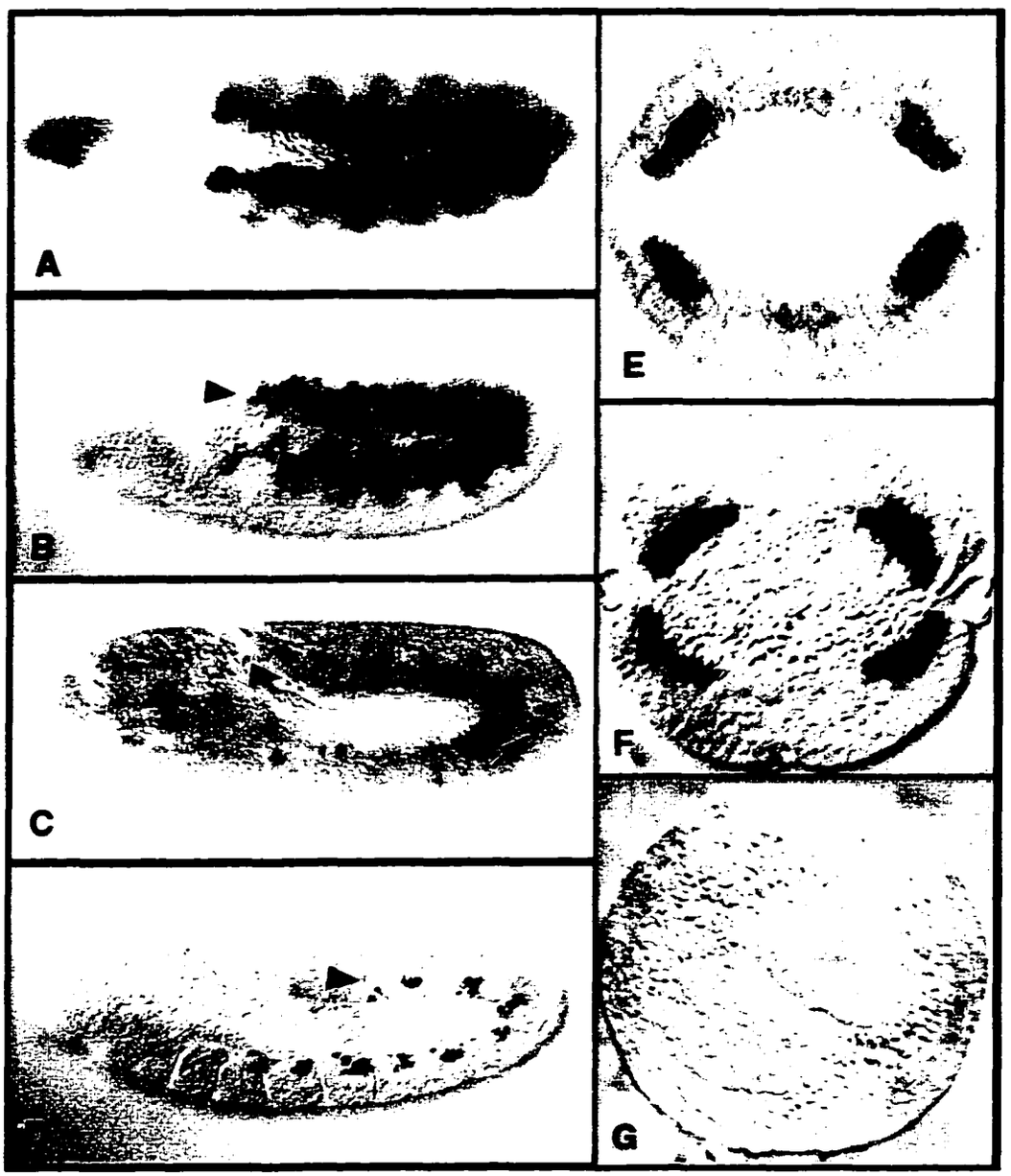


Figure 6

Figure 7. Strategy and constructs to test the insulator hypothesis.

(A) A model construct to test the existence of an insulator. The sequence is inserted between the *even-skipped* stripe2 and stripe3 enhancers. Two reporter genes, *lacZ* and *white*, are transcripted toward the opposite direction. If there is an insulator, then the *lacZ* reporter gene should be only expressed in the stripe3 while the *white* reporter gene should be only expressed in the stripe2. If the inserted sequence does not contain an insulator, then either of the reporter genes will be expressed in both stripe2 and stripe3. (B) The constructs to test the hypothesis that there is an insulator in the fragment B+C (see Figure 5). *tin*-D-#8 and #9 which cover the fragments B+C were used for the cloning. The relative position and direction of the resulted constructs were indicated. The *lacZ* expression patterns driven by these constructs were summarized in the column on the right.

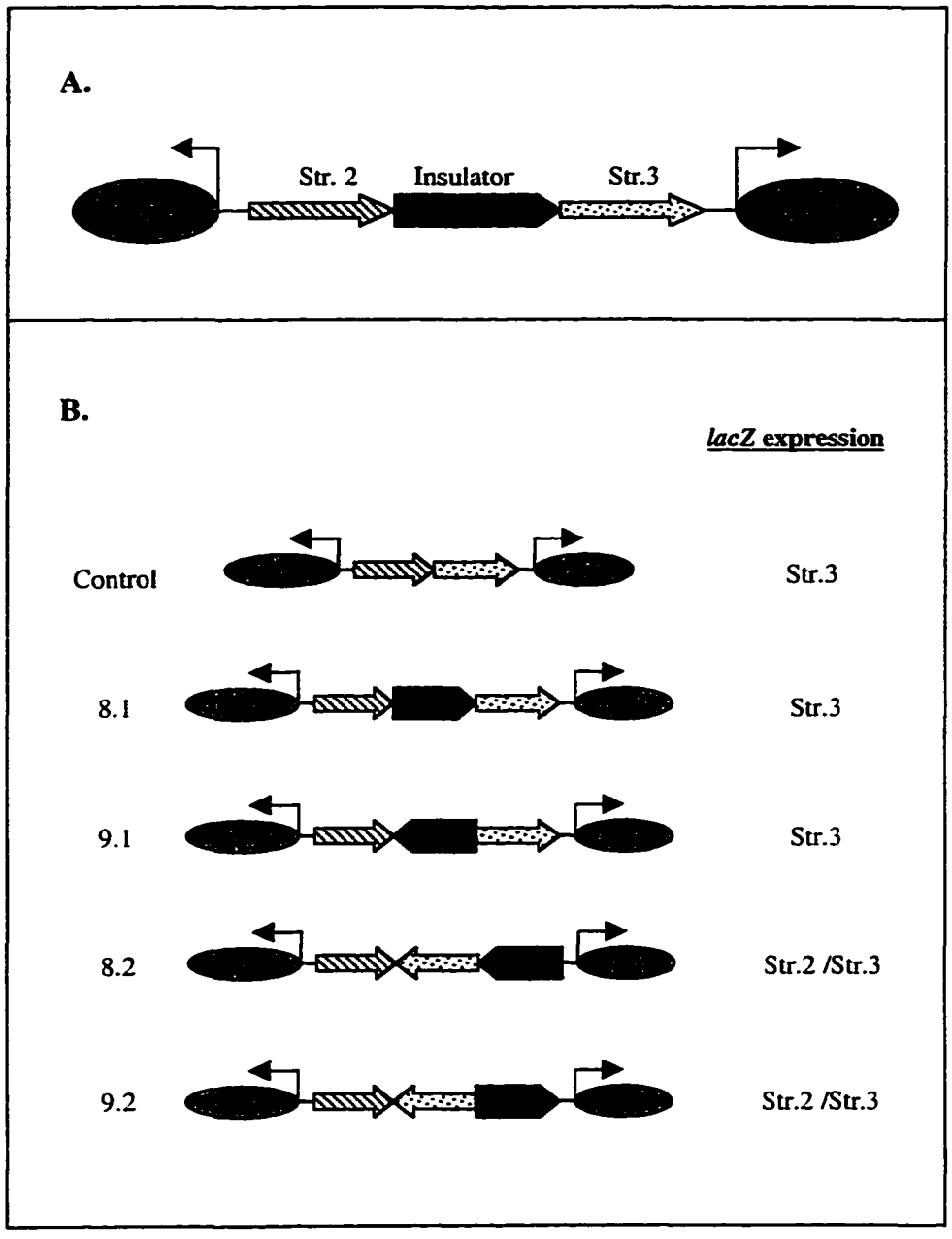


Figure 7

Figure 8. Sequence alignment of *tin*-D enhancer elements from *Drosophila melanogaster* and *Drosophila virilis*.

Identical sequences are boxed and shaded. *tin*-D from both species were sequenced by Zhizhang Yin. The nucleotide number starts at the 5' end of the *tin*-D element, which corresponds to residue #59 in the previous numbering system used in Figure 5.

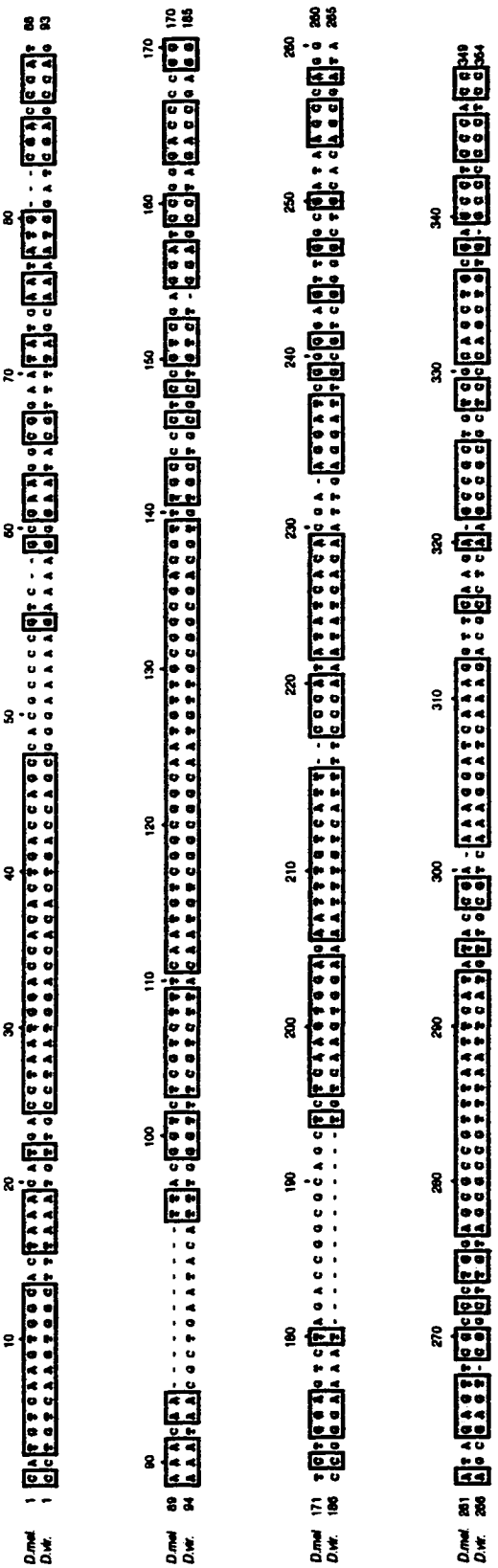


Figure 8

techniques in the lab, the *tin*-B element was first labeled as a probe to test the possibility that GST-fused Twist protein could bind to *tin*-B. It had been shown previously that the phase I expression of *tinman* in the whole presumptive mesoderm depends on *twist* (Bodmer et al., 1990). Twist belongs to the basic helix-loop-helix (bHLH) domain transcription factor family whose members recognize E-box sequences (CANNTG). Visual inspection identified three E-boxes which were located inside the *tin*-B element (Yin et al., 1997). Indeed, GST-fused Twist protein protects all three E-boxes *in vitro* (Figure 9). Zhizhang Yin in our lab went on to provide *in vivo* evidence that mutation of the three E-boxes disrupt the *tin*-B activity while a combination of the three E-boxes can mimic *tin*-B activity (Yin et al., 1997). Therefore, it is highly likely that Twist initiates the phase I expression of *tinman* by directly binding to the three E-boxes in the *tin*-B element.

GST-fused proteins of Twist and Snail, a second essential protein for mesoderm formation were then extracted to perform footprinting assays with ^{32}P-labeled *tin*-D as a probe. No protection sites were detected for either GST-fused Twist or Snail, although the GST-fused Twist protein was functional as shown by its capacity to bind to the *tin*-B element. Schnurri is a zinc-finger containing protein which is implicated in downstream events of the Dpp signal (Arora et al., 1995; Grieder e al., 1995; Staehling-Hampton et al., 1995). Two protection sites were detected in *tin*-D when GST-fused Schnurri protein

was used in the binding reaction (Figure 10A). This result confirms the prediction based on its sequence that Schnurri is a DNA binding protein.

GST-fused Tinman proteins was extracted in order to perform footprinting assays with ^{32}P-labeled *tin*-D as a probe. Two protection sites were identified (Figure 10B), which include two identical TCAAGTGG sequences that correspond to the consensus DNA binding sequences for NK homeo domain transcription factors (Chen and Schwartz 1995; Tsao et al., 1994). The two sub-elements within the protected regions were then named D1a and D1b. This result suggested that Tinman protein from its phase I expression may regulate phase II expression of *tinman* by directly binding to the *tin*-D element. Indeed, both *tinman* mRNA and *tin*-D activity were strongly reduced in *tinman* mutant embryos (see Xu et al., 1998).

2. D1, D3 and D6, three essential sub-elements in *tin*-D

Six internal deletions of *tin*-D of ~30 bp each were then designed to confirm the importance of D1a and D1b and to identify other essential sub-elements (Figure 11). Two internal deletions, *tin*-D-ΔD3 & *tin*-D-ΔD6, strongly reduced the *tin*-D activity in the dorsal mesoderm (Figure 12 A&B), indicating that they are essential for *tin*-D activity. When the presumed autoregulatory sub-elements D1a and D1b were deleted, the resulting construct *tin*-D-ΔD1 drove much weaker *lacZ* reporter expression in the dorsal mesoderm. This *in vivo* result supports the notion that autoregulation contributes to full-

level induction by the Dpp signal. Unexpectedly, *tin*-D-ΔD1 also drove strong *lacZ* expression in the dorsal ectoderm (Figure 12C). This observation suggests that *tin*-D-ΔD1 can still respond to the Dpp signal, although without tissue specificity. The induced reporter expression is extended from the Dpp target tissue, the dorsal mesoderm, to the Dpp signaling tissue, the dorsal ectoderm. One possible mechanism to explain the normally tissue specific induction is that there is a repressor in the dorsal ectoderm that can bind to D1 and prevent *tin*-D expression in the dorsal ectoderm. In the dorsal mesoderm, the phase I expressed Tinman protein may compete with the repressor and release this repression.

Three of the other internal deletions, *tin*-D-ΔD2, *tin*-D-ΔD4 and *tin*-D-ΔD5, did not affect the *tin*-D driven *lacZ* reporter gene expression (data not shown).

3. D3 is the minimal Dpp-responsive element, while D1 is able to confer tissue specificity

To test whether the essential sub-elements D1, D3 and D6 were sufficient to mimic the *tin*-D activity in the dorsal mesoderm, constructs with multiple copies and combinations of these sub-elements were generated. The nucleotides covered by these sub-elements are shown in Figure 11. Notably, the construct $(D3)_5$+$(D6)_4$ which contains five copies of D3 and four copies of D6 drives a similar reporter expression pattern as *tin*-D-ΔD1

(Figure 12, compare D with C), although the expression levels were lower in both dorsal mesoderm and dorsal ectoderm. With an even lower expression level, the $(D3)_5$ construct which contains five copies of D3 alone also drove a similar *lacZ* expression pattern (Figure 12E). Conversely, four copies of D6 alone failed to drive reporter expression. These results defined the 32 bp D3 sub-element as the minimal Dpp-responsive element which is sufficient to respond to the Dpp signal, although without tissue specificity. To test whether the D1 sub-element was sufficient to modify this basal level of induction into a tissue specific induction, four copies of D1 sub-elements were cloned into the $(D3)_5$ construct to generate $(D1)_4$+$(D3)_5$. As had been expected from our previous data, the dorsal ectoderm expression driven by $(D3)_5$ was suppressed, while the dorsal mesoderm expression was strongly enhanced. The combination of the D1 and D3 sub-elements thus reconstitutes the specific *tin*-D activity in the dorsal mesoderm (Figure 12F). Four copies of the autoregulatory sub-elements D1 by itself failed to drive *lacZ* reporter gene activity (data not shown), indicating that *tinman* alone does not function as a transcriptional activator.

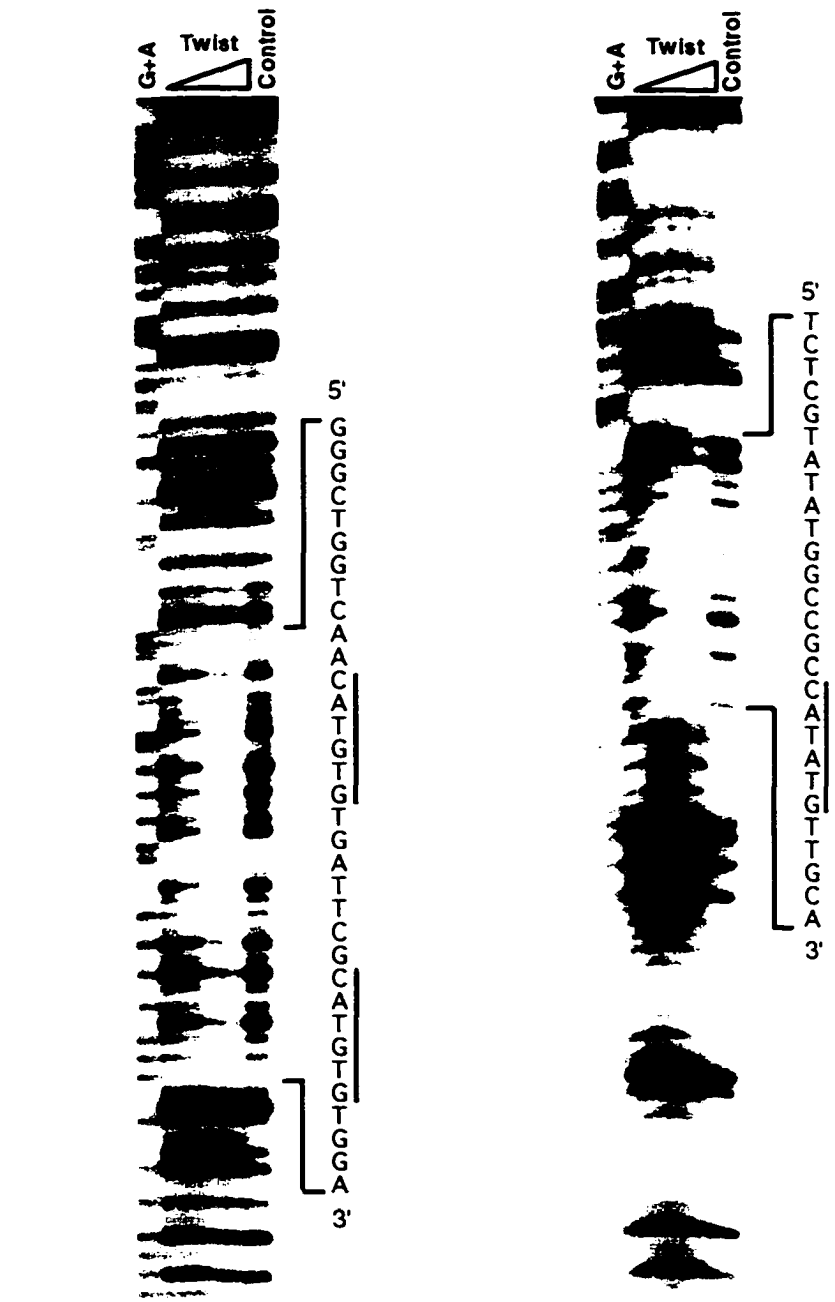
Cloning and characterization of *Medea*

1. Yeast one-hybrid screening

The yeast one-hybrid screening system was used to identify transcription factors that bind to the D1 or D3 sub-elements. One-hybrid screening is an *in vivo* genetic assay used

Figure 9. DNaseI footprinting analysis with GST-fused Twist on the *tin*-B enhancer element.

The 374bp *tin*-B element was labeled as two probes: *tin*-B-A and *tin*-B-B (for detailed process, see Materials and Methods). Chemical G+A reaction was shown to indicate the sequence ladder. 0.2, 0.4, 0.8 and 1.6ug GST-fused Twist proteins were used for the protection assays, as been indicated by wedges. 1.6ug GST tag protein alone was used for the control lanes. The sequences of the protected regions are listed. The three E-boxes (CANNTG) are marked by lines.



Probe: tin-B-A **tin-B-B**

Figure 9

Figure 10. DNaseI footprinting analysis with GST-fused Tinman and His-tagged Schnurri on the *tin*-D enhancer element.

(A) Two Schnurri protection sites were detected. *tin*-D element was labeled as a probe. A long run was performed to show the two protected sites in one gel. (B) Two Tinman protection sites were detected. The 349bp *tin*-D element was digested into two fragments to be labeled as probes: *tin*-D-A and *tin*-D-B. Chemical G+A reaction was shown to indicate the sequence ladder, which was numbered according to Figure 8. ~1ug GST-fused Tinman proteins were used for the protection assays. No proteins was added for the control lanes. The sequences of the protected regions were listed. The consensus TCAAGTG sequences are marked by lines. The results are summarized in Figure 33.

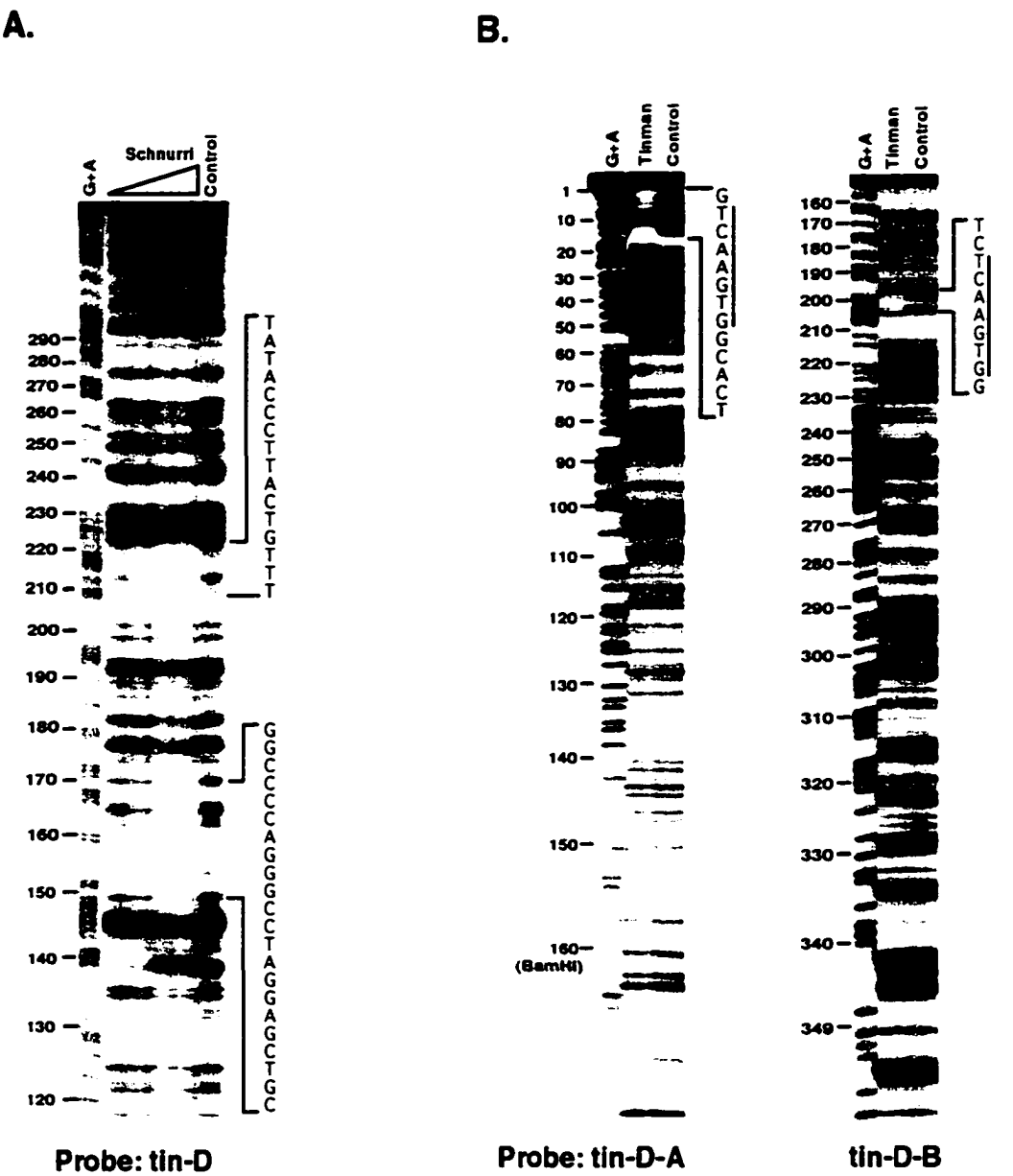


Figure 10

Figure 11. Map of internal deletions and oligonucleotides used for multiple copy constructs of essential sub-elements within the *tin*-D element.

tin-D-ΔD1 deletes both nucleotides 1-12 and 197-204, which cover two Tinman binding sites. *tin*-D-ΔD2 deletes nucleotides 16-47. *tin*-D-ΔD3 deletes nucleotides 111-135, which cover two CAATGT repeats and two GC-rich motifs. *tin*-D-ΔD4 deletes nucleotides 205-229, which cover one of the two Schnurri protecting sites. *tin*-D-ΔD5 deletes nucleotides 244-312. *tin*-D-ΔD6 deletes nucleotides 313-349. The deletions in *tin*-D-ΔD1, *tin*-D-ΔD3 and *tin*-D-ΔD6 disrupt the dorsal mesoderm expression driven by *tin*-D (see Figure 12), indicating the existence of essential sub-elements. Three oligonucleotides were then designed to generate multi-copy constructs for the purpose of both *in vivo* reporter gene expression and yeast one-hybrid screening (see Materials and Methods). The nucleotides in *tin*-D element that are covered by these three sub-elements are indicated by black boxes. For the exact sequences, see Materials and Methods. The sub-element D1 covers both Tinman binding sites: nucleotides 3-14 (D1a) and 194-205 (D1b). The sub-element D3 covers nucleotides 108-139. The sub-element D6 covers nucleotides 320-349.

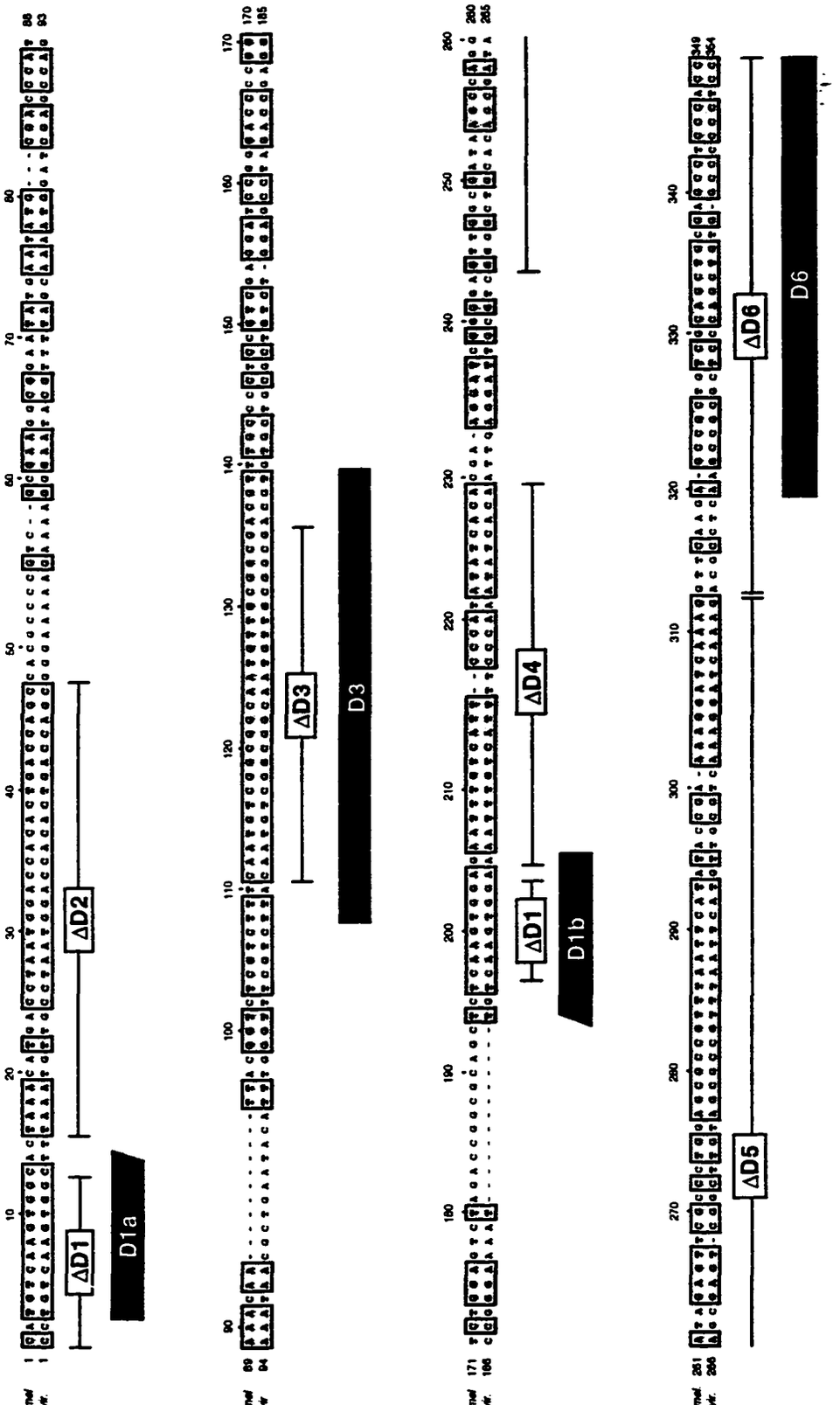


Figure 11

Figure 12. The minimal Dpp-responsive subelement D3 and the tissue specific sub-element D1.

