Many forms of fish, bird, and beast
Brought forth an Infant form
Where was a worm before

Blake
First Book of Urizen

The intimacy of evolution and genetics is beautifully illustrated by homoeotic mutants, mutants in which one part of an organism is transformed into a semblance of another part.

An example of a homoeotic mutant is the fruit fly (Drosophila) shown in Fig. 1 in which the antennae are transformed into legs. Such mutations remind us that the organization of pattern can be altered by relatively minor genetic changes, and that small changes in regulatory functions may be more important in the evolution of multicellular organisms than small changes in structural components. Homoeotic mutations are thought to identify 'master switch' genes that control the selection of developmental pathways. An important point is that pattern, and not merely the synthesis of specialized products, is regulated. The structural components of an arm are probably nearly identical to those of a leg; it is the organization of the two that differs.

Drosophila have bodies with about five head segments, three thoracic segments (each with a pair of legs) and eight abdominal segments. The segments can be seen at the earliest stages of differentiation in the embryo and also in the larvae and in adults. The homoeotic genes affect the identities of body segments; other classes of genes affect the number of seg-

Molecules and puzzles from the antennapedia homolectic gene complex of Drosophila

Matthew P. Scott

How are loci that control the number of body segments related to loci that regulate segmental identity? In the Antennapedia complex of Drosophila, the two kinds of loci are close together, and molecular analysis has revealed shared and distinguishing features.

© 1985, Elsevier Science Publishers B.V., Amsterdam 0168-9525/85/$02.00
The Antennapedia homoeotic gene complex

My purpose here is to summarize what has been learned about the molecular organization of the ANT-C. The genetic aspects of the ANT-C are reviewed in Ref. 7 and the developmental effects of ANT-C mutations upon embryos are described in Ref. 8. Of the ANT-C loci, Antp (Antennapedia, a homoeotic locus in which dominant mutations transform antennae into legs, Fig. 1, and loss-of-function mutations transform legs into antennae) and fze (fushi tarazu, 'segment deficient', a segmentation locus active in the zygote, recessive mutation of which results in embryos with half the normal number of body segments) have been studied most extensively by molecular techniques. While these two loci will be the focus of this discussion, a few points about the ANT-C as a whole are worth mentioning. Loci that are normally active in the head (e.g. pb, proboscipedia, in which recessive mutations turn the labial palps into legs or antennae) tend to be at the left end of the complex, while loci such as Ser (Sex combs reduced, in which mutations turn first legs into second legs) and loss-of-function mutations in Antp that normally act in the thorax are at the right-hand end of the ANT-C as it is usually represented. In the BX-C most of the genes are in the order of the body segments they affect, so the order of the genes in both complexes may be important. The ANT-C is located in two fairly heavily stained polytene bands, 84A4,5 and 84B1,2, on the right arm of chromosome 3 (Fig. 2); the proximity of the ANT-C to the centromere results in low frequencies of recombination. In the ANT-C, a rough estimate of recombination frequency is 0.001% per 1000 base pairs. This is in contrast to, for example, the white locus near the tip of the X chromosome, where the frequency is about 0.002% per 1000 base pairs. It appears that the size of the ANT-C will be roughly 300 kilobase pairs (kb), which is similar to the length of the sum of the thoracic and abdominal regions of the BX-C.

The Antp locus: a complex locus within the complex

The Antp locus inhabits polytene band 84B1,2, and spans about 103 kb of DNA (Fig. 3). A major unsolved question is why Antp needs so much DNA - the finished transcripts are only 3.5 and 5.0 kb long. The details of transcription have not yet been worked out but at least five exons have been identified through their homologies with cDNA clones. The exons are distributed as shown in Fig. 3. Since the exons at +100, +112 and +200 all hybridize to both of the major transcripts, it is likely that differential RNA processing gives rise to the different RNA species. However, the possibility that DNA rearrangements also occur has not been ruled out. Two different types of cDNA clones have been described; one type hybridizes to the exons at +100, +112, +170 and +200 while the second type hybridizes to the same two exons at +100 and +112 and to an exon at +130, but not to the exons at +170 and +200. Since the cDNA libraries used were prepared from total embryonic RNA including nuclear RNA, it is possible that some cDNA clones represent processing intermediates rather than finished RNA species.

