Cell, Vol. 78, 181–189, July 29, 1994, Copyright © 1994 by Cell Press

Homeobox Genes: Their Function in Drosophila Segmentation and Pattern Formation

Review

Peter A. Lawrence* and Gines Morata†
*Medical Research Council Laboratory of Molecular Biology
Hills Road
Cambridge CB2 2QH
England
†Centro de Biologia Molecular
Consejo Superior de Investigaciones Cientificas
Facultad de Ciencias C-X
Universidad Autonoma de Madrid
28049 Madrid
Spain

A Brief History

Long before the discovery of the homeobox (McGinnis et al., 1984; Scott and Weiner, 1984), Drosophila workers knew that homeotic genes are special. In the late 1940s, Ed Lewis began a study of mutations that produced almost magical transformations; he made flies with four wings instead of two, with halteres instead of wings, and from then until now he has worked on the bithorax complex (BX-C), a group of ruling genes that help design the fly. It was Lewis who discovered the vitally important fact that the wild-type function of each homeotic gene is restricted to a specific region of the developing insect. He and others showed that the developmental pathway followed by each cell depends on the set of BX-C genes active within it. By studying genetic mosaics, these researchers showed that each cell differentiates autonomously (Lewis, 1963; Morata and Garcia-Bellido, 1976). Thus, homeotic genes like the BX-C are responsible for determination, the internal molecular and genetic state of a cell that makes it stably different from other cells.

The role of homeotic genes in determination was amply discussed in the 1970s, especially after the discovery of developmental compartments, precise anatomical regions produced by all the descendants of a small set of founder cells (Garcia-Bellido et al., 1973; Lawrence, 1973). Compartments and their invariant boundaries were mapped by cell lineage experiments using genetically marked clones of cells. Garcia-Bellido (1975) and coworkers noted that some of the homeotic mutations transformed domains that were exactly coextensive with compartments, and this led to the idea that a set of controlling or selector genes act in the founder cells of compartments to specify the body parts they will construct. A key characteristic of selector genes is that they act as binary switches, with loss of function producing one transformation in which the gene is active and gain of function tending to produce the opposite transformation (in the places where the gene is normally inactive). Garcia-Bellido's hypothesis (1975) linked genetics and cell determination with cell lineage and anatomy. Many selector genes were later shown to encode homeodomain proteins.

Lewis' seminal paper (1978) analyzed the BX-C in embryos, showed that many elements (about one per segment) determine the middle and posterior body parts, and described their deployment and function. Meanwhile, another cluster of selector genes, the *Antennapedia* (*Antp*) complex (ANT-C) was being defined by Kaufman and coworkers (see Kaufman, 1983). This complex acts more anteriorly in the embryo; together, the BX-C and the ANT-C specify much of the body plan or, more specifically, parasegments 1–14, which we call the trunk.

In the mid-1980s, Morata and colleagues (Sanchez-Herrero et al., 1985) undertook mutagenesis to determine the number of separate genes in the BX-C and, rather surprisingly, found only three; it then became clear that most of the genetic elements named by Lewis did not produce their own proteins but were due to mutations that alter patterns of transcription of the three genes. It was at this point that the homeobox was discovered, and just three homeoboxes were found in the BX-C, one in each gene.

The homeobox binds DNA and the homeoproteins are transcription factors (Desplan et al., 1985; Hoey and Levine, 1988; Levine and Hoey, 1988; Thali et al., 1988); this indicated that the proteins would regulate other genes. The concept of determination could now become less abstract and more molecular. Fishing with homeobox probes led to the discovery of new selector genes and to a better understanding of previously known genes, such as *engrailed* (*en*). Similar fishing expeditions have identified homeoboxes in all animal groups in which they have been looked for and have contributed massively to the new and growing science of comparative molecular genetics, a more objective version of comparative embryology and anatomy, subjects that were popular in the first half of this century.

We divide this review into two main sections. The first describes the function of genes of the BX-C and ANT-C (we call these the primary homeotic genes), that is, how they specify development of regions of the embryo and fly and how they act in relation to cell lineage and segmentation. We also discuss some aspects of mechanism. In the second section, we summarize recent work that begins to explain how the homeobox genes, in specifying compartments, set up the machinery to build the wing, an example of a complex pattern.

For a description of the genes themselves, genetics, and other matters not pertaining to function, please see other reviews (for example, Duncan, 1987; Kaufman, 1983; Kaufman et al., 1990; McGinnis and Krumlauf, 1992). Also kindly note that this review is limited in length and in the number of references and that many relevant topics and primary sources therefore have had to be omitted.

Homeobox Genes: Their Functions in Cell Determination

Definition of Compartments in the Embryo

The maternal genes, such as the homeobox gene *bicoid* and the zygotic gap genes, together set up gradients of

positional information that position the expression of pair-rule genes, including the seven stripes of fushi tarazu (ftz) and the intervening stripes of even skipped (eve) proteins (Nüsslein-Volhard, 1991; Lawrence, 1992). Both *ftz* and *eve* are homeobox genes. Soon after the blastoderm stage, it is these stripes that help allocate the cells to the 14 parasegments, with each parasegment being founded by all the cells that lie between the anterior boundaries of adjacent ftz and eve stripes. Since the sets of founding cells form polyclones that construct precisely defined parts of at least the ectoderm of the larva and adult, the parasegments constitute the first compartments to be formed. At around the same time, ftz and eve activate *en* in each parasegment, the anterior borders of the stripes of *en* expression defining the parasegment borders. The posterior borders of the *en* stripes are initially labile (Vincent and O'Farrell, 1992) and then become fixed; once so defined, the sets of cells that express *en* originate the posterior compartments. The trunk of the embryo thereby becomes a chain of alternating posterior and anterior sets of cells, a parasegment comprising one posterior and one anterior set (Martinez-Arias and Lawrence, 1985).

