Combinatorial patterning mechanisms in the Drosophila embryo

Vivek S. Chopra and Mike Levine

Advance Access publication date 3 August 2009

Abstract

The classical concept of the morphogen gradient proposes that small differences in the levels of a signalling molecule or transcription factor are responsible for producing a continuous spectrum of distinctive cellular identities across a naive field of cells. In this review, we discuss how the Dorsal gradient controls the dorsal-ventral patterning of the early Drosophila embryo. This gradient extends from the ventral midline of the embryo into dorso-lateral regions, encompassing a cross-sectional field of approximately 20 cells. There is no evidence that these cells acquire distinctive identities due to subtle changes in the nuclear concentrations of the Dorsal protein. Rather, a variety of evidence suggests that the Dorsal gradient generates just three primary thresholds of gene activity. High levels activate gene expression in the presumptive mesoderm, while intermediate and low levels activate gene expression in the ventral and dorsal neurogenic ectoderm, respectively. We discuss how these primary readouts of the gradient establish localized domains of cell signalling, which work in a combinatorial manner with transcriptional networks to produce complex patterns of gene expression and tissue differentiation.

Keywords: DV patterning; cis-regulatory elements; Drosophila; enhancers; repressors

FORMATION OF THE DORSAL GRADIENT

The Dorsal protein is related to mammalian NF-κB, a critical transcription factor implicated in a variety of processes including innate immunity, inflammation and B-lymphocyte differentiation [1–3]. The Dorsal mRNA is maternally expressed in the developing oocyte, and it is possible that low levels of the protein are also synthesized within the ovary. However, an inhibitory protein, Cactus (related to mammalian IκB), retains Dorsal within the cytoplasm of mature oocytes and early embryos [4–6].

After fertilization, an extracellular serine protease cascade leads to the proteolytic processing of pro-Spaetzle into an active ligand [7–9]. pro-Spaetzle is distributed throughout the perivitelline matrix surrounding the unfertilized egg and early embryo.

This processing event is probably triggered by the synthesis of one or more serine proteases [9]. These enzymes are encoded by maternal mRNAs present in the unfertilized egg. After fertilization they produce proteins with signal peptides that result in their secretion into the perivitelline matrix.

Although distributed throughout the extracellular matrix, the resulting serine proteases (Gastrulation defective, Snake and Easter) are active only in ventral regions of the matrix, where there is a presumed modification ‘mark’ produced by the prior action of a localized extracellular heparan sulfotransferase encoded by Pipe [9, 10].

The activation of the serine protease cascade at the extracellular mark results in the localized cleavage of pro-Spaetzle into the active Spaetzle ligand [7, 11–14].
During the first 90 min after fertilization, activated Spaetzle interacts with the transmembrane Toll receptor distributed throughout the plasma membrane of the early embryo. It is thought that Spaetzle is distributed in an extracellular gradient, with peak levels of the active ligand located in ventral regions and lower levels in more lateral regions [7, 14, 15]. Several models have been suggested to account for the formation of this gradient, but the exact details remain elusive.

The localized activation of the Toll receptor leads to the translocation of the Dorsal protein from the cytoplasm to the nucleus of syncitial-stage embryos. The activated receptor leads to the induction of the cytoplasmic Pelle kinase, which in turn results in the modification and proteasome-mediated degradation of the Cactus inhibitor [16–20]. This degradation releases Dorsal from the cytoplasm to the nucleus. Thus, an extracellular gradient of the Toll ligand Spaetzle leads to the formation of a Dorsal nuclear gradient [18, 21–23]. There is evidence for a linear correlation between the levels of activated Toll receptor and the amount of Dorsal protein that enters nuclei [18, 24].

**DORSAL GRADIENT THRESHOLDS**

The Dorsal nuclear gradient is first formed 90 min after fertilization, shortly after cleavage nuclei enter the cortical regions of the syncitial embryo. The gradient persists during cellularization, but rapidly disappears during gastrulation and the initial phases of germ band elongation. Thus, the Dorsal gradient is active for <2 h interval during embryogenesis, but nonetheless plays a decisive role in launching an elaborate gene regulatory network that governs dorsal–ventral (DV) patterning [25, 26].