The molecular structure of Antp suggests that there is room for genetic complexity and the diverse array of different mutations confirms this expectation (Table 1). Different dominant mutations may turn antennae into legs (Antp), dorsal head into dorsal thorax (AntpDorsal), or second and third legs into first legs (AntpLegs). The dominant mutations cause abnormal activity of the Antp locus and are therefore referred to as 'neomorphic' or 'gain of function' alleles. All of the dominant mutations are related because they are each recessive lethals as well and they fail to complement each other. Thus, the abnormal activity of the dominant alleles is accompanied by the loss of their ability (in most cases) to provide the normal Antp functions. A single functional Antp allele is sufficient for normal development; Antp normally functions in the thorax and perhaps the abdomen, but not in the head. Therefore, at least some of the dominant alleles apparently cause Antp to be active in the head where the locus would normally be silent.

---

Fig. 1. A fly of the genotype Antp/CxD. The antennae are transformed into legs.
As if the dominant alleles are not confusing enough, the alleles that are completely recessive are actually the source of the genetic observations that lead to the formal designation of *Antp* as 'complex'. Certain recessive lethal alleles exhibit a complementation pattern as follows: allele A fails to complement B or C, while B complements C (Ref. 7). Therefore, it seems that type B or C mutations affect only a subset of the functions of *Antp*, while a type A allele affects all of the functions. A molecular analogue, which may or may not apply to *Antp*, is that a type A mutation would be in an exon that is included in all *Antp* products, while a type B or C mutation would be in an exon that is incorporated into only a subset of the *Antp* products.

The task of mapping the *Antp* locus onto the chromosome walk was made easier by the fact that most of the dominant *Antp* alleles are associated with cytologically visible chromosome rearrangements. The breakpoints are distributed over a region of about 60 kb, from +199 near the 5’ exon to about +140 (Refs. 9, 11). Therefore, abnormal gene activity results from mutations that split the locus within the transcription unit. The dominant mutations are fairly rare, perhaps because only breakpoints in precisely defined places within *Antp* can lead to a dominant effect. Alternatively, perhaps certain kinds of DNA sequences must be brought into juxtaposition with the *Antp* DNA for a dominant phenotype to result.

The dominant mutations that are associated with chromosome rearrangements may offer clues to how *Antp* functions. Is it the 5’ part of the split transcription unit or the 3’ part, or the combination of the two, that is responsible for the mutant phenotype? This question has been answered for one allele, *AntpB*, in an elegant genetic experiment. The *AntpB* allele, which causes the antennae to be transformed into legs, is associated with a small cytologically visible inversion with one breakpoint within *Antp* at +140 (Ref. 9). About three-fifths of *Antp* is therefore displaced away from the remaining 3’ 40 kb. Kaufman et al. isolated *Drosophila* in which the two ends of the *Antp* inversion (and therefore the 5’ and 3’ parts of *Antp*) had segregated from each other following recombination within the inverted region. The result

---

**Fig. 2.** A drawing of the right arm of polytene chromosome three in the vicinity of the Antennapedia Complex (ANT-C). The complementation groups in and near the complex are shown below and the open boxes indicate deficiency mutations that were used in ordering the loci. The genetic analyses are the work of Thomas Kaufman and his colleagues at Indiana University and the figure was generously provided by T. Kaufman. The part of the chromosome that has been cloned as a series of overlapping pieces in phage vectors is indicated above the chromosome. Symbols for genetic loci: Tpl = triple-lethal, aT = alpha tubulin gene, roe = roughened eye, tn = rotund, dsex = doublesex, twr = twisted bristles roughened eye, cel = cell lethal, pb = proboscipedia, mum = multimorph, zen = zerknullt, Dfd = Deformed, Scr = Sex comb reduced, ftz = fushi tarazu, Antp = Antennapedia, elb = embryo-larval boundary lethal, dco = discs overgrown, lpl = larval pupal lethal, mab = malformed abdomen, ecl = eclosion defective. Scr lesions are located on the physical map on both sides of ftz, hence its placement in two positions. pb, Dfd, Scr and Antp are homeotic loci, ftz is a pair rule segmentation locus and several of the other loci affect early development.
Table 1. Mutations affecting the Antennapedia and ftz loci of Drosophila

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antp&lt;sup&gt;B&lt;/sup&gt;</td>
<td>antennae → second thoracic segment: legs</td>
</tr>
<tr>
<td>Antp&lt;sup&gt;Ca&lt;/sup&gt;</td>
<td>dorsal head → dorsal thorax (sometimes eye → wing)</td>
</tr>
<tr>
<td>Antp&lt;sup&gt;Ca&lt;/sup&gt;</td>
<td>second and third legs → first legs</td>
</tr>
</tbody>
</table>

Deficiency (Df) in Antp

- Hemizygous: (Df<sup>+</sup>) no phenotype
- Homozygous: lethal, second and third thoracic segments → first thoracic segments

