102. S. Green and P. Chambon, *Nature (London)* **325**, 75 (1987).
103. R. Miesfeld, P. J. Godowski, B. A. Maler, K. R. Yamamoto, *Science* **236**, 423 (1987).
104. S. M. Hollenberg, V. Giguere, P. Segui, R. M. Evans, *Cell* **49**, 39 (1987).
105. R. Brent and M. Ptashne, *ibid.* **43**, 729 (1985).
106. I. A. Hope and K. Struhl, *ibid.*, p. 177.
107. D. R. Smith, I. J. Jackson, D. D. Brown, *ibid.* **37**, 645 (1984).
108. J. Ma and M. Ptashne, *ibid.* **48**, 847 (1987).
109. J. M. Sugano and R. G. Roeder, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8511 (1986); Y. Suzuki *et al.*, *ibid.*, p. 9522; K. Gorski, M. Carneiro, U. Schibler, *Cell* **47**, 767 (1986).

Homeo Boxes in the Study of Development

WALTER J. GEHRING

The body plan of *Drosophila* is determined to a large extent by homeotic genes, which specify the identity and spatial arrangement of the body segments. Homeotic genes share a characteristic DNA segment, the homeo box, which encodes a defined domain of the homeotic proteins. The homeo domain seems to mediate the binding to specific DNA sequences, whereby the homeotic proteins exert a gene regulatory function. By isolating the normal *Antennapedia* gene, fusing its protein-coding sequences to an inducible promoter, and reintroducing this fusion gene into the germline of flies, it has been possible to transform head structures into thoracic structures and to alter the body plan in a predicted way. Sequence homologies suggest that similar genetic mechanisms may control development in higher organisms.

ORGANISMS DEVELOP ACCORDING TO A PRECISE DEVELOPmental program that specifies their body plan in great detail and also determines the sequence and timing of the developmental events. This developmental information is stored in the nucleotide sequences of the DNA. The question of how the one-dimensional sequence information stored in the DNA is converted into the three-dimensional structure of an embryo, or four-dimensional formation if we also include time, is the fundamental problem of developmental biology. Structural genes have been identified that specify the molecular building blocks from which the organism is constructed. The developmental program consists of a precise spatial and temporal pattern of expression of these structural genes that forms the basis of development. Normal development requires the coordinate expression of thousands of structural genes in a concerted fashion. Since independent control of the individual structural genes would lead to chaotic development, we might predict that there are controlling genes that regulate the activity of groups of structural genes coordinately. Such genes would presumably be arranged hierarchically or form a controlling network that ensures the proper timing of the developmental events and generates the proper spatial pattern. However, it proved to be difficult to find the controlling genes that specify the architecture, the body plan. Candidates for such developmental controlling genes were first identified as homeotic mutations in *Drosophila* as early as 1915 (*1*), but their molecular analysis had to await the advent of DNA technology. Homeotic mutations transform certain parts or an entire body segment into the corresponding structures of another body segment, thereby changing the architecture of the organism. *Drosophila* belongs to the dipteran insects that have only one pair of wings. However, in certain homeotic mutants, like those of the bithorax complex, the third thoracic segment becomes transformed into a second thoracic segment with a second pair of wings. This dramatic change in the architecture also may reflect evolutionary history, since the diptera evolved from more primitive insects that had four wings. Such homeotic mutations were found mainly in insects and other arthropods whose body is subdivided into typical segments along the anteroposterior body axis. However, they may also exist in vertebrates including humans.

The first homeotic genes were cloned in the absence of any biochemical information about their gene products by "chromosome walking" and by microdissection of bands from giant polytene chromosomes (*2*). The structural analysis of the *Antennapedia* (*Antp*) gene led to the discovery of the homeo box (Fig. 1), a small DNA segment of approximately 180 bp, that is characteristic for homeotic genes (*3, 4*). The significance of the homeo box homology was demonstrated by isolating previously unknown homeotic genes from *Drosophila* with the homeo box as a probe (*3*); perhaps more importantly, sequences homologous to the homeo box have been isolated from higher organisms including vertebrates (*5*), mammals, and humans (*6*). This might provide an entry point to cloning the genes that control development in higher organisms, on the basis of their partial homology to the *Drosophila* homeo box.

Comparative Anatomy of the Homeo Box

Analysis of the DNA sequences of the various homeo boxes shows that these sequences are highly conserved during evolution, whereas the flanking sequences differ considerably among different genes. The various homeo boxes share the same open reading frame, which extends into the flanking sequences and so indicates that the homeo box encodes a particular domain of the homeotic proteins, the homeo domain. A first hint with regard to the function of the homeo domain came from comparative protein sequence analysis, which revealed a small but significant degree of homology to the yeast mating–type proteins MAT a1 and MAT α 2 (*7*). These proteins are known to control cellular differentiation into mating-types a or α, or into spores, that is, into the three cell types that yeasts can form (*8*). They are sequence-specific DNA-binding

The author is professor of Developmental Biology and Genetics, Biozentrum, University of Basel, Switzerland.