A combination of genetic methods and whole-genome assays suggests that the Dorsal gradient directly regulates over 50 different target genes in the cellularizing embryo [27, 28]. It also indirectly regulates another ~10–20 genes engaged in the patterning of the Dorsal ectoderm in response to a Dpp signalling gradient. Most of the direct Dorsal target genes are also regulated by Twist and/or Snail, which are sequence-specific transcription factors encoded by ‘immediate early’ target genes of the Dorsal gradient. Whole-genome ChIP-chip assays have identified most of the regulatory DNAs that are controlled by Dorsal along with Twist and/or Snail [26–30].

Approximately 50 direct Dorsal target genes display at least six different patterns of gene expression. However, there is reason to believe that these six patterns are produced by just three ‘primary’ threshold readouts of the Dorsal gradient. High levels of the gradient activate ‘type 1’ target genes such as twist, snail, heartless and NF-Yc. Intermediate levels activate ‘type 2’ genes in ventral regions of the neurogenic ectoderm, while low levels activate ‘type 3’ target genes throughout the presumptive neurogenic ectoderm. In principle, type 2 and type 3 genes can be activated by high levels of the Dorsal gradient within the presumptive mesoderm, but the associated regulatory DNAs contain Snail repressor binding sites that inhibit expression in these regions [27–29, 31] (Figure 1).

The computational analysis of approximately 25 direct Dorsal target enhancers identified just two classes of Dorsal binding sites: high-affinity sites and low-affinity sites. Most of the type 1 enhancers, which mediate gene expression in the presumptive mesoderm containing the highest levels of the Dorsal gradient, contain low-affinity binding sites (Figure 2A). In contrast, at least a subset of the Dorsal binding sites contained within type 2 and type 3 enhancers correspond to high-affinity sites (Figure 2C and E). There is no clear difference in the quality of the Dorsal binding sites contained in the type 2 and type 3 enhancers, even though the type 3 genes are activated by lower levels of the Dorsal gradient than type 2 genes (Figure 2C–F). That is, the dorsal limits of the type 3 expression patterns extend at least five cell diameters beyond the limits of the type 2 expression patterns [25, 31–34].

What accounts for the distinctive type 2 and type 3 expression patterns? Conventional views of morphogen gradients might predict that the binding affinities of the Dorsal recognition sequences would be a key determinant of these different thresholds. However, as mentioned above, there is no significant difference in the Dorsal binding affinities of type 2 and type 3 enhancers. Instead, distinct type 2 and type 3 expression patterns appear to depend on different combinatorial codes of gene expression. All of the type 2 enhancers contain linked binding sites for Dorsal and Twist [25, 31, 33] (Figure 2C). Twist is a bHLH activator that probably forms heterodimeric complexes with Daughterless (Da), a ubiquitous, maternal bHLH activator related to...
mammalian E12/E47 [35–37]. Twist (and Twist/Da) is distributed in a steep and narrow gradient, with a rapid loss of protein across several cell diameters in the ventral-most regions of the presumptive neurogenic ectoderm [36, 38, 39]. Linked Dorsal and Twist binding sites appear to foster cooperative protein–protein interactions leading to efficient occupancy of the sites in the ventral neurogenic ectoderm, where there are low levels of the Dorsal

**Figure 1:** The nuclear Dorsal gradient creates various thresholds of gene expression. The nuclear gradient of Dorsal lead to concentration dependent activation of 60–70 target genes in the dorsal–ventral axis of the early embryo. The expression domains of these genes are represented in the cross-section through an early embryo (arranged as dorsal side up and ventral side down). Blue circles represent different levels of Dorsal expression in the nuclei from high (filled circle) to intermediate (light blue shades) to no Dorsal (white circles). At least six different expression patterns are believed to be generated. High Dorsal in the nucleus activate type I genes such as sna in the presumptive mesoderm. Type 2 genes like vnd are activated by intermediate levels of Dorsal gradient in the presumptive neuroectoderm. The Type IA and 2A expression profile shown by sim and ind depend on Notch and EGF signalling and high and intermediate Dorsal levels, respectively. The lowest level of Dorsal gradient generates two type 3 kind of responses as depicted for sog and zen. The similar low levels of Dorsal that is required for expression of sog represses zen. Thus, nuclear Dorsal gradient generates three basic transcription responses that produces a total of six different patterns of gene expression along the DV axis. This figure was adapted from review by Stathopoulos and Levine [44].