Primary Homeobox Genes and the Diversification of Segments

We consider here only the 14 trunk parasegments that make the posterior head, thorax, and abdomen and ignore the terminal structures (the anterior head and telson), which are less well understood. However, three putative selector genes have been identified that act in the terminalia, and two of them, *empty spiracles* (*ems*) and *orthodenticle*, are homeobox genes (Cohen and Jürgens, 1991). Each of the 14 parasegments in the trunk is initially identical, but they diversify by the action of eight primary homeobox genes. Five of these genes belong to the ANT-C and three to the BX-C, and together they amount to the homeotic cluster, a genetic system that has been conserved in the evolution of all animals examined so far. For simplicity, we discuss here those genes active in the thorax and abdomen; these are *Sex combs reduced* (*Scr*) and *Antp* from the ANT-C and *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*), and *Abdominal-B* (*Abd-B*) from the BX-C. These genes are activated imprecisely by the concentration of gap gene proteins so that, for example, *Ubx* is turned on between certain values of the *hunchback* gradient (Struhl et al., 1992), giving a broad band of *Ubx* expression near the middle of the embryo. Later, the limits of expression of the homeobox genes respond to ftz and eve (Müller and Bienz, 1992), sharpen, and come into register with the anterior boundaries of the parasegments; for example, the anterior boundary of *abd-A* coincides with the anterior boundary of parasegment 7.

The diversity of the body pattern depends on the deployment of genes of the ANT-C and the BX-C, which act alone or in combination. There are rules to their deployment: each gene is activated at a particular parasegment boundary and then is expressed posteriorly from there. In Figure 1 we summarize the core domains of expression and action of the genes in the ectoderm; note that, in the absence of trans-interactions, the anterior as well as the posterior limits of these domains coincide with parasegment boundaries. Separate promoter elements drive the expression in the two mesoderms; for example, *abd-A* is transcribed from parasegments 7–12 in the somatic mesoderm and from parasegments 8–12 in the visceral mesoderm. In the ectoderm, particular combinations of these genes make binary code words that specify segmental identities; other combinations of genes, nonsense code words (Struhl, 1982), give a mishmash pattern in which the cuticular elements are variously and inappropriately chosen.

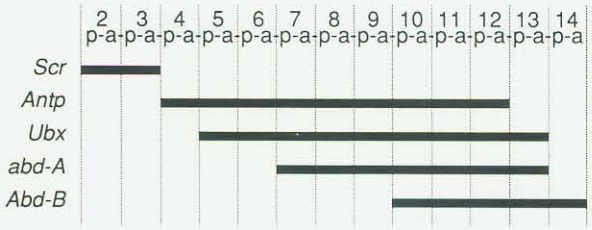

Figure 1. The Primary Homeotic Genes: Core Domains in the Epidermis

Domains of expression of each of the homeobox genes are given in the absence of the posteriorly acting genes; for example, we show the *Antp* pattern of expression in the absence of the three BX-C genes and the *Ubx* pattern in the absence of *abd-A* and *Abd-B*. The *Scr* pattern applies only during early development, and there are some complications (Kaufman et al., 1990). Abbreviations: p, posterior; a, anterior.

What is the ground pattern? That is, how would the trunk or a single parasegment differentiate if it lacked all these selector genes? The combination Scr^- $Antp^-$ Ubx^- $abd\text{-}A^-$ $Abd\text{-}B^-$ removes all homeotic information in the thoracic and abdominal metameres; also, the other elements of the ANT-C (such as *labial* [*lab*]) are not derepressed there. The outcome in the cuticle is a pattern consisting of mainly thoracic elements with, in the posterior compartments, some cephalic structures (Struhl, 1983). Most likely these latter are a complication due to derepression of some selector genes that are normally active in the head. The primordial metameric pattern is therefore probably thoracic (with legs on each segment), meaning that in evolution and, to some extent in development, every metamere starts with the same thoracic ground plan but diverges as more and more homeobox genes are activated. Study of the patterns produced by combinations of mutants in the primary homeobox genes suggests that, in the wild type, the genes act somewhat step by step, simplified as follows: step 0, $Antp^-$ Ubx^- $abd\text{-}A^-$ $Abd\text{-}B^-$, ground; step 1, $Antp^+$ Ubx^- $abd\text{-}A^-$ $Abd\text{-}B^-$, T2; step 3, $Antp^+$ Ubx^+ $abd\text{-}A^-$ $Abd\text{-}B^-$, T3 and A1; step 4, $Antp^+$ Ubx^+ $abd\text{-}A^+$ $Abd\text{-}B^-$, A2–A4; step 5, $Antp^+$ Ubx^+ $abd\text{-}A^+$ $Abd\text{-}B^+$, A5–A8.

These combinations are insufficient to account for all the different segments, such as A3 and A4, each of which has its own pattern. The origins of these subtle but significant differences in design are not fully understood. Part of the answer may lie in varying levels of the products of selector genes; an example concerns parasegment 5 (T3), which can be transformed into parasegment 6 (A1) simply by adding extra Ubx protein (with a heat shock construct). If level is the right explanation, this makes an exception to the general rule that selector genes act as binary switches.

Another part of the answer is that some gene products dominate completely or partially over others. These mechanisms make differing contributions in particular cases: consider the T3 segment in the embryo or adult in which both *Antp* at a high level and *Ubx* at a low level are expressed. If either gene is removed, the pattern is altered; it follows that, in the wild type, both genes must combine to make the T3 pattern. Note that this cannot be done by the genes acting in a mosaic pepper-and-salt fashion because individual denticles, made by single cells in T3, are of characteristic T3 identity, which neither *Ubx* nor *Antp* can specify alone. In another slightly different example, although Ubx and abd-A proteins coexist in parasegments 7 and 8 and both make some contribution to the pattern, removing them one at a time shows that the abd-A protein is by far the most important.

We find ourselves describing the realms of action of the different genes sometimes in terms of parasegments and sometimes in terms of segments. The earliest (and we think the most fundamental) modules of development are parasegments. This is well established for the primary homeotic genes, whose anterior limits coincide, cell by cell, with parasegment boundaries and whose posterior limits also tend to do so (Figure 1), although there are some complications (Kaufman et al., 1990). Note that it is not just that the homeotic genes are expressed in and required in parasegmental domains, but that it is clear that the homeotic genes specify parasegments rather than segments. For example, if the only gene active in an embryo is *Ubx*, the embryo develops as a chain of parasegments 6, with alternating compartments of posterior T3 and anterior A1s.