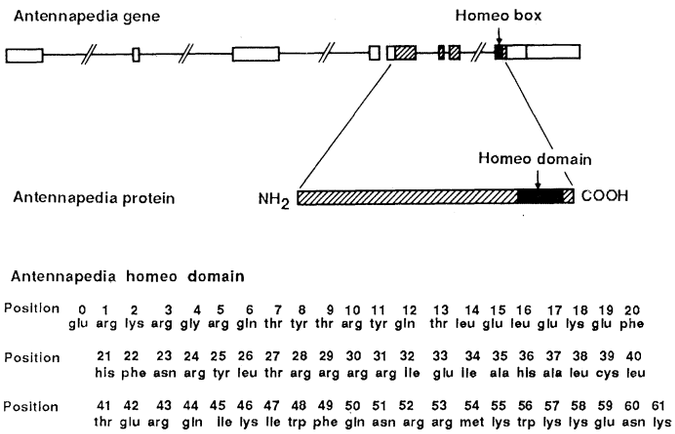

Fig. 1. Structural organization of the *Antp* gene and location of the homeo box. The gene spans more than 100 kb and consists of eight exons indicated by blocks, separated by introns (thin lines) that are not drawn to scale. The protein coding region is indicated by crosshatched bars. The homeo box (black bar) is a highly conserved DNA segment of 186 bp located in the eighth exon. It encodes the homeo domain that is located near the carboxyl terminus of the Antennapedia protein. The amino acid sequence of the homeo domain is indicated at the bottom.

proteins that bind to regulatory sequences of the genes that they regulate (*9*). Since the MAT genes control the basic decisions with regard to cell differentiation and also control sizable groups of other genes, they are considered to be master control genes. For MAT α 2 it has recently been shown that the region that is homologous to the homeo domain contains the sequence-specific DNA-binding domain of the protein (*10*). The homology between the MAT and the homeotic genes suggests that they share a similar functional role in gene regulation. Weak homologies to prokaryotic gene regulatory proteins have also been found (*7, 11*). These proteins are characterized by a helix-turn-helix motif, in which the so-called recognition helix directly binds to specific nucleotides in the DNA (*12*).

Sequence comparison among the 16 *Drosophila* genes from which the homeo box has been sequenced reveals several interesting features (Fig. 2). Nine amino acids in the homeo domains are absolutely invariant, which is also true if we include another 22 homeo domains of other species that range from sea urchin to human. If we determine the most commonly found amino acid at any given position, we find that in 57 out of 60 positions the most frequently found amino acid corresponds to the one found in the *Antp* gene, which therefore represents the consensus sequence most closely. The most highly conserved area corresponds to the putative recognition helix and the three adjacent amino acids toward the carboxyl terminus. This area includes four invariant amino acids, three of which are also found in the MAT genes of yeast. This is in marked contrast to the prokaryotic gene regulatory proteins (Fig. 3), which show much more variation in the recognition helix, reflecting the fact that each of the prokaryotic proteins binds to a different operator sequence in the DNA. This marked difference indicates that the homeotic proteins are more closely related to each other and suggests that they may recognize and bind to similar DNA sequences, albeit with different affinities.

There are several indications that the homeotic genes encode gene regulatory proteins that bind to specific DNA sequences. Antibody-localization studies have shown that all homeo domain–containing proteins tested so far accumulate in the nucleus, which is compatible with their proposed functional role (*13–18*). In vitro DNA-binding studies have provided tentative evidence for sequence-specific DNA binding of fusion proteins in which the homeo domain is fused to β-galactosidase (*19*). By means of a truncated *Antp* polypeptide that includes the homeo domain, footprint analysis demonstrated sequence-specific binding to the first *Antp* exon (*20*), in the absence of a carrier protein like β-galactosidase. In an independent study overlapping footprints were found with a full-length *Ultrabithorax* (*Ubx*) polypeptide (*21*). Since the *Antp* and *Ubx* proteins have the same putative recognition helix, this finding is consistent with the notion that these proteins may recognize similar DNA sequences. However, these studies are hampered by the fact that the proteins used in these experiments are produced by expression vectors in *Escherichia coli*, and therefore, they are not modified as are the native proteins in *Drosophila* (*22*). More definitive evidence has to come from in vitro experiments with proteins isolated from *Drosophila* combined with in vivo studies.