Shown are cross-sections of embryos with *lacZ* staining driven by either internally deleted *tin*-D elements (Panel A, B and C) or combinations of multiple-copies of *tin*-D sub-elements (Panel D, E and F). The internal deletion of the sub-element D3 (A) or D6 (B) dramatically decreased the *lacZ* reporter gene expression in the dorsal mesoderm. Arrowheads indicate the residual staining in the dorsal mesoderm. (C) The internal deletion of the sub-element D1, which covers two Tinman binding sites, resulted in a strong reduction in the dorsal mesoderm and an ectopic expression in the dorsal ectoderm. (D) The combination of five copies of sub-element D3 and four copies of sub-element D6 drove *lacZ* reporter gene expression strongly in the dorsal ectoderm and weakly in the dorsal mesoderm. (E) Five copies of the sub-element D3 alone drove expression in a pattern similar to (C) and (D), although the expression was much weaker. The expression in the dorsal mesoderm was almost undetectable. (F) The combination of four copies of sub-element D1 and five copies of sub-element D3 drove *lacZ* reporter gene expression only in the dorsal mesoderm, reconstituting the *tin*-D activity. For a schematic summary of these constructs, see Figure 24. (ms) mesoderm. (ec) ectoderm.

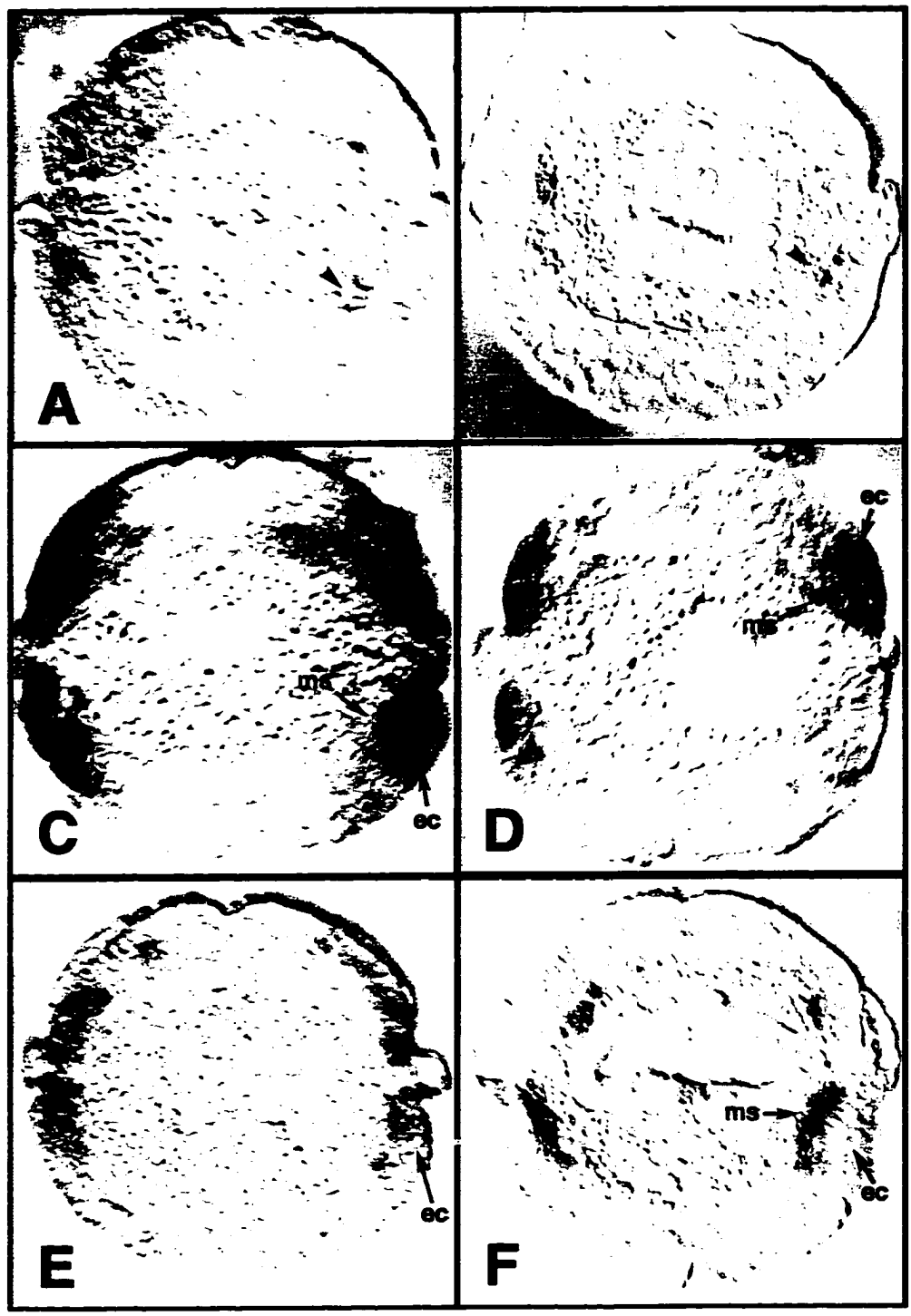


Figure 12

for isolating novel genes encoding proteins that bind to a *cis*-regulating element. It is based on the finding that many transcription factors contain separable DNA-binding and transcription activation domains. The *cis*-regulatory element is inserted in the 5' promoter region of a reporter plasmid and the construct is then integrated into a yeast genome to generate a reporter strain. A library of cDNAs fused with sequences encoding the GAL4 transcriptional activation domain is transformed into the reporter strain. If the cDNA insert encodes a *cis*-regulatory element binding peptide, the fused GAL4 transcription activation domain will activate the reporter gene. Theoretically, both transcription activators and repressors could be isolated by the yeast one-hybrid assay. (for details of the procedure, see Materials and Methods).

When using four copies of D1 as a bait, the only two positive clones that met all of the screening criteria encoded Tinman protein sequences containing the homeo domain fused in frame with the GAL4 activation domain. This result further supports the idea that D1 represents functional Tinman binding site and also indicates that the screening system works well.

As the D3 sub-element is the minimal Dpp responsive element, three different libraries were extensively screened to search for the Dpp downstream mediators. After re-transformation tests and sequencing/BLAST searches, 55 positive clones remained, which are summarized in Figure 13. *Deformed Epidermal Activation Factor-1 (DEAF-1)* was the

only transcription factor that appeared in all three libraries. The other positive clones were unique to a specific library, indicating that either the screens were not saturating or that each library has some unique content. Clone 3.15 from Library #3 drew immediate attention because the 3' sequence of the insert was highly homologous with the MH2 domain of *DPC4/Smad4*. It then became the major focus of further study. The possibility that *DEAF-1* and other potential positive factors including *Adh distal factor-1 (Adf-1)*, *HMG-D* and *ccf-β* are involved in *tin*-D enhancer binding and regulation is discussed later in this study. The sequences of the inserts from clones 1.3, 1.34, 1.69, 2.16 and 2.20 do not have any similarities with any known sequences in Genbank. *In situ* hybridization staining indicated that the former three clones are not expressed in the mesoderm. The mRNA expression patterns for the latter two clones could not be detected during early embryogenesis (data not shown). For clone 1.34, a stop codon in frame with the GAL4 activation domain was detected only ten amino acids from the start site of the insert. Because of the above reasons, these five clones were not further discussed in this study.

2. *Medea*, a *Drosophila Smad4* homologue

Full-length sequencing of the insert from clone 3.15 revealed a typical N-terminal MH1 domain and a C-terminal MH2 domain, which are the landmarks for Smad proteins. Additional cDNAs were isolated by Zhizhang Yin from an embryonic cDNA library with 3.15 as a probe. These clones belonged to two different versions, which were represented

by cDNAs #14 and #29. Subsequent sequence comparisons indicated that the clone 3.15 corresponded to the *Medea* gene, which was identified during the same time independently by other labs using either a genetic approach (Hudson et al., 1998; Wisotzkey et al., 1998) or a degenerate PCR approach (Das et al., 1998). Compared with the genomic sequence from Dr. E. Ferguson's lab, the two cDNA clones represent two alternative splicing products (#14 was named *Medea* $A_{1\text{-}692}$ and #29 was named *Medea* B, see Figure 15). The full-length *Medea* (or *Medea A in* Figure 15) sequence was reconstituted by adding the C-terminal sequences from *Medea* B to *Medea* $A_{1\text{-}692}$. The deduced Medea protein sequence is shown in Figure 14. *Medea* B lacks an alternative exon corresponding to amino acids 231-304 in *Medea* A (Figure 15, boxed region in Figure 14). The insert of plasmid 3.15 corresponds to Medea $A_{1\text{-}681}$. It contains a C-terminal truncation at a NotI site which was used to clone the inserts for library #3. The Medea protein contains a less conserved linker region between the conserved MH1 domain and the MH2 domain, which is significantly longer than the other Smad proteins and unusually Gln-rich. The sequence of Medea is more closely related to vertebrate DPC4/Smad4 and the *Caenorhabiditis elegans* Sma4 than the pathway-specific Smad proteins. Moreover, Medea lacks the C-terminal SSXS motif which is required for the pathway-specific Smad proteins to be phosphorylated and activated by receptors (Figure 14; see Macias-Silva et al., 1996; Kretzschmar et al., 1997). Therefore, we concluded that *Medea* represents the common-mediator *Smad* in *Drosophila*.

3. Medea binds to DNA through its MH1 domain, while the MH2 domain has an inhibitory effect

The fact that I cloned the *Medea* gene through yeast one-hybrid screening implied that Medea can directly bind to DNA. Taking the advantage of the yeast one-hybrid system, the location of the DNA binding domain in the Medea protein was determined by generating a series of deletion constructs fused with the GAL4 activation domain. Using $(D3)_5$/*lacZ* as the reporter, it was found that the full-length Medea B protein and small C-terminal truncated versions of Medea A (Medea $A_{1\text{-}692}$, Medea $A_{1\text{-}681}$) contain some DNA binding activity (Figure 15). The removal of the conserved MH2 domain resulted in an increase of the reporter gene activity for both of the alternative products (Medea $A_{1\text{-}505}$, and Medea $B_{1\text{-}432}$). Further deletion of the linker region did not affect reporter gene activity (Medea $A_{1\text{-}303}$, Medea $A_{1\text{-}265}$). However, a 10 amino acid deletion into the conserved MH1 domain lowered reporter gene expression to background levels (Medea $A_{1\text{-}178}$). The linker region or the MH2 domain only drove background levels of reporter gene expression (Medea $A_{167\text{-}505}$, Medea $A_{506\text{-}692}$). In summary, the MH1 domain functions as the DNA-binding domain for Medea protein, while the MH2 domain has an inhibitory effect. Similar results were also obtained for the Mad protein (data not shown), which is consistent with the conclusions of Kim et al., 1997.

Figure 13. Summary of the results of the yeast one-hybrid screening with five copies of D3 as a bait.

Three different libraries were used. All of the listed clones have passed the re-transformation test and have been sequenced from both 5' and 3' ends. BLAST search was performed to identify the homologous sequences in GenBank.

Library #1: Vector: λACT
Embryo stage: 0-6hr.
Complexity: $1X10^7$

- DEAF-1 (Deformed Epidermal Activation Factor-1), three identical clones cover full-length DEAF-1 cDNA in reverse orientation
- 1.3, one clone, no homology with known sequences
- 1.34, four identical clones, one clone with in frame deletion, no homology with known sequences
- 1.69, one clone, no homology with known sequences

Library #2: Vector: pGAD10
Embryo stage: 0-16hr.
Complexity: $2X10^6$

- DEAF-1, thirteen identical clones in correct orientation, six identical clones in reverse orientation, all cover the full-length DEAF-1 cDNA
- Adf-1 (ADH Distal Factor 1), five identical clones, cover full-length Adf-1 cDNA
- HMG-D, one clone, covers the full-length HMG-D cDNA
- 2.16, one clone, no homology with known sequences
- 2.20, one clone, no homology with known sequences

Library #3: Vector: pGADNOT
Embryo stage : 3-12hr.
Complexity: $2X10^5$

- DEAF-1, one clone, covers full-length DEAF-1 cDNA in correct orientation
- 3.15, eight identical clones, cover C-terminal truncated form of Medea cDNA
- 3.19, nine identical clones, cover the ccf-β cDNA

Figure 13

Figure 14. Deduced protein sequence of Medea and alignment with other Smad proteins.

The nucleotide sequence of Medea was a combination of the sequences from two cDNAs products (#14 and #29, corresponding to Medea A_{1-692} and Medea B in Figure 15). The deduced protein sequence was aligned with DPC4 from human, Sma4 from *C. elegans* and Mad from *Drosophila melanogaster*. Note the two highly conserved regions: the N-terminal MH1 domain and the C-terminal MH2 domain. The linker region in between is not conserved. Medea has a significantly longer Glutamine-rich linker region. Boxed amino acids belong to an alternative exon which only exists in Medea A but not in the alternative product Medea B. The GenBank accession number for Medea A_{1-692} is AF019753, for Medea B is AF019754.

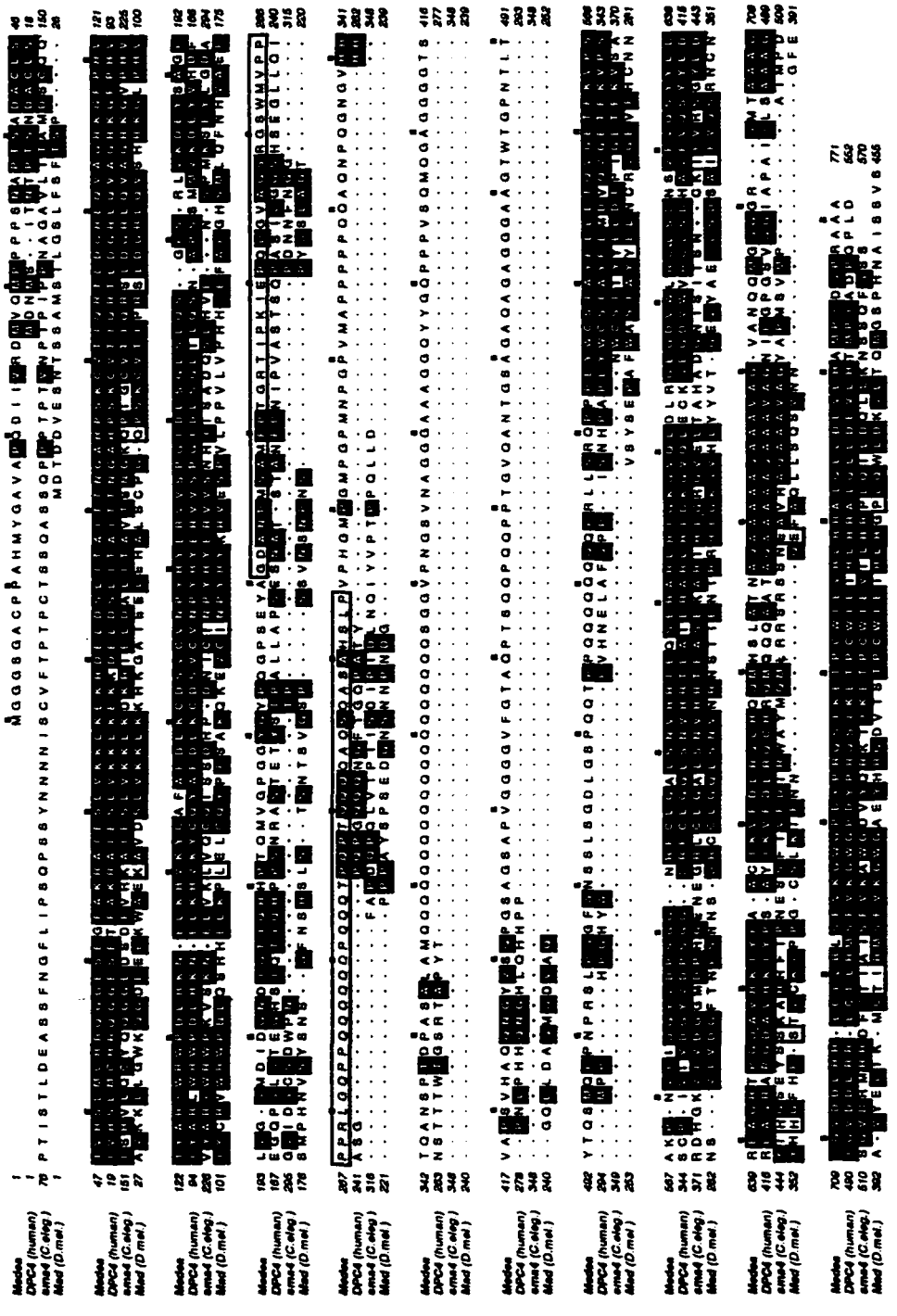


Figure 14

Figure 15. Determination of the MH1 domain as the DNA binding domain of Medea protein.

All of the constructs were fused at their 5'-end in frame with the GAL4 activation domain. They were transformed into the yeast $(D3)_5$/*lacZ* reporter strain that was used for the yeast one-hybrid screening. Their DNA binding abilities were reflected in the quantitative measurements of the β-galactosidase activities. The numbering of the Medea B constructs refers to the corresponding numbers of amino acids in Medea A constructs. The shadowed boxes are conserved MH1 and MH2 domains. Medea $A_{1\text{-}692}$ and Medea B correspond to cDNA clone #14 and #29, respectively.

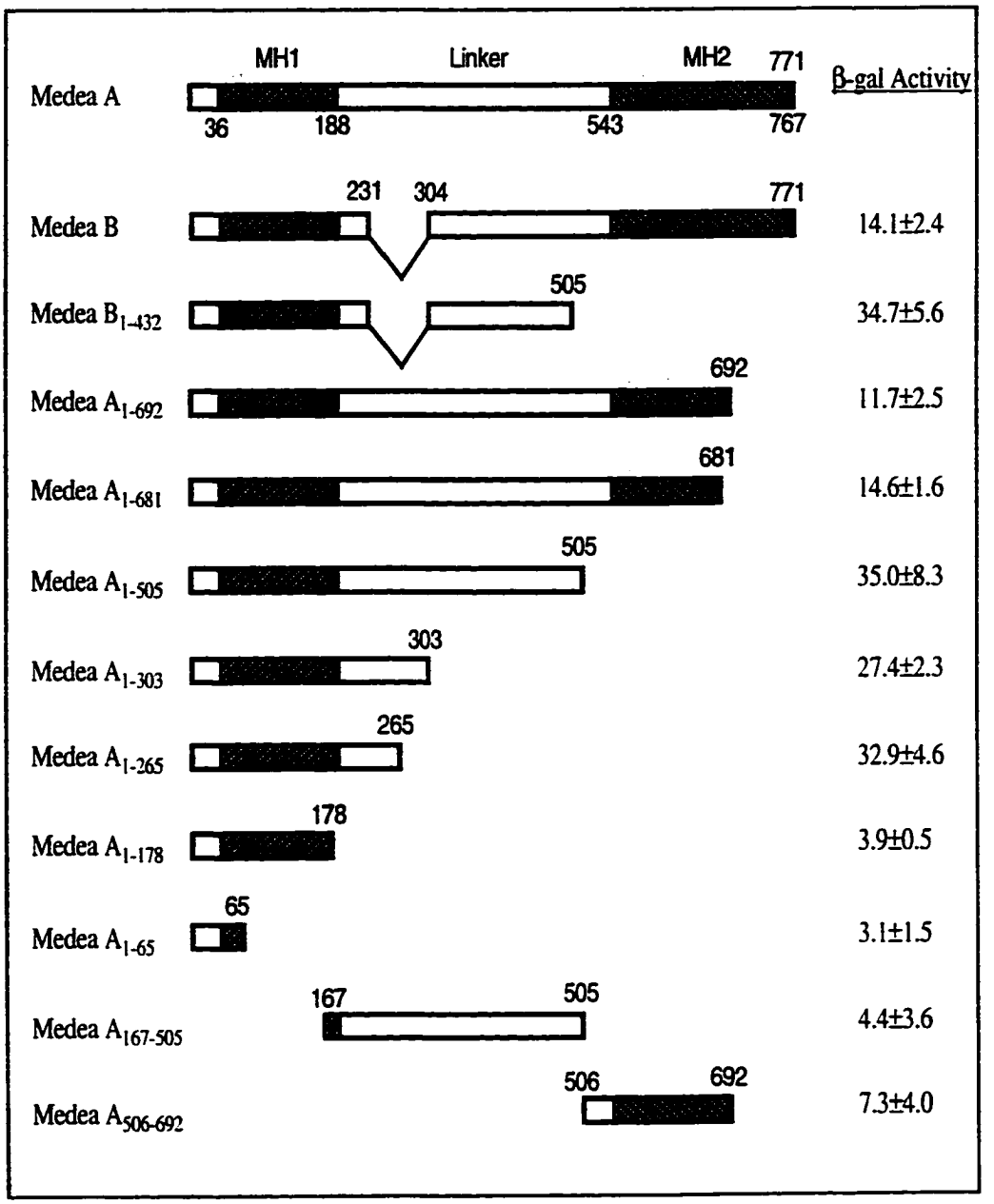


Figure 15

Analysis of Mad/Medea binding in the *tin*-D element

1. There are multiple binding sites for Smad proteins within the *tin*-D element

To further characterize the DNA binding activities of the Smad proteins, sequences encoding the *Medea* and *Mad* MH1 domains or MH1+Linker were cloned into the pGEX vector to generate bacterially expressed GST-fusion proteins (See Materials and Methods). DNaseI footprinting assays were used to identify Medea/Mad binding sites in the *tin*-D element. Surprisingly, at least seven protection sites were detected (Figure 16). The GST-fused MH1 domain and MH1+Linker generated the same protection pattern for both Medea and Mad (data not shown). The M1, M2 and M4 sites were protected by both GST-fused Medea and Mad proteins. Compared with the GST-fused Medea, GST-fused Mad protein generated a larger footprint in the M2 site and protected additional sites such as M3, M5, M6 and M7. Notably, the protection site M2 overlaps with the D3 sub-element, which was used as a bait to identify *Medea*. M6 and M7 overlap with the D6 sub-element, which is essential to *tin*-D activity. To test the possibility that a lower concentration of GST-fused Mad would generate a protection pattern similar to the one observed for GST-fused Medea, serial dilutions of the Mad protein were added to the binding reaction. However, the bands in the protected region just became uniformly weaker (Figure 16, central panel), indicating that there is a difference in binding specificity between Mad and Medea.

2. Medea/Mad bind to both D3 and D6 sub-elements

Gel shift assays were then performed to characterize the ability of Medea/Mad to bind the D3 and D6 sub-elements. When D3 was labeled as a probe, a retarded band was observed when either Medea or Mad was added. These bands could be competed specifically by excess amounts of unlabeled D3 oligo (Figure 17, lane 1,2 and 3 for Medea, lane 14, 15 and 16 for Mad). When oligonucleotide E6, which covers the D6 sub-element, was labeled as a probe, GST-fused Mad protein generated a specific shifted band, as would be expected from the footprinting results (data not shown). GST-fused Medea also bound specifically to E6 (lane 30, 31&32), although Medea protein could only generate hypersensitive bands in the DNaseI footprinting assays (Figure 16). When D3up, a sequence which is located just upstream of D3 within the M2 footprint (see Figures 16 and 18), was labeled as a probe, both Medea and Mad protein generated specific shifted bands (Figure 17, lane 27, 28 and 29 for Medea, data not shown for Mad). Based on this observation, the protection site M2 was further divided into two protection sites: M2A and M2B.

3. There are binding affinity variations between different binding sites and between Medea/Mad.

As shown in Figure 16, Medea protects M4 much stronger than M5, M6 and M7. The difference in binding affinities between the protection sites were further analyzed by gel shift assays. When the unlabeled D3up sequence was added as a competitor to the

Medea binding reaction with D3 as a probe, it competed even better than unlabeled D3 itself (Figure 17, compare lane 8&9 to lane 2&3). In contrast, similar amounts of unlabeled E6 sequence competed less efficiently (compare lane 10&11 to lane 2&3). Therefore the binding affinities for GST-fused Medea proteins are relative high for D3up, intermediate for D3 and low for E6. For GST-fused Mad, the binding affinities were shown to be relative high for D3, intermediate for E6 and low for D3up (compare lane 21&22, 23&24 to lane 15&16).

4. Medea/Mad recognize GC-rich motifs, but not the CAATGT motif.

Visual inspection of the D3 sub-element sequence revealed two GC-rich stretches and two CAATGT repeats. To test which of the two motifs contributes to Smad protein binding, oligonucleotides were designed to mutate these two 6bp motifs respectively (d3g..c and d3c..t, for sequence see Figure 18). The annealed oligos were then added to the binding reactions as specific competitors. d3g..c, the D3 derivative with mutations in GC-rich motifs, failed to compete for binding (Figure 17, lane 4&5). d3c..t, the D3 derivative with mutations in CAATGT motifs, still competed equally well as D3 itself. These results suggested that the GST-fused Medea protein recognize GC-rich motifs but not the CAATGT motifs. Similar results were observed for GST-fused Mad (lane 17, 18, 19 and 20). When the two GC-rich motifs in E6 were mutated (e6, for sequence see Figure 18), the resulting e6 oligo lost its ability to compete with D3 or E6 for Smad protein binding (lane 12&13, 25&26, 37&38). Indeed, all of the Medea/Mad binding

Figure 16. DNaseI footprinting analysis with GST-fused Medea, Mad and Tinman proteins on the *tin*-D enhancer element.

The *tin*-D element was labeled as two probes for the DNaseI footprinting assay: *tin*-D-A and *tin*-D-B (see Materials and Methods). The numbers at the left of the panels correspond to the nucleotides in Figure 11. The brackets indicate the position of the essential sub-elements which were indicated as black boxes in Figure 11. ~2ug GST-fused Medea or Mad were used except in the central panel, where decreasing amounts of Mad were used (The numbers are volumes of Mad proteins which correspond to 0.05, 0.25, 0.5 and 1ug). The GST-fused $Medea_{MH1}$ or the $Medea_{MH1+Linker}$ generated the same DNA protection pattern, which was also true for the GST-fused Mad. ~1ug GST-fused Tinman protein were used. The protected regions were summarized at the right side of the panels. Oval circles represent Medea/Mad binding sites while the rectangle-boxes represent Tinman binding sites. There are 7 protected sites detected for Medea/Mad. Notably, four of the protected sites (M2 and M3, M6 and M7) overlap with the essential sub-elements D3 and D6, respectively.