Nevertheless, when we look at later development, especially of the imaginal discs, the homeotic genes seem to act as if their units of function were not the parasegment, but the compartment or segment. The *Antp*, *Scr*, and *en* mutations produce segmental, rather than parasegmental, transformations: in the case of viable mutations of *en*, the posterior wing is transformed toward anterior wing (the same segment) and not to anterior A1 (the same parasegment) (see Lawrence and Morata, 1976; Struhl, 1982). One possibility is that the larva is initially built in parasegments but, once parasegments are subdivided, the adult body plan is then specified in compartmental or segmental units.

Finally, we do not wish to give the impression that homeobox genes necessarily have a single function and domain throughout development, for there are many instances to the contrary. *lab* is active in head development, but is reused later in the endoderm, where it specifies differentiation of a special type of gut cell (Hoppler and Bienz, 1994). The homeobox gene *ems* is an amusing example, for, in the young embryo, it acts to direct head development, and yet later on it works at the other end to specify part of the posterior spiracles or filzkörper (Jones and McGinnis, 1993).

Elaboration of Pattern

After the initial allocation to compartments, the cells proliferate and the parasegments become subdivided further with the help of different homeobox genes. At its simplest,

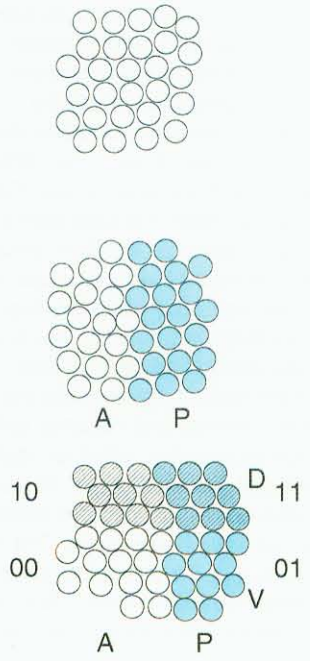

Figure 2. The Model of Compartition

A subset of a group of equivalent cells is allocated by positional cues and, as a result, a homeobox selector gene is activated in the posterior set (blue). Later, the process is repeated: another homeobox selector gene is activated in the dorsal sets (hatched circles), and the result is four cell identities specified by simple binary codes (adapted from Garcia-Bellido et al., 1979). Abbreviations: A, anterior; P, posterior; D, dorsal; V, ventral.

compartition of cells can occur serially, with, at each step, the subdivision of a set of cells into two, followed by the activation of a selector gene in one daughter set only, the difference between the determined state or cell identities in the two sets depending entirely on the selector gene being on in one and off in another. This makes a binary code that is extended at each compartition step (Figure 2; see also *apterous* [*ap*], which is discussed below).

Even though compartition has been most studied in the ectoderm, it probably affects the other germ layers, too. There are two types of mesoderm, the somatic that forms the muscles, heart, and fat body and the visceral, a sheath of mainly muscle cells that enwraps the gut endoderm. The visceral mesoderm may be a separate germ layer (Lawrence, 1992). The two mesoderms have different patterns of expression of the main homeobox genes; the pattern in the visceral mesoderm is simpler and nonoverlapping. There are two homeobox genes, curiously named *bagpipe* and *tinman*, that help segregate the two mesoderms; in the absence of *bagpipe*, the visceral mesoderm cells form somatic muscles (Aspiazu and Frasch, 1993). As in the ectoderm, there are specific sets of founder cells that form compartments in the mesoderm and there may be selector genes that determine the development of these. The same principles probably apply elsewhere: the *serpent* gene is needed for the determination of endoderm, the precursor of the midgut. In *serpent*$^-$, the prospective endodermal cells are transformed to ectoderm. The mo-

lecular structure of the gene is not yet known, nor is it known whether ectopic expression of *serpent* would specify adventitious endoderm, as would be expected if *serpent* were a true selector gene (Reuter, 1994).

The combination of selector genes also determines cell affinities; it has long been known that differently determined cells tend to sort out when mixed (Garcia-Bellido, 1966). In vivo this can mean that cells from adjacent compartments do not intermingle but form smooth and precisely positioned boundaries at the interface. For example, in the thorax, cells marked in early development (but not late) make clones that cross between the larval epidermis and the adult leg disc (Meise and Janning, 1993). These imaginal disc cells remain diploid while the surrounding larval cells cease mitosis and become polyploid; thus, there is a lineage segregation, albeit an atypical one, between the adult and the larval cells. At the same time, *Distal-less*, another homeobox gene, is switched on in the adult leg primordia and off elsewhere. The disc cells expressing *Distal-less* span the preexisting parasegment border (Cohen, 1990); the result is cells of four identities: disc anterior, disc posterior, larval anterior, and larval posterior.

Cell Autonomy

If it is true that cell identity is dependent on the homeobox genes and if this identity is to be propagated through cell divisions, it follows that compartments and domains of homeobox gene action must generally be coextensive. Suppose that daughter cells could acquire a differing cell identity from their mother cell, for example, that a cell expressing *Ubx* could divide and one daughter lose *Ubx* expression. That errant cell would then try to make thorax within the abdomen, which would not be a good thing. It is therefore essential for orderly development that the state of homeotic gene activity, once it has been definitively established, is faithfully propagated through cell lineage. It is for this reason that clones of cells mutant for homeotic genes (and therefore transformed to another cell identity) are cell autonomous, meaning they do not influence adjacent cells to adopt their own identity. It is also why, for example, *Ubx*-expressing cells do not (and presumably cannot) recruit nearby cells to turn on the *Ubx* gene. If this principle did not generally apply, cell identity and lineage compartments would fall out of register and errant cells would disturb development.

This conclusion raises a question: as some of the target genes of the selectors may be secreted proteins, could not the effect of selector gene expression spread into other cells? We think it can (an example is the effect of *Ubx* on the expression of *lab* in the endoderm); however, we believe that, in general, selector gene expression will not feed back nonautonomously on itself, for the reasons given above (but for a possible exception, see Thuringer and Bienz, 1993). It is relevant that, although the long-term culture experiments carried out by Hadorn and colleagues are best remembered for the occasional transdetermination that occurs, it is the extraordinary stability of the determined state that Hadorn himself often emphasized.