The homeotic proteins are considerably larger than just the homeo domain and have other domains with additional functions. Are these other domains also conserved during evolution? To answer this question we must compare amino acid sequences from homologous genes of different organisms. Figure 4 illustrates the engrailed homologs in *Drosophila* (*23*), honeybee (*24*), and mouse (*25*). In each of these three species a pair of closely related genes was found, and their homologies were not confined to the homeo domain but also extended considerably into the flanking regions, particularly toward the carboxyl termini of the proteins. This extended homology indicates that in addition to the homeo domain other parts of the homeotic proteins were conserved during an evolutionary period of more than 500 million years. In the case of the *Deformed* gene (*Dfd*) of *Drosophila* and its homologs in frogs and humans (*26*), conserved sequences are also found at the amino terminus of the protein, which indicates that large parts of the homeotic proteins are conserved during evolution, including several domains encoded by different exons. These structural homologies among the homeotic proteins provide additional support for the idea that the homeo box–containing genes serve a function in higher vertebrates similar to that served in *Drosophila*.

At the moment, genetic evidence from mutants is lacking for

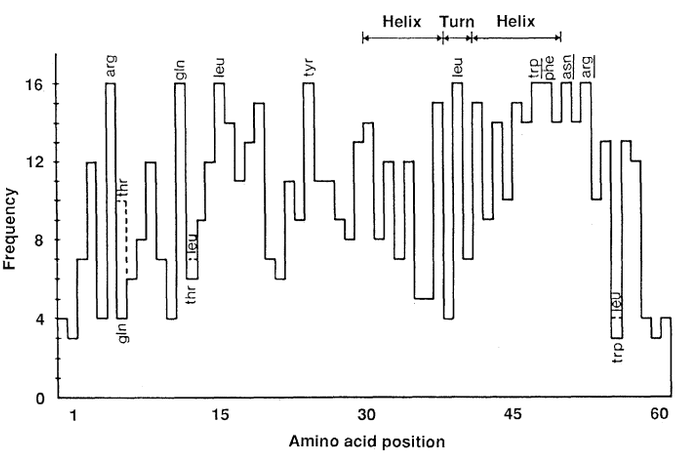

Fig. 2. Comparison of the amino acid sequences encoded by 16 homeo boxes in *Drosophila*. The amino acid frequency is given for positions 0 to +61 of the homeo domains, which include 62 amino acids. At nine positions all homeo boxes encode the same amino acid that is indicated. The underlined amino acids are also invariant in the MAT genes a1, α 2, and P1 of yeasts. The putative homology to the helix-turn-helix motif is indicated. In all but three positions indicated by dotted lines, the most frequent amino acid is the one found in the *Antp* gene. On the basis of the sequence characteristics, three subfamilies can be distinguished so far, the Antennapedia subfamily (the largest), the engrailed subfamily [includes *engrailed* and *invected* genes (*23*)], and the paired subfamily, which in addition to *paired* contains at least two other known genes (*36*) that are not included in this figure.

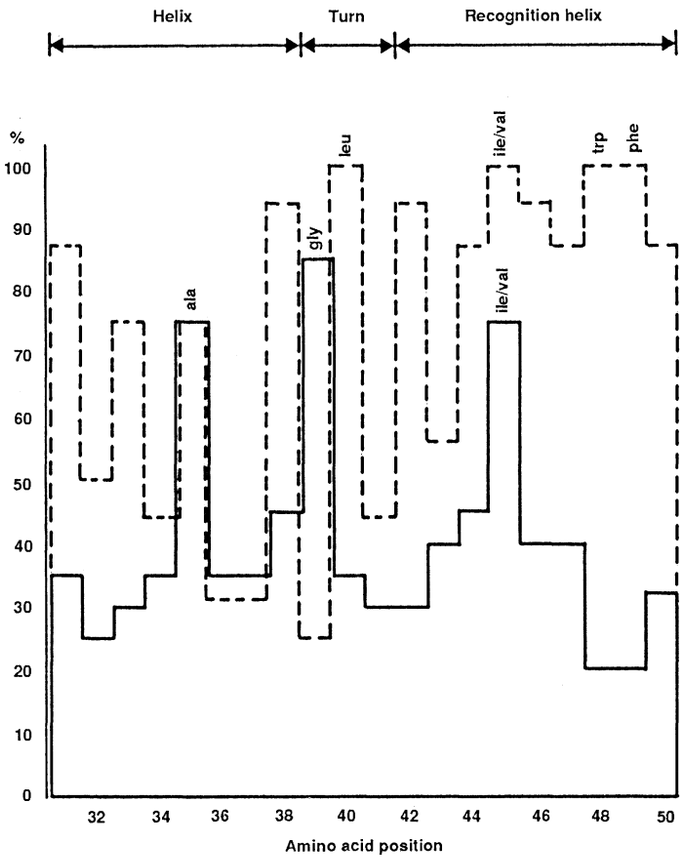

Fig. 3. Comparison of the helix-turn-helix motif of 20 prokaryotic DNA-binding proteins (−) with 16 homeo-domain sequences of *Drosophila* (---). The prokaryotic proteins include the repressor and *cro* protein of bacteriophages λ, P22 and 434, CAP, *Fnr* protein, *lac* repressor, *gal* repressor, λ CII and P22 CI protein, the tetracycline repressors of Tn10 and pSC101, *Trp* repressor, H-inversion protein, Tn3 resolvase, γ-δ resolvase, arabinose-C protein, and the *Lex* repressor (see *12*). The most conserved amino acids are alanine in the first helix, glycine in the turn, and isoleucine or valine in the recognition helix, as indicated by the peaks. The alanine and the isoleucine or valine are similarly conserved in the homeo domains, but the glycine is found only in the turn of the two homeo domains of engrailed and invected, which form a different subfamily of homeo boxes (*23*) and other exceptional cases. The leucine in the turn is invariant in the *Drosophila* proteins and in all 22 homeo domains known from other organisms. The degree of conservation in the putative recognition helix of the homeo domains is considerably higher than in the prokaryotic proteins.