**Figure 2:** Cis-regulatory code for DV patterning. (A) High levels of Dorsal activate type I genes such as sna and restrict the pattern to presumptive mesoderm. The type I enhancers contain low affinity Dorsal and/or Twist binding sites. (B) The type IA enhancers containing gene sim contain high affinity Dorsal site along with Su(H) site that depends on Notch signalling. This is expressed in the mesectoderm regions. (C) Type 2 enhancers containing genes such as vnd, are activated by intermediate levels of Dorsal gradient and low levels of Twist. All the known type 2 enhancers contain fixed arrangement of Dorsal and Twist binding sites, i.e. an optimal Dorsal site is closely linked to an asymmetric Twist site that is in convergent orientation relative to Dorsal site (black arrow). This cooperative arrangement of binding sites leads to stable expression in ventral regions of neurogenic ectoderm. (D) EGF signalling is induced by intermediate levels of Dorsal and is important for induction of type 2A gene ind in the neurogenic ectoderm. The ind enhancer contains high affinity Dorsal sites and ETS sites that mediate EGF signalling. Two classes of type 3 enhancers are activated by low levels of Dorsal gradient (E) activation and (F) repression. Optimal Dorsal binding sites are linked to Zelda binding sites within the intronic enhancer of sog (E) and leads to activation throughout neurogenic ectoderm. Same low levels of Dorsal also represses zen and it contains optimal Dorsal binding sites and an AT-rich motif (type 3A) that binds one or more ubiquitous factors that converts Dorsal into a silencer via the recruitment of Groucho corepressor (F).
gradient and severely limiting levels of Twist [25, 31, 38, 40, 41]. The extensive reliance of type 2 enhancers on critical Twist binding sites is probably responsible for the restricted expression of type 2 genes in ventral regions of the neurogenic ectoderm.

In contrast, at least some of the type 3 enhancers, particularly those associated with the sog locus, contain linked binding sites for Dorsal and a ubiquitous, maternal zinc-finger transcription factor called Zelda [42] (Figure 2E). In essence, type 3 enhancers have Zelda binding sites in place of the Twist sites seen in type 2 enhancers [43]. It is possible, but currently unproven, that Dorsal and Zelda cooperatively interact to ensure efficient occupancy of the Dorsal binding sites in Dorsal regions of the presumptive neurogenic ectoderm where there are diminishing levels of the Dorsal gradient.

Altogether, the current evidence suggests that there are just two basic classes of Dorsal recognition sequences, low-affinity sites and high-affinity sites. These two classes of binding sites distinguish activation by high versus intermediate or low levels of the Dorsal gradient. Differential patterns of type 2 and type 3 Dorsal target genes result from distinctive combinatorial codes. Dorsal + Twist define the type 2 response while Dorsal + Zelda define the type 3 response. Many of the type 1 enhancers contain both Dorsal and Twist binding sites. However, expression is restricted to ventral regions containing peak levels of the Dorsal gradient since type 1 enhancers contain low-affinity Dorsal binding sites that are not tightly linked to Twist sites, as seen in type 2 enhancers [25, 26, 43, 44].

**CONVERSION OF TYPE 3 ENHANCERS INTO SILENCERS**

The preceding arguments account for just three of the six DV expression patterns seen for Dorsal target genes. A fourth pattern, localized expression in the presumptive dorsal ectoderm, employs the same logic as that described for the type 3 response. Dorsal target genes like dpp and zen are repressed by the same low levels of the Dorsal gradient that activate sog and other type 3 genes. However, the regulatory DNAs associated with dpp and zen do not contain linked Dorsal and Zelda binding sites. Instead, they contain linked binding sites for Dorsal and an AT-rich sequence motif [45] (Figure 2F). Several proteins might interact with this motif, including Cut and Dead Ringer [46, 47], but it is not clear which of these (or other) factors are critical for the activities of the dpp and zen regulatory DNAs.

There is evidence that protein(s) bound to the AT-motif interact with Dorsal at neighboring sites, and cause a conformational change in the Dorsal protein so that it interacts with the corepressor protein, Groucho [48, 49] (Figure 2F). Normally, Dorsal functions as an activator, but when it interacts with proteins bound at the AT-motif, it works as a repressor by recruiting Groucho through exposure of a cryptic repression peptide, FxLxxIL [45, 50]. In principle, the replacement of Zelda activator sites with AT-rich repressor elements would convert type 3 enhancers into type 3 silencers, which restrict gene expression to the Dorsal ectoderm due to repression in ventral and lateral regions by high and low levels of the Dorsal gradient.