Propagation of Cell Identity

Once the founder cells of a compartment have been allocated, the selector (often a homeobox) gene is switched on in a process that has two steps, an initial step that activates the gene and then an independent step that maintains the state of the gene permanently, propagating it through cell divisions (Morata and Lawrence, 1977). Propagation is sometimes dependent on an autoregulatory loop in which the homeodomain protein binds to its own promoter (Schier and Gehring, 1992) but also relies on general activators such as *brahma* (Tamkun et al., 1992) and *trithorax* (Ingham, 1985; Breen and Harte, 1992). Where the gene products are not needed, they must be repressed, and this is the work of the *Polycomb* (*Pc*) group of genes (Lewis, 1978; Jürgens, 1985; see Paro, 1990). There are many genes in this group, and their products probably form a complex with the chromatin, whose state can be propagated through cell divisions. It is often thought that *Pc* and other genes in the group repress only members of the main homeotic cluster, but we wonder whether this is so. It has been shown that *en* is repressed by *Pc* (Busturia and Morata, 1988) and that some gap genes are also subject to members of the *Pc* group (Pelegri and Lehmann, 1994). Do the genes that are repressed by the *Pc* group have much in common? Do they include persistent but nonhomeobox transcription factors, such as *twist*?

Another unresolved question concerns oogenesis: genetic studies (and lampbrush chromosomes of amphibia) indicate that most genes are turned on during oogenesis, as if by default (see Lawrence, 1992). There may therefore have to be specific mechanisms during oogenesis to turn off homeobox selector genes, as well as others, whose products must be locally restricted during early development, products that would derail pattern formation if they were to stray (Lawrence et al., 1983).

Clustering and Colinearity in the Homeobox Gene Complexes

A hallmark of the primary homeotic genes is that in all animals they are clustered and their order is conserved, a conservation that would therefore appear to be important for function. Nevertheless, in Drosophila the clustering is not essential; for example, the complex can function if split (Struhl, 1984). *Ubx* can be put on another chromosome, where it works well, and there are other similar examples from both the BX-C and the ANT-C.

Why is it that the order of the genes on the chromosome is colinear with the patterns of expression in the body, with the most proximal gene on the chromosome having the most anterior role and the next gene along being expressed in the next-most anterior region? This rule is impressive, for it holds true from *lab* on the left side of the ANT-C to *Abd-B* on the right side of the BX-C; it also applies to vertebrates and nematodes and not only orders the protein coding parts of the genes, but also the disposition of the 3′ regulatory elements. This question we cannot answer; we merely do our duty by posing it yet again!

There is another aspect to colinearity discovered by Struhl (1983): when all the homeotic genes are derepressed, the larval cuticle pattern is principally determined by the more posteriorally acting protein present, as if the

effects of the more anterior genes were overruled. At first this was thought to be entirely due to transcriptional regulation, and indeed the more posterior gene products do repress transcription of the more anterior genes (Hafen et al., 1984; Struhl and White, 1985). However, there is a more powerful effect, and this is posttranscriptional: when homeodomain proteins are strongly and universally expressed under a heat shock promoter, they produce phenotypic changes, but only anterior to their normal domain (Gibson and Gehring, 1988; Gonzales-Reyes and Morata, 1990; Mann and Hogness, 1990). Posterior to their normal domains, even extra amounts of the protein do not alter the cuticle pattern—because other homeodomain proteins are expressed in these regions and they phenotypically suppress action of the ectopic protein (Gonzales-Reyes and Morata, 1990). Phenotypic suppression follows a hierarchical rule; the more posteriorly acting gene products override the more anterior ones. In vertebrates there is a similar phenomenon called posterior prevalence: when several homeobox genes are active, the more posterior gene is most effective (Duboule, 1991). When these same homeobox genes are expressed in Drosophila, the hierarchy still obtains (Bachiller et al., 1994). The mechanism is not known but is probably due to competition for binding sites on target genes. Although phenotypic suppression has been found in artificial circumstances, it tells us something important about normal development. One way of changing pattern during evolution would be to tinker with binding sites, to vary the impact of different homeodomain proteins on target genes. As new genes appeared by tandem duplication, their products could compete with preexisting homeodomain proteins and, in some places, could take over their domains completely. Therefore, when we now look at flies, it is not so surprising that, here and there, genes are expressed but have no function. One example is *proboscipedia*, a homeobox gene from the ANT-C that is strongly expressed in the embryo; however, null mutations of this gene have no embryonic phenotype (Kaufman et al., 1990).

Homeotic Target Genes

Given that homeodomain proteins bind to DNA and regulate other genes, it is clearly important to find whether they produce their effects on cell identity and pattern by binding to a small number of master genes that are at the upper end of a cascade or whether they bind to and modulate a large number of housekeeping genes.

The task of identifying target genes has proved difficult. One approach is to map where homeotic proteins bind to the polytene chromosomes, but it is hard to know which sites are significant and whether many are undetected. Another strategy is to fish in chromatin using antibodies; this has identified a few genes whose pattern of expression fits with expectations (Gould et al., 1990). However, it is disquieting that one direct binding site for Ubx protein turned out to be retroposon 412 (Brookman et al., 1992), of which there are many copies in the genome; it is unlikely that this element has anything to do with Ubx function. Another approach is to find enhancer trap patterns of expression that depend on a particular homeotic gene, such as *Antp* (Wagner-Bernholz et al., 1991). Using these and other methods, a number of target genes have been identified (see Botas, 1993).

One case of a target gene concerns the visceral mesoderm. Bienz and coworkers (Immergluck et al., 1990) discovered that *Ubx* is expressed in the visceral mesoderm in parasegment 7, where it is necessary for the expression of *decapentaplegic* (*dpp*) in the same parasegment. The Ubx protein may bind directly to *dpp* regulatory sequences to drive expression (Capovilla et al., 1994). The dpp protein is secreted and is needed for the normal expression of *lab* in the adjacent endoderm (Immergluck et al., 1990; Panganiban et al., 1990).