higher organisms. However, the cloning of numerous homeo box–containing genes in the mouse might make it possible to assign a given gene to a mutation. Several proposals have been made concerning the phenotype of putative mouse homeotic mutants (*27*), but so far no mutations in homeo box–containing genes have been identified. The mouse mutation *rachiterata*, which leads to the transformation of the seventh cervical vertebra into a first thoracic vertebra with ribs (*28*), fits the phenotypic definition of a homeotic mutation rather closely, since it leads to a segmental transformation. However, the different mode of mammalian development may also result in other kinds of phenotypic changes.

The Genetic Control of *Drosophila* Development

The *Drosophila* egg is roughly ellipsoidal in shape and clearly polarized so that the future axes of the embryo—the anteroposterior, dorsoventral, and left-right axes—can be distinguished. The early stages of *Drosophila* development are peculiar in that after fertilization the zygote nucleus first undergoes a rapid series of synchronous nuclear divisions that are not accompanied by cytokinesis. This leads to the formation of a syncytium in which 512 nuclei are located in a common cytoplasm. The actual cleavage, resulting in the subdivision into cells, is delayed. First, a small group of nuclei moves into the cortical cytoplasm at the posterior pole of the egg and becomes surrounded by plasma membranes to form the pole cells that later give rise to the germ cells. Most of the remaining nuclei colonize the cortical cytoplasm at the periphery of the egg, where they undergo another four division cycles until the actual cleavage of the cytoplasm occurs, and plasma membranes separating the nuclei are formed. The resulting monolayer of cells around the periphery of the egg is called the blastoderm. Earlier embryological experiments and the analysis of genetic mosaics have demonstrated that the dividing "cleavage" nuclei are equivalent and totipotent, but when the blastoderm cells are formed they only give rise to certain body segments (*29*). The basic architecture of the embryo is laid down at the blastoderm stage and is reflected by the fate map [Fig. 5 (*30*)]. Since the nuclei are totipotent before they reach the periphery, the cortical cytoplasm is thought to contain determinants or positional information that allow the nuclei to differentiate according to their position (rather than by virtue of their origin). The body plan essentially consists of an anteroposterior series of equal-sized segments that subsequently acquire different identities, that is, different structures and functions. There are six head, three thoracic, and about ten abdominal segments. The number of head and abdominal segments is known only approximately, since in the course of development several segments fuse to form the head and the posterior end of the fly, whereas the segmental organization of the thorax and the anterior abdomen is clearly retained. In order to generate the body plan, the polar pattern of the egg must be converted first into a periodic pattern of segments, and, in a second conceptual step, the periodic pattern must be converted into a sequential pattern according to which the different segments are arranged. Not only must the segments acquire different identities, but also the proper sequence, so that for example the second thoracic segment (T2) is formed between T1 and T3.

Three classes of genes have been identified that are involved in the specification of the body plan: (i) maternal-effect genes that specify egg polarity and the spatial coordinates of the egg and the future embryo (*31*); (ii) the segmentation genes, which determine the number and polarity of the body segments (*32*); and (iii) the homeotic genes, which determine the identity and sequence of the body segments (*33*). Although most of the homeo box–containing genes have been found in the homeotic class, several genes with a homeo box have also been found among the coordinate and the segmentation genes.

Determinants of Positional Information

There has been a long controversy concerning the question of whether there are localized cytoplasmic determinants in the egg that determine the fate of the early blastomeres or whether the positional information in the egg consists of morphogenetic gradients operating by long-range signals. A number of embryological experiments make it very unlikely that the *Drosophila* egg is a mosaic of strictly localized, qualitatively distinct determinants (*34*). However, recent studies of the genes *bicoid* (*bcd*) and *caudal* (*cad*) indicate that probably both mechanisms, precise cytoplasmic localization and gradients of morphogens, may operate. *Bicoid* mutants (*35*) show a strict maternal effect; irrespective of the paternal genotype, mutant