Recessive Antp<sup>-</sup> alleles

- Hemizygous (Antp<sup>-/+</sup>): no phenotype
- Homozygous: lethal, same as Antp deficiency. However, some alleles complement to give viable wild-type flies

Deficiency for ftz

- Hemizygous (Df<sup>+</sup>): no phenotype
- Homozygous: lethal, half the usual number of segments form

Dominant effect: posterior balancer organs, the halteres (vestigial wings on the third thoracic segment) transformed into posterior wing tissue

- Hemizygous (ftz<sup>R3</sup>/Df): lethal, some segmental boundaries missing but not as many as with a ftz null allele (Ref. 12)

was that the dominant phenotype of antenna to leg conversion was associated with the 3′ portion of Antp<sup>B</sup>. Therefore, the 3′ 40 kb of Antp, juxtaposed to novel chromosomal sequences, is sufficient to transform the antennae. No phenotype was associated with the 5′ portion of the disrupted gene.

A simple model for what could be happening is that the dominant Antp alleles are gene fusions in which a 3′ part of Antp capable of directing the formation of thoracic structures comes under the control of a promoter or other regulatory element that normally causes gene expression in the head (or anywhere). Abnormal expression of the 3′ part of Antp in the head (apparently 40 kb is enough) would result in the observed antenna-to-leg transformation. This model predicts that novel RNA and perhaps protein species are encoded by the dominant alleles, and that different alleles would encode different products. It also predicts that Antp dominant alleles will be expressed in the head, where the wild-type gene is inactive.

The (extra sex combs) Sc<sup>x</sup> alleles of Antp, which transform second and third legs into first legs, are diverse in molecular structure. The Sc<sup>x</sup> phenotype is reminiscent of the embryonic phenotype of Antp recessive lethal mutants. The homozygous Antp<sup>-</sup> mutants die as embryos but exhibit thoracic segmental transformations (T2 → T1, occasionally also T3 → T1) that are analogous to the Sc<sup>x</sup> dominant effects.<sup>3</sup> Therefore, the abnormal activity of Sc<sup>x</sup> alleles<sup>17</sup> results in a phenotype that mimics the loss of Antp function. The dominant alleles may be antimorphic, that is, they may interfere in a non-lethal way with the activity of the wild-type Antp allele that is present on the other homologue, giving rise to a phenotype resembling partial loss of Antp function. One of the two Sc<sup>x</sup> alleles, Sc<sup>x<sup>W</sup></sup>, is associated with a 50 kb inversion from +85 to +135, so that the 3′ part of Antp<sup>B</sup> is inverted.<sup>6</sup> The only change yet detected in the DNA of the original Sc<sup>x</sup> allele is an insertion of about 3 kb just downstream of the 5′ exon<sup>6</sup>. How the molecular structure of Sc<sup>x</sup> and Sc<sup>x<sup>W</sup></sup> relate to the phenotype to or antimorphic effects is not yet clear, although it is possible that in Sc<sup>x<sup>W</sup></sup> Drosophila the 50 kb inversion results in the production of antisense Antp RNA, which could interfere with some function of the normal RNA.

Transcription of a 103 kb gene probably takes about one hour<sup>18</sup>. One might predict that genes expressed very early in development would require shorter transcription time and have a shorter length. An example consistent with this idea is the fushi tarazu (ftz) gene<sup>2,19</sup>, an ANT-C gene (Fig. 3) which has a 1.95 kb transcription unit and a single 150 bp intron. The ftz gene functions only at the blastoderm stage, as has been demonstrated by mosaic analysis<sup>4</sup>, by the temperature-sensitive period of a temperature-sensitive allele<sup>8</sup>, by Northern blots of ftz RNA<sup>12,19</sup> and most dramatically by in situ hybridization of ftz probes to frozen sections of embryos at different stages of development<sup>14</sup>. It will be interesting to learn whether other genes that need to be expressed at an early stage also tend to have short transcription units.

**ftz: a segmentation locus among the homoeotic loci in the ANT-C**

It is now common to refer to genes that govern the number and polarity of segments as 'segmentation genes' and to distinguish them from the homoeotic genes that regulate segment identity. Neither class of gene is 'tissue specific', since genes in each class affect, for example, both neural tissue and epidermis.