One important observation is that some, but not all, of the target genes so far identified have turned out to be transcription factors; examples are the homeobox genes *ems*, *Distal-less*, and *ap* and the zinc finger gene *spalt major* (see Botas, 1993). These genes must have their own target genes, favoring the idea that at least some of the primary homeotic genes act through a cascade, the first step of which might be acting on a lower or secondary rank of homeobox genes, which would come into operation later, perhaps as compartments are subdivided. Take *ap*, for example: *ap* is only activated in the dorsal imaginal discs of the T2 and T3 segments, so *ap* would be induced by *Antp*, but repressed by abdominal genes such as *abd-A*. As we have seen, the functions of *Ubx* and *ap* are equivalent in that both genes specify cell identity and differ only in their rank; *ap* functions later and within a subset of the cells of the *Ubx* domain. Another example is *ems*, which is activated by Abd-B; both homeoproteins then act together to specify the development of filzkörpers (Jones and McGinnis, 1993). Again, this suggests that the generation of subpatterns within the realms of action of the primary homeotic genes can depend on the action of secondary homeotic genes.

Cofactors

When analyzed in vitro, homeodomain proteins show little specificity in their DNA binding, which is odd since they have such specific effects on anatomy in vivo. Part of this specificity may be due to cofactors that modulate homeotic function. The best candidate is the homeobox protein extradenticle (exd), which alters the action of several homeotics, such as *Antp*, *Ubx*, *abd-A*, and some others in the head, without inactivating their products or affecting their patterns of expression. The exd protein may also interact with other homeodomain proteins (Peifer and Wieschaus, 1990). In exd^+ embryos, the Ubx protein specifies A1 pattern, whereas in exd^- embryos, the outcome is A3 pattern. This is puzzling since, normally, A3 identity is conferred by the abd-A protein. It would be novel and unsettling to suggest that the Ubx and abd-A proteins are very similar and do the same things in different places, and it is only in combination with the exd protein that they can act distinctly, but this is what the data suggest.

Another possible cofactor is the *teashirt* gene, which encodes a zinc finger protein and acts in the trunk (Roder

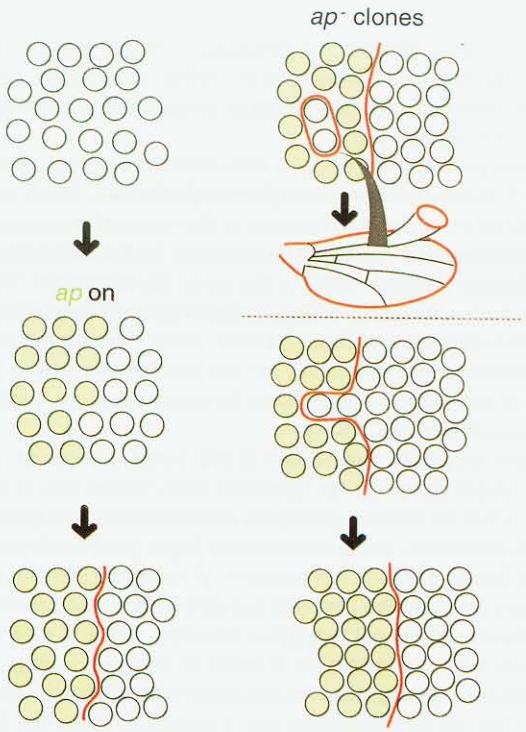

Figure 3. The *ap* Experiments

(Left) Normal development. The *ap* gene is first activated within the imaginal disc in the dorsally located cells (green). As a result, a compartment boundary, probably with structural and physiological attributes, is formed at the interface (red line).

(Right) Two examples of *ap* clones. The upper one is surrounded by *ap*-expressing dorsal cells, forms an entire boundary, and grows out as an extension of the wing. The lower clone is in contact with ventral cells that do not express *ap* and joins with them as the boundary straightens (probably owing to cell affinities, Garcia-Bellido, 1966), and normal development follows (adapted from Blair, 1993; Diaz-Benjumea and Cohen, 1993; Williams et al., 1993).

et al., 1992). The teashirt protein acts without much altering the expression of the homeotic genes; it also cooperates with Scr to make the distinctions between the labial and T1 segments. However, *teashirt* has an independent function; in the absence of the main homeotic genes, it transforms the segments of the ground pattern into a series of head segments.

Cell Identity and Pattern Formation

In the following section, we give examples of how homeobox genes work in the wing to make blocks of cells of differing identities, explain in outline how these blocks develop in different directions, and describe how the interfaces between them become compartment borders that can act as engines to drive detailed pattern formation (see Figures 3–5).

ap, *the Perfect Selector Gene?*

During larval growth, some of the discs are further subdivided, this time into dorsal and ventral, by a homeobox gene called *ap* (Blair, 1993; Diaz-Benjumea and Cohen, 1993). *ap* appears to be the long-lost selector gene for the dorsal and ventral wing compartments (Garcia-Bellido, 1975); it is expressed in the dorsal imaginal discs (wing and haltere) and is first activated in the second larval instar when it appears in some 100 cells, corresponding to about half the cells in each disc. A sharp boundary forms between expressing and nonexpressing cells (Williams et al., 1993). This boundary is exactly colinear with the dorsoventral compartment boundary, defined many years ago by experiments with clones of marked cells: if induced after the second instar, clones never cross over a line at the perimeter of the wing. This lineage boundary is associated with pattern; there are rows of bristles and other elements that develop along both sides of it like riverine vegetation in a desert.

If the *ap* gene is removed from a dorsal cell, it becomes a ventral one. If such a cell is near the dorsoventral compartment boundary, it becomes subsumed into the ventral group, but if further away, it forms an independent clone of cells of perfect ventral identity. Clones like this make interfaces with the cells surrounding them, that is, new boundaries between *ap*-expressing and *ap*-nonexpressing cells (Figure 3). Each interface alone is sufficient to induce a new border, including all the pattern, the bristles, and the gene expression associated with such a dorsoventral compartment boundary. There is evidence that cells of the compartment boundary control growth within the compartment (Lawrence and Morata, 1976); in *ap* flies, the wing is absent, presumably because there is no outgrowth without the dorsoventral boundary. Further, when an ectopic boundary encloses a clone of mutant, now-ventral cells in the dorsal compartment, the clone extends out from the wing surface like an extra winglet (Figure 3). The take-home message is a starkly simple one: the difference between the two classes of cells is entirely due to *ap*. Simply apposing two populations of cells of different identity can build a boundary that becomes a pattern element itself, a point of reference for further pattern formation and an orchestrator of growth (Blair, 1993; Diaz-Benjumea and Cohen, 1993). This same principle is taken further in our next example, *hedgehog* (*hh*).