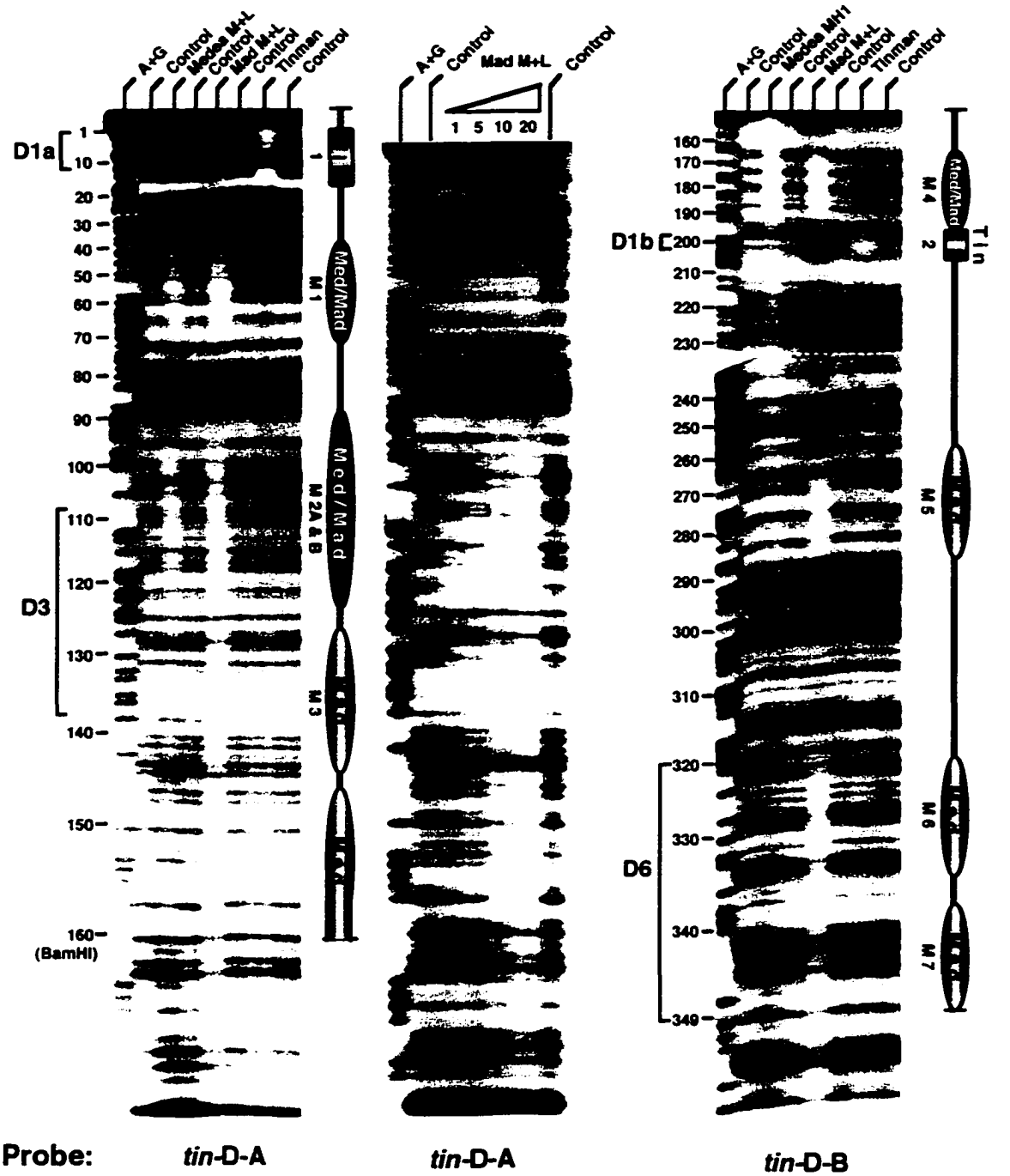


Figure 16

Figure 17. Gel mobility shift analysis with GST-fused Medea and Mad proteins on the ^{32}P-labeled sub-elements in *tin*-D.

The sequences of the probes were listed in Figure 18. D3up and D3 divided the protected site M2 into two binding sites: M2A and M2B. ~20ng of GST-fused $Medea_{MH1}$ and ~50ng GST-fused Mad_{MH1} were used for the assays. Unlabeled specific competitors were added in 10 folds and 100 folds molecular excess for Medea, or 100 folds and 300 folds molecular excess for Mad in the binding reaction, as have been indicated by the wedges. d3g..c and e6 have mutated GC-rich motifs in sub-element D3 and D6, respectively, while d3c..t has mutated CAATGT motifs in the D3 sub-element. Cross competition experiment were also done to compare the binding affinities between D3, D3up, E6 and e6.

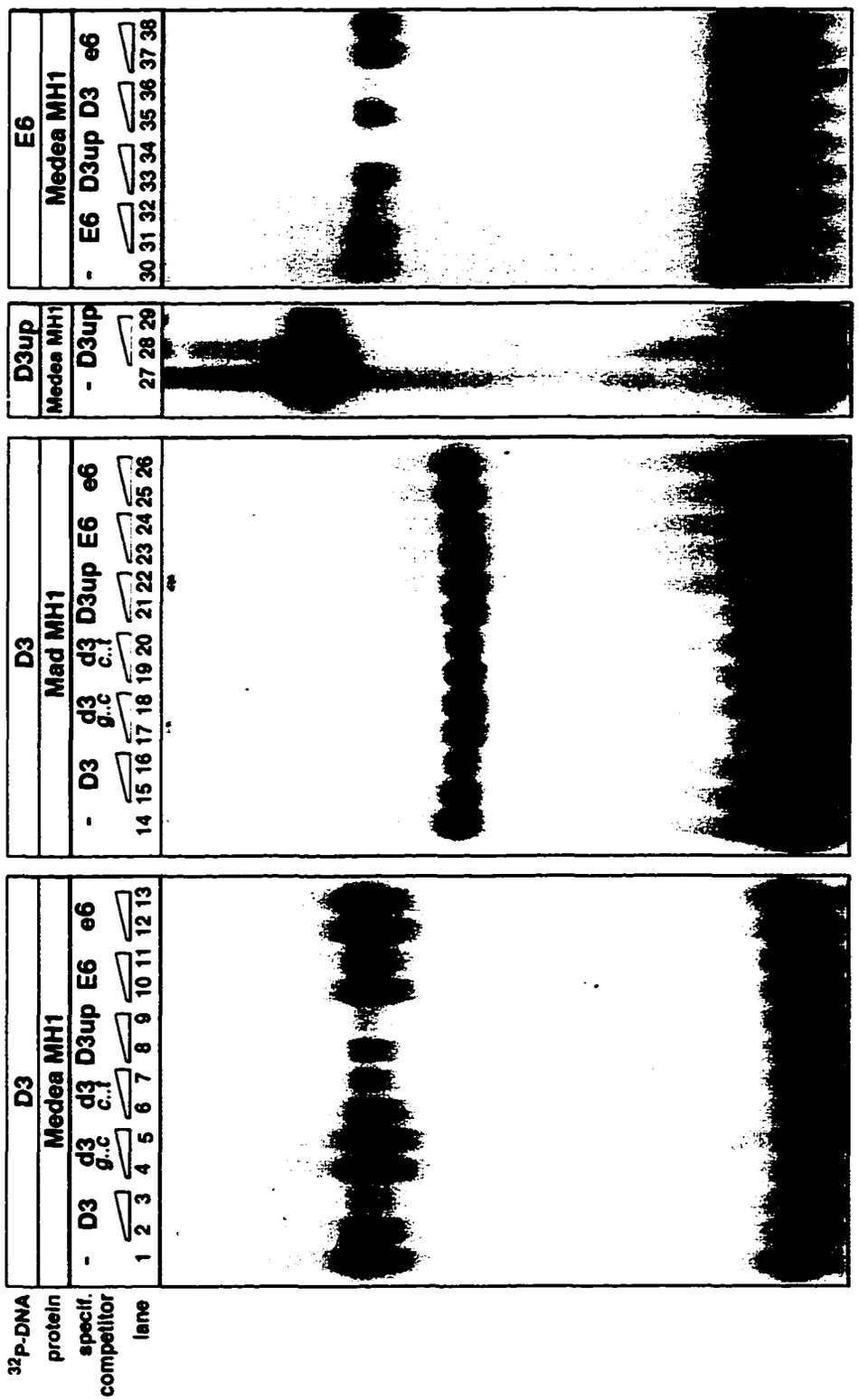


Figure 17

Figure 18. Summary of the footprinting results and the oligonucleotides used for gel shift assays.

Black boxes represent two sites protected by GST-fused Tinman. Oval circles represent sites protected by GST-fused Medea/Mad proteins. The sequences of the oligonucleotides labeled as probes for the gel shift assays were listed. The shaded regions are nucleotides that overlap with the *tin*-D element, while the nucleotides outside the shaded region are designed for the convenience of end labeling and cloning. The mutated nucleotides are underlined. Of note, E6 and e6 include both the D6 sub-element and the second Tinman binding site.

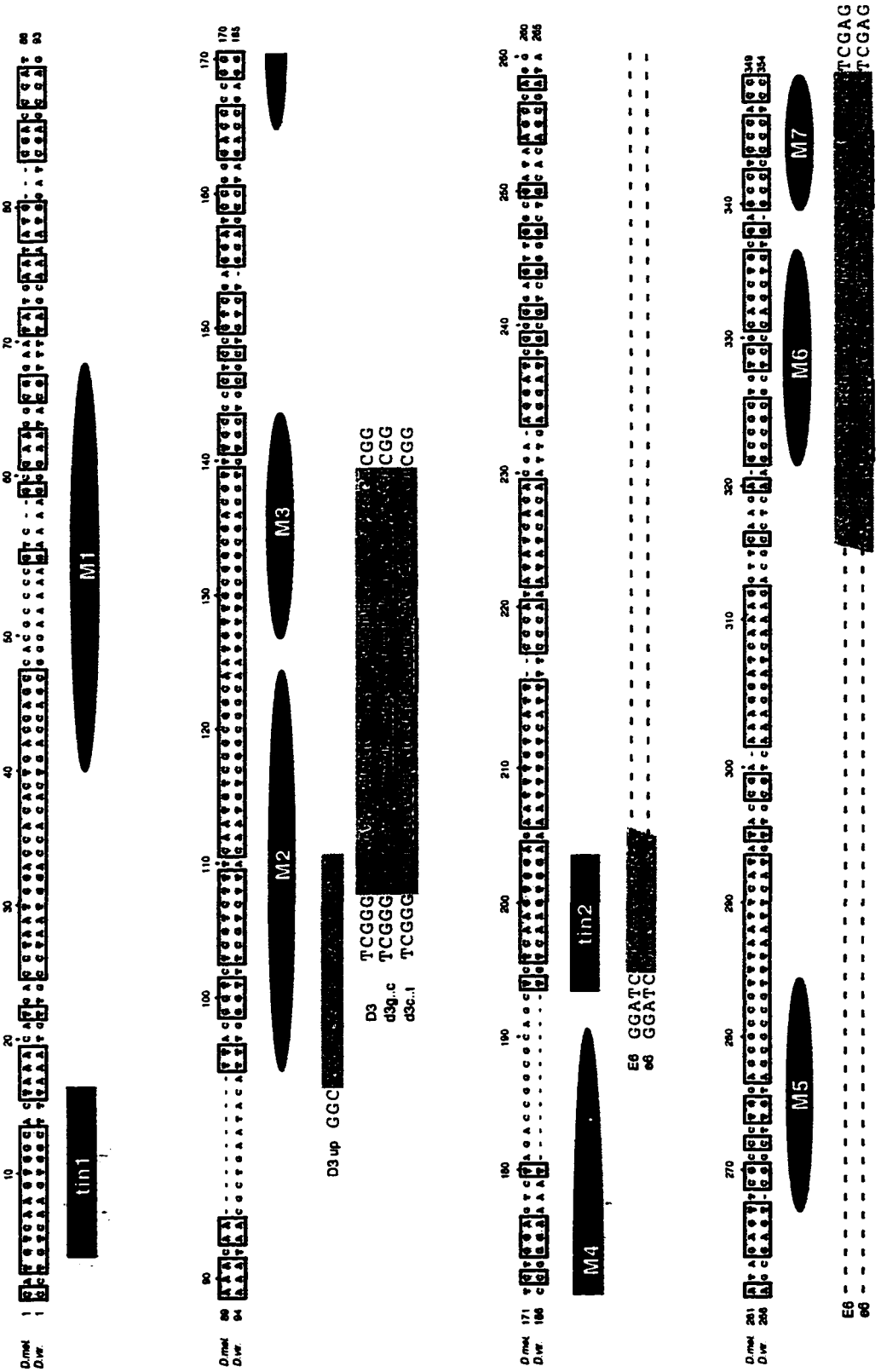


Figure 18

Figure 19. Sequence alignment of Medea/Mad binding sites.

These binding sites were identified by DNaseI footprinting assays with GST-fused Medea/Mad proteins (Figure 16). M2 is divided into M2A and M2B based on the results from the gel shift assays (Figure 17). A GC-rich consensus sequence is deduced, which is consistent with the results from the competition experiments in gel shift assays.

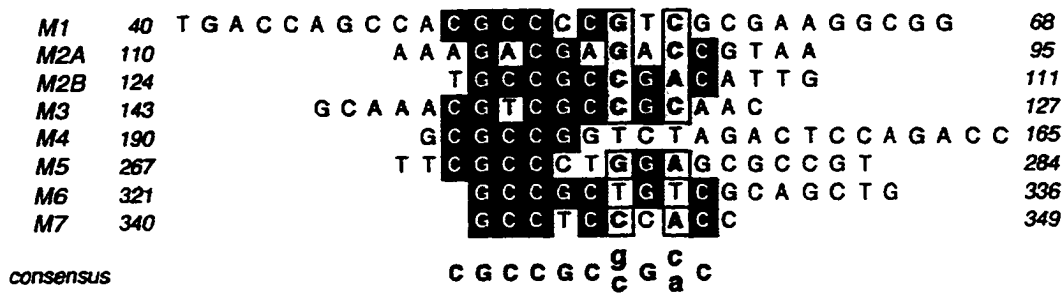


Figure 19

sites that we identified so far from *tin*-D element contain GC-rich motifs. Figure 19 shows an alignment of these binding sites and a derived consensus sequence.

***dSmad2*, a new member of *Drosophila Smad* family, does not bind to *tin*-D**

From the *Drosophila* EST project, a clone which contains sequence similarity with members of the *Smad* gene family was cloned in our lab. The deduced protein sequence contains a characteristic N-terminal MH1 domain and a C-terminal MH2 domain with a less conserved linker region in between. The C-terminal tail of the new Smad protein contains the SSXS motif. The amino acids in the L3 loop is arginine (R)and threonine (T) instead of histidine (H) and aspartate (D), which suggests that the new Smad is a *Smad2* homologue in *Drosophila* and was therefore named *dSmad2* (Lo et al., 1998). The protein sequence of the dSmad2 MH1 domain is shown in Figure 20. The MH1 domain and the MH1+linker region were cloned into the pGEX vector to generate GST-fused proteins to test whether this novel member of the Smad family has the ability to bind to DNA. When the D3 sub-element was labeled as a probe, even large amounts of either version could not generate any specific shifted bands in gel shift assays (data not shown).

Recently, two Smad binding motifs that differ from GC-rich motifs were identified. The CAGA box was derived from promoter analysis of the TGF-β target gene, *human plasminogen activator inhibitor-1 (PAI-1)* (Dennler et al., 1998). A 8bp palindromic

sequence, GTCTAGAC, was defined as the optimal binding element for Smad proteins (SBE) based on an *in vitro* PCR selection strategy (Zawel et al., 1998). Both Medea and Mad can bind to the CAGA box (Figure 21, lanes 1 and 2) or the SBE motif (Figure 21, lanes 5 and 6) when oligonucleotides containing either these two elements were labeled as probes, respectively. However, the GST-fused dSmad2 failed to bind to either of these motifs (Figure 21, lanes 3 and 7). *In vitro* selection experiments indicated that also Smad2 proteins in vertebrates failed to bind to DNA (Zawel et al., 1998). Thus it seems that not all of the Smad family members contain DNA binding activity.

The study of the crystal structure of the MH1 domain of Smad3 revealed a novel DNA-binding motif (Shi et al., 1998). The DNA contacting region is a 11-residue β hairpin structure that is embedded in the major groove of DNA. The position of this β hairpin structure is marked by the shaded amino acids in Figure 20. In dSmad2, 9 out of the 11 residues from this region differ in composition to those in the other Smad subtypes. In vertebrate Smad2 proteins, there are two insertions which do not exist in other Smads (Figure 20). The 30 amino acid insertion (stretch 2) is located just before the β hairpin structure, which could interfere with DNA binding activity. Therefore, dSmad2 and the vertebrate Smad2 may have adopted two different evolutionary paths to disrupt the DNA-binding activity of the MH1 domain.

To test whether the 9 residue change in the β hairpin structure is the major reason for the

lack of DNA binding activity of dSmad2, these residues were mutated back by *in vitro* mutagenesis. Results from gel shift assays indicated that the GST-fused protein of this mutated version of the dSmad2 restored DNA-binding activity weakly when CAGA boxes and SBE were labeled as probes (Figure 21, lanes 4 and 8). Even in the presence of larger amounts of the dSmad2 protein, the shifted bands were still much weaker than those with Mad/Medea. This mutated version of dSmad2 failed to generate a specific band when D3 was labeled as a probe (data not shown). It seems there are other structural features that decrease dSmad2 DNA binding in addition to the 9-residue change.

Medea/Mad binding sites in D3 and D6 are required for induction by Dpp

To test whether the Medea/Mad binding sites identified *in vitro* are functional *in vivo*, the ~6bp GC rich motifs in the D3/D6 sub-elements that are required for Medea/Mad binding were mutated. For cloning convenience, the *lacZ* activity of these mutated *tin*-D derivatives was tested in the context of *tin*-D*, a truncated version of *tin*-D (Figure 22, see Materials and Methods for cloning details). *tin*-D* lacks residues 144-194 and 205-315 which include the two Medea/Mad binding sites M4/M5 and the two Schnurri binding sites. *tin*-D* still drives reporter gene expression in the dorsal mesoderm (Figure 23A). When the two 6bp GC rich motifs in D3 were mutated (*tin*-D*-d3g..c, for sequence see Figure 21), dorsal mesoderm expression pattern of the *tin*-D* was ablated (Figure 23B). Similar disruption of the *tin*-D* activity was also observed when the two

Figure 20. Sequence alignment of the MH1 domain of dSmad2 with other Smad proteins.

Only the conserved MH1 domains of the Smad proteins are shown. Boxed amino acids are conserved between Smad proteins. The most conserved residues are listed below the alignments. The shaded 11 residues correspond to the position of the β-hairpin structure which is used to contact with DNA (Shi et al., 1998). 9 out of these 11 amino acids in the dSmad2 protein (#69-#77) are not conserved. Bracketed regions indicate the two stretches of amino acids that are only present in various Smad2 proteins. Stretch 2 is located right before the β-hairpin structure and could interfere with the DNA binding activity of Smad2.

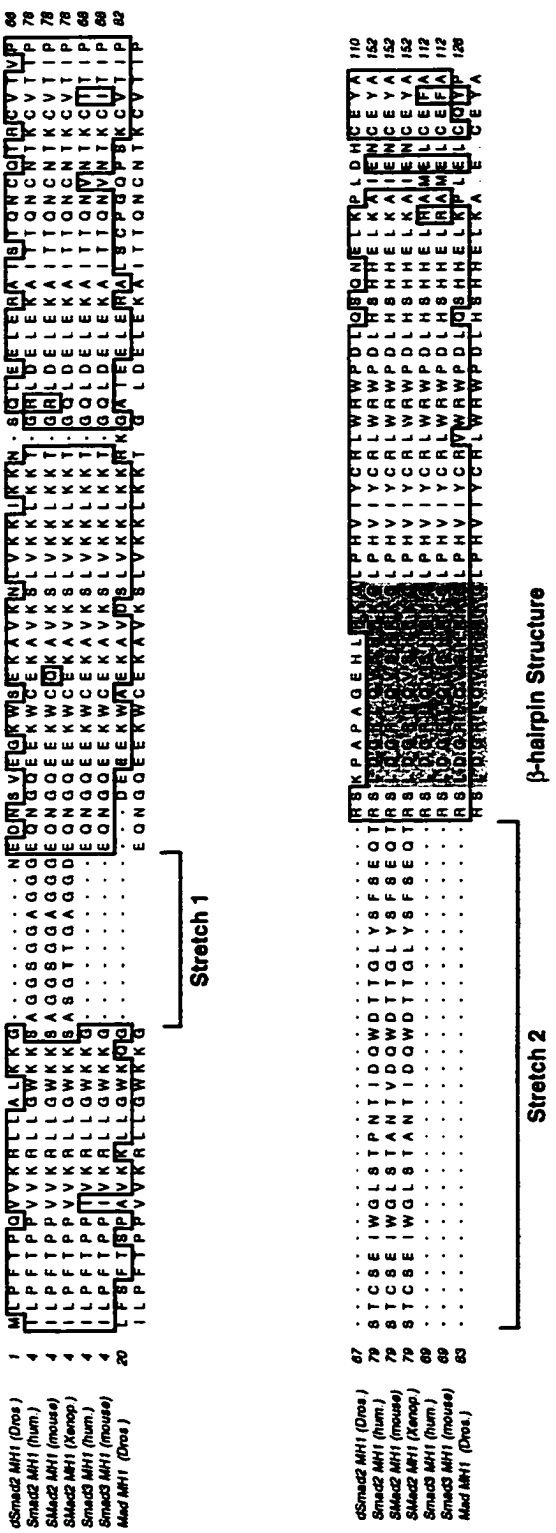


Figure 20

Figure 21. Gel mobility shift analysis of the DNA binding activity of wild-type and mutated dSmad2 proteins.

Oligonucleotides containing either CAGA box or Smad optimal binding site SBE element were labeled as probes (for sequences, see Materials and Methods). ~10ng of GST-fused $Medea_{MH1}$, ~25ng GST-fused Mad_{MH1}, ~200ng GST-fused $dSmad2_{MH1+L}$ (represented by dSmad2 for lanes 3 and 7) or $dSmad2_{MH1+L}$(mut.) (represented by mut. for lanes 4 and 8) were used for the binding reaction. The gel was overexposed for four days to show the weak bands generated by the GST-fused $dSmad2_{MH1+L}$(mut.). The signals in lanes 3 and 7 were generated by spill-over from lanes 2 and 6 when the same tip was used during the loading procedure.

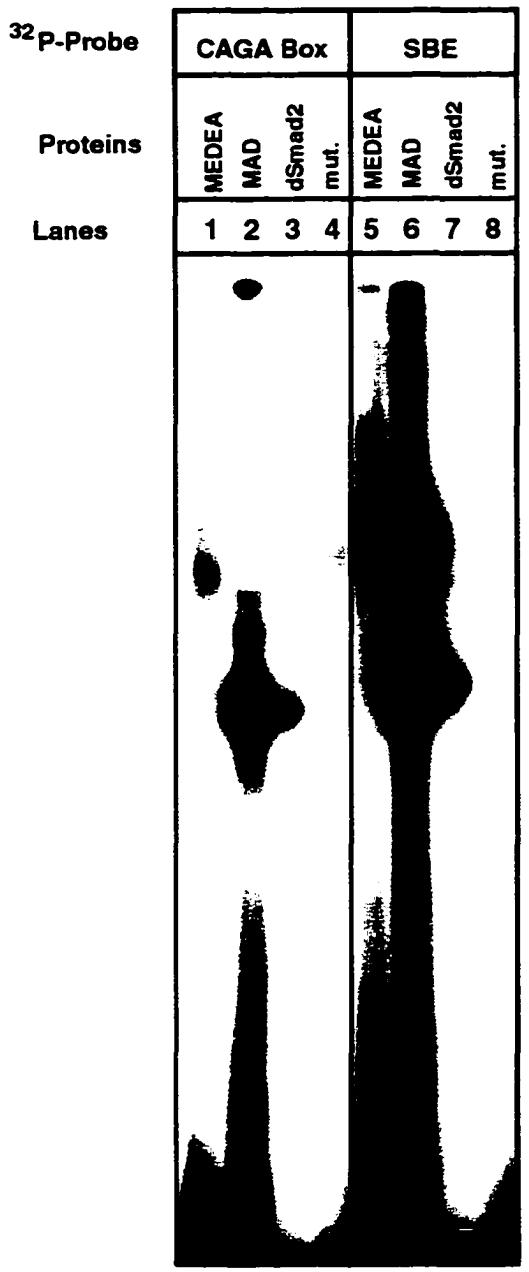


Figure 21

GC-rich motifs in D6 (*tin*-D*-d6) or all four GC-rich motifs (*tin*-D*-d3,d6) were mutated (data not shown, see Figure 24 for a schematic summary). Together, these data demonstrate that these Medea/Mad binding sites are required for *tin*-D activity.

Moreover, when the two CAATGT motifs in the D3 sub-element were mutated (*tin*-D*-d3c.t, see Figure 22 for sequence), *tin*-D* activity was also abolished (Figure 23C). Previous gel shift data had shown that the CAATGT motif does not contribute to the Medea/Mad binding (Figure 17, lane 6&7, 19&20). This suggests that there is an unknown transcription factor which recognizes this motif and is required for *tin*-D activity.

One interesting difference between the expression pattern of *tin*-D and *tin*-D* is the ectopic expression in the amnioserosa (Figure 23A), suggesting that there is a repressor element for this tissue located in the deleted region (residues 144-194 and 205-315). The GC-rich motifs in the D3 and D6 sub-elements are required for the amnioserosa expression (Figure 23B), while the CAATGT motifs are not (Figure 23C, arrows).

The above results and the results from the other *tin*-D derivatives generated in this study are summarized schematically in Figure 24. Listed are Tinman binding sites, Medea/Mad binding sites and the CAATGT motifs (Figure 24).

Figure 22. Sequences of the constructs used to test the *in vivo* functions of Mad/Medea binding sites in *tin*-D.

The mutations of the Medea/Mad binding sites were tested based on the *tin*-D* construct, a version of *tin*-D that lacked nucleotides 144-194 and 205-315. The first part of these constructs came from the *tin*-D-#1 construct, covering nucleotides 1-100, which is shown by the first shaded box. The second part of *tin*-D* series came from a set of oligonucleotides, covering nucleotides 94-143, which is shown by the second shaded sequences. The third part of the *tin*-D* series came from another set of oligonucleotides, covering nucleotides 195-204 and 316-349, which are shown by the third and fourth shaded sequences. The fact that the third and fourth shaded sequences came from the same nucleotides is indicated by the slashed shaded boxes. Underlined sequences are mutations introduced to disrupt either the GC-rich motifs or the CAATGT motifs. Boxed and unshaded sequences in the oligos are nucleotides that were lost in the final *tin*-D* constructs after cloning. Tinman and Medea/Mad binding sites are shown for comparison.

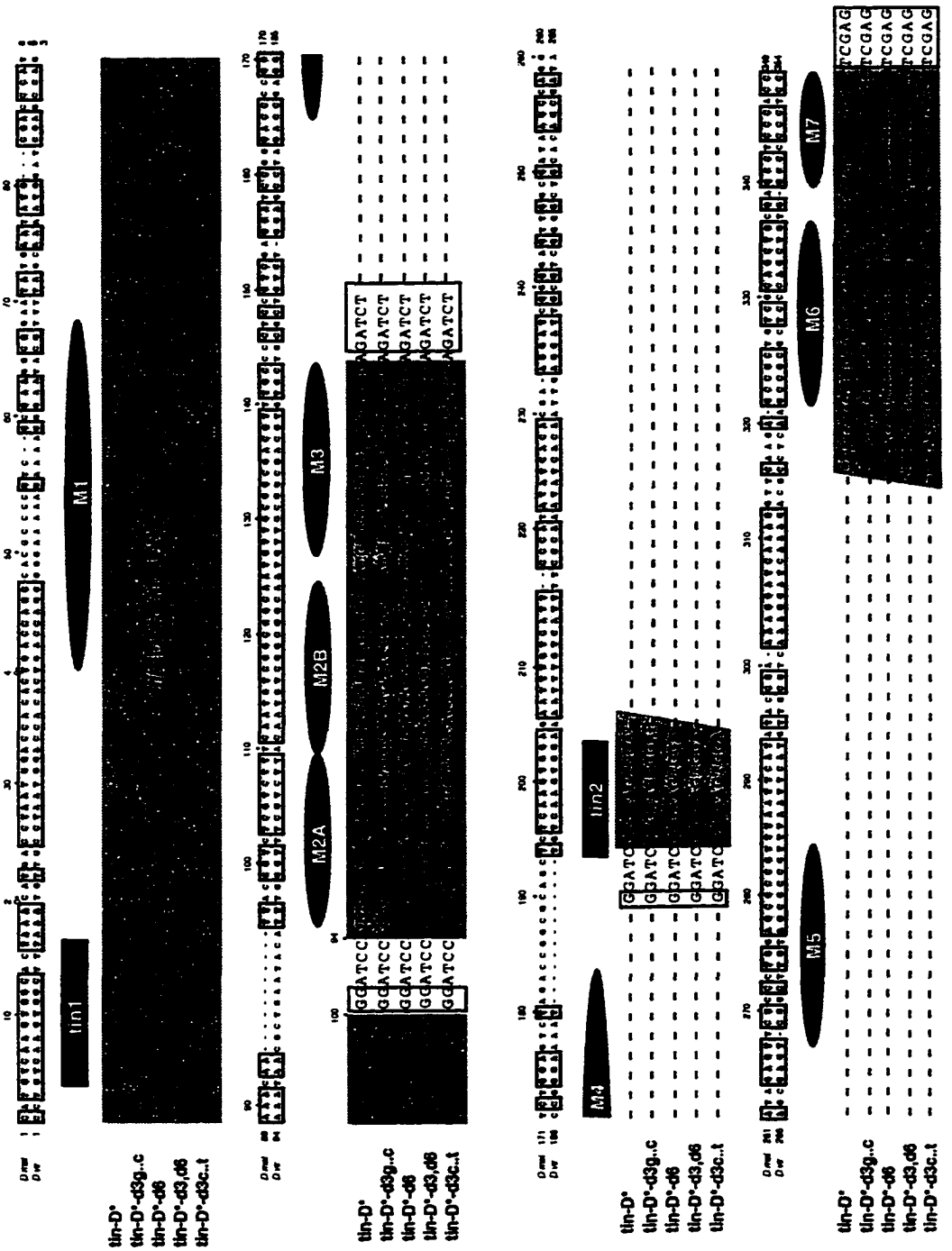


Figure 22

Figure 23. *in vivo* activity of the *tin*-D* and *tin*-D* derivatives that carry nucleotides exchanges.