The ftz segmentation locus was discovered during work designed to saturate the ANT-C with recessive lethal mutations.<sup>20</sup> The phenotype of ftz is that half the usual number of segments develop, a lethal event in
homoyzgous embryos. This places ftz in the 'pair-rule' class of segmentation mutants, a class which also includes genes such as paired, even-skipped, odd-skipped and so forth. The names convey the effects of mutations in the pair-rule loci on the segmentation pattern. Since segmentation genes other than ftz are not (so far as is known) in clusters of homoeotic genes, ftz seems remarkable in being 30 kb downstream from Antp (Fig. 3). Does this mean that ftz is a closet homoeotic? One mutant allele of ftz, ftzRp1, found by Ian Duncan, has a recessive phenotype in which fewer than half the segmental boundaries are missing, suggesting reduced but not eliminated ftz function. The unusual attribute of ftzRp1 is that it also has a dominant effect: the posterior halteres are transformed into posterior wing, and there are slight changes in the adult abdominal segmentation as well. The phenotype is reminiscent of the (postbithorax) pbx phenotype of the bithorax complex, but pbx is recessive. The ftzRp1 mutation shows that an abnormal ftz allele is capable of inducing a homoeotic transformation, but this does not mean that ftz is a normal homoeotic gene. Rather, the phenomenon may provide us with a first example of an interactive link between the homoeotic loci and the segmentation loci.

Because ftz is a small locus, more is known about the structure of the gene and its products than is known about the giant Antp and Ubx loci. The gene was located by mapping an insertion mutation (a 4.9 kb piece of DNA was inserted) and a translocation breakpoint (associated with ftzRp1). The gene has been sequenced and the primary structure of the encoded protein predicted. The RNA splicing pattern was deduced by comparing the sequence of a cDNA clone to the genomic sequence. The encoded protein, 413 amino acids long, has an unusual composition. The amino terminal part (about 250 amino acids) is more than 10% proline and 10% tyrosine and ends at the point where the single intron interrupts the gene. The next part of the protein, a domain of 60 amino acids, is about 30% basic amino acids and is not rich in either proline or tyrosine. The C-terminal part is again 10% proline and 10% tyrosine and is also about 15% glutamine. The ftzRp1 translocation results in replacement of the C-terminal part with a few novel amino acids, so that a truncated protein is encoded. Since the mutant phenotype of ftzRp1/ftz- embryos is less extreme than that of ftz/ftz- embryos, the truncated protein probably does have partial function.

In situ hybridization to tissues

The availability of molecular clones had made it possible to observe the distribution of ftz and Antp transcripts using in situ hybridization to sectioned tissue. The technique reveals the accumulation of transcripts and therefore reflects the balance between synthesis and turnover. The results for ftz were especially striking. The body segments first appear in the embryo about 5 hours of development, but ftz, which regulates the segmental pattern, is first active at the blastoderm stage 3 hours earlier. The blastoderm-stage embryo is initially a syncytium of nuclei lying along the surface of the embryo. Later, cell membranes form around the nuclei to give a monolayer of cells that appear to be homogeneous (except for the germ line or 'pole' cells which look different). In situ hybridization revealed that ftz transcripts are present even before the cells have formed and that they are localized in stripes around the embryo. The stripes appear to correspond to the primordia of alternate segments of the embryo. Both ectodermal and mesodermal progenitor cells contain ftz transcripts. On the basis of these results, it appears that the nuclei in syncytial blastoderm-stage embryos become aware of their spatial arrangement and a subset of them respond by synthesizing ftz products. An interesting subtlety is that initially transcripts are not arranged in well-defined stripes, but are instead more uniform. The striped pattern appears subsequently. The gradual emergence of the final pattern suggests that regulatory interactions occur at the initial time of ftz expression. Presumably the striped pattern of ftz transcripts, and quite likely similar patterns of expression of other segmentation genes, is the way in which the segmental differentiation process begins. More precise segmental identities would arise in response to the differential expression of homoeotic genes such as Antp.

Antp transcripts are found in embryos, larvae and pupae. The segmental distribution of the RNAs corresponds to the segments that genetic studies show require Antp function. It is not yet known whether the two major Antp transcripts differ in their spatial distribution. They do vary in relative amount during development. An unexpected finding is that Antp transcripts, like Ubx transcripts, are most abundant in the neural tissue. The neural ganglia are segmented, so the activity of Antp in neural cells is not surprising but why the highest concentration of RNA is in neural cells is not known. Antp transcripts are found early in development in abdominal segments but the abdominal segments do not have detectable Antp RNA at later stages. As in the case of ftz, the location of the transcripts becomes progressively more restricted. The absence of BX-C function, which transforms the abdominal segments into thoracic segments, also results in stable and abundant accumulation of Antp transcripts in the transformed abdominal segments. Thus, BX-C gene activity is capable of affecting the location of Antp transcripts.