Cell Identity and Pattern: hh

In the 1970s, although compartments had been discovered and their correlation with the realms of action of the homeotic selector genes noted, it was not at all clear how they are used to build intricate patterns. One conjecture was that gradients of positional information, possibly of a morphogen, would be established in register with the compartments, meaning that the gradient boundaries would coincide with the compartment boundaries (Crick and Lawrence, 1975). This would allow a single mechanism of positional information to be used in many compartments, the different responses to it arising from the unique identities of the cells in each compartment. The model of positional information, which derives from grafting experiments on other insects, makes the scalar and the vector at any point in the gradient landscape define both the type of differentiation and the polarity of the cells at that point (see Lawrence, 1992; Sampedro and Lawrence, 1993). The main virtue of a gradient hypothesis over others is that it offers a single mechanism that engenders both detailed pattern and cell polarity.

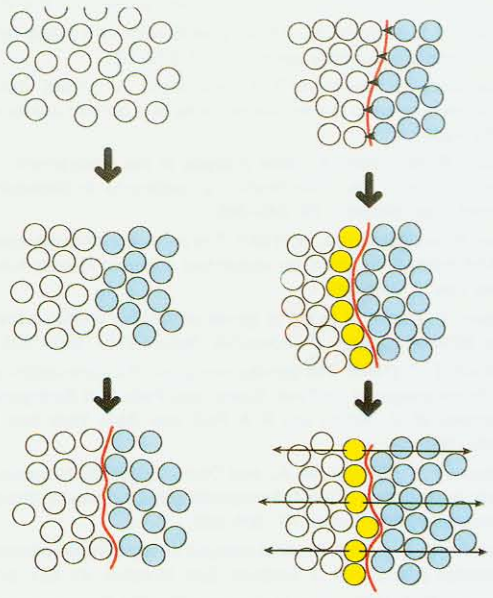

Figure 4. Model for Wing Disc Pattern Formation

The *en* gene (blue) specifies posterior identity; its absence specifies anterior. A boundary (red line) forms at the interface between anterior and posterior cells. The posterior cells then secrete the hh protein of short range (arrowheads), to which only cells of anterior identity can respond. This turns on the *dpp* gene in anterior cells near the boundary (yellow), and these act as a source of morphogen that has a longer range and provides detailed positional information to both anterior and posterior compartments. Adapted from work by Basler and Struhl (1994) and Tabata and Kornberg (1994).

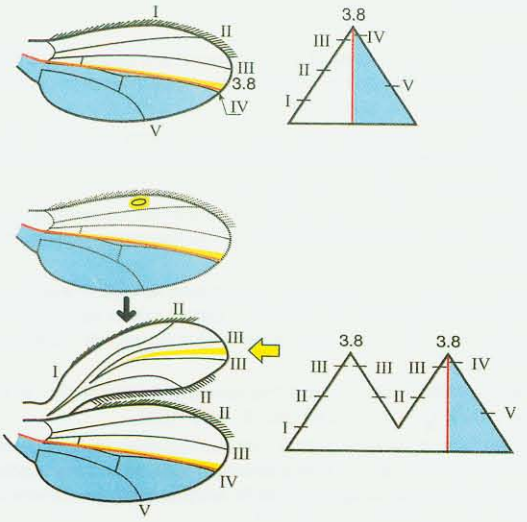

Figure 5. Result of Ectopic *hh* Expression in the Anterior Compartment

(Above) The wild-type wing, with the veins given their traditional numbers and the posterior compartment shown in blue. The yellow streak is the source of morphogen (possibly dpp protein) that, in this model (see Figure 4), emanates from anterior cells that are close to the anteroposterior compartment boundary (red line). The cells at that position are at the peak of a morphogen gradient that we have arbitrarily called 3.8. At the right, a cross section through the gradient is shown, the height (scalar) at each point specifying the pattern elements formed at that point and the slope (vector) specifying the local polarity of the cells. Cell division is limited by the steepness of the gradient (see Lawrence, 1992). The landscape is given a mirror-image symmetry because the *en* phenotype suggests it (Crick and Lawrence, 1975). In the posterior compartment, the cells, because they have a different identity, at least partly due to the homeobox gene *en*, respond to the same positional values differently from cells of anterior identity; for example, they make the posterior veins IV and V, rather than anterior veins II and III.

(Below) A clone of *hh*-expressing cells is induced in the anterior compartment of the developing wing (showed faintly). It becomes a new source of morphogen (yellow) that specifies the positional information scalar 3.8. The gradient landscape shown on the right organizes pattern and extra growth that produce a bifurcated and partially duplicated wing, as shown below. The yellow arrow points to the position where the *hh*-expressing cells were detected in the experiment. Adapted from Basler and Struhl (1994), with added speculations.

Several groups have been studying the function of the *hh* gene (Basler and Struhl, 1994; Tabata and Kornberg, 1994), and the results give insight into how gradients might be set up and registered with the compartment boundaries. The model is as follows: in the developing wing disc, there are four populations of cells (anterodorsal, anteroventral, posterodorsal, and posteroventral) whose identities are determined by a binary code involving the two homeobox genes *en* and *ap*. Directly or indirectly, the *en* gene ensures that all posterior cells express *hh*. Hh is a secreted protein that can be sensed by nonexpressing cells nearby. Cells of posterior identity do not respond to hh protein, because not responding to it is part of their nature, as specified by *en*; by contrast, all anterior cells can respond: it is an integral part of their anterior identity (Figure 4). The beauty of this simple mechanism is that it is only those anterior cells near the boundary that can react to *hh*; the number of cells responding will depend on the effective range of hh protein (this is unknown but is probably not more than a few cell diameters) (Basler and Struhl, 1994). The pattern is thereby elaborated, and now another step can follow: these newly specified cells turn on another gene, *dpp*, that encodes a secreted transforming growth factor β–like protein (Padgett et al., 1987). This may act as a longer-range morphogen, forming a mirror-image gradient with its source just anterior to the compartment border, a gradient that can pattern both the anterior and the posterior compartments (Figure 4). If an ectopic source of hh protein is made in the anterior wing, it induces a new gradient source of dpp in the surrounding cells. This causes major changes in the pattern and polarity of the wing, changes whose detailed features can be predicted by the gradient model, following the same rules sketched out above. This is explained in Figure 5.