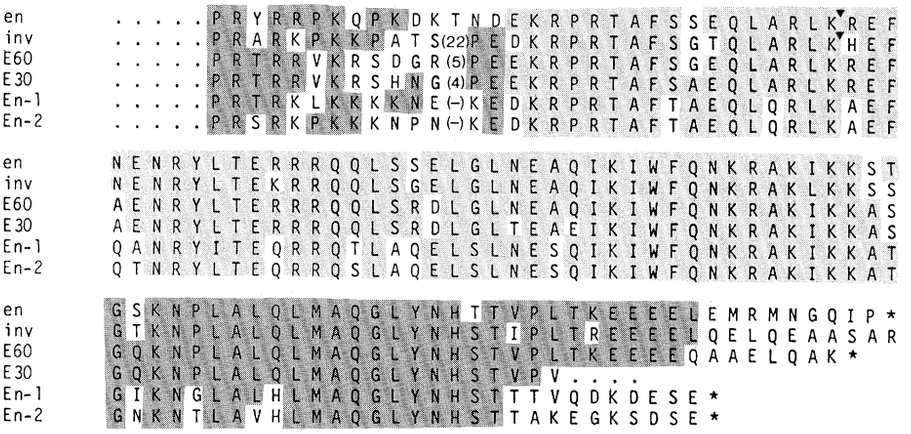

Fig. 4. Comparison of the amino acid sequences encoded by the *engrailed* genes of *Drosophila*, honeybee and mouse. Labels are en, engrailed (*23*); inv, invected of *Drosophila* (*23*); E60 and E30, homologous sequences of the honeybee (*24*); En-1 and En-2, homologous sequences of the mouse (*25*). The most frequently found amino acids among the six sequences are indicated in lighter shaded regions for the homeo domain and in darker shaded regions for the flanking sequences that show homology. The amino terminal sequences are not available (....); insertions and deletions are indicated in parentheses. The arrowheads indicate the location of an intron in the two *Drosophila* genes. The stars refer to the carboxyl termini of the proteins.

mothers produce defective eggs that give rise to mutant embryos lacking head and thorax. The missing segments are partly replaced by terminal abdominal segments, a replacement that corresponds to a homeotic transformation. Injection of cytoplasm from the anterior pole of wild-type eggs into mutant eggs results in the formation of embryos that resemble the wild-type to the extent that the three thoracic segments and almost complete heads are formed (*35*). The activity dependent on bcd^+ is high at the anterior pole of the donor embryo and declines sharply in the posterior direction. Anterior structures can be induced by injecting anterior wild-type cytoplasm at any position along the anteroposterior axis of *bcd* embryos, and the pattern reorganizes accordingly. Since there is a strong decrease in the inducing activity when cytoplasm is injected progressively toward the posterior end, an antagonistic suppressive activity emanating from the posterior pole may be present. These data are compatible with the hypothesis that *bcd* may encode an anterior determinant that specifies the coordinates of the future embryo by interaction with other determinants.

The *bcd* gene has been cloned on the basis of cross homology with *paired* (*prd*), a segmentation gene that encodes a characteristic histidine-proline repeat (*36*). Both *bcd* and *prd* have a homeo box that is considerably different from the consensus *Antp* sequence. In situ hybridization of a *bcd* complementary DNA (cDNA) probe to ovarian and embryonic sections revealed a striking accumulation of the *bcd* transcripts at the anterior pole of the egg in the cortical cytoplasm (*36*). The transcripts are first detected in the nurse cells, which in the course of oogenesis feed their contents (including macromolecules and cellular organelles) into the oocyte. The *bcd* transcripts become anchored in the anterior cortical cytoplasm. The mechanism of targeting these RNA molecules is not known, but it probably involves preferential binding to prelocalized components in the anterior cortical cytoplasm. Such prelocalized receptors might be anchored in the plasma membrane of the oocyte and/or associated with the cytoskeleton. After fertilization, during the early nuclear divisions, the *bcd* transcripts form a concentration gradient along the anteroposterior axis with the highest concentration at the anterior pole. These data are consistent with the idea that *bcd* specifies a determinant of anterior positional information. The RNA is precisely localized as predicted for a cytoplasmic determinant and subsequently forms a concentration gradient that may exert a long-range effect. Gradient models for positional information are usually described in terms of a localized source and a long-range signal (*37*), but in this case "anchor" is a more specific term than source, and signal can be substituted by determinant. Nothing is known about the anchor, but in molecular terms it has to bind or sequester *bcd* RNA. With respect to the determinant, the *bcd* protein itself is certainly a good candidate. To my knowledge, this protein has not been characterized yet, but the fact that it has a homeo domain suggests that it is a DNA-binding protein. Since the early *Drosophila* embryo is a syncytium, DNA-binding proteins may function directly as determinants and exert long-range effects by diffusion in the cytoplasm and localized accumulation in the nuclei. At later stages of development and in other organisms where the positional information has to be transmitted across cell membranes, different transmitter substances are likely to be involved, but DNA-binding proteins may still be involved as second messengers to the nucleus.