Shown are cross sectioned embryos stained with *lacZ* antibody. (A) *tin*-D* drives *lacZ* reporter gene expression in the dorsal mesoderm at a level comparable to that of *tin*-D (see Figure 6F). In addition, there is ectopic expression in the amnioserosa. (B) The nucleotide exchanges in *tin*-D*-d3g..c disrupt the dorsal mesoderm expression, suggesting that the Medea/Mad binding sites are essential for *tin*-D *in vivo* activity. (C) The nucleotides exchange in *tin*-D-d3c..t also disrupt the *lacZ* expression in the dorsal mesoderm, suggesting the functional importance of the CAATGT motifs. However, the ectopic expression in the aminoserosa is still observed. (ms) mesoderm. (as) amnioserosa.

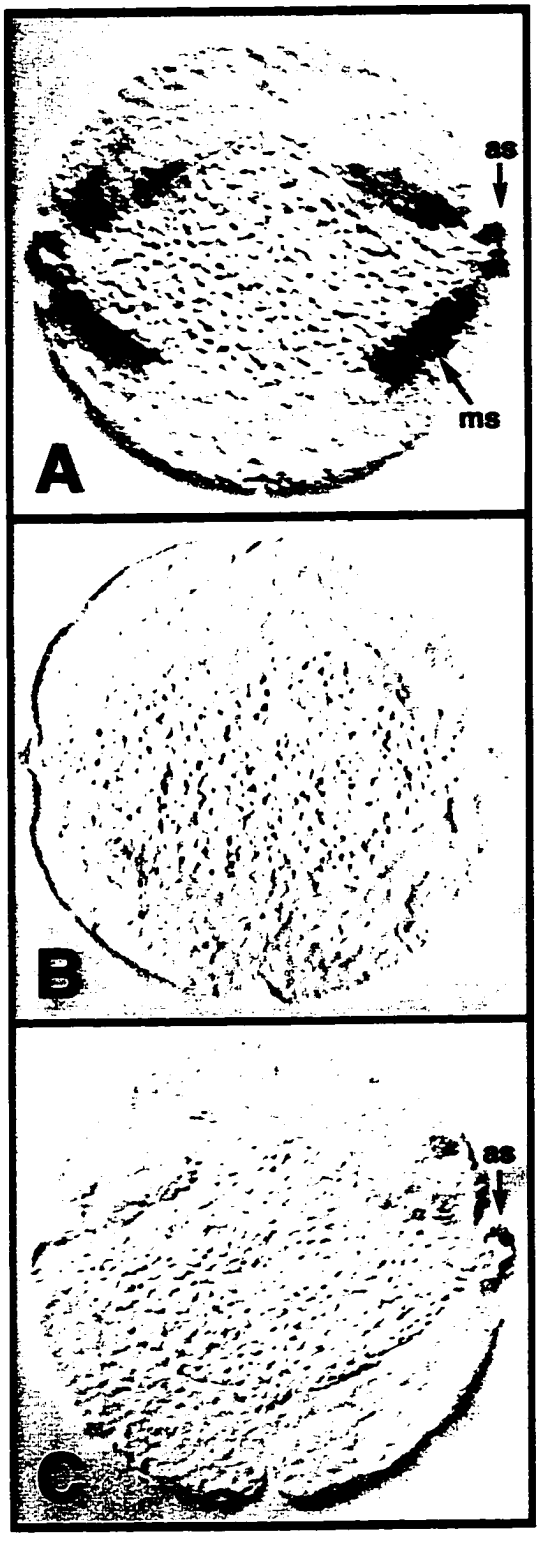


Figure 23

Figure 24. Schematic summary of the *in vivo* activities of *tin*-D derivatives in transgenic embryos.

The *tin*-D element is simplified as a bar with two *tinman* binding sites as black boxes, eight Medea/Mad binding sites as oval circles and two CAATGT motifs as triangles. The essential Medea/Mad binding sites are underlined and in bold. Internal deletions, combinations of multiple copy sub-elements and the *tin*-D* series containing the nucleotide exchanges are also listed. Dashed lines indicate the deletions. Opened circles represent nucleotides changes. The *lacZ* reporter expression pattern driven by these constructs in *Drosophila* embryos are summarized at the right side in three columns, representing dorsal mesoderm, dorsal ectoderm and amnioserosa. The figure number that showed the embryo staining picture are listed when available. (d.ms) dorsal mesoderm. (d.ec) dorsal ectoderm. (as) amnioserosa. (n.s.) not shown.

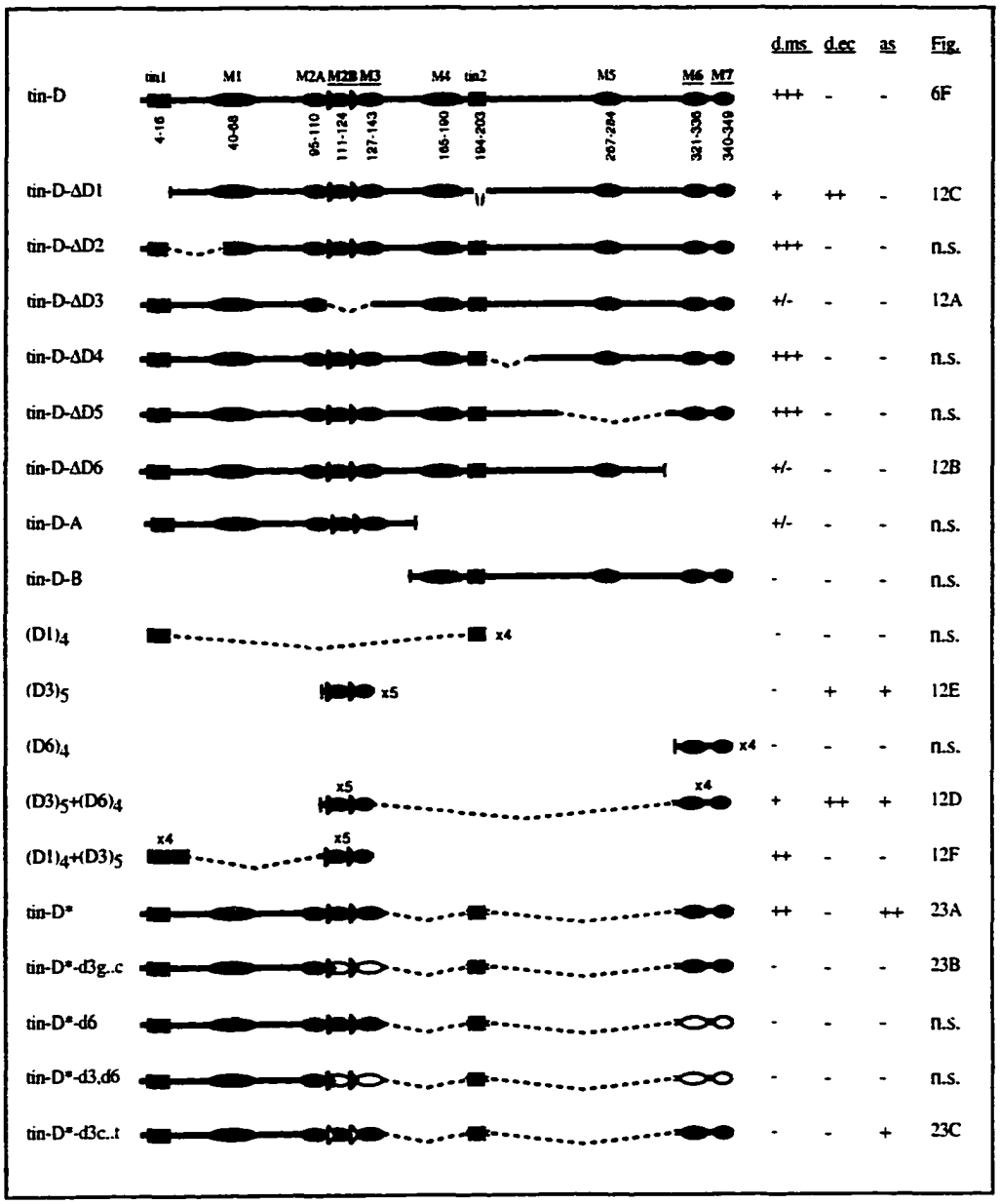


Figure 24

In collaboration with Dr. E. Ferguson's lab, embryos derived from germline mutations in *Medea* alleles were collected. In *Medea* mutant background, *tin*-D-*lacZ* activity is virtually extinguished (see Xu et al., 1998). In addition, the injection of the mRNA encoding constitutively active Dpp receptors ($tkv^{Q\text{-}D}$) into these mutant embryos failed to induce ectopic expression of the *tin*-D activity in the mescderm, in contrast with the embryos containing wild type *Medea* alleles (Xu et al., 1998). These observations strongly suggest that *Medea* is directly required to transduce the signal from the Dpp receptor *tkv* to the Dpp target element *tin*-D.

I-D_L, a Dpp-responsive element from the *mef2* promoter, also contains multiple Medea/Mad binding sites and CAATGT motifs

Drosophila mef2 is a member of the MADS box supergene family of transcription factors (Olson et al., 1995). Genetic studies have demonstrated that its function is required for the terminal differentiation of the three major types of musculature, namely somatic, visceral and cardiac (Bour et al., 1995; Lilly et al., 1995; Ranganayakulu et al., 1995). Like that of the *tinman* gene, the expression of *mef2* is first observed in the whole presumptive mesoderm, which is then restricted to the dorsal mesoderm (Lilly et al., 1994; Nguyen et al., 1994; Taylor, 1995). At late stages, *mef2* expression is restricted to somatic muscles, visceral musculatures and the heart, while it is excluded from non-myogenic mesodermal tissues.

The restriction of *mef2* to the dorsal mesoderm is also a Dpp-dependent event (Nguyen and Xu, 1998). I-D_L, a 460bp *cis*-regulatory enhancer element which drives reporter gene expression in the dorsal mesoderm was isolated from the 5' flanking region of the *mef2* gene. Sequence examination of the I-D_L element uncovered three GC-rich stretches and two CAATGT motifs (Figure 25A, boxed sequences). To test the possibility that these GC-rich sequences could bind Medea/Mad proteins, three oligonucleotides covering these sequences (GC-1, GC-2 and GC-3, see Figure 25A) were labeled as probes to perform gel shift assays. All three oligos bound both GST-fused Medea (Figure 25B, lane 1,4,8) and Mad proteins (Figure 25C, lane 1,5 and 9). The shifted bands could be competed by excess amounts of the unlabeled oligos (for Medea, see Figure 25B, lane 2,5,9; for Mad, see Figure 25C, lane 2,6 and 10). In contrast, GC-1mut/GC-2mut/GC-3mut derivatives of the oligos, which contain mutations in the GC-rich motifs, or the unrelated oligo D1 (for sequences, see Materials and Methods), could not compete for binding (for Medea see Figure 25B, lanes 3,6,7,10, 11; for Mad see Figure 25C, lanes 3,4,7,8,11,12). These results indicate that the bindings are specific and that Medea/Mad recognize the GC-rich motifs.

DNase I footprinting assays were then performed to confirm these results and to identify other sites protected by Medea/Mad proteins in the I-D_L element. Three protected regions overlapped with the GC-1, GC-2 and GC-3 oligos, although the Medea protection was much weaker in the GC-3 region compared with the Mad protection

Figure 25. Medea/Mad can bind to the GC-rich sequences in the Dpp-responsive element from the *Drosophila mef2* promoter region.

(A) Sequence of the I-D_L element from *mef2*. Boxed are three GC-rich sub-elements (GC-1 to 3) that were labeled as probes for gel shift assays. Bracketed sequence are protected sites by GST-fused Medea/Mad proteins identified in DNaseI footprinting assays (see Figure 26). Dashed bracket indicates additional nucleotides protected by GST-Mad but not by GST-Medea. MM-1 and MM-2 are two more protected sites by Mad/Medea besides GC1-3. CAATGT motifs are circled in ovals. (B) Gel shift assays with the GST-fused $Medea_{MH1}$ protein. ~50 ng proteins were used. 100 folds molecular excess of unlabeled oligonucleotides were added to the binding reactions as specific competitors. (C) Gel shift assays performed with GST-fused Mad_{MH1} proteins. 300 folds molecular excess of unlabeled oligonucleotides were added to the binding reactions as specific competitors. All of the three probes, GC-1, GC-2 and GC-3 are bound specifically by both Medea and Mad. The GC-rich motifs contribute to the binding. For the sequences of the competitors, see Materials and Methods. The oligonucleotide D1 which contains two Tinman binding sites was included as a control of non-specific competitor.

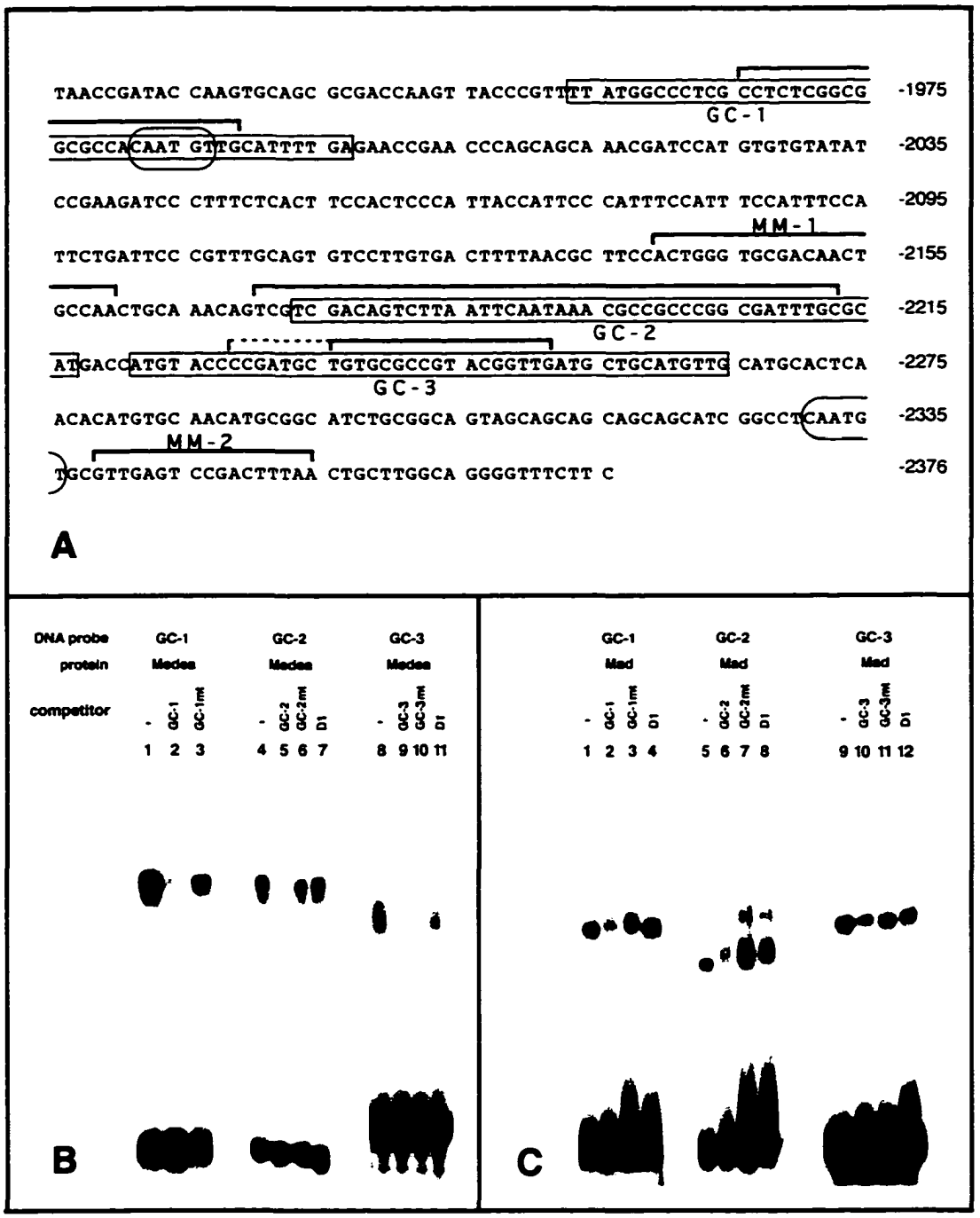
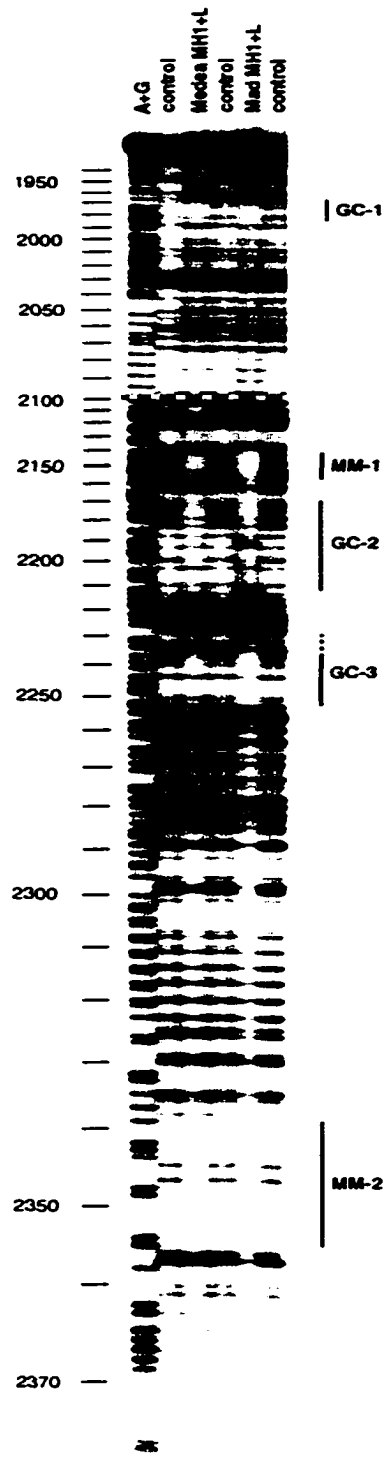


Figure 25

Figure 26. DNaseI footprinting assays using the full-length I-D_L element from *mef2* as a probe.

~2ug GST-fused Medea or Mad were used. No proteins was added for control lanes. GST-fused $Medea_{MH1}$ and $Medea_{MH1+Linker}$ protected the same region in the I-D_L element, which was also true for GST-fused Mad (data not shown). The protected sites are summarized at the left side of the panel, whose sequences were covered by the brackets in Figure 25A. MM-1 and MM-2 are two additional binding sites identified that are not related to the GC-rich consensus derived from the *tin*-D element study.



Probe: ***D-mef2*** **I-D$_L$**

Figure 26

(Figure 26). Two additional protection sites (MM-1 and MM-2) were detected, neither of which was GC-rich (for sequence, see Figure 25A), confirming that the GC-rich motif is not the only recognition site for Medea/Mad.

Is a *Drosophila* FAST-1 homologue the CAATGT Binding factor?

Three pieces of evidence in this study suggested that there is a CAATGT binding factor which functions together with Smad proteins to activate the *tin*-D element. First, the binding of Medea/Mad on their own is not sufficient to stimulate the transcription, as was indicated by the construct containing four copies of the sub-element D6 (see Figure 24). Both CAATGT motifs and the Medea/Mad binding sites are required for the minimal Dpp-response, as was reflected by the construct $(D3)_5$. Second, CAATGT motifs are required for the tin-D* activity *in vivo* (Figure 23), although they do not contribute to the Smad binding (Figure 17). Third, the CAATGT motifs are also detected in the Dpp-responsive element from *mef2* promoter (Figure 25A).

Xenopus FAST-1 was cloned through the yeast one-hybrid screening using the activin-responsive element (ARE) from the *Mix.2* gene as a bait. The AAATGT repeats in the ARE were considered to be the xFAST-1 recognition sites (Chen, X et al. 1996). Moreover, xFAST-1 functions as a partner for Smad proteins to transcriptionally activate *activin* downstream target genes (Chen, X. et al., 1996, 1997). Because of the similarity with the AAATGT motif, we suspected that the factor binding to the CAATGT motif

may be a *Drosophila* homologue of *xFAST-1*. To collect more evidence for this hypothesis, GST-fused xFAST-1 protein was extracted to perform gel shift assays with the *Drosophila tin*-D3 sub-element as a probe. Shifted bands were observed which could be competed by excess amounts of unlabeled D3 oligonucleotides (Figure 30). The D3 derivative which contains mutations in the CAATGT motif (d3c..t) failed to compete with the binding. These results confirm that xFAST-1 does have the ability to bind to the *Drosophila* CAATGT motif in the *tin*-D3 sub-element, which could support of our hypothesis that there is a *Drosophila FAST-1* homologue.

Several attempts to clone the *Drosophila FAST-1* homologue through a degenerate PCR approach were unsuccessful. Two other tries by Z.Yin in this lab through a low stringency cDNA library hybridization approach also failed. A recent study on the binding motifs for human *FAST-1* indicated that the optimal binding sites are TGT(G/T)(T/G)ATT rather than AAATGT (Zhou et al., 1998). The ARE element from the *mix.2* promoter contains a perfect match (TGTGTATT) to this consensus sequence, which overlaps with the TGT tri-nuleotides in one of the two AAATGT repeats. It is likely that the TGTGTATT sequence instead of AAATGT is actually responsible for the xFAST-1 binding. In *tin*-D, the sequences follow the CAATGT motifs are TGTCGGCG and TGTTGCGG, which do not match perfectly to the FAST optimal binding sequence. These studies on the hFAST-1 binding sites and the negative results from our screens, make our hypothesis less likely.

Other potentially involved transcription factors in Dpp-mediated *tinman* induction: *HMG-D, ccf-β, Adf-1, DEAF-1* and *schnurri*

During the yeast one-hybrid screening using $(D3)_5$ as a bait, several other positive clones were identified (see Figure 13). Preliminary experiments were performed to test their potential as sub-element D3 binding transcription factors and their possible involvement in mesoderm induction.

1. *HMG-D*

High-mobility-group (HMG) proteins are a class of abundant non-histone chromosomal proteins which are thought to play important roles in chromosomal structure and organization. During *Drosophila* early embryogenesis, *HMG-D* is thought to organize a condensed state of chromatin during the first six nuclear cleavage cycles to facilitate rapid nuclear cycles (Ner and Travers, 1994). Later, its chromatin organization role is taken over by histone H1 protein. HMG-D preferentially binds to DNA containing the dinucleotide TG (Churchill et al., 1995). In addition to their general role as chromatin organizers, HMG proteins also have been shown to regulate specific gene transcription by binding and bending the enhancer region (Falvo et al., 1995). It was demonstrated that HMG1 could interact with homeotic proteins and enhance their DNA binding activity (Zappavigna et al., 1996). However, another group provided evidence showing that HMGI-C and HMGY inhibit the function of homeotic proteins by interfering with their DNA binding activity (Arlotta et al., 1997).

During the one-hybrid screening, *HMG-D* was identified from library #2. *HMG-D* was also isolated in a yeast two-hybrid screening when using Tinman as a bait by S. Zaffran in this lab. These data suggest that the HMG-D protein can bind to the *tin*-D element and interact with the bound Tinman protein. In addition, the expression pattern of HMG-D includes the mesoderm (Stroumbakis and Tolias, 1994), which allows for the possibility that HMG-D is involved in the restriction of *tinman* expression to the dorsal mesoderm.

To characterize its DNA binding ability, *HMG-D* was subcloned into the pGEX-vector to generate GST-fused proteins. A retarded band was observed when the D3 sub-element was used as a probe in gel shift assays (Figure 27, lane1). This band could be competed by excess amounts of unlabeled D3 (lane 2), or even better by a D3 derivative which contains mutations in the GC-rich motif (d3g..c, lane 4). Interestingly, mutation of the CAATGT motif in D3 disrupts its ability to compete (d3c..t, lane 3). These results suggested that HMG-D recognizes ~30bp oligonucleotide sequences preferentially, if not specifically. To test the possibility that HMG-D specifically recognizes the CAATGT motif, the oligo GC-1, which is derived from *mef2* Dpp-responsive element I-D_L and also contains CAATGT motifs, was used as a specific competitor. GC-1 competed efficiently for the specific binding of HMG-D to the D3 probe (compare lane 6 with 5). However, in contrast with d3c..t, mutation of the CAATGT motif in the GC-1 oligo from *mef2* (GC-1c..t) did not disturb its ability to compete with the HMG-D binding (lane 7). Moreover, the GC-2 oligo, which does not contain a CAATGT motif, also competed

efficiently with the HMG-D binding (lane 9). These results argued against that HMG-D specifically binds to the CAATGT motif. In agreement with this notion, DNaseI footprinting assays failed to generate any specific protection sites for HMG-D when *tin-D* was labeled as a probe (data not shown). However, HMG-D does prefer to bind certain ~30 bp DNA sequence, which was further confirmed by the observation that comparable excess amounts of D1 oligonucleotide could not compete with the specific binding of the D3 probe (lane 8). In summary, HMG-D prefers to bind ~30bp sequences like D3, GC-1, GC-1c..t, GC2 rather than d3c..t and D1. How this binding preference contributes to the *in vivo* function of HMG-D remains unclear.

Possible interactions between the Tinman and HMG-D proteins were also characterized by gel shift assays. When D3 was labeled as a probe, HMG-D generated a specific shifted band (Figure 28A, lane 1, band indicated by an arrowhead), as shown before. With the addition of GST-fused Tinman, which by itself can not bind to D3 (data not shown), a supershifted band was observed in addition to the HMG-D binding band (Figure 28, lane 2, band B1). The derivatives of Tinman proteins containing either the N-terminus (aa. 1-124) or the C-terminus (aa. 301-416) alone failed to generate any supershifted bands (lane 3 and 4), while the derivative that deletes the 10 amino acids around the N-terminal TN domain maintained the ability to generate a supershifted band (band B2). As has been expected from the size difference of the proteins, the band B2 is slightly lower than B1 (lane 5). The reason for the band B2 being stronger than B1 could

be due to the higher specific activity of the tin-ΔTN protein extracts. These results confirmed that Tinman can physically interact with HMG-D, and also indicated that the TN domain in the Tinman protein is not required for this interaction. Further experiments are needed to define the location of the interaction domains within the two proteins.

When D1 was labeled as a probe, GST-fused Tinman full-length protein (FL) generated two specific shifted bands (Figure 28B, lane 6, Band B3 and B4). As the D1 oligo contains two Tinman binding sites (Figure 11) , it is likely that B4 represents a Tinman monomer while B3 represents two molecules of Tinman protein bound to one D1 DNA molecule. Band B3 became weaker when the amounts of the Tinman protein were decreased (compare lane 8 with 6, 7). At the lowest amounts of Tinman, the addition of the GST-fused HMG-D protein strongly enhanced the band B3 (lane 9). Doubling the amounts of HMG-D made the band B3 even stronger and generated more complexes which were stuck in the well (lane 10). Interestingly, further doubling the amounts of the HMG-D disrupted the B3 and B4 bands generated by Tinman binding, and much more complexes remained in the well (lane 11). In addition, an HMG-D binding band (indicated by the arrowhead) and a new shifted band were observed (B5). The band B5 could represent a supershifted band consisting of HMG-D and Tinman protein. The highest amounts of the GST-fused HMG-D protein alone generated one shifted band, as indicated by the arrowhead (lane 12). Similar results were obtained when Tinman-CT, a truncated

version of Tinman protein (aa. 301-416) containing the C-terminal homeo domain, was used instead of the full-length Tinman protein. Two shifted bands was also observed (lane 13, B6 and B7). B6 was weakened with decreasing amounts of the Tinman-CT protein (lane 14, 15). The addition of HMG-D enhanced the strength of B6 dramatically (lane 15, 16). Higher amounts of the HMG-D protein disrupted the bands B6 and B7, and some complexes in the well but no additional band were observed (lane 18). In summary, low amounts of HMG-D protein enhances the DNA-binding activity of Tinman protein. The Tinman-CT (aa. 301-416) is sufficient for this interaction. In contrast, higher amounts of HMG-D may interfere with the DNA binding of the Tinman homeo domain, possibly through the DNA-binding activity of HMG-D. The results in this study resemble the seemingly conflicting results from the previous reports about the effects of HMG proteins on the binding activity of the homeotic proteins. Here, we show that HMG-D proteins can physically interact with the NK family of the homeo box-containing proteins and then enhance or disrupt their DNA binding activity, depending on the amounts of HMG-D.