Homologs of *hh* have been identified in vertebrates, and there is evidence from the limb bud that the protein may be organizing pattern there somewhat as it does in the Drosophila wing (Riddle et al., 1993).

Acknowledgments

We thank Julie Ahringer, Mariann Bienz, Ana Busturia, Mark Bretscher, Matthew Freeman, Ben Lewin, Jürg Müller, and Gary Struhl for commenting on this manuscript and for their advice. G. M. acknowledges financial support from the Dirección General de Investigación Cientifica y Téchnica and the Fundación Ramón Areces.

References

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

Bachiller, D., Macias, A., Duboule, D., and Morata, G. (1994). Conservation of a functional hierarchy between mammalian and insect *Hox/HOM* genes. EMBO J. *13*, 1930–1941.

Basler, K., and Struhl, G. (1994). Compartment boundaries and the control of *Drosophila* limb pattern by the hedgehog protein. Nature *368*, 208–215.

Blair, S. (1993). Mechanisms of compartment formation: evidence that non-proliferating cells do not play a critical role in defining the D/V lineage restriction in the developing wing of *Drosophila*. Development *119*, 339–351.

Botas, J. (1993). Control of morphogenesis and differentiation by *HOM/Hox* genes. Curr. Opin. Cell Biol. *5*, 1015–1022.

Breen, T., and Harte, P. (1992). *trithorax* regulates multiple homeotic genes in the *bithorax* and *Antennapedia* complexes and exerts different tissue-specific, parasegment-specific and promoter-specific effects on each. Development *117*, 119–134.

Brookman, J., Toosy, A., Shashidhara, L., and White, R. (1992). The 412 transposon and the development of gonadal mesoderm in *Drosophila*. Development *116*, 1185–1192.

Busturia, A., and Morata, G. (1988). Ectopic expression of homeotic genes caused by the elimination of the *Polycomb* gene in *Drosophila* imaginal epidermis. Development *104*, 713–720.

Capovilla, M., Brandt, M., and Botas, J. (1994). Direct regulation of *decapentaplegic* by *Ultrabithorax* and its role in Drosophila midgut morphogenesis. Cell *76*, 461–475.

Cohen, S. (1990). Specification of limb development in the *Drosophila* embryo by positional cues from segmentation genes. Nature *343*, 173–177.

Cohen, S., and Jürgens, G. (1991). *Drosophila* headlines. Trends Genet. *7*, 267–272.

Crick, F. H. C., and Lawrence, P. A. (1975). Compartments and polyclones in *Drosophila* development. Science *189*, 340–347.

Desplan, C., Theis, J., and O'Farrell, P. H. (1985). The *Drosophila* developmental gene *engrailed* encodes a sequence specific DNA binding activity. Nature *318*, 630–635.

Diaz-Benjumea, F., and Cohen, S. (1993). Interaction between dorsal and ventral cells in the imaginal disc direct wing development in Drosophila. Cell *75*, 741–752.

Duboule, D. (1991). Patterning in the vertebrate limb. Curr. Opin. Genet. Dev. *1*, 211–216.

Duncan, I. (1987). The *bithorax* complex. Annu. Rev. Genet. *102*, 49–70.

Garcia-Bellido, A. (1966). Pattern reconstruction by dissociated imaginal disc cells of *Drosophila melanogaster*. Dev. Biol. *14*, 278–306.

Garcia-Bellido, A. (1975). Genetic control of wing disk development in *Drosophila*. Ciba Foundation Symp. *29*, 161–182.

Garcia-Bellido, A., Ripoll, P., and Morata, G. (1973). Developmental compartmentalization of the wing disc of *Drosophila*. Nature *245*, 251–253.

Garcia-Bellido, A., Lawrence, P. A., and Morata, G. (1979). Compartments in animal development. Sci. Am. *241*, 102–110.

Gibson, G., and Gehring, W. J. (1988). Head and thoracic transformations caused by ectopic expression of *Antennapedia* during *Drosophila* development. Development *102*, 657–675.

Gonzales-Reyes, A., and Morata, G. (1990). The developmental effect of overexpressing a Ubx product in Drosophila embryos is dependent on its interactions with other homeotic products. Cell *61*, 515–522.

Gould, A., Brookman, J., Strutt, D., and White, R. (1990). Targets of homeotic gene control in *Drosophila*. Nature *348*, 308–312.

Hafen, E., Levine, M., and Gehring, W. (1984). Regulation of *Antennapedia* transcripts by the *bithorax* complex of *Drosophila*. Nature *307*, 287–289.

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

Hoppler, S., and Bienz, M. (1994). Specification of a single cell type by a Drosophila homeotic gene. Cell *76*, 689–702.

Immergluck, K., Lawrence, P. A., and Bienz, M. (1990). Induction across germ layers in Drosophila mediated by a genetic cascade. Cell *62*, 261–268.

Ingham, P. W. (1985). A clonal analysis of the requirement for the *trithorax* gene in the diversification of segments in *Drosophila*. J. Embryol. Exp. Morphol. *89*, 349–365.

Jones, B., and McGinnis, W. (1993). The regulation of *empty spiracles* by *Abdominal-B* mediates an abdominal segment identity function. Genes Dev. *7*, 229–240.

Jürgens, G. (1985). A group of genes controlling spatial expression of the *bithorax* complex of *Drosophila*. Nature *316*, 1533–1535.

Kaufman, T. C. (1983). The genetic regulation of segmentation in *Drosophila melanogaster*. In Time, Space, and Pattern in Embryonic Development, W. R. Jeffrey and R. A. Raff, eds. (New York: Alan Liss), pp. 365–383.

Kaufman, T. C., Seeger, M. A., and Olsen, G. (1990). Molecular and genetic organization of the *Antennapedia* gene complex of *Drosophila melanogaster*. Adv. Genet. *27*, 309–362.

Lawrence, P. A. (1973). A clonal analysis of segment development in *Oncopeltus* (*Hemiptera*). J. Embryol. Exp. Morphol. *30*, 681–691.