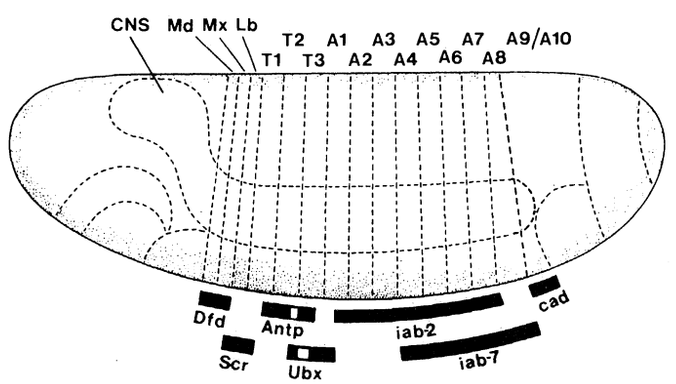

Fig. 5. Schematic representation of the fate map of the *Drosophila* embryo at the blastoderm stage. Anterior is to the left and dorsal on the top. The segments in which the various homeotic genes are preferentially expressed are indicated by black bars. The primordia of the central nervous system (CNS) and the body segments are indicated (Md, mandibulary; Mx, maxillary; Lb, labial segment, T1–T3, thoracic; A1–A10, abdominal segments). *Dfd, Deformed; Scr, Sex combs reduced; Antp, Antennapedia; Ubx, Ultrabithorax; iab-2* and *iab-7, infraabdominal 2* and *7; and cad, caudal.*

An anteroposterior concentration gradient of transcripts has been found for the *cad* gene, which was isolated on the basis of homeo box homology (*38*). However, in contrast to *bcd*, the *cad* transcripts are first uniformly distributed throughout the late oocyte and the mature egg. The concentration gradient with its peak at the posterior pole forms only after fertilization. During blastoderm formation a zygotic transcript different in size from that in the unfertilized egg is synthesized and accumulates in a single stripe in the posteriormost abdominal segments (*38*). Immunolocalization studies indicate that the transcript gradient is preceded by a concentration gradient of the *cad* protein (*17, 18*). This protein is not detectable in the oocyte when the RNA is uniformly distributed, indicating that the maternal *cad* RNA is not translated until after fertilization. The mechanism by which the protein gradient forms is not known. The protein appears first in the cytoplasm, but soon

after (at the eighth nuclear division) it accumulates in a graded pattern in the nuclei as they begin to migrate to the cortex (Fig. 6). At the blastoderm stage a single belt of cells expresses *cad* protein near the posterior end of the embryo. The idea that the *cad* protein gradient has functional significance is supported by the analysis of *cad* mutants (*17, 18*), which in contrast to *bcd* mutants are lethal when homozygous and cause the absence of structures of the posteriormost segments. In flies lacking both the maternal and zygotic *cad* products synthesized by the embryo after fertilization, a spectrum of graded segmental defects is observed that roughly correlates with graded distribution of the *cad* protein. However, the analysis of *cad* expression in another maternal-effect mutant *Bicaudal* (*BicD*) indicates that *cad* does not encode a primary spatial determinant. *Bicaudal* affects the anteroposterior polarity of the embryo such that the anterior end of the embryo is replaced by a second posterior end, leading to a double abdomen phenotype arranged in mirror symmetry (*39*). As shown in Fig. 6, the *cad* protein gradient is essentially abolished in such symmetrical embryos, and at the blastoderm stage two symmetrical belts of zygotic *cad* expression are detected. This indicates that *BicD* is epistatic and affects a primary spatial determinant to which *cad* responds. Since the *cad* protein first accumulates in the pole cell nuclei and later in the germ cells of both sexes, it may not only specify the posteriormost abdominal segments, but it may also be one of the germ-cell determinants.

Segmentation Genes

During blastoderm formation the polar pattern specified by determinants of positional information is converted into a periodic pattern of repeating body segments. On the basis of mutations about 20 genes have been identified (*32*) that specify the number and polarity of the segments. These genes have been subdivided into three classes according to their phenotype: (i) the gap mutants, which delete groups of adjacent segments; (ii) the pair-rule mutants, which cause pattern deletions in alternating segments; and (iii) the segment polarity mutants that cause pattern defects in every segment. Most of these mutants are zygotic lethals (in essential genes expressed after fertilization), but some gap genes show a maternal-effect and thus are also expressed during oogenesis. Four of the segmentation genes cloned so far, *fushi tarazu* (*11, 40*), *evenskipped* (*41*), *paired* (*36*), and *engrailed* (*23*) contain a homeo box, and the gap gene *Krüppel* has homology to the transcription factor TFIIIA of *Xenopus* (*42*). Thus, these genes may encode gene regulatory proteins or transcription factors binding to specific DNA and/or RNA sequences. The presence of a homeo box in segmentation genes may appear unusual but, at least for *fushi tarazu*, it has been shown that mutations in this gene can also produce a homeotic phenotype (*43*), strengthening the correlation between the homeo box and homeosis.