2. *ccf–β*

Clone 3.19 from library #3 represents 9 positive clones which share an identical insert. The sequence of the 1.5kb insert and the deduced protein are shown in Figure 29. Interestingly, the 680-nucleotide 3'-UTR is identical to that of a previously cloned gene, the *chromosome and centrosome factor (ccf)*. The gene identified from the clone 3.19 was

therefore renamed *ccf–β*. Ccf is a Gln-rich protein that is associated with more than 100 sites on polytene chromosomes (Kodjabachian et al., 1998). *ccf* was reported to have essential functions in cell proliferation and mitotic chromosome condensation. A possible explanation for the fact that *ccf–β* shares the same 3'-UTR with *ccf* is that they are alternative splicing products. To confirm its DNA binding activity *in vitro*, the *ccf–β* insert was subcloned into the pGEX vector to extract bacterially expressed GST-fusion proteins. Gel shift assays indicated that the GST-fused Ccf–β protein can bind to the D3 sub-element specifically (Figure 30). Competition experiments suggested that the GC-rich motifs, but not the CAATGT motifs, contribute to the binding (Figure 30, compare competition with d3c..t and d3g..c). DNaseI footprinting assays uncovered two protection sites in the *tin*-D-A element (Figure 31), the stronger of which overlapped with the D3 sub-element. The sequences of the protected sites are summarized in Figure 33. These *in vitro* binding results indicate that *ccf–β* is a sequence-specific DNA binding factor, in contrast with *ccf*, which is a general DNA binding factor (Kodjabachian et al., 1998).

The *ccf* gene is located in chromosome region 82E. The *Drosophila* line Df(3R) 6-7 contains a deficiency deleting the chomosome region 82D3-82F, which includes the *ccf* gene and presumably also *ccf-β*. To test whether *ccf-β* is involved in *tinman* restriction, Tinman expression levels in homozygous Df(3R) 6-7 deficient embryos were examined

by antibody staining. The phase I Tinman expression in the whole mesoderm was not affected (data not shown). However, the phase II tinman expression in the dorsal mesoderm was strongly reduced and narrowed (Figure 32, compare B with A). The pericardial cells which derive from the dorsal mesoderm and whose specification depends on *tinman* were almost completely ablated, as indicated by Even-skipped antibody staining (Figure 32, compare D&F with C&E). These results suggested that there are one or more genes located in the 82D3-82F region which are required for normal induction of *tinman* by Dpp during phase II of its expression. Based on our *in vitro* data, it is possible that this gene is *ccf-β*, but further evidence is required to confirm this hypothesis.

3. *DEAF-1*

Deformed epidermal autoregulatory factor-1 (DEAF-1) was the gene that appeared as the most frequent one during yeast one-hybrid screening. Most of the clones contained inserts in the reverse orientation. This phenomenon has been reported previously (Yu et al., 1997). The explanation could be that these clones used the yeast 3'UTR as a transcriptional promoter and generated a transcriptionally active protein without fusing with the GAL4 transcription activating domain. *DEAF-1* was originally cloned as a co-factor for the homeotic gene *deformed* (Gross and McGinnis, 1996). DEAF-1 binds to the neighboring region of the *deformed* autoregulatory binding sites in a homeotic responsive element, possibly contributing to the functional specificity of the Deformed

protein. DEAF-1 protein is ubiquitously expressed during *Drosophila* early embryogenesis, which would allow it to be involved in mesoderm patterning.

To characterize its DNA binding activity with respect to *tin*-D sequences, the *DEAF-1* cDNA was cloned into the pGEX vector to extract bacterially expressed GST-fusion protein. Gel shift assays with D3 as a probe indicated a specific binding (Figure 30). Competition experiments suggested that the GC-rich motifs, but not the CAATGT motifs, contributed to the binding (Figure 30, compare competition with d3c..t and d3g..c). DNaseI footprinting assays uncovered two protection sites, one of them overlapping with the D3 sub-element (Figure 31). The sequences of the two sites protected by DEAF-1 are summarized in Figure 33.

Tinman expression in embryos homozygous for a deficiency that deletes the DEAF-1 gene was determined by antibody staining. The phase II expression of Tinman in the dorsal mesoderm appeared to be normal. The Even-skipped stained pericardial cells were also not affected (data not shown). Because of the strong maternal expression of *DEAF-1*, these negative results do not necessarily mean that *DEAF*-1 is not involved in dorsal mesoderm induction. Further studies require the generation of embryos with both maternally and zygotically mutant *DEAF-1* genetic backgrounds.

4. *Adf-1*

Adh distal factor-1 (Adf-1) was cloned from library #2. Adf-1 is a Myb-like transcription factor that contains a TAF-binding motif (Cutler et al., 1998). The *Adf-1* gene was originally cloned by biochemical purification of proteins that could specifically bind to upstream recognition elements of the *Drosophila alcohol dehydrogenase* gene (*Adh*) (England et al., 1992). *Adf-1* binding sites have been found in several *Drosophila* promoters (England et al., 1990). Adf-1 protein is detected in the nucleus of most cells during early embryogenesis, which would allow it to have a function in mesoderm patterning (England et al., 1992).

A GST-fused version of the Adf-1 protein was also generated to perform gel shift assays with D3 as a probe. Specific binding was observed and the competition experiments suggested that the GC-rich motifs rather than the CAATGT motifs contributed to the binding (Figure 30). Tinman expression in *Adf-1* mutant embryos remains to be analyzed.

5. *schnurri*

Using a candidate gene approach, two protection sites were identified for Schnurri protein (Figure 10B). However, when one of the protection sites was deleted (construct tin-D-ΔD4), tin-D activity remained to be normal (staining not shown, see Figure 24). Even when both of the protected sites were deleted (construct tin-D*), the remaining tin-D fragment still drove strong expression in the dorsal mesoderm (Figure 23A). Moreover,

in *schnurri* mutant embryos, *tinman* expression in the dorsal mesoderm is only slightly reduced (Yin and Frasch, 1998). Therefore, it is less likely that *schnurri* plays an essential function in *dpp* signaling to restrict *tinman* tc the dorsal mesoderm.

The binding sites in the *tin*-D element protected by all factors identified in this study are summarized schematically in Figure 33.

Figure 27. *in vitro* analysis of the DNA binding specificity of GST-fused HMG-D protein.

^{32}P labeled D3 oligonucleotide was labeled as a probe. ~150ng GST fused HMG-D protein was used for each lane. Specific competitors were added in 2000 folds molecular excess to the binding reaction. D3, d3c..t and d3g..c are oligonucleotides derived from *tin*-D sub-element D3. GC-1, GC-1-c..t and GC-2 are oligonucleotides derived from the *mef2* I-D_L elements. GC-2 does not contain a CAATGT motif. The sub-element D1 contains two Tinman binding sites. CAATGT motifs are mutated in either d3c..t or GC-1c..t oligonucleotides. Although d3c..t disrupted the ability of the sub-element D3 to compete with the HMG-D binding, the oligonucleotide GC-1c..t competes equally well as GC-1. D1 could not compete for binding, supporting the notion that certain DNA sequences are preferred to be recognized by HMG-D.

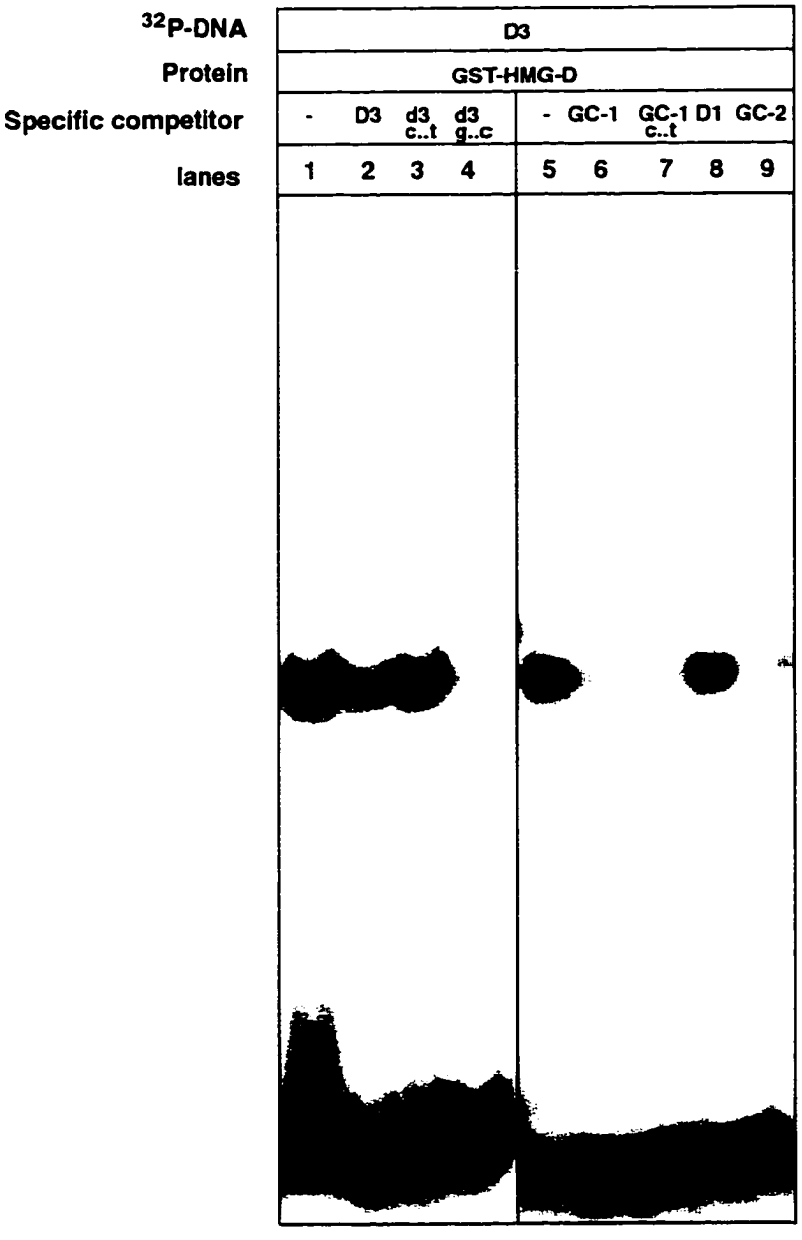


Figure 27

Figure 28. Gel shift analysis of the interactions between GST-fused HMG-D and Tinman proteins.

(A) Oligonucleotide D3 which contains two Medea/Mad binding sites but no Tinman binding sites was labeled as a probe. ~300ng GST-fused HMG-D proteins were added in each lane. ~500ng GST-fused Tinman full length or derivatives were then added individually into each of the binding reactions. (FL) full-length Tinman. (CT) Tinman-CT (aa. 301-416). (NT) Tinman-NT (aa. 1-124). (ΔTN) Tinman derivatives with the deleted TN domain (deletes aa. 29-42). (B) Oligonucleotide D1 which contains two Tinman binding sites was labeled as a probe. Free probes are not shown. Lanes 6/13, 7/14 and 8/15: 150ng, 75ng and 20ng GST-fused full-length Tinman/Tinman-CT proteins were added. Lanes 9/16, 10/17 and 11/18: 50ng, 100ng and 200ng GST-fused HMG-D proteins were added, together with 20ng GST-Tinman/Tinman-CT. Lane 12/19: 200ng GST-fused HMG-D proteins were added in the absence of GST-fused Tinman/Tinman-CT. Arrowhead indicates the position of the shifted band generated by HMG-D. (B 1) Position of the super-shifted band generated by Tinman-FL and HMG-D. (B2) Position of the super-shifted band generated by Tinman- ΔTN and HMG-D. (B3 and B4) Position of the shifted bands generated by Tinman. B4 may represent two Tinman molecules bound per molecule of the probe. (B5) Position of the additional shifted band generated by GST-fused Tinman and HMG-D. (B6 and B7) Position of the shifted bands generated by Tinman-CT. B6 may represent two Tinman-CT molecules bound per molecule of the DNA probe.

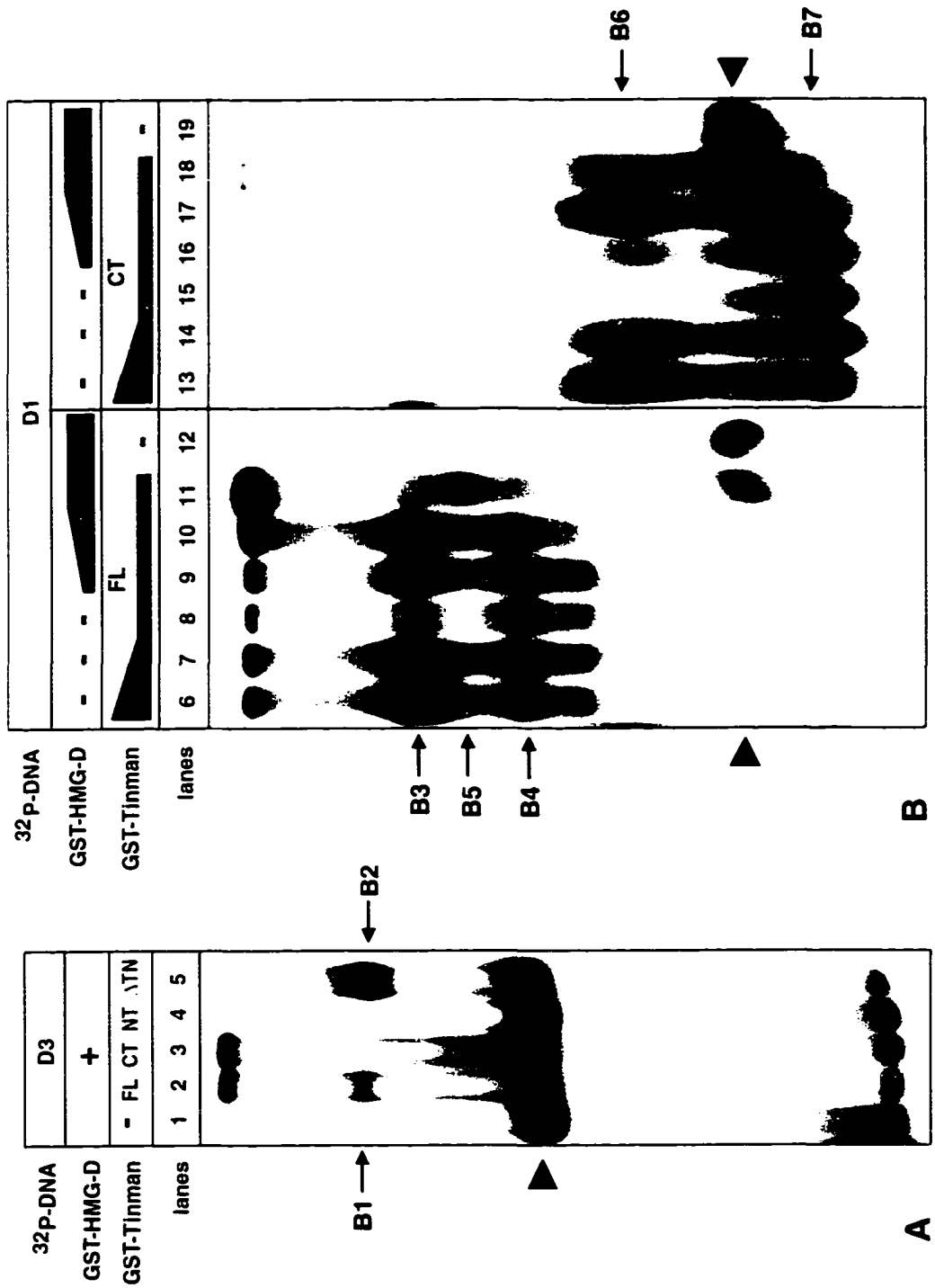


Figure 28

Figure 29. Partial cDNA sequences of the *Drosophila ccf-β* gene.

Listed below is the deduced amino acid sequence which encodes a glutamine-rich protein. The DNA sequence starts from the BamHI site that was used for fusion with the GAL4 activation domain. The transcriptional and translational start sites have not been determined. Shaded sequences are shared with the 3'-UTR of the *ccf* gene.

```
1       GGATCCGAAG GGGTTCGGTC ACAGTGACGG CGCACCACAG CCAGCAGCAG CAGCAACTGC
           I  R  R   G  S  V   T  V  T   A  H  H  S   Q  Q  Q   Q  Q  L>
71      AGCAGCAGCA ACACCACCAG CAGCAGCAGC AACACAGCCA CCAACAGCAG CAGCAACATC
        Q  Q  Q  Q   H  H  Q   Q  Q  Q   Q  H  S  H   Q  Q  Q   Q  Q  H>
121     TCCTGTCCAG TGTGACGATC ACGCCCAATT TCCATCCGGC GCAGCAGCAG CACCACCACC
        L  L  S  S   V  T  I   T  P  N   F  H  P  A   Q  Q  Q   H  H  H>
181     AACCGATGCG CGGCCACCAG CAGCAGCATC CACAGACAAC TGCCGGCAAC ATGGTTGCCC
        Q  P  M  R   G  H  Q   Q  Q  H   P  Q  T  T   A  G  N   M  V  A>
251     AAAACAACAG TAACAACCAC AGCAACGGAA GCAATCCGCT GCAGCAGCAA CAGCACATGG
        Q  N  N  S   N  N  H   S  N  G   S  N  P  L   Q  Q  Q   Q  H  M>
311     CGCAACAGGT AGCCGTCAAG CATACGCCCC ATTCGCCCGG CAAGAAGACG CGTGGCGAGA
        A  Q  Q  V   A  V  K   H  T  P   H  S  P  G   K  K  T   R  G  E>
361     ACAAAAATGC CGCAAGGTGT ACGGCATGGA GAAACGCGAC CAGTGGTGCA CCCAGTGCCG
        N  K  N  A   A  R  C   T  A  W   R  N  A  T   S  G  A   P  S  A>
421     GTGGAAGAAG GCGTGCAGTC GCTTCGGCGA CTGAAAGCGA CAGAAGTCGT CGACGGAGGC
        G  G  R  R   R  A  V   A  S  A   T  E  S  D   R  S  R   R  R  R>
481     GGCAAACCAG GCGTCGGCGC ACCAGTCCAA GGCAGCAACA GCAGCATCAT CAGCAGCATT
        R  Q  T  R   R  R  R   T  S  P   R  Q  Q  Q   Q  H  H   Q  Q  H>
541     ATCATCATCA GCATCAGCAA CAGCAACACC CAACCCAGAA GCAATATCAG CAACTCCCAG
        Y  H  H  Q   H  Q  Q   Q  Q  H   P  T  Q  K   Q  Y  Q   Q  L  P>
601     CCCGGCAGCA ACATCCTCAA CGTCCTCTAA AATGACCACC ACAGTGGTGC GTTTGGCGGC
        A  R  Q  Q   H  P  Q   R  P  L   K  *
661     CGAGGTGGCT GCCGGCAAGC AGCTACCCTC GTCGTTGACC ATCAAAACCA ACAAGCGCTA
721     CGGTGGCCCG GT
781
841
901
961
1021
1081
1141
1201
1261
1321
1381                                    AATTTGAAGC GAGATATGCC AAAATTTAAA
1441    AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAG CGGCCGCCAT ATGGTCGACC
1501    AATTCCAGAT CTATGAAT
```

Figure 29

Figure 30. Analysis of the DNA binding activities of GST-fused Adf-1, DEAF-1, Ccf-β and xFAST-1 proteins using gel shift assays.

^{32}P labeled D3 sub-element was labeled as a probe. ~20ng GST fused Adf-1, ~400ng GST-fused DEAF-1, ~50ng GST fused ccf-β and ~100ng GST-fused Xenopus FAST-1 proteins were used. Specific competitors were added in ~1000 folds of molecular excess for Adf-1 and DEAF-1, ~300 and ~1500 folds of molecular excess for Ccf-β, ~600 and ~1500 folds of molecular excess for xFAST-1. d3c..t have mutated the CAATGT motifs in D3. d3g..c have mutated GC-rich motifs in D3. For sequences of these oligos, see Figure 18.

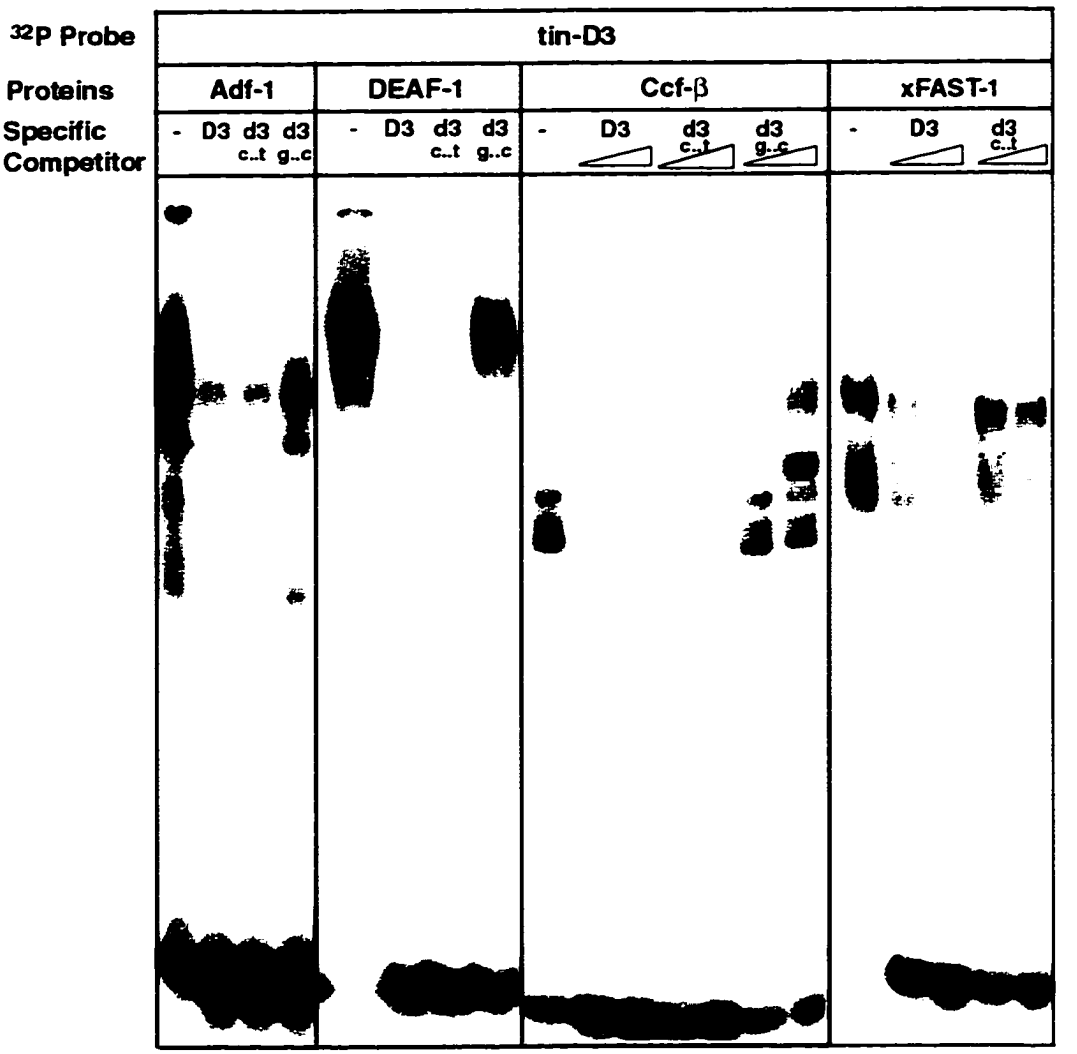


Figure 30

Figure 31. DNaseI footprinting analysis with the GST-fused DEAF-1 and Ccf-β proteins on the *tin*-D enhancer element.

^{32}P labeled *tin*-D-A was labeled as a probe. ~ 2ug GST fused proteins were used for the binding reactions. No proteins was added for the control lanes. Two DEAF-1 protection sites and two Ccf-β protection sites are detected. They are marked as bars at the right side of the figure. The position of the D3 sub-element is marked by a bracket at the left side of the figure. The sequences of the protected sites are summarized in Figure 33.

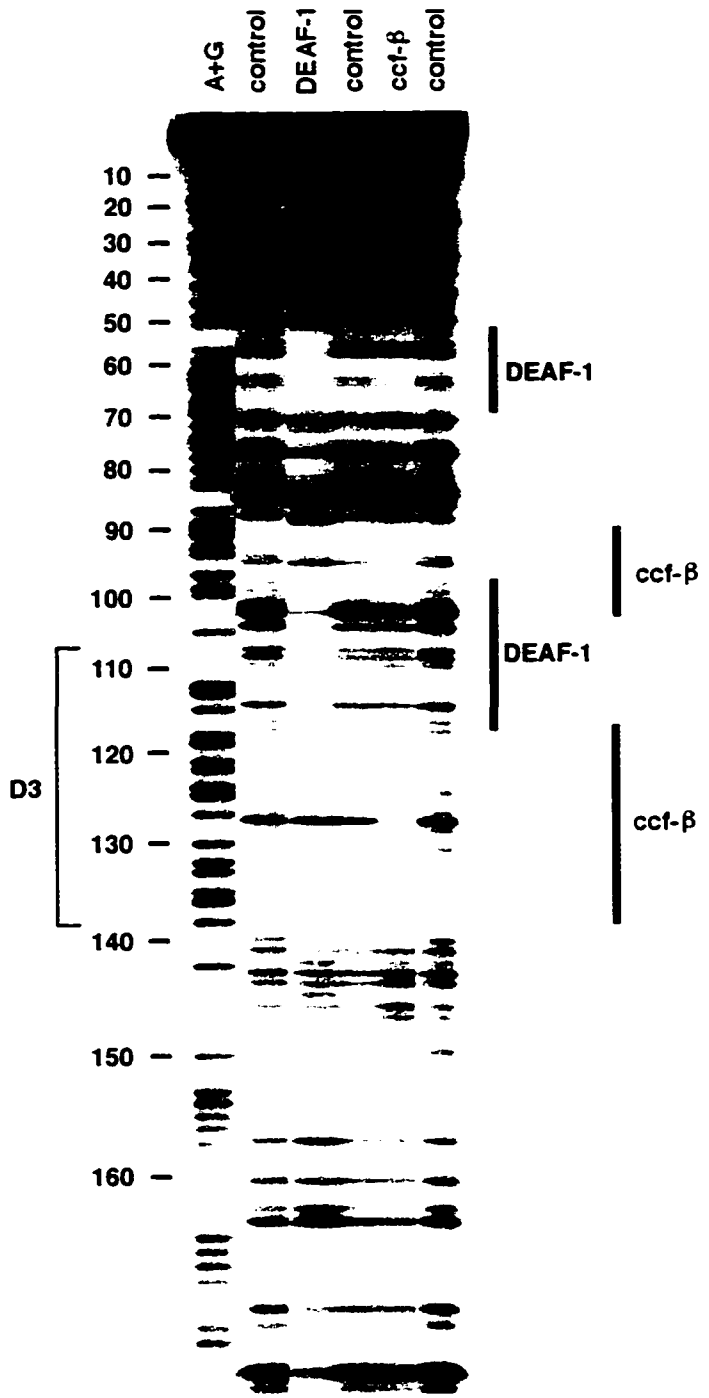


Figure 31

Figure 32. Phenotypes in embryos homozygous for the deficiency *Drosophila* line Df(3R) 6-7.

The 82D-82F region is deleted in Df(3R) 6-7, which deletes the *ccf* gene and likely *ccf-β*. Panels A,C,E are wild type controls, while panels B,D,F are homozygous deficiency embryos. (A&B) Ventral-lateral views of stage 10 embryos stained with Tinman antibody. The dorsal mesoderm expression of the Tinman is strongly reduced in the mutant embryos. Panel (A) is kindly provided by Dr. Manfred Frasch. (C&D) Lateral views of stage 11 embryos stained with Even-skipped antibody. 11 clusters of pericardial cells were stained in wild type embryos (C). In a Df(3R) 6-7 deficiency embryos (D), there is only a single cluster of pericardial cells left. A row of ectopic Even-skipped positive cells with unknown identity appear at the posterior end of the embryos. (E&F) Dorsal views of stage 13 embryos stained with Even-skipped antibody. No pericardial cells are observed in this Df(3R) 6-7 deficiency embryo (F). Arrowheads indicate the position of the dorsal mesoderm. Arrows indicate the staining in the pericardial cells.