**DV Patterning Network Explains the Remaining Expression Patterns**

The two remaining expression patterns are exemplified by the regulation of two critical neurogenic genes, sim and ind. sim encodes a bHLH-PAS transcription factor that controls the differentiation of the specialized ventral midline of the Drosophila ventral nerve cord [51]. ind encodes a homeodomain transcription factor that controls the development of intermediate neurons in lateral regions of the nerve cord [52].

Sim is expressed in a single line of cells on both sides of the presumptive mesoderm. The Dorsal gradient establishes this highly refined expression pattern through the localized deployment of the Notch signalling pathway. Dorsal (and Twist) activate snail expression in the ventral mesoderm. Snail represses the expression of Tom, a repressor of neuralized, which encodes a ubiquitin ligase required for the trafficking of the Notch ligand Delta [53, 54]. As a result of this double negative gate (Snail represses Tom, which inhibits Notch signalling via repression of neuralized), the mesoderm triggers Notch signalling and the activation of sim expression in the ventral-most cells of the presumptive neurogenic ectoderm [55]. The sim enhancer contains a series of low-affinity Dorsal binding sites, Twist sites and Suppressor of Hairless [Su(H)] binding sites, which mediate Notch signalling [25]. Thus, the combination of Dorsal, Twist and Notch
signalling is essential for the localized sim expression pattern (Figure 2B).

Similar logic applies to the localized expression of ind in lateral regions of the presumptive nerve cord. ind expression is induced by EGF signalling emanating from ventral regions of the neurogenic ectoderm [56]. EGF signals arise from two sources. First, Dorsal (and Twist) directly activates the type 2 target genes vein and rho, which encode an EGF signalling molecule and a membrane protease that releases EGF, respectively [57, 58]. Second, sim activates the expression of spitz in the ventral midline. spitz encodes a membrane-bound EGF signalling ligand that also contributes to the activation of ind in lateral regions [59, 60]. The ind enhancer contains binding sites for Dorsal and Ets-containing activator(s), along with binding sites for at least two repressors, Vnd and Snail, which inhibit ind expression in the ventral neurogenic ectoderm and ventral mesoderm, respectively (Figure 2D).

**CONCLUSIONS**

The Dorsal gradient generates six distinct patterns of gene expression across the DV axis of the early Drosophila embryo. The in-depth analysis of a large collection of DV regulatory DNAs suggests that Dorsal does not function as a classical morphogen gradient. It is not sufficient to produce a spectrum of expression patterns, for example, by interacting with target enhancers containing different numbers, arrangements or affinities of Dorsal binding sites. Instead, Dorsal functions in a highly combinatorial manner to direct different patterns of gene expression. Dorsal works with Twist, Zelda, Su(H) and other sequence-specific transcription factors to produce localized DV expression patterns.

The only obvious role for diverse Dorsal binding affinities is seen for the differential expression of type 1 and type 2 enhancers, which are activated by high and intermediate levels of the Dorsal gradient, respectively. Both classes of enhancers contain Dorsal and Twist binding sites. However, type 1 enhancers tend to contain low-affinity Dorsal binding sites, and the Dorsal and Twist sites are disordered. In contrast, type 2 enhancers contain high-affinity Dorsal sites that are tightly linked to Twist sites. This combinatorial view of Dorsal gradient thresholds is in keeping with the general rules of transcriptional control, and explains the precision and reproducibility of the DV patterning process.

---

**Key Points**

- The nuclear concentrations of Dorsal protein controls the dorsal–ventral patterning in the early Drosophila embryo.
- The Dorsal gradient generates three primary thresholds of gene activity via combinatorial action of various transcription factors.
- High levels of nuclear Dorsal activate gene expression in the presumptive mesoderm, intermediate levels activate genes in ventral neurogenic ectoderm and low levels activate gene expression in the Dorsal neurogenic ectoderm.
- The primary readouts of the gradient establish localized domains of cell signalling, which work in a combinatorial manner with inputs from transcriptional networks to produce complex patterns of gene expression and tissue differentiation.

---

**References**


46. Iwahara J, Clubb RT. Solution structure of the DNA binding domain from Dead ringer, a sequence-specific AT-rich interaction domain (ARID). *Proc Natl Acad Sci USA* 1996;93:6084–94.


60. Golembo M, Raz E, Shilo BZ. The Drosophila embryonic midline is the site of Spitz processing, and induces activation of the EGF receptor in the ventral ectoderm. *Development* 1996;122:3363–70.