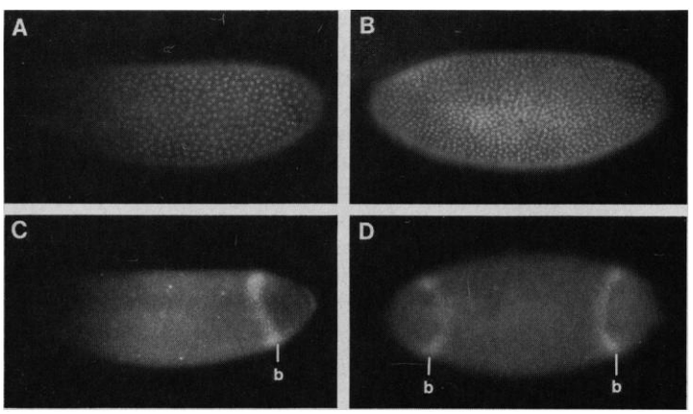

Fig. 6. Immunolocalization of the *caudal* protein in early wild type (A, C) and *Bicaudal-D* (B, D) embryos. [Adapted from (*18*) with permission of Cell Press.] (**A**) In the syncytial blastoderm of the wild type the protein forms an anteroposterior concentration gradient with its high point at the posterior end (on the right). The protein clearly accumulates in the nuclei. (**B**) In the *Bicaudal* mutant embryo, which is mirror symmetric, the gradient is abolished. (**C**) At later blastoderm stages a belt (b) of cells that accumulates *caudal* protein is formed at a position that corresponds to the posteriormost abdominal segments. At the corresponding stage of the *Bicaudal* embryo (**D**) two symmetrically arranged stripes are formed reflecting the symmetry of the mutant embryo.

Fushi tarazu (*ftz*) is a Japanese designation and means "not enough segments," which refers to the fact that *ftz*$^-$ mutants have only half the number of segments (*44, 45*). Molecular analysis of *ftz* revealed that during blastoderm stages its transcripts accumulate in a regular pattern of seven stripes corresponding to those sections on the fate map that are missing in *ftz*$^-$ mutants (*46*). The *ftz* transcripts are targeted to the cortex and precisely localized between the nuclei and plasma membrane at the periphery of the egg. All pair-rule genes analyzed so far show a corresponding pattern of stripes in different phases, reflecting the periodic pattern of the body segments and the fate map (*23, 41, 47, 48*). In *ftz* the stripes form prior to cellularization when the nuclei have entered the cortical cytoplasm, which suggests that the nuclei receive positional information from the cortical cytoplasm and respond by expressing or not expressing *ftz*$^+$ depending on their position. Consistent with this interpretation is the finding that the pattern of *ftz* stripes is mirror symmetrical in *BicD* embryos (*49*), which have a reversed polarity in the anterior half of the embryo and give rise to a double abdomen. This led us to postulate that the nuclei might have a sensor for positional information, presumably associated with the *ftz* gene, capable of interacting with determinants of positional information to generate the periodic pattern of stripes (*50*). This hypothesis was tested by germline transformation with the *P* transposon as a vector (*51*). These transformation experiments showed that the functional *ftz* gene consists of a relatively small transcription unit of 1.9 kb and a large control region, comprising approximately 6.1 kb of 5′ flanking sequences, which is required for proper expression and for the rescue of *ftz*$^-$ recipients. Subsequently, the large 5′ flanking region was fused to bacterial β-galactosidase protein-coding sequences in order to test whether this putative control region has the capacity to generate a *ftz*-type of stripe pattern with a foreign reporter gene. Indeed, in these transformed embryos, β-galatosidase is expressed in the same "zebra" pattern as the *ftz* protein (Fig. 7). At later embryonic stages, β-galactosidase (*51*) and the authentic *ftz* protein in normal embryos (*14*) are also expressed in specific neuronal precursors of the central nervous system in a precise pattern indicating that gene activity is regulated at the level of the single cell with remarkable precision. In contrast to the zebra pattern (which has double-segment periodicity) the pattern in the nervous system is repeated in every segment. By deleting and recombining different sequences in the 5′ flanking region at least three controlling elements were identified: an upstream element, most distal to the gene, a neurogenic element required for expression in the nervous system, and a "zebra element" located within 600 bp from the start of transcription (*51*). By examining *ftz* expression in various mutants by means of in situ hybridization, fluorescent antibody staining (*52*), and histochemical staining for β-galactosidase (*53*), it can be shown that the generation of the zebra pattern requires not only the activity of maternal-effect genes specifying positional information but also zygotic segmentation genes. For example, in *hairy*$^-$ embryos the stripes of *ftz* expression become much broader and tend to fuse, which indicates

that in the wild-type embryo the normal zygotic *hairy* gene is required for proper expression of *ftz* and that it directly or indirectly represses *ftz* in the areas between the stripes. This shows that the generation of the pattern also depends on the interaction between the nuclei. Recent experiments with β-galactosidase fusion genes suggest that the *hairy* gene product acts on the zebra element (*53*). The analysis of the upstream element in various fusion constructs indicates that it is an enhancer required for the expression in stripes and that the *ftz* protein itself may interact autocatalytically with its own enhancer (*53*). This is consistent with preliminary in vitro binding studies in which full-length *ftz* protein produced in *E. coli* binds to the same DNA fragment in which the enhancer is localized (*22*). The data raise the possibility in general that homeotic proteins are transcription factors binding to enhancers.