The major finding from the in situ hybridization studies is that the accumulation of RNA encoded by a regulatory gene is limited to tissues in which the gene has been shown to function. An alternative would have been, for example, that a homoeotic gene is expressed everywhere but its activity is restricted by the presence or absence of modulating cofactors.

The homoeodomain

An early surprise was the finding that an Antp cDNA clone hybridized to the ftz locus. Further investigation revealed cross-hybridization between ftz protein-coding DNA and DNA within the 3' exons of Antp and Ubx (Fig. 2). DNA sequencing revealed that the 60 amino acid middle-domain of the ftz protein is nearly identical to a predicted protein sequence encoded in the 3' exons of Ubx and Antp. In addition, similar cross-hybridizing sequences are found elsewhere in the ANT-C and BX-C, including positions that probably correspond to the Deformed gene of the ANT-C and the infraabdominal-2 gene of
the BX-C. The 60 amino acid protein sequence is commonly referred to as the homoeodomain; the DNA sequence as the homeobox. The ftz homoeodomain is shown in Fig. 4. Whether all the genes that contain the sequence are homeotic (ts/ftz) is an unsettled issue. Also, only some of the homeotic genes appear to encode a homoeodomain, although this could be a result of our inability to detect poorly-conserved coding sequences by DNA hybridization. About seven homoeodomain coding sequences have been found in D. melanogaster — fewer than the number of homeotic loci (at least 15). The homoeodomain is also encoded by genes found in higher vertebrates, including humans. Two copies of the coding sequence have been found to be within 3 Vb of each other in the human genome, suggesting that development-regulating gene clusters also exist in humans.

The homoeodomain is highly basic, which suggests that it may be employed in RNA or DNA interactions. More specific support for this idea comes from the homology of part of the homoeodomain with the DNA binding domains of certain bacterial proteins. Three bacterial DNA-binding proteins have been crystallized and in each case the part of the protein that contacts the DNA is composed of two α-helical sections with a sharp turn between them. Certain amino acids (or types of amino acid) are required in particular framework positions in the structure. For example, small residues such as glycine are needed in the tight turn between the helices. Non-framework residues are involved in DNA sequence recognition and therefore differ in proteins of different function. The homology between a 20 amino acid part of the homoeodomain and the bacterial helix-turn-helix structure is in the framework residues.

Since there is quite a lot of variation between the putative DNA binding domains of different bacterial proteins, it is hard to say whether the homoeodomain (and yeast mating-type) homologies with the helix-turn-helix structure are significant. A clue that shows one of the critical framework amino acids to be important for ftz function comes from a temperature-sensitive (ts) allele of ftz. In the ts allele, an alanine that is one of the most conserved amino acids in the proposed helix-turn-helix structure is changed into a valine. The structural homology and the ts allele suggest that the possibility of the homoeodomain being a DNA binding sequence is worth investigating, as are alternatives such as an RNA-binding function. The recent observation of Ubx-encoded protein(s) in the nuclei of imaginal discs and embryonic neural cells is consistent with DNA (or RNA) binding function for homeoetic gene products. DNA binding activity is one way that the homoeotic proteins could influence the expressions of other genes. However, the central question (really a cell biology problem) is how gene activity results in the spatial localization of structures. A DNA-binding model is a long way from explaining how bristles are always found in the same places.

**Puzzles and prospects**

E. Lewis proposed that the different homeotic genes of the BX-C evolved by a process of duplication and divergence. As yet, the homoeodomain is the only molecular evidence that different homeotic genes are related structurally. It is increasingly clear, however, that the homoeotic genes are part of an interacting network of genes that together create the normal pattern of development. Such regulatory genes are plausible targets of mutations that lead to new patterns and, eventually, to new species. Genetic analysis is the key to identifying the components of the regulatory gene hierarchy. Molecular analysis supplies the fine details of structure and sensitive probes for gene products. With the tools that are currently being prepared, it will soon be possible to study the molecular mechanisms through which the regulatory genes exert their influence. Both genetic and molecular techniques will help to reveal how various genes interact. We need to know why genes are clustered (history or function?), why genes as different as Antp and ftz encode proteins that share a domain, the reason for giant transcription units, how the position-specific differences in homeolectic gene expression are controlled and how these differences lead to a choice of a differentiation pathway. The extraordinary evolutionary conservation of the homoeodomain suggests that at least some of the answers have ancient roots and may have very general importance.

**References**


Acknowledgements
I would like to thank Drs Thomas Kaufman, Ian Duncan, Allen Laughon and Sean Carroll for thoughtful discussions relevant to this review.

M. P. Scott is at the Department of Molecular, Cellular and Developmental Biology, University of Colorado at Boulder, Campus Box 347, Boulder, CO 80309, USA.