Lawrence, P. A. (1992). The Making of a Fly: The Genetics of Animal Design (Oxford: Blackwell Scientific Publications).

Lawrence, P. A., and Morata, G. (1976). Compartments in the wing disc of *Drosophila*: a study with the *engrailed* gene. Dev. Biol. *50*, 321–337.

Lawrence, P. A., Johnston, P., and Struhl, G. (1983). Different requirements for homeotic genes in the soma and germ line of Drosophila. Cell *35*, 27–34.

Levine, M., and Hoey, T. (1988). Homeobox proteins as sequence specific transcription factors. Cell *55*, 537–540.

Lewis, E. B. (1963). Genes and developmental pathways. Am. Zool. *3*, 33–56.

Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. Nature *276*, 565–570.

Mann, R., and Hogness, D. S. (1990). Functional dissection of Ultrabithorax proteins in Drosophila melanogaster. Cell *60*, 597–610.

Martinez-Arias, A., and Lawrence, P. A. (1985). Parasegments and compartments in the *Drosophila* embryo. Nature *313*, 639–642.

McGinnis, W., and Krumlauf, R. (1992). Homeobox genes and axial patterning. Cell *68*, 283–302.

McGinnis, W., Levine, M., Hafen, E., Kuroiwa, A., and Gehring, W. J. (1984). A conserved DNA sequence in homeotic genes of the *Drosophila Antennapedia* and *bithorax* complexes. Nature *308*, 428–433.

Meise, M., and Janning, W. (1993). Cell lineage of larval and imaginal thoracic anlagen cells of *Drosophila melanogaster*, as revealed by single cell transplantations. Development *118*, 1107–1120.

Morata, G., and Garcia-Bellido, A. (1976). Developmental analysis of some mutants of the *bithorax* system of *Drosophila*. Roux's Arch. Dev. Biol. *179*, 125–143.

Morata, G., and Lawrence, P. A. (1977). Homeotic genes, compartments and cell determination in *Drosophila*. Nature *265*, 211–216.

Müller, J., and Bienz, M. (1992). Sharp anterior boundary of homeotic gene expression conferred by the *fushi tarazu* protein. EMBO J. *11*, 3653–3661.

Nüsslein-Volhard, C. (1991). Determination of the embryonic axes of *Drosophila*. Development *113* (Suppl. 1), 1–10.

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 gene family. Nature *325*, 81–84.

Panganiban, G., Reuter, R., Scott, M., and Hoffmann, F. (1990). A *Drosophila* growth factor homolog, *decapentaplegic*, regulates homeotic gene expression within and across germ layers during midgut morphogenesis. Development *110*, 1041–1050.

Paro, R. (1990). Imprinting a determined state into the chromatin of *Drosophila*. Trends Genet. *6*, 416–421.

Peifer, M., and Wieschaus, E. (1990). Mutations in the *Drosophila* gene *extradenticle* affect the way specific homeo domain proteins regulate segmental identity. Genes Dev. *4*, 1209–1223.

Pelegri, F., and Lehmann, R. (1994). A role of *Polycomb* group genes in the regulation of gap gene expression in *Drosophila*. Genetics *136*, 1341–1353.

Reuter, R. (1994). The gene *serpent* has homeotic properties and specifies endoderm versus ectoderm within the *Drosophila* gut. Development *120*, 1123–1135.

Riddle, R., Johnson, R., Laufer, E., and Tabin, C. (1993). *Sonic hedgehog* mediates the polarizing activity of the ZPA. Cell *75*, 1401–1416.

Roder, L., Vola, C., and Kerridge, S. (1992). The role of *teashirt* in trunk segmental identity in *Drosophila*. Development *115*, 1017–1033.

Sampedro, J., and Lawrence, P. A. (1993). *Drosophila* development: after the first three hours. Development *119*, 971–976.

Sanchez-Herrero, E., Vernos, I., Marco, R., and Morata, G. (1985). Genetic organization of *Drosophila bithorax* complex. Nature *313*, 108–113.

Schier, A. F., and Gehring, W. J. (1992). Direct homeodomain–DNA interaction in the regulation of the *fushi tarazu* gene. Nature *356*, 804–807.

Scott, M. P., and Weiner, A. J. (1984). Structural relationships among genes that control development: sequence homology between the *Antennapedia*, *Ultrabithorax* and *fushi tarazu* loci of *Drosophila*. Proc. Natl. Acad. Sci. USA *81*, 4115–4119.

Struhl, G. (1982). Genes controlling segmental specification in the *Drosophila* thorax. Proc. Natl. Acad. Sci. USA *79*, 7380–7384.

Struhl, G. (1983). Role of *esc*$^{+}$ gene in ensuring the selective expression of segment specific homeotic genes in *Drosophila*. J. Embryol. Exp. Morphol. *76*, 297–331.

Struhl, G. (1984). Splitting the *bithorax* complex of *Drosophila*. Nature *308*, 454–457.

Struhl, G., and White, R. (1985). Regulation of the *Ultrabithorax* gene by other *bithorax* complex genes. Cell *43*, 507–519.

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

Tabata, T., and Kornberg, T. (1994). Hedgehog is a signaling protein with a key role in patterning Drosophila imaginal discs. Cell *76*, 89–102.

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C., and Kennison, J. A. (1992). *brahma*: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SW12. Cell *68*, 561–572.

Thali, M., Müller, M. M., DeLorenzi, M., Mathias, P., and Bienz, M. (1988). *Drosophila* homeotic genes encode transcriptional activators similar to OTF-2. Nature *336*, 598–600.

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

Vincent, J. P., and O'Farrell, P. H. (1992). The state of *engrailed* expression is not clonally transmitted during early Drosophila development. Cell *68*, 923–931.

Wagner-Bernholz, J., Wilson, C., Gibson, G., Schuh, R., and Gehring, W. (1991). Identification of target genes of the homeotic gene *Antennapedia* by enhancer detection. Genes Dev. *5*, 2467–2480.

Williams, J. A., Paddock, S. W., and Carroll, S. B. (1993). Pattern formation in a secondary field: a hierarchy of regulatory genes subdivides the developing *Drosophila* wing disc into discrete subregions. Development *117*, 571–584.