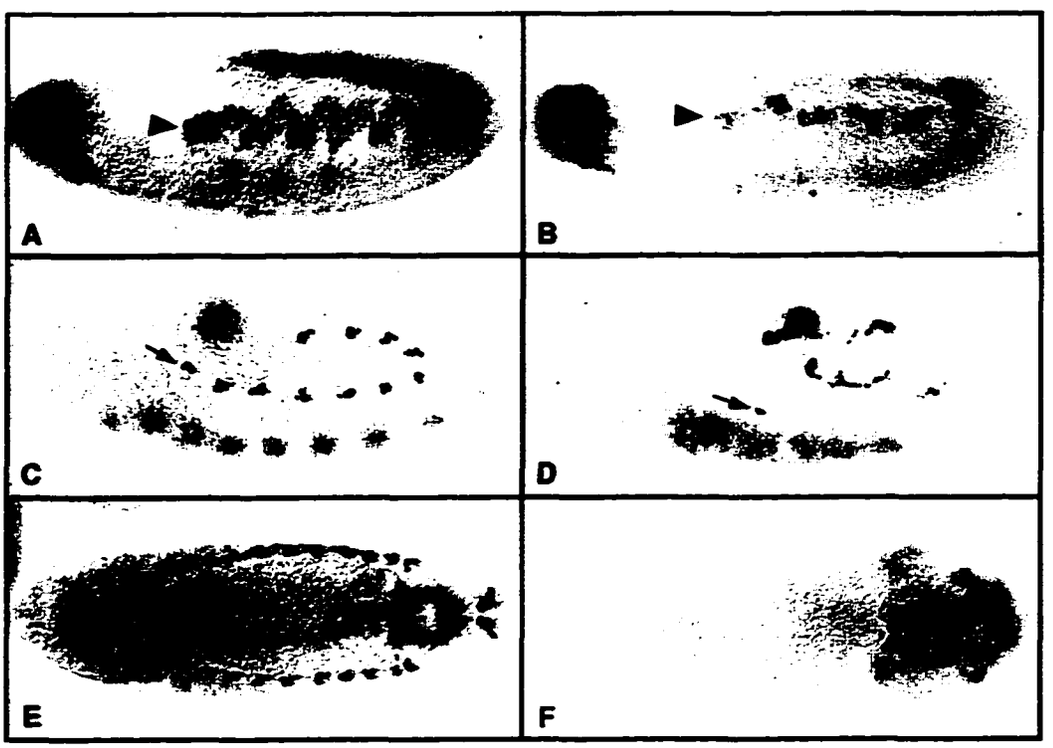


Figure 32

Figure 33. Summary of the DNA binding sites of potentially functional transcription factors in *tin*-D.

Two Tinman protection sites are indicated as black boxes. Eight Medea/Mad binding sites are indicated as oval circles. Two DEAF-1, two Ccf-β and two Schnurri protein protection sites are indicated as gray boxes.

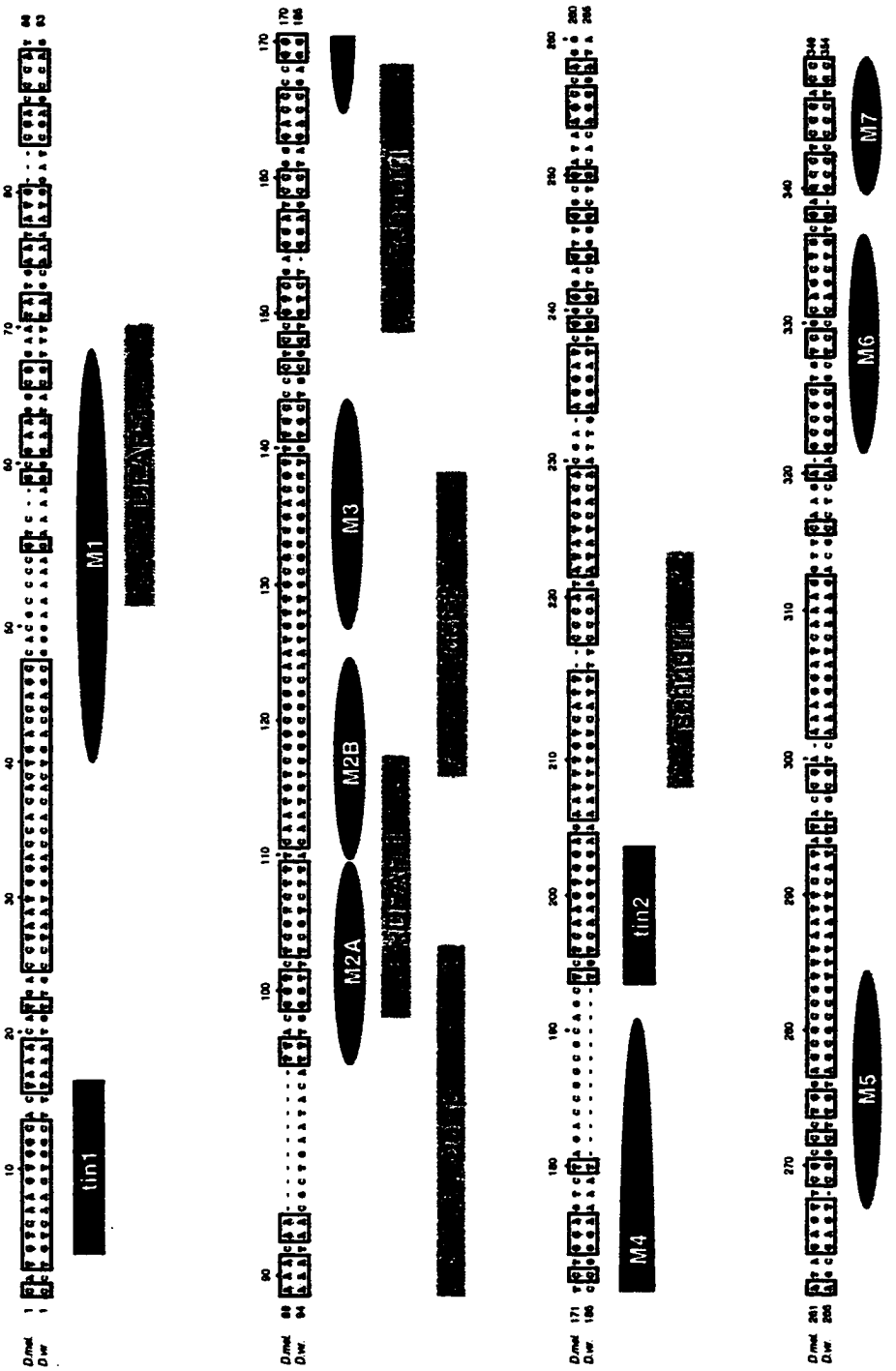


Figure 33

DISCUSSION

Molecular mechanisms of dorsal mesoderm formation in *Drosophila*

The dissection of the *tin*-D element has provided insights into the molecular mechanisms of the establishment of the dorsal-ventral polarity in *Drosophila* mesoderm. This work has demonstrated that induction through *tin*-D involves the combinatorial activities of two distinct sub-elements, D1 and D3 (and the related D6). Through binding to the sub-element D1, the phase I-expressed Tinman protein provides the mesoderm-intrinsic information from a cell-autonomous gene hierarchy. An important component of this hierarchy is *twist*, which initiates the phase I expression of *tinman* in the whole mesoderm by directly binding to the three E-boxes in the *tin*-B element (Yin et al., 1997). In contrast to D1, the sub-element D3 (together with D6) receives the inputs from the exogenous Dpp signals which restrict *tin*-D activity only to the dorsal parts of the embryo, including both the Dpp signaling tissue and the target tissue. In a cooperative way, these two inputs intersect to generate a temporal-spatial specific phase II expression of *tinman* specifically in the dorsal mesoderm. The restricted expression of Tinman in the dorsal mesoderm then contributes to the determination of specific cell fates in this region, thus establishing its dorsal-ventral polarity.

It appears that the temporal-spatial and tissue specificity results from a combination of antagonistic and synergistic mechanisms, which may act as follows (see also Figure 34): In the dorsal ectoderm, a perhaps ubiquitous repressor binds to D1 and prevents *tinman* induction in the Dpp signaling tissue. In the dorsal mesoderm, the Tinman protein generated during phase I expression competes with the repressor for the DNA binding sites and thus releases the repression. In addition, Tinman functions in synergy with the Dpp signaling to enhance the transcriptional activation of *tinman* in the dorsal mesoderm. Further support for the notion that Tinman can function in synergy with Dpp signaling pathway came from ectopic expression experiments done by Manfred Frasch (see Xu et al., 1998). When *tinman* was ectopically expressed in transverse stripes in the ectoderm using the UAS-GAL4 binary system, *tin*-D activity was induced only in the dorsal ectoderm, but not in ventral regions. This observation confirmed that *tinman* autoregulation by itself is not sufficient to activate *de novo tin*-D activity. However, in synergy with Dpp in the dorsal ectoderm, Tinman can compete with the presumed repressor and allow ectopic *tin*-D expression there. Part of the synergy may come from the observed physical interaction between Tinman and Smad proteins (S. Zaffran et al., unpublished data).

This mechanistic interpretation is consistent with the isolation of Tinman and the Smad gene product Medea as binding factors for D1 and D3, respectively. The reason that no candidate for the repressor as a D1-binding factor was identified from the screening could

be three-fold. First, the repressor could bind to D1 with much lower affinity than Tinman. Second, the library #1 used for this screening might not contain a functional clone for the repressor. This could be due to a biased amplification of the library, where only five 150-mm plates instead of one hundred fifty plates were used to culture the bacteria (see Materials and Methods). Another possibility is that all clones in the library include strong repressor domain. Third, the screening scale may not have been large enough.

Molecular mechanisms of Dpp signaling during *Drosophila* mesoderm induction

Genetic evidence has been provided in this study to show that *tkv* is the type I receptor that mediates the Dpp signal to induce *tinman* expression to the dorsal mesoderm. It would be interesting to test whether the other known type I receptor, *sax*, is also involved in this inductive event. This result would provide an additional clue to whether Scw, a second BMP4 homologue that is expressed uniformly during embryogenesis (Arora et al., 1994), functions in mesoderm induction. Recent reports have indicated that *sax* mediates the signaling of Scw instead of Dpp (Neul and Ferguson, 1998; Nguyen et al., 1998). It has been shown that *tinman* expression is strongly reduced in *scw* mutant embryos (Yin and Frasch, 1998). As to the type II receptor, the *punt* gene product is the only one identified for the Dpp ligand (Letsou et al., 1995; Ruberte et al., 1995), and thus is very likely to be involved in dorsal mesoderm induction.

The data from this study support the notion that Smads are the major signal transducers of the Dpp signals. The cloning of the *Medea* defines the common-mediator *Smad* in *Drosophila*. Genetic evidence indicated that *Medea* is involved in dorsal mesoderm induction (Xu et al., 1998). For pathway-restricted Smads, Mad is the only known Smad protein that contains a Dpp/BMP receptor-recognizing MH2 domain in *Drosophila*. Together with the fact that Mad is expressed in the mesoderm during stage 8-13 (Yin Z. et al., unpublished data), it is proposed that *Mad* is the pathway-restricted *Smad* which is involved in *tinman* restriction to the dorsal mesoderm. *dSmad2*, on the other hand, represents another member of the pathway-restricted *Smads*, which possibly transduces signals from an activin homologue in *Drosophila*. A candidate may be a recently identified activin-β (Kutty et al., 1998).

The mechanisms of the signal transduction pathway for Dpp is conserved with that of members of the vertebrate TGF-β superfamily. It has been shown that Mad protein is a monomer in the cytoplasm before receiving signals from Dpp (Inoue et al., 1998). Upon Dpp stimulation, Mad is phosphorylated and translocated into the nucleus (Newfeld et al., 1997; Maduzia and Padgett, 1997). Medea can physically interact with Mad to form the heteromeric Mad/Medea complexes (Wisotzkey et al., 1998).

Our data have suggested that Medea/Mad complexes mediate Dpp signaling by directly binding to the *tin*-D element through their MH1 domains. Four out of the eight protected

sites are essential for *tin*-D activity (M2B, M3, M6 and M7), while the binding sites M4 and M5 seem dispensable (Figure 24; M1 and M2A were not tested). An alternative explanation of the data is that the fully functional response of *tin*-D requires a minimal number of Medea/Mad binding sites. As Smad proteins tend to form oligomers through their MH2 domains upon stimulation, it is proposed that each of the released MH1 domain could bind to one of the protection sites in the *tin*-D element. A Smad/DNA complex is thus formed which probably triggers a structural reorganization of the chromatin to allow optimal transcriptional regulation (see Figure 34).

In addition to the Smad binding sites, CAATGT motifs are essential for the Dpp signaling. However, the CAATGT binding factor has not been identified despite extensive yeast one hybrid screens during this study. A reason could be that the DNA binding capacity of this factor requires ligand stimulation or protein-protein interactions. Modified yeast one-hybrid screens could be designed. One possibility could be to include Medea and/or Mad in the reporter strain to mimic Dpp signaling.

Several other DNA-binding factors have been identified in this study, which could possibly be involved in the Dpp signaling pathway. *DEAF-1*, *Adf-1*, *ccf–β* and *HMG-D* were identified through yeast one hybrid screening using D3 as a bait. The first three of these factors bind to specific sequences within *tin*-D, while HMG-D protein binds with lower sequence preference. However, HMG-D can interact with the homeodomain

protein Tinman and affect its DNA binding activity. It is thus proposed that HMG-D can modulate gene regulation by providing an optimal environment for transcription factors. *ccf–β* is a novel gene which shares the 3'-UTR with a previously published gene, *ccf* (Kodjabachian et al., 1998). A deletion of chromosomal regions around the *ccf/ccf–β* locus results in a strong reduction of dorsal *tinman* expression and a severe reduction of the pericardial cells which are derived from the dorsal mesoderm, which could be consistent with a function of *ccf–β* in Dpp-mediated *tinman* induction. Further experiments are needed to determine whether this effect is due to a deletion of the *ccf–β* gene. Whether DEAF-1 and Adf-1 are involved in the dorsal mesoderm formation or Dpp signaling remains to be investigated.

Synergistic mechanisms in Dpp signaling: A comparison between dorsal mesoderm induction and visceral mesoderm/endoderm induction in *Drosophila*

Another well studied Dpp-induced event is the subdivision of the *Drosophila* endoderm into regionally distinct sections by the adhering visceral mesoderm during later stages of embryogenesis (Bienz, 1997). Here, an inductive cascade is initiated by the homeotic gene *Ultrabithorax (Ubx)* which is expressed in a single parasegment in the visceral mesoderm. *Ubx* activates another homeotic gene, *labial*, in the neighboring endoderm cells through two extracellular signals: Dpp and Wingless, a Wnt homologue (Hoppler and Bienz, 1995). Through an indirect autoregulation loop, Dpp and Wingless also signal

Figure 34. Model of the tissue specific induction mechanism by which Dpp restricts *tinman* expression to the dorsal mesoderm.

tin-D element was simplified as a curved bar with Tinman binding sites as boxes, Medea/Mad binding sites as oval circles and CAATGT motifs as triangles. Dpp is secreted from the dorsal ectoderm and function toward both dorsal mesoderm and the signaling tissue itself. Medea/Mad mediate the Dpp signal by directly binding to multiple sites in *tin*-D element through their MH1 domains (represented by M). Through their MH2 domains, Smad proteins form hetero-oligomer and thus could change the chromosome structure of the enhancer region. There is an unknown transcription factor (represented by ?) recognizing the CAATGT sites which is also required for the *Dpp* signaling. In dorsal ectoderm, a repressor (represented by R) occupies the Tinman binding sites and prevents *tinman* ectopic expression in the Dpp signaling tissue. In dorsal mesoderm, the phase I-expressed Tinman protein competes with the repressor and releases the repression. Tinman could function in synergy with the Dpp signal, probably through physical interactions with Medea/Mad, to guarantee tissue-specific Dpp induction.

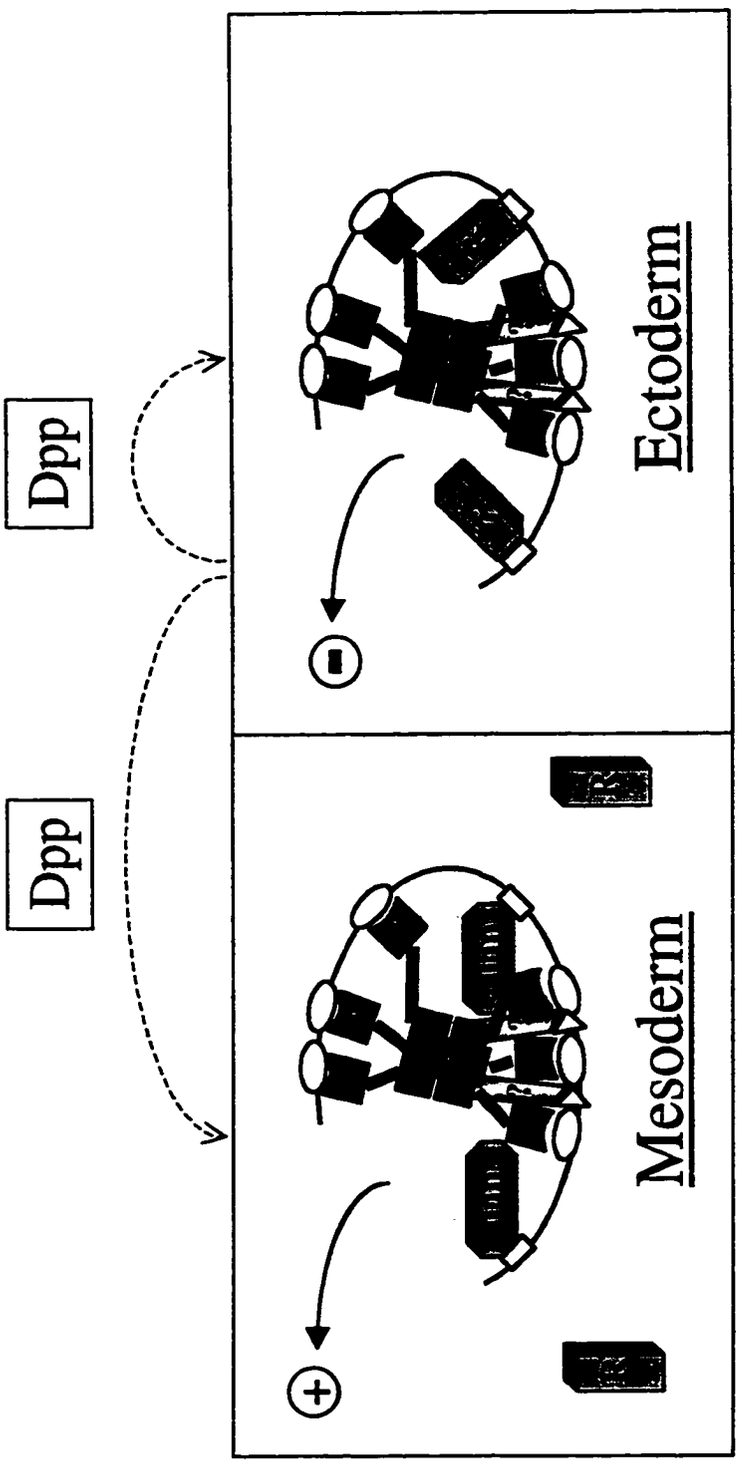


Figure 34

within the visceral mesoderm to maintain the expression of *Ubx* (Thuringer and Bienz, 1993).

Dpp responsive elements have been identified from *Ubx* and *labial* promoter regions. The 270bp B enhancer from *Ubx* mediates the *wg*- and *dpp*-dependent transcription in the visceral mesoderm (Thuringer et al., 1993). This enhancer contains a *wg* response sequence (WRS) and a *dpp* response sequence (DRS), which are located next to each other. The two signal response elements together, but not on their own, are sufficient to drive reporter gene expression in the visceral mesoderm. The DRS contains a CRE (cAMP response element)-like motif and two adjacent GC-rich Mad binding sites. Both of these motifs are required for full-level induction by Dpp (Kim et al., 1997; Szuts et al., 1998). The 546bp lab550 element from the upstream region of the *labial* gene is activated in endoderm cells in response to the Dpp signal (Tremml and Bienz, 1992). Similar to the regulation in the *tin*-D element, autoregulation is involved in *labial* induction through this enhancer. Labial binds cooperatively with another homeodomain protein Extradenticle (EXD) to a 45 bp essential element (nucleotide 48-92 in the 546bp element). Similar to the D1+D3 combination of sub-elements in *tin*-D, this 45 bp element could essentially reproduce lab550 activity, although its response to the Dpp signal is not completely accurate (Grieder et al., 1997). A faithful response to the Dpp signal requires the co-existence of both this essential element and a second Dpp-responsive element (nucleotide 92-546) (Grieder et al., 1997).

Different transcription factors are used to interpret Dpp signal during dorsal mesoderm induction and endoderm induction. For example, Schnurri is required for the Dpp signaling in endoderm induction (Arora et al., 1995; Grieder et al., 1995; Staehling-Hampton et al., 1995). In contrast, data from this study and from Yin and Frasch, 1998 indicated that Schnurri does not play an essential role in dorsal mesoderm induction. The CAATGT motif, which is essential for dorsal *tinman* induction, was not detected in endodermal Dpp-responsive elements. Therefore, CAATGT might be targeted by a stage-specific factor that is only involved in the Dpp inductive events during dorsal mesoderm induction by Dpp, which includes the induction of *tinman* and *mef2*.

Similar to the finding from this study, Mad also mediates the Dpp signal to the DRS in the B enhancer from the *Ubx* promoter (Kim et al., 1997; Szuts et al., 1998). Although the lab550 element contains several Mad binding sites, it remains to be investigated whether Smad proteins mediate Dpp signaling to the lab550 element *in vivo*. In addition to Smad proteins as signal transducers, several novel transcription factors were determined to be Dpp downstream factors during endoderm induction. The CRE binding proteins (CREB) have been proposed as candidates for such factors. The CRE-like sequence represents a Dpp-responsive target (Eresh et al., 1997), which is activated indirectly by a secondary signal, *vein* (Szuts et al., 1998). Vein is a secreted molecule that is activated by Dpp and may function through the EGF receptor (EGFR) to stimulate the CRE binding proteins (CREB). In addition, DRS receives direct Dpp signal through

Mad-binding sites and both pathways are thought to intertwine to guarantee a sharp and accurate response to Dpp (Szuts et al., 1998). Interestingly, *D-fos*, whose vertebrate homologues can form heterodimers with CREB proteins, may also be involved in *dpp* signaling during the endoderm induction (Riese et al., 1997). It appears to function as an "immediate-early gene" that is in turn inducing the "late genes" such as *labial* (Bienz, 1997). In the lab550 element, EXD protein might mediate part of the Dpp signaling to the 45 bp essential element, since it has been shown that the translocation of EXD from the cytoplasm to the nucleus in the endoderm is controlled by Dpp and Wingless (Mann and Abu-Shaar, 1996).

During dorsal mesoderm induction, a combination of Smad proteins and Tinman autoregulatory proteins is required for tissue specific activation of the Dpp-responsive element *tin*-D. Synergistic mechanisms are also found during inductive events in the visceral mesoderm/endoderm. In the B enhancer from the *Ubx* promoter, the cooperation between Lymphocyte enhancer binding factor 1 (LEF-1), which mediates the Wingless signal by binding to WRS, is required together with Dpp signaling to DRS for transcriptional stimulation in the midgut (Riese et al., 1997). Within DRS, the above mentioned secondary EGFR pathway synergistically interacts with the primary Dpp signaling pathway (Szuts et al., 1998). Likewise, in the lab550 element the synergy between the essential 45 bp element and the second Dpp-responsive element (nucleotide 92-546) guarantees full and correct Dpp responsiveness (Grieder et al., 1997).

Based on the results from this study, an antagonistic mechanism is proposed to explain the tissue specificity of the Dpp signaling during dorsal mesoderm induction. Similarly, it would be interesting to ask why Dpp and Wingless induce the B enhancer of *Ubx* specifically in the visceral mesoderm and the lab550 enhancer only in the neighboring endoderm, but not vice versa. Like the $(D3)_5$ construct in *tin*-D element, four copies of CRE in the *Ubx* B enhancer drives reporter gene ectopically in the endoderm, but not in the visceral mesoderm (Eresh et al.,1997). It is thus possible that, similar to the situation in *tin*-D, a repressor could bind to the *Ubx* B enhancer to prevent ectopic expression and restrict Dpp induction of *Ubx* to the visceral mesoderm.

Smad as DNA binding factors for TGF-β signal transduction

It has been suspected that Smad proteins could function as transcription factors since the identification of the transcription activator activity in MH2 domains of Smad2/4 (Liu et al., 1996). During the course of my studies, Kim et al. demonstrated that bacterially expressed *Drosophila* Mad fusion protein can bind to the *vestigial* promoter region, as well as *Ubx* and *labial* enhancers, through its MH1 domain (Kim et al., 1997). The binding consensus sequence derived is similar to what we have concluded from the *tin*-D element study, which is a GC-rich motif (GCCGnCGc from Kim et al., as compared to CGCCGC(G/C)G(C/A)C from our data). This novel biochemical property of Smad proteins was later confirmed with vertebrate homologues. Smad4 was determined to be

able to bind to the promoter region of the p3TP-Lux luciferase reporter, a widely used TGF–β responsive artificial promoter construct (Yingling et al., 1997). The DNA binding activity of Smad3 and Smad4 were also detected in the promoter region of another TGF–β target gene, *human plasminogen activator inhibitor-1 (PAI-1)* (Dennler et al., 1998). Based on their data, a CAGA box motif rather than a GC-rich motif was concluded to be the recognition site for Smad proteins. The *in vitro* DNA binding specificity of Smad proteins were determined later using a PCR/selection strategy over a random pool of oligonucleotides (Zawel et al., 1998). A 8bp palindromic sequence (SBE), GTCTAGAC, was identified as the optimal binding site for both Smad3 and Smad4, which is reminiscent of the CAGA motif. Crystal structure analysis indicated that a 11 residue β-hairpin structure within the MH1 domain is used to contact the SBE DNA (Shi et al., 1998). Three of the eleven residues (Arg-74, Gln-76 and Lys-81 of Smad3) form direct hydrogen bonds with three distinct nucleotides. Two out of these nucleotides are guanidines (Shi et al., 1998), which may explain why MH1 domain can recognize GC-rich motifs in addition to the SBE or CAGA box. Indeed, we have shown that Medea/Mad can bind to the latter two motifs in gel shift assays (Figure 21). Two Medea/Mad protection sites from the *tin*-D element (M2A and M4, see Figure 19) actually contain SBE motifs and CAGA boxes. Based on these observations, we propose that Smads are not strictly sequence specific DNA binding proteins. Consistent with this notion, the sequences of the two Mad/Medea binding sites (MM-1 and MM-2) identified from the I-D_L element do not match any of the above consensus sequences.

The presence of a DNA binding activity in Smad MH1 domains suggested that the MH1 domain could function as an effector domain for Smad signaling, in addition to the MH2 domain. Like STAT proteins, Smads receive the signal from the receptors, then translocate into the nucleus and directly bind to promoter regions of the target genes. Before ligand stimulation, the intra-molecular interaction between MH1 and MH2 domains seems to inhibit both the transcription activator function at the MH2 domain and the DNA binding activity at the MH1 domain. Upon ligand stimulation, hetero-oligomerization through MH2 domains may release both of the inhibitions. Based on this hypothesis, a novel mechanism is proposed here to explain why MH2 domain alone is constitutively active (Baker and Harland, 1996). It is conceivable that this MH2 domain could hetero-oligomerize with endogenous Smads and release the inhibition generated from the intra-molecular interaction. The resulting Smad complex could use the MH1 domains of the endogenous Smads to bind to the target promoter DNA, while the combined MH2 domains stimulate transcription of the target gene.

Implications on how the simple Smad pathway decodes complex TGF–β signals

The results from this study provide clues on how the seemingly simple Smad pathway interprets a variety of TGF-β signals. As has been mentioned before, the pathway-restricted Smads are responsible for part of the specificity. This notion was supported by the finding that the dSmad2 in *Drosophila* could not bind to DNA, in contrast to

Medea and Mad. The variability in this biochemical property might provide additional flexibility for Smad proteins to interpret pleiotropic functions of TGF–β molecules. This variability could represent an important functional divergence among Smad proteins during evolution, as has been indicated by the fact that Smad2 and dSmad2 have adopted two different ways to disrupt their DNA binding activity (see Results).

The existence of multiple Smad binding sites with divergent binding affinities in the promoter region of the target genes could contribute to the specificity of TGF-β signaling. We have identified multiple Mad/Medea binding sites in two *dpp*-responsive elements: *tin*-D from *tinman* and I-D_L from *mef2*. In the *PAI-1* promoter, three CAGA boxes were identified (Dennler et al., 1998). At least four tandem copies of the Smad optimal binding element (SBE) were required to restore the responsiveness to TGF–β signals (Zawel et al., 1998). Moreover, the binding affinities of these binding sites are variable. The high-affinity binding sites are not necessarily the ones that are essential, or perhaps even functional, *in vivo*. For example, the binding site M4 in *tin*-D contains an optimal binding site and a CAGA box (see Figure 19) and the binding affinity of Medea for this protection site is much higher than for those of M6 &7 (see footprinting results in Figure 16). However, *in vivo* testing of the *tin*-D derivatives indicated that M6&7 are essential for *tin*-D activity, while M4 (together with M5) is dispensable (see Figure 24, compare constructs tin-D-ΔD6 and tin-D*). One potential advantage of the presence of multiple

Smad binding sites with variable affinities could be the capability to sense different concentrations of Smad proteins in the nucleus. This characteristic of a target promoter could contribute to the interpretation of a graded TGF-β signal, which plays important roles in various patterning events for TGF-β members as morphogens.

As has been discussed before, the synergistic interaction between Medea/Mad and Tinman determines the tissue specificity of Dpp induction during mesoderm formation. Functional synergy with other transcription factors could be one of the major mechanisms for identical Smad complexes to activate different downstream genes (Derynck et al., 1998). FAST-1 in *Xenopus* is the first co-factor identified for Smad proteins (Chen, X. et al., 1996). The cloning of the mouse homologue FAST-2 resulted in the identification of a synergistic regulation of the *goosecoid* gene, one of the earliest known activin-inducible targets in the dorsal-anterior mesoderm (Blum et al., 1992). In this promoter, Smad proteins function in synergy with FAST-2 by binding to a site which is adjacent to a FAST-2 binding site (Labbe et al., 1998). Although Smad2 itself can not bind to DNA, it recruits Smad4 into the hetero Smad complex to bind to DNA and activate the promoter. Smad3 blocks the induction possibly by competing with Smad4 for the DNA binding site (Labbe et al., 1998). These findings from the cell culture system provide an additional example on how the variability in DNA binding activities of Smad proteins contributes to diverse patterns of transcriptional responses.

Several other co-factors have recently been identified for Smad proteins. The Sp1 transcription factor can interact with Smad proteins and form a DNA-binding complex to regulate either the human *p21/WAF1/Cip1* promoter (Moustakas et al., 1998) or the human *type VII collagen* gene (*COL7A1*) promoter (Vindevoghel et al., 1998). The bHLH transcriptional factor TFE3 is required for the maximal activation of the TGF–β induced *PAI-1* promoter, which is mediated by the DNA binding of Smad3/Smad4 (Hua et al., 1998). Synergistic interaction with other transcription factors can provide endless possibilities for Smads to transduce TGF-β signals to specific target genes. The Smad protein complexes could then interact and cooperate with bridging coactivators such as CBP (CREB binding protein) and p300 to contact the basal transcription machinery (Feng et al., 1998; Janknecht et al., 1998).

Conserved molecular mechanisms in mesoderm induction

Mesoderm induction has been studied extensively in *Xenopus laevis* (Sive 1993; Kessler 1994). Frog embryos are large and easy to dissect into pieces that can be assayed for their inductive potential or their responses to inductive signals. In contrast to *Drosophila* dorsal mesoderm induction, the formation of the *Xenopus* mesoderm is induced by the neighboring endoderm, and the dorsal-ventral polarity is reversed (for review, see Ferguson, 1996). However, many of the underlying molecular mechanisms are highly conserved. In *Drosophila*, Dpp, or more efficiently Scw, is antagonized by Short

gastrulation (Sog) (Neul and Ferguson, 1998; Nguyen et al., 1998). Dpp (and Scw) activity is thus restricted to the dorsal ectoderm (Francois et al., 1994) and induces the formation of dorsal mesoderm in the underlying mesoderm cells. Whether Scw is directly involved in mesoderm induction still needs further investigation. In *Xenopus*, *BMP4*, a *dpp* homologue, is antagonized by *chordin*, a *sog* homologue, to promote ventral fates of the mesoderm (Jones et al., 1996). Human BMP protein can mimic Dpp and establish dorsal-ventral polarity in *Drosophila* (Padgett et al., 1993) while Dpp can promote ventral cell fates in *Xenopus* (Holley et al., 1995). Injection of *sog* mRNA promotes ventral fates in *Drosophila* and dorsal fates in *Xenopus* (Schmidt et al., 1995), whereas injection of *chordin* mRNA, which promotes dorsal fates in *Xenopus* (Sasai et al., 1994), promotes ventral fates in *Drosophila* (Holley et al., 1995). Therefore, the dorsal-ventral patterning within the mesoderm is specified by a system of genes that has been conserved over 500 million years between arthropods and chordates. The downstream signaling pathway is also conserved. *Drosophila* Mad can function in BMP-4 signaling pathway and induces ventral mesoderm formation in *Xenopus* embryos (Newfeld et al., 1996), while human Smad4 protein can rescue the dorsal-ventral patterning defects in *Medea* mutant embryos (Hudson et al., 1998).

The function of *tinman* in heart formation is also conserved between *Drosophila* and vertebrates. The *Drosophila* heart is a simple tubular structure which is located at the dorsal midline (Bodmer and Frasch, 1998). As has been mentioned before, the heart is

completely ablated in *tinman* mutant embryos, hence the name *tinman*. *tinman* is represented by a family of homologous genes in vertebrates (Evans et al., 1995; Harvey, 1996). Two of the vertebrate representatives are *Nkx 2.5* and *Nkx 2.3*, both of which show early and specific expression during cardiac specification (Tonissen et al., 1994; Buchberger et al., 1996; Chen, J. et al., 1996; Shiojima et al., 1996). Targeted disruption in the murine *Nkx2.5* gene resulted in abnormal heart morphogenesis (Lyons et al., 1995). Three mutations in human *Nkx2.5* gene were recently found to cause human congenital heart disease (Schott et al., 1998). Similarly in the frog embryo, dominant negative versions of Nkx 2.5 and Nkx 2.3 ablated heart formation (Fu et al., 1998; Grow et al., 1998). In addition to the conserved function of the *tinman* gene, the regulation appears also to be conserved. Heart formation in vertebrates is also an inductive event that is comparable to dorsal mesoderm induction in *Drosophila*. In chicken, the cell fates of the heart and Nkx 2.5 expression in the cardiogenic mesoderm were demonstrated to be induced by the adjacent anterior endoderm (Schultheiss et al., 1995). *BMP-2*, a *dpp* homologue, turned out to be one of major inducers in this event (Lough et al., 1996; Schultheiss et al., 1997; Andree et al. 1998).

In summary, the molecular mechanisms we uncovered here in *Drosophila* dorsal mesoderm induction may provide clues to *Xenopus* mesoderm induction as well as heart formation in vertebrates and other TGF−β induced events. Thus *Drosophila* has served

as an excellent model system to study this type of inductive event and the signaling pathways involved in it.

REFERENCES

Abdollah, S., Macias, S.M., Tsukazaki, T., Hayashi, H., Attisano, L., and Wrana, J.L. (1997). TGFbetaRI phosphorylation of Smad2 on Ser465 and Ser467 is required for Smad2-Smad4 complex formation and signaling. J Biol Chem *272*, 27678-27685.

Andree, B., Duprez, D., Vorbusch, B., Arnold, H.H., and Brand, T. (1998). BMP-2 induces ectopic expression of cardiac lineage markers and interferes with somite formation in chicken embryos. Mech.Dev. *70*, 119-131.

Arlotta, P., Rustighi, A., Mantovani, F., Manfioletti, G., Giancotti, V., Tell, G., and Damante, G. (1997). High mobility group I proteins interfere with the homeodomains binding to DNA. J Biol Chem *272*, 29904-29910.

Arora, K., Levine, M.S., and O'Connor, M.B. (1994). The *screw* gene encodes a ubiquitously expressed member of the TGF-beta family required for specification of dorsal cell fates in the *Drosophila* embryo. Genes Dev. *8*, 2588-2601.

Arora, K., Dai, H., Kazuko, S.G., Jamal, J., O'Connor, M.B., Letsou, A., and Warrior, R. (1995). The *Drosophila schmurri* gene acts in the Dpp/TGF beta signaling pathway and encodes a transcription factor homologous to the human MBP family. Cell *81*, 781-790.

Attisano, L., Wrana, J.L., Lopez, C.F., and Massague, J. (1994). TGF-beta receptors and actions. Biochim.Biophys.Acta *1222*, 71-80.

Attisano, L. and Wrana, J.L. (1998). Mads and Smads in TGF beta signalling. Curr.Opin.Cell Biol *10*, 188-194.

Azpiazu, N. and Frasch, M. (1993). *tinman* and *bagpipe*: two homeo box genes that determine cell fates in the dorsal mesoderm of *Drosophila*. Genes Dev. *7*, 1325-1340.

Azpiazu, N., Lawrence, P.A., Vincent, J.P., and Frasch, M. (1996). Segmentation and specification of the *Drosophila* mesoderm. Genes Dev. *10*, 3183-3194.

Baehrecke, E.H. (1997). *who* encodes a KH RNA binding protein that functions in muscle development. Development *124*, 1323-1332.

Baker, J.C. and Harland, R.M. (1996). A novel mesoderm inducer, Madr2, functions in the activin signal transduction pathway. Genes Dev. *10*, 1880-1889.

Baker, J.C. and Harland, R.M. (1997). From receptor to nucleus: the Smad pathway. Curr.Opin.Genet.Dev. *7*, 467-473.

Baylies, M.K., Martinez, A.A., and Bate, M. (1995). *wingless* is required for the formation of a subset of muscle founder cells during *Drosophila* embryogenesis. Development *121*, 3829-3837.

Beer, J., Technau, G.M., and Campos-Ortega, J.A. (1987). Lineage analysis of transplanted individual cells in embryos of *Drosophila melanogaster*. Commitment and proliferative capabilities of mesodermal cells. Roux's Arch.Dev.Biol. *196*, 220-230.

Bienz, M. (1994). Homeotic genes and positional signalling in the *Drosophila* viscera. Trends.Genet. *10*, 22-26.

Bienz, M. (1997). Endoderm induction in *Drosophila*: the nuclear targets of the inducing signals. Curr.Opin.Genet.Dev. *7*, 683-688.

Blum, M., Gaunt, S.J., Cho, K.W., Steinbeisser, H., Blumberg, B., Bittner, D., and De-Robertis, E.M. (1992). Gastrulation in the mouse: the role of the homeobox gene *goosecoid*. Cell *69*, 1097-1106.

Bodmer, R., Jan, L.Y., and Jan, Y.N. (1990). A new homeobox-containing gene, *msh-2*, is transiently expressed early during mesoderm formation of *Drosophila*. Development *110*, 661-669.

Bodmer, R. (1993). The gene *tinman* is required for specification of the heart and visceral muscles in *Drosophila*. Development *118*, 719-729.

Bodmer, R. and Frasch, M. (1999). Genetic Determination of *Drosophila* Heart Development. In Heart Development. R.P. Harvey and N. Rosenthal, eds. (Orlando, FL: Academic Press), pp. 65-89.

Boulay, J.L., Dennefeld, C., and Alberga, A. (1987). The *Drosophila* developmental gene *snail* encodes a protein with nucleic acid binding fingers. Nature *330*, 395-398.

Bour, B.A., O'Brien, M.A., Lockwood, W.L., Goldstein, E.S., Bodmer, R., Taghert, P.H., Abmayr, S.M., and Nguyen, H.T. (1995). *Drosophila MEF2*, a transcription factor that is essential for myogenesis. Genes Dev. *9*, 730-741.

Brown, N.H. and Kafatos, F.C. (1988). Functional cDNA libraries from *Drosophila* embryos. J Mol Biol *203*, 425-437.

Brummel, T.J., Twombly, V., Marques, G., Wrana, J.L., Newfeld, S.J., Attisano, L., Massague, J., O'Connor, M.B., and Gelbart, W.M. (1994). Characterization and relationship of Dpp receptors encoded by the *saxophone* and *thick veins* genes in *Drosophila*. Cell *78*, 251-261.

Buchberger, A., Pabst, O., Brand, T., Seidl, K., and Arnold, H.H. (1996). Chick *NKx-2.3* represents a novel family member of vertebrate homologues to the *Drosophila* homeobox gene *tinman*: differential expression of *cNKx-2.3* and *cNKx-2.5* during heart and gut development. Mech.Dev. *56*, 151-163.

Cai, H. and Levine, M. (1995). Modulation of enhancer-promoter interactions by insulators in the *Drosophila* embryo. Nature *376*, 533-536.

Carcamo, J., Weis, F.M., Ventura, F., Wieser, R., Wrana, J.L., Attisano, L., and Massague, J. (1994). Type I receptors specify growth-inhibitory and transcriptional responses to transforming growth factor beta and activin. Mol Cell Biol *14*, 3810-3821.

Casal, J. and Leptin, M. (1996). Identification of novel genes in *Drosophila* reveals the complex regulation of early gene activity in the mesoderm. Proc.Natl.Acad.Sci U.S.A. *93*, 10327-10332.

Chanut, F. and Heberlein, U. (1997). Role of decapentaplegic in initiation and progression of the morphogenetic furrow in the developing *Drosophila* retina. Development *124*, 559-567.

Chen, C.Y. and Schwartz, R.J. (1995). Identification of novel DNA binding targets and regulatory domains of a murine tinman homeodomain factor, *nkx-2.5*. J Biol Chem *270*, 15628-15633.

Chen, J.N. and Fishman, M.C. (1996). Zebrafish *tinman* homolog demarcates the heart field and initiates myocardial differentiation. Development *122*, 3809-3816.

Chen, X., Rubock, M.J., and Whitman, M. (1996). A transcriptional partner for Mad proteins in TGF-β signaling. Nature *383*, 691-696.

Chen, X., Weisberg, E., Fridmacher, V., Watanabe, M., Naco, G., and Whitman, M. (1997). Smad4 and FAST-1 in the assembly of activin-response factor. Nature *389*, 85-89.

Chen, Y., Lebrun, J.J., and Vale, W. (1996). Regulation of transforming growth factor beta- and activin-induced transcription by mammalian Mad proteins. Proc.Natl.Acad.Sci U.S.A. *93*, 12992-12997.

Chen, Y.G., Hata, A., Lo, R.S., Wotton, D., Shi, Y., Pavletich, N., and Massague, J. (1998). Determinants of specificity in TGF-beta signal transduction. Genes Dev. *12*, 2144-2152.

Churchill, M.E., Jones, D.N., Glaser, T., Hefner, H., Searles, M.A., and Travers, A.A. (1995). HMG-D is an architecture-specific protein that preferentially binds to DNA containing the dinucleotide TG. EMBO J *14*, 1264-1275.

Cutler, G., Perry, K.M., and Tjian, R. (1998). Adf-1 is a nonmodular transcription factor that contains a TAF-binding Myb-like motif. Mol Cell Biol *18*, 2252-2261.

Das, P., Maduzia, L.L., Wang, H., Finelli, A.L., Cho, S.H., Smith, M.M., and Padgett, R.W. (1998). The *Drosophila* gene *Medea* demonstrates the requirement for different classes of *Smads* in *dpp* signaling. Development *125*, 1519-1528.

Dennler, S., Itoh, S., Vivien, D., ten-Dijke, P., Huet, S., and Gauthier, J.M. (1998). Direct binding of Smad3 and Smad4 to critical TGF beta-inducible elements in the promoter of human *plasminogen activator inhibitor-type 1* gene. EMBO J *17*, 3091-3100.

Derynck, R., Gelbart, W.M., Harland, R.M., Heldin, C.H., Kern, S.E., Massague, J., Melton, D.A., Mlodzik, M., Padgett, R.W., Roberts, A.B., Smith, J., Thomsen, G.H., Vogelstein, B., and Wang, X.F. (1996). Nomenclature: vertebrate mediators of TGFbeta family signals. Cell *87*, 173

Derynck, R. and Feng, X.H. (1997). TGF-beta receptor signaling. Biochim.Biophys.Acta *1333*, F105-F150

Derynck, R., Zhang, Y., and Feng, X.H. (1998). Smads: transcriptional activators of TGF-β responses. Cell *95*, 737-740.

Dohrmann, C., Azpiazu, N., and Frasch, M. (1990). A new *Drosophila* homeo box gene is expressed in mesodermal precursor cells of distinct muscles during embryogenesis. Genes Dev. *4*, 2098-2111.

England, B.P., Heberlein, U., and Tjian, R. (1990). Purified *Drosophila* transcription factor, Adh distal factor-1 (Adf-1), binds to sites in several *Drosophila* promoters and activates transcription. J Biol Chem *265*, 5086-5094.

England, B.P., Admon, A., and Tjian, R. (1992). Cloning of *Drosophila* transcription factor Adf-1 reveals homology to Myb oncoproteins. Proc.Natl.Acad.Sci U.S.A. *89*, 683-687.

Eresh, S., Riese, J., Jackson, D.B., Bohmann, D., and Bienz, M. (1997). A CREB-binding site as a target for *decapentaplegic* signalling during *Drosophila* endoderm induction. EMBO J *16*, 2014-2022.

Evans, S.M., Yan, W., Murillo, M.P., Ponce, J., and Papalopulu, N. (1995). *tinman*, a *Drosophila* homeobox gene required for heart and visceral mesoderm specification, may be represented by a family of genes in vertebrates: *XNkx-2.3*, a second vertebrate homologue of *tinman*. Development *121*, 3889-3899.

Falvo, J.V., Thanos, D., and Maniatis, T. (1995). Reversal of intrinsic DNA bends in the *IFN beta* gene enhancer by transcription factors and the architectural protein HMG I(Y). Cell *83*, 1101-1111.

Feng, X.H., Zhang, Y., Wu, R.Y., and Derynck, R. (1998). The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are coactivators for Smad3 in TGF-beta-induced transcriptional activation. Genes Dev. *12*, 2153-2163.

Ferguson, E.L. and Anderson, K.V. (1992). Decapentaplegic acts as a morphogen to organize dorsal-ventral pattern in the *Drosophila* embryo. Cell *71*, 451-461.

Ferguson, E.L. (1996). Conservation of dorsal-ventral patterning in arthropods and chordates. Curr.Opin.Genet.Dev. *6*, 424-431.

Francois, V., Solloway, M., O'Neill, J.W., Emery, J., and Bier, E. (1994). Dorsal-ventral patterning of the *Drosophila* embryo depends on a putative negative growth factor encoded by the *short gastrulation* gene. Genes Dev. *8*, 2602-2616.

Frasch, M. (1995). Induction of visceral and cardiac mesoderm by ectodermal Dpp in the early *Drosophila* embryo. Nature *374*, 464-467.

Frasch, M. and Nguyen, H.T. (1999). Genetic control of mesoderm patterning and differentiation during *Drosophila* embryogenesis. In Advances in Developmental Biochemistry, Vol.5. Wassarman P.M., ed. (Stamford, CT: JAI Press Inc.),

Fu, Y., Yan, W., Mohun, T.J., Evans, S.M. (1998). Vertebrate tinman homologues XNkx2-3 and XNkx2-5 are required for heart formation in a functionally redundant manner. Development *125*(22):4439-49

Gajewski, K., Kim, Y., Lee, Y.M., Olson, E.N., and Schulz, R.A. (1997). D-mef2 is a target for Tinman activation during *Drosophila* heart development. EMBO J *16*, 515-522.

Glise, B. and Noselli, S. (1997). Coupling of Jun amino-terminal kinase and Decapentaplegic signaling pathways in *Drosophila* morphogenesis. Genes Dev. *11*, 1738-1747.

Goto, S. and Hayashi, S. (1997). Specification of the embryonic limb primordium by graded activity of Decapentaplegic. Development *124*, 125-132.

Graff, J.M., Bansal, A., and Melton, D.A. (1996). Xenopus Mad proteins transduce distinct subsets of signals for the TGF beta superfamily. Cell *85*, 479-487.

Grau, Y., Carteret, C., and Simpson, P. (1984). Mutations and chromsomal rearrangements affecting the expression of *snail*, a gene involved in embryonic patterning in *Drosophila melanogaster*. Genetics *108*, 347-360.

Grieder, N.C., Nellen, D., Burke, R., Basler, K., and Affolter, M. (1995). Schnurri is required for *Drosophila* Dpp signaling and encodes a zinc finger protein similar to the mammalian transcription factor PRDII-BF1. Cell *81*, 791-800.

Grieder, N.C., Marty, T., Ryoo, H.D., Mann, R., and Affolter, M. (1997). Synergistic activation of a *Drosophila* enhancer by HOM/EXD and DPP signaling. EMBO J *16*, 7402-7410.

Gross, C.T. and McGinnis, W. (1996). DEAF-1, a novel protein that binds an essential region in a Deformed response element. EMBO J *15*, 1961-1970.

Grow, M.W., Krieg, P.A. (1998). Tinman function is essential for vertebrate heart development: elimination of cardiac differentiation by dominant inhibitory mutants of the *tinman*-related genes, XNkx2-3 and XNkx2-5. Dev Biol *204*(1):187-96

Hahn, S.A., Schutte, M., Hoque, A.T., Moskaluk, C.A., da-Costa, L.T., Rozenblum, E., Weinstein, C.L., Fischer, A., Yeo, C.J., Hruban, R.H., and Kern, S.E. (1996). DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1.. Science *271*, 350-353.

Harvey, R.P. (1996). NK-2 homeobox genes and heart development. Dev.Biol *178*, 203-216.

Hata, A., Lo, R.S., Wotton, D., Lagna, G., and Massague, J. (1997). Mutations increasing autoinhibition inactivate tumour suppressors Smad2 and Smad4. Nature *388*, 82-87.

Hayashi, H., Abdollah, S., Qui, Y., Cai, J., Xu, Y., Grinnel, B., Richardson, M., Topper, J., Gimbrone, M., Wrana, J., and Falb, D. (1997). The Mad-related protein Smad7 associates with the TGF-β receptor and functions as an antagonist of TGF-β signaling. Cell *89*, 1165-1173.

Heldin, C.H., Miyazono, K., and ten-Dijke, P. (1997). TGF-beta signalling from cell membrane to nucleus through Smad proteins. Nature *390*, 465-471.

Hogan, B.L., Blessing, M., Winnier, G.E., Suzuki, N., and Jones, C.M. (1994). Growth factors in development: the role of TGF-beta related polypeptide signalling molecules in embryogenesis. Dev.Suppl. 53-60.

Hogan, B.L. (1996). Bone morphogenetic proteins in development. Curr.Opin.Genet.Dev. *6*, 432-438.

Holley, S.A., Jackson, P.D., Sasai, Y., Lu, B., De-Robertis, E.M., Hoffmann, F.M., and Ferguson, E.L. (1995). A conserved system for dorsal-ventral patterning in insects and vertebrates involving *sog* and *chordin*. Nature *376*, 249-253.

Holley, S.A., Neul, J.L., Attisano, L., Wrana, J.L., Sasai, Y., O'Connor, M.B., De-Robertis, E.M., and Ferguson, E.L. (1996). The Xenopus dorsalizing factor *noggin* ventralizes *Drosophila* embryos by preventing DPP from activating its receptor. Cell *86*, 607-617.

Hoppler, S. and Bienz, M. (1995). Two different thresholds of *wingless* signalling with distinct developmental consequences in the *Drosophila* midgut. EMBO J *14*, 5016-5026.

Hou, X.S., Goldstein, E.S., and Perrimon, N. (1997). *Drosophila* Jun relays the Jun amino-terminal kinase signal transduction pathway to the Decapentaplegic signal transduction pathway in regulating epithelial cell sheet movement. Genes Dev. *11*, 1728-1737.

Hua, X., Liu, X., Ansari, D.O., and Lodish, H.F. (1998). Synergistic cooperation of TFE3 and Smad proteins in TGF-β-induced transcription of the *plasminogen activator inhibitor-1* gene. Genes Dev. *12*, 3084-3095.

Hudson, J.B., Podos, S.D., Keith, K., Simpson, S.L., and Ferguson, E.L. (1998). The *Drosophila Medea* gene is required downstream of *dpp* and encodes a functional homolog of human *Smad4*. Development *125*, 1407-1420.

Imamura, T., Takase, M., Nishihara, A., Oeda, E., Hanai, J., Kawabata, M., and Miyazono, K. (1997). Smad6 inhibits signalling by the TGF-beta superfamily. Nature *389*, 622-626.

Inoue, H., Imamura, T., Ishidou, Y., Takase, M., Udagawa, Y., Oka, Y., Tsuneizumi, K., Tabata, T., Miyazono, K., and Kawabata, M. (1998). Interplay of signal mediators of Decapentaplegic (DPP): molecular characterization of Mothers against *dpp*, Medea and Daughters against *dpp*. Molecular Biology of the Cell *9*, 2145-2156.

Jagla, K., Stanceva, I., Dretzen, G., Bellard, F., and Bellard, M. (1994). A distinct class of homeodomain proteins is encoded by two sequentially expressed *Drosophila* genes from the 93D/E cluster. Nucleic.Acids.Res. *22*, 1202-1207.

Janknecht, R., Wells, N.J., and Hunter, T. (1998). TGF-beta-stimulated cooperation of Smad proteins with the coactivators CBP/p300. Genes Dev. *12*, 2114-2119.

Jiang, J. and Struhl, G. (1996). Complementary and mutually exclusive activities of *decapentaplegic* and *wingless* organize axial patterning during *Drosophila* leg development. Cell *86*, 401-409.

Jones, C.M. and Smith, J.C. (1995). Inductive signals. Revolving vertebrates. Curr.Biol *5*, 574-576.

Jones, C.M., Dale, L., Hogan, B.L., Wright, C., and Smith, J. (1996). Bone morphogenetic protein-4 (BMP-4) acts during gastrula stages to cause ventralization of *Xenopus* embryos. Development *122*, 1545-1554.

Kawahata, M., Inoue, H., Hanyu, A., Imamura, T., and Miyazono, K. (1998). Smad proteins exist as monomers *in vivo* and undergo homo- and hetero-oligomerization upon activation by serine/threonine kinase receptors. EMBO J *17*, 4056-4065.

Kessler, D.S. and Melton, D.A. (1994). Vertebrate embryonic induction: mesodermal and neural patterning. Science *266*, 596-604.

Kim, J., Johnson, K., Chen, H.J., Carroll, S., and Laughon, A. (1997). *Drosophila* Mad binds to DNA and directly mediates activation of *vestigial* by Decapentaplegic. Nature *388*, 304-308.

Kim, Y. and Nirenberg, M. (1989). *Drosophila NK*-homeobox genes. Proc.Natl.Acad.Sci U.S.A. *86*, 7716-7720.

Kodjabachian, L., Delaage, M., Maurel, C., Miassod, R., Jacq, B., and Rosset, R. (1998). Mutations in *ccf*, a novel *Drosophila* gene encoding a chromosomal factor, affect progression through mitosis and interact with Pc-G mutations. EMBO J *17*, 1063-1075.

Kosman, D., Ip, Y.T., Levine, M., and Arora, K. (1991). Establishment of the mesoderm-neuroectoderm boundary in the *Drosophila* embryo. Science *254*, 118-122.

Kretzschmar, M., Liu, F., Hata, A., Doody, J., and Massague, J. (1997). The TGF-beta family mediator Smad1 is phosphorylated directly and activated functionally by the BMP receptor kinase. Genes Dev. *11*, 984-995.

Kretzschmar, M. and Massague, J. (1998). Smads: mediators and regulators of TGF-beta signaling. Curr.Opin.Genet.Dev. *8*, 103-111.

Kutty, G., Kutty, R.K., Samuel, W., Duncan, T., Jaworski, C., and Wiggert, B. (1998). Identification of a new member of transforming growth factor-beta superfamily in *Drosophila*: the first invertebrate *activin* gene. Biochem Biophys.Res.Commun. *246*, 644-649.

Labbe, E., Silvestri, C., Hoodless, P.A., Wrana, J., and Attisano, L. (1998). Smad2 and Smad3 positively and negatively regulate TGFβ-dependent transcription through the forkhead DNA-binding protein FAST2. Molecular Cell *2*, 109-120.

Lagna, G., Hata, A., Hemmati, B.A., and Massague, J. (1996). Partnership between DPC4 and Smad proteins in TGF-beta signalling pathways. Nature *383*, 832-836.

Lai, Z.C., Fortini, M.E., and Rubin, G.M. (1991). The embryonic expression patterns of *zfh-1* and *zfh-2*, two *Drosophila* genes encoding novel zinc-finger homeodomain proteins. Mech.Dev. *34*, 123-134.

Lecuit, T., Brook, W.J., Ng, M., Calleja, M., Sun, H., and Cohen, S.M. (1996). Two distinct mechanisms for long-range patterning by Decapentaplegic in the *Drosophila* wing. Nature *381*, 387-393.

Lecuit, T. and Cohen, S.M. (1997). Proximal-distal axis formation in the *Drosophila* leg. Nature *388*, 139-145.

Leptin, M. (1991). *twist* and *snail* as positive and negative regulators during *Drosophila* mesoderm development. Genes Dev. *5* , 1568-1576.

Letsou, A., Arora, K., Wrana, J.L., Simin, K., Twombly, V., Jamal, J., Staehling, H.K., Hoffmann, F.M., Gelbart, W.M., Massague, J., and et, a. (1995). *Drosophila* Dpp signaling is mediated by the *punt* gene product: a dual ligand-binding type II receptor of the TGF beta receptor family. Cell *80*, 899-908.

Lilly, B., Galewsky, S., Firulli, A.B., Schulz, R.A., and Olson, E.N. (1994). D-MEF2: a MADS box transcription factor expressed in differentiating mesoderm and muscle cell lineages during *Drosophila* embryogenesis. Proc.Natl.Acad.Sci U.S.A. *91*, 5662-5666.

Lilly, B., Zhao, B., Ranganayakulu, G., Paterson, B.M., Schulz, R.A., and Olson, E.N. (1995). Requirement of MADS domain transcription factor D-MEF2 for muscle formation in *Drosophila*. Science *267*, 688-693.

Liu, F., Hata, A., Baker, J.C., Doody, J., Carcamo, J., Harland, R.M., and Massague, J. (1996). A human Mad protein acting as a BMP-regulated transcriptional activator. Nature *381*, 620-623.

Liu, F., Pouponnot, C., and Massague, J. (1997). Dual role of the Smad4/DPC4 tumor suppressor in TGFbeta-inducible transcriptional complexes. Genes Dev. *11*, 3157-3167.

Liu, X., Sun, Y., Constantinescu, S.N., Karam, E., Weinberg, R.A., and Lodish, H.F. (1997). Transforming growth factor beta-induced phosphorylation of Smad3 is required for growth inhibition and transcriptional induction in epithelial cells. Proc.Natl.Acad.Sci U.S.A. *94*, 10669-10674.

Lo, P.C. and Frasch, M. (1997). A novel KH-domain protein mediates cell adhesion processes in *Drosophila*. Dev.Biol *190*, 241-256.

Lo, R.S., Chen, Y.G., Shi, Y., Pavletich, N.P., and Massague, J. (1998). The L3 loop: a structural motif determining specific interactions between Smad proteins and TGF-beta receptors. EMBO J *17*, 996-1005.

Lough, J., Barron, M., Brogley, M., Sugi, Y., Bolender, D.L., and Zhu, X. (1996). Combined BMP-2 and FGF-4, but neither factor alone, induces cardiogenesis in non-precardiac embryonic mesoderm. Dev.Biol *178*, 198-202.

Luer, K., Urban, J., Klambt, C., and Technau, G.M. (1997). Induction of identified mesodermal cells by CNS midline progenitors in *Drosophila*. Development *124*, 2681-2690.

Lyons, I., Parsons, L.M., Hartley, L., Li, R., Andrews, J.E., Robb, L., and Harvey, R.P. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene *Nkx2-5*. Genes Dev. *9*, 1654-1666.

Macias, S.M., Abdollah, S., Hoodless, P.A., Pirone, R., Attisano, L., and Wrana, J.L. (1996). MADR2 is a substrate of the TGFbeta receptor and its phosphorylation is required for nuclear accumulation and signaling. Cell *87*, 1215-1224.

Maduzia, L.L. and Padgett, R.W. (1997). *Drosophila* Mad, a member of the Smad family, translocates to the nucleus upon stimulation of the dpp pathway. Biochem Biophys.Res.Commun. *238*, 595-598.

Maggert, K., Levine, M., and Frasch, M. (1995). The somatic-visceral subdivision of the embryonic mesoderm is initiated by dorsal gradient thresholds in *Drosophila*. Development *121*, 2107-2116.

Maniatis, T., Fritsch, E.F., and Sambrook, S. (1982). Molecular Cloning: A laboratory manual (Cold Spring Harbor Laboratory Press, New York).

Massague, J. (1990). The transforming growth factor-beta family. Annu.Rev Cell Biol *6*, 597-641.

Massague, J. (1992). Receptors for the TGF-beta family. Cell *69*, 1067-1070.

Massague, J. (1996). TGFbeta signaling: receptors, transducers, and Mad proteins. Cell *85*, 947-950.

Moustakas, A. and Kardassis, D. (1998). Regulation of the human *p21/WAF1/Cip1* promoter in hepatic cells by functional interactions between Sp1 and Smad family members. Proc.Natl.Acad.Sci U.S.A. *95*, 6733-6738.

Nakao, A., Imamura, T., Souchelnytskyi, S., Kawabata, M., Ishisaki, A., Oeda, E., Tamaki, K., Hanai, J., Heldin, C.H., Miyazono, K., and ten-Dijke, P. (1997a). TGF-beta receptor-mediated signalling through Smad2, Smad3 and Smad4. EMBO J *16*, 5353-5362.

Nakao, A., Afrakhte, M., Moren, A., Nakayama, T., Christian, J.L., Heuchel, R., Itoh, S., Kawabata, M., Heldin, N.E., Heldin, C.H., and ten-Dijke, P. (1997b). Identification of Smad7, a TGFbeta-inducible antagonist of TGF-beta signalling. Nature *389*, 631-635.

Nakayama, T., Snyder, M.A., Grewal, S.S., Tsuneizumi, K., Tabata, T., and Christian, J.L. (1998). Xenopus Smad8 acts downstream of BMP-4 to modulate its activity during vertebrate embryonic patterning. Development *125*, 857-867.

Nellen, D., Affolter, M., and Basler, K. (1994). Receptor serine/threonine kinases implicated in the control of *Drosophila* body pattern by *decapentaplegic*. Cell *78*, 225-237.

Nellen, D., Burke, R., Struhl, G., and Basler, K. (1996). Direct and long-range action of a DPP morphogen gradient. Cell *85* , 357-368.

Ner, S. and Travers, A.A. (1994). HMG-D, the *Drosophila melanogaster* homologue of HMG 1 protein, is associated with early embryonic chromatin in the absence of histon H1. EMBO J *13*, 1817-1822.

Neul, J.L. and Ferguson, E.L. (1998). Spatially restricted activation of the SAX receptor by SCW modulates DPP/TKV signaling in *Drosophila* dorsal-ventral patterning. Cell *95*, 483-494.

Newfeld, S.J., Chartoff, E.H., Graff, J.M., Melton, D.A., and Gelbart, W.M. (1996). *Mothers against dpp* encodes a conserved cytoplasmic protein required in DPP/TGF-beta responsive cells. Development *122*, 2099-2108.

Newfeld, S.J., Mehra, A., Singer, M.A., Wrana, J.L., Attisano, L., and Gelbart, W.M. (1997). Mothers against *dpp* participates in a DPP/TGF-beta responsive serine-threonine kinase signal transduction cascade. Development *124*, 3167-3176.

Nguyen, H.T., Bodmer, R., Abmayr, S.M., McDermott, J.C., and Spoerel, N.A. (1994). *D-mef2*: a *Drosophila* mesoderm-specific MADS box-containing gene with a biphasic expression profile during embryogenesis. Proc.Natl.Acad.Sci U.S A. *91*, 7520-7524.

Nguyen, H.T. and Xu, X. (1998). *Drosophila mef2* expression during mesoderm development is controlled by a complex array of *cis*-acting regulatory modules. Dev.Biol *204*, 550-566.

Nguyen, M., Park, S., Marques, G., and Arora, K. (1998). Interpretation of a BMP activity gradient in *Drosophila* embryos depens on synergistic signaling by two type I receptors, SAX and TKV. Cell *95*, 495-506.

Olson, E.N., Perry, M., and Schulz, R.A. (1995). Regulation of muscle differentiation by the MEF2 family of MADS box transcription factors. Dev.Biol *172*, 2-14.

Padgett, R.W., St.-Johnston, R.D., and Gelbart, W.M. (1987). A transcript from a *Drosophila* pattern gene predicts a protein homologous to the *transforming growth factor-β* family. Nature *325*, 81-84.

Padgett, R.W., Wozney, J.M., and Gelbart, W.M. (1993). Human BMP sequences can confer normal dorsal-ventral patterning in the *Drosophila* embryo. Proc.Natl.Acad.Sci U.S.A. *90*, 2905-2909.

Padgett, R.W., Das, P., and Krishna, S. (1998a). TGF-beta signaling, Smads, and tumor suppressors. Bioessays *20*, 382-390.

Padgett, R.W., Cho, S.H., and Evangelista, C. (1998b). Smads are the central component in transforming growth factor-beta signaling. Pharmacol.Ther. *78*, 47-52.

Penton, A., Chen, Y., Staehling, H.K., Wrana, J.L., Attisano, L., Szidonya, J., Cassill, J.A., Massague, J., and Hoffmann, F.M. (1994). Identification of two bone morphogenetic protein type I receptors in *Drosophila* and evidence that Brk25D is a decapentaplegic receptor. Cell *78*, 239-250.

Penton, A. and Hoffmann, F.M. (1996). Decapentaplegic restricts the domain of *wingless* during *Drosophila* limb patterning. Nature *382*, 162-164.

Pignoni, F. and Zipursky, S.L. (1997). Induction of *Drosophila* eye development by *decapentaplegic*. Development *124*, 271-278.

Raftery, L.A., Twombly, V., Wharton, K., and Gelbart, W.M. (1995). Genetic screens to identify elements of the *decapentaplegic* signaling pathway in *Drosophila*. Genetics *139*, 241-254.

Ranganayakulu, G., Zhao, B., Dokidis, A., Molkentin, J.D., Olson, E.N., and Schulz, R.A. (1995). A series of mutations in the D-MEF2 transcription factor reveal multiple functions in larval and adult myogenesis in *Drosophila*. Dev.Biol *171*, 169-181.

Ranganayakulu, G., Schulz, R.A., and Olson, E.N. (1996). Wingless signaling induces *nautilus* expression in the ventral mesoderm of the *Drosophila* embryo. Dev.Biol *176*, 143-148.

Rao, Y., Vaessin, H., Jan, L.Y., and Jan, Y.N. (1991). Neuroectoderm in *Drosophila* embryos is dependent on the mesoderm for positioning but not for formation. Genes Dev. *5*, 1577-1588.

Riese, J., Tremml, G., and Bienz, M. (1997). *D-Fos*, a target gene of Decapentaplegic signalling with a critical role during *Drosophila* endoderm induction. Development *124*, 3353-3361.

Riesgo, E.J. and Hafen, E. (1997). *Drosophila* Jun kinase regulates expression of decapentaplegic via the ETS-domain protein Aop and the AP-1 transcription factor DJun during dorsal closure. Genes Dev. *11* , 1717-1727.

Riggins, G.J., Thiagalingam, S., Rozenblum, E., Weinstein, C.L., Kern, S.E., Hamilton, S.R., Willson, J.K., Markowitz, S.D., Kinzler, K.W., and Vogelstein, B. (1996). Mad-related genes in the human. Nat.Genet. *13*, 347-349.

Roth, S., Stein, D., and Nusslein-Volhard, C. (1989). A gradient of nuclear localization of the Dorsal protein determines dorsalventral pattern in the *Drosophila* embryos. Cell *59*, 1189-1202.

Ruberte, E., Marty, T., Nellen, D., Affolter, M., and Basler, K. (1995). An absolute requirement for both the type II and type I receptors, *punt* and *thick veins*, for *dpp* signaling in vivo. Cell *80*, 889-897.

Rusch, J. and Levine, M. (1996). Threshold responses to the dorsal regulatory gradient and the subdivision of primary tissue territories in the *Drosophila* embryo. Curr.Opin.Genet.Dev. *6*, 416-423.

Sasai, Y., Lu, B., Steinbeisser, H., and De-Robertis, E.M. (1995). Regulation of neural induction by the Chd and Bmp-4 antagonistic patterning signals in Xenopus. Nature *376*, 333-336.

Savage, C., Das, P., Finelli, A.L., Townsend, S.R., Sun, C.Y., Baird, S.E., and Padgett, R.W. (1996). Caenorhabditis elegans genes *sma-2*, *sma-3*, and *sma-4* define a conserved family of *transforming growth factor beta* pathway components. Proc.Natl.Acad.Sci U.S.A. *93*, 790-794.

Schmidt, J., Francois, V., Bier, E., and Kimelman, D. (1995). *Drosophila short gastrulation* induces an ectopic axis in Xenopus: evidence for conserved mechanisms of dorsal-ventral patterning. Development *121*, 4319-4328.

Schott, J.J., Benson, D.W., Basson, C.T., Pease, W., Silberbach, G.M., Moak, J.P., Maron, B.J., Seidman, C.E., and Seidman, J.G. (1998). Congenital heart disease caused by mutations in the transcription factor NKX2-5. Science *281*, 108-111.

Schultheiss, T.M., Xydas, S., and Lassar, A.B. (1995). Induction of avian cardiac myogenesis by anterior endoderm. Development *121*, 4203-4214.

Schultheiss, T.M., Burch, J.B., and Lassar, A.B. (1997). A role for bone morphogenetic proteins in the induction of cardiac myogenesis. Genes Dev. *11*, 451-462.

Seider, F., Bock, E., and Krause, G. (1940). Die Organisation des insekteneies. Naturwissenschaften *28*, 434-447.

Sekelsky, J.J., Newfeld, S.J., Raftery, L.A., Chartoff, E.H., and Gelbart, W.M. (1995). Genetic characterization and cloning of *mothers against dpp*, a gene required for decapentaplegic function in *Drosophila melanogaster*. Genetics *139*, 1347-1358.

Shi, Y., Hata, A., Lo, R.S., Massague, J., and Pavletich, N.P. (1997). A structural basis for mutational inactivation of the tumour suppressor Smad4. Nature *388*, 87-93.

Shi, Y., Wang Y-F., Jayaraman, L., Yang, H., Massague, J., and Pavletich, N.P. (1998). Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-β signaling. Cell *94*, 585-594.

Shiojima, I., Komuro, I., Mizuno, T., Aikawa, R., Akazawa, H., Oka, T., Yamazaki, T., and Yazaki, Y. (1996). Molecular cloning and characterization of human cardiac homeobox gene *CSX1*. Circ.Res. *79*, 920-929.

Shishido, E., Higashijima, S., Emori, Y., and Saigo, K. (1993). Two FGF-receptor homologues of *Drosophila*: one is expressed in mesodermal primordium in early embryos. Development *117*, 751-761.

Simpson, P. (1983). Maternal-zygotic gene inetractions during formation of the dorsalventral pattern in *Drosophila* embryos. Genetics *105*, 615-632.

Sive, H.L. (1993). The frog prince-ss: a molecular formula for dorsoventral patterning in Xenopus. Genes Dev. *7*, 1-12.

Sluss, H.K. and Davis, R.J. (1997). Embryonic morphogenesis signaling pathway mediated by JNK targets the transcription factor JUN and the TGF-beta homologue decapentaplegic. J Cell Biochem *67*, 1-12.

Souchelnytskyi, S., Tamaki, K., Engstrom, U., Wernstedt, C., ten-Dijke, P., and Heldin, C.H. (1997). Phosphorylation of Ser465 and Ser467 in the C terminus of Smad2 mediates interaction with Smad4 and is required for transforming growth factor-beta signaling. J Biol Chem *272*, 28107-28115.

St.-Johnston, R.D. and Gelbart, W.M. (1987). Decapentaplegic transcripts are localized along the dorsal-ventral axis of the *Drosophila* embryo. EMBO J *6*, 2785-2791.

Staehling, H.K., Hoffmann, F.M., Baylies, M.K., Rushton, E., and Bate, M. (1994). *dpp* induces mesodermal gene expression in *Drosophila*. Nature *372*, 783-786.

Staehling, H.K., Laughon, A.S., and Hoffmann, F.M. (1995). A *Drosophila* protein related to the human zinc finger transcription factor PRDII/MBPI/HIV-EP1 is required for dpp signaling. Development *121*, 3393-3403.

Stroumbakis, N.D. and Tolias, P.P. (1994). Localized maternal and zygotic expression of the gene encoding *Drosophila* HMG D. Biochim.Biophys.Acta *1218*, 245-249.

Szuts, D., Eresh, S., and Bienz, M. (1998). Functional interwining of Dpp and EGFR signaling during *Drosophila* endoderm induction. Genes Dev. *12*, 2022-2035.

Takeda, K., Ichijo, H., Fujii, M., Mochida, Y., Saitoh, M., Nishitoh, H., Sampath, T.K., and Miyazono, K. (1998). Identification of a novel *bone morphogenetic protein*-responsive gene that may function as a noncoding RNA. J Biol Chem *273*, 17079-17085.

Taylor, M.V. (1995). Muscle development. Making *Drosophila* muscle. Curr.Biol *5*, 740-742.

Terracol, R. and Lengyel, J.A. (1994). The *thick veins* gene of *Drosophila* is required for dorsoventral polarity of the embryo. Genetics *138*, 165-178.

Thisse, B., Stoetzel, C., Gorostiza, T.C., and Perrin, S.F. (1988). Sequence of the *twist* gene and nuclear localization of its protein in endomesodermal cells of early *Drosophila* embryos. EMBO J *7*, 2175-2183.

Thuringer, F. and Bienz, M. (1993). Indirect autoregulation of a homeotic *Drosophila* gene mediated by extracellular signaling. Proc.Natl.Acad.Sci U.S.A. *90*, 3899-3903.

Thuringer, F., Cohen, S.M., and Bienz, M. (1993). Dissection of an indirect autoregulatory response of a homeotic *Drosophila* gene. EMBO J *12*, 2419-2430.

Tonissen, K.F., Drysdale, T.A., Lints, T.J., Harvey, R.P., and Krieg, P.A. (1994). *XNkx-2.5*, a Xenopus gene related to *Nkx-2.5* and *tinman*: evidence for a conserved role in cardiac development. Dev.Biol *162*, 325-328.

Tremml, G. and Bienz, M. (1992). Induction of *labial* expression in the *Drosophila* endoderm: response elements for *dpp* signalling and for autoregulation. Development *116*, 447-456.

Tsao, D.H., Gruschus, J.M., Wang, L.H., Nirenberg, M., and Ferretti, J.A. (1995). The three-dimensional solution structure of the NK-2 homeodomain from *Drosophila*. J Mol Biol *251*, 297-307.

Tsuneizumi, K., Nakayama, T., Kamoshida, Y., Kornberg, T.B., Christian, J.L., and Tabata, T. (1997). *Daughters against dpp* modulates *dpp* organizing activity in *Drosophila* wing development. Nature *389*, 627-631.

Vincent, S., Ruberte, E., Grieder, N.C., Chen, C.K., Haerry, T., Schuh, R., and Affolter, M. (1997). DPP controls tracheal cell migration along the dorsoventral body axis of the *Drosophila* embryo. Development *124*, 2741-2750.

Vindevoghel, L., Kon, A., Lechleider, R.J., Uitto, J., Roberts, A.B., and Mauviel, A. (1998). Smad-dependent transcriptional activation of human type VII collagen gene (COL7A1) promoter by *transforming growth factor-beta*. J Biol Chem *273*, 13053-13057.

Wharton, K.A., Ray, R.P., and Gelbart, W.M. (1993). An activity gradient of *decapentaplegic* is necessary for the specification of dorsal pattern elements in the *Drosophila* embryo. Development *117*, 807-822.

Whitman, M. (1998). Smads and early developmental signaling by the TGFβ superfamily. Genes Dev. *12*, 2445-2462.

Wieser, R., Wrana, J.L., and Massague, J. (1995). GS domain mutations that constitutively activate T beta R-I, the downstream signaling component in the TGF-beta receptor complex. EMBO J *14*, 2199-2208.

Wilson, P.A., Lagna, G., Suzuki, A., and Hemmati, B.A. (1997). Concentration-dependent patterning of the Xenopus ectoderm by BMP4 and its signal transducer Smad1. Development *124*, 3177-3184.

Wisotzkey, R.G., Mehra, A., Sutherland, D.J., Dobens, L.L., Liu, X., Dohrmann, C., Attisano, L., and Raftery, L.A. (1998). *Medea* is a *Drosophila* Smad4 homolog that is differentially required to potentiate DPP responses. Development *125*, 1433-1445.

Wrana, J.L., Attisano, L., Wieser, R., Ventura, F., and Massague, J. (1994). Mechanism of activation of the TGF-beta receptor. Nature *370*, 341-347.

Wu, R.Y., Zhang, Y., Feng, X.H., and Derynck, R. (1997). Heteromeric and homomeric interactions correlate with signaling activity and functional cooperativity of Smad3 and Smad4/DPC4. Mol Cell Biol *17*, 2521-2528.

Xie, T., Finelli, A.L., and Padgett, R.W. (1994). The *Drosophila saxophone* gene: a serine-threonine kinase receptor of the TGF-beta superfamily. Science *263*, 1756-1759.

Xu, X., Yin, Z., Hudson, J.B., Ferguson, E.L., and Frasch, M. (1998). Smad proteins act in combination with synergistic and antagonistic regulators to target Dpp responses to the *Drosophila* mesoderm. Genes Dev. *12*, 2354-2370.

Yamamoto, N., Akiyama, S., Katagiri, T., Namiki, M., Kurokawa, T., and Suda, T. (1997). Smad1 and smad5 act downstream of intracellular signalings of BMP-2 that inhibits myogenic differentiation and induces osteoblast differentiation in C2C12 myoblasts. Biochem Biophys.Res.Commun. *238*, 574-580.

Yin, Z., Xu, X.L., and Frasch, M. (1997). Regulation of the *twist* target gene *tinman* by modular *cis*-regulatory elements during early mesoderm development. Development *124*, 4971-4982.

Yin, Z. and Frasch, M. (1998). Regulation and function of *tinman* during dorsal mesoderm induction and heart specification in *Drosophila*. Dev.Genet. *22*, 187-200.

Yingling, J.M., Das, P., Savage, C., Zhang, M., Padgett, R.W., and Wang, X.F. (1996). Mammalian *dwarfins* are phosphorylated in response to *transforming growth factor beta* and are implicated in control of cell growth. Proc.Natl.Acad.Sci U.S.A. *93*, 8940-8944.

Yingling, J.M., Datto, M.B., Wong, C., Frederick, J.P., Liberati, N.T., and Wang, X.F. (1997). Tumor suppressor Smad4 is a *transforming growth factor beta*-inducible DNA binding protein. Mol Cell Biol *17*, 7019-7028.

Yu,Y. (1996). Transcriptional regulation of the *Drosophila fushi tarazu* gene by its proximal enhancer. Dissertation, Mount Sinai Medical Center, CUNY.

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

Zaffran, S., Astier, M., Gratecos, D., and Semeriva, M. (1997). The *held out wings (how) Drosophila* gene encodes a putative RNA-binding protein involved in the control of muscular and cardiac activity. Development *124*, 2087-2098.

Zappavigna, V., Falciola, L., Citterich, M.H., Mavilio, F., and Bianchi, M.E. (1996). HMG1 interacts with HOX proteins and enhances their DNA binding and transcriptional activation. EMBO J *15*, 4981-4991.

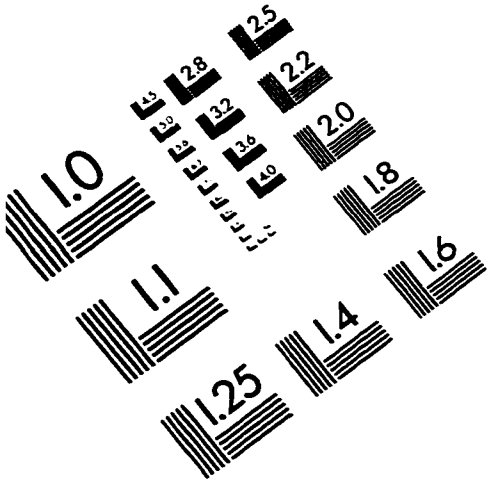
Zawel, L., Dai, J.L., Buckhaults, P., Zhou, S., Kinzler, K.W., Vogelstein, B., and Kern, S.E. (1998). Human Smad3 and Smad4 are sequence-specific transcription activators. Mol Cell *1*, 611-617.

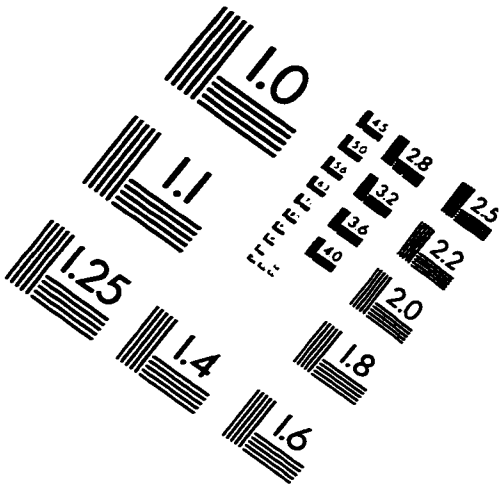
Zhang, Y., Feng, X., We, R., and Derynck, R. (1996). Receptor-associated Mad homologues synergize as effectors of the TGF-beta response. Nature *383*, 168-172.

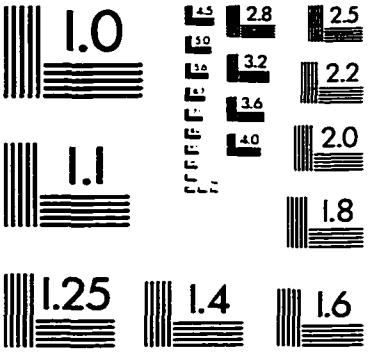
Zhang, Y., Musci, T., and Derynck, R. (1997). The tumor suppressor Smad4/DPC 4 as a central mediator of Smad function. Curr.Biol *7*, 270-276.

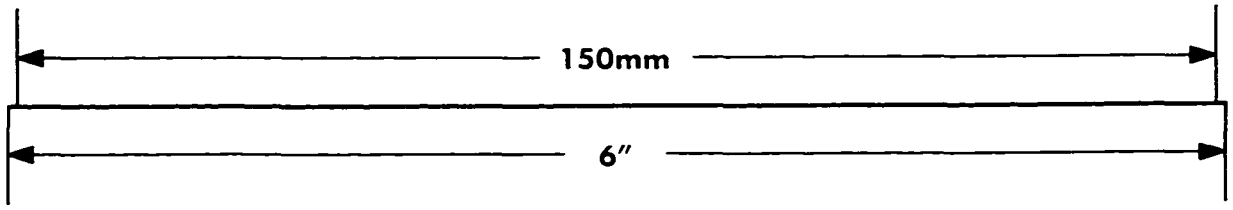
Zhou S., Zawel, L., Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1998). Characterization of human *FAST-1*, a TGF-β and Activin signal transducer. Mol Cell *2*, 121-127.

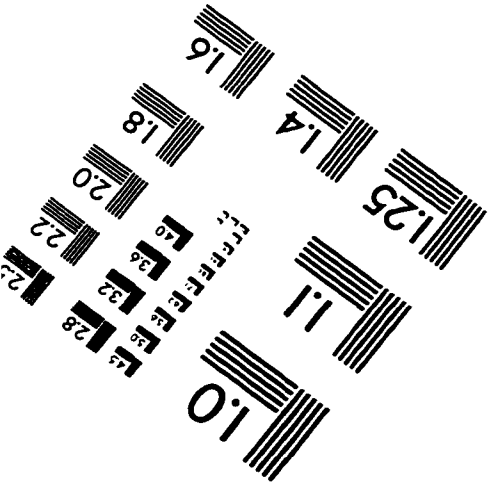
IMAGE EVALUATION
TEST TARGET (QA-3)

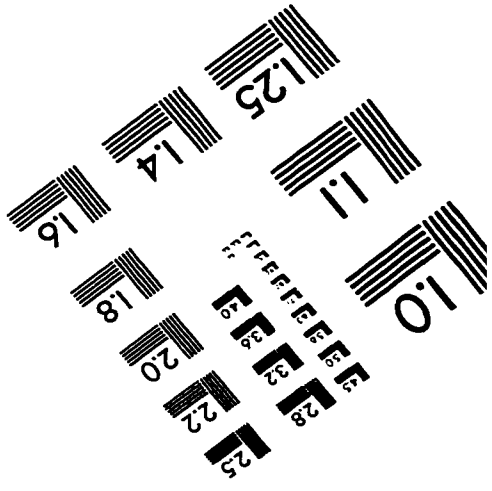












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989