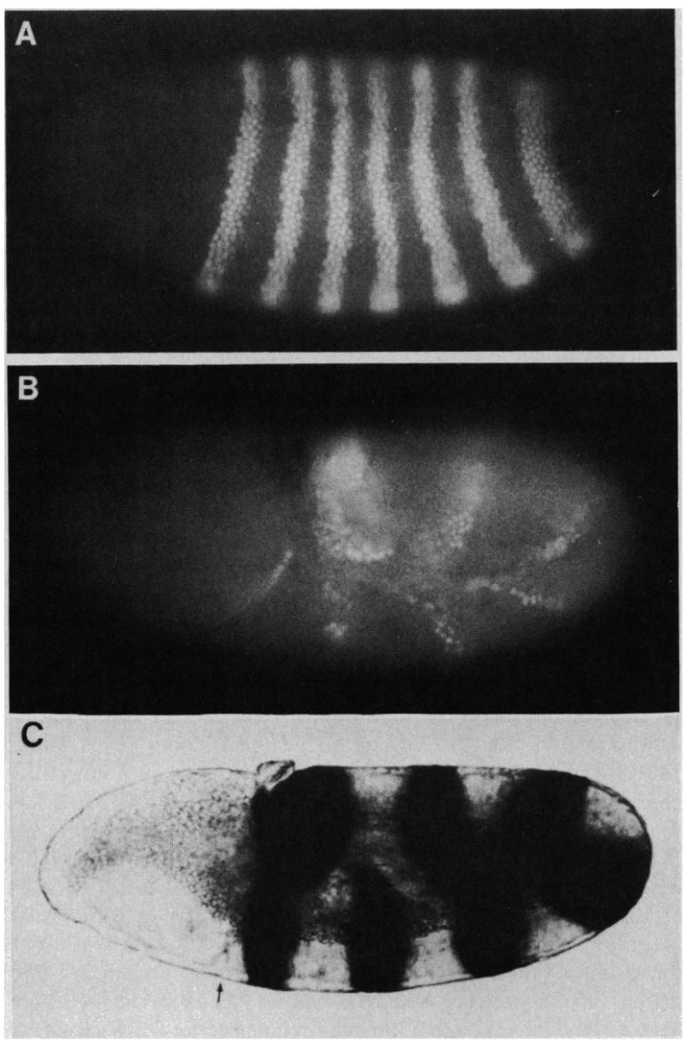

Fig. 7. The spatial pattern of expression of the *ftz* gene. (**A**) Immunolocalization of the *ftz* protein by fluorescent antibody techniques reveals a precise pattern of seven stripes (belts) around the blastoderm embryo. Anterior is to the left and dorsal on the top. The labeling is nuclear and reflects the fate map (Fig. 5). (**B**) Later, at the stage of germ-band extension, the ventral germ band extends around the posterior pole so that the posteriormost segments reach an anterior dorsal position. The seven stripes are clearly visible although weaker than in (A) (*22*). (**C**) Pattern of expression of β-galactosidase in a transformant embryo carrying 6.1 kb of 5′ flanking sequences of the *ftz* gene fused to the protein coding sequence of bacterial β-galactosidase at the same stage as in (B). β-Galactosidase is expressed in the same pattern of seven stripes as the *ftz* protein. The higher levels of β-galactosidase activity are due to higher stability of the enzyme (*51*).

Homeotic Genes

Conceptually the periodic pattern of body segments generated by the segmentation genes has to be converted into a sequential pattern of diverse segments. The diversity and sequence of segments is controlled by homeotic genes (*33*). On the basis of his extensive genetic analysis of the bithorax complex, Lewis (*54*) has proposed a combinatorial model that assumes that each body segment is specified by a unique combination of homeotic genes that are expressed in this particular segment. According to this model, the smallest number of homeotic genes is required in the second thoracic segment (T2), which is considered to be the ground state (or prototype segment), and progressively more genes have to be activated in the more posterior segments. The genes of the bithorax complex specify the posterior thoracic and abdominal segments of the fly (*54*), whereas a separate cluster of homeotic genes, the Antennapedia complex, controls the anterior thoracic and the head segments (*55*). The homeotic genes were originally thought to be activated according to a gradient of positional information laid down in the egg by maternal genes. However, subsequent genetic experiments suggested that, in addition, there are regulatory interactions among the homeotic genes themselves (*56*).

Homeo box homology and chromosomal walking have allowed the cloning of a series of homeotic genes that are expressed in a spatially restricted manner along the anteroposterior axis (Fig. 5). The area of preferential expression of a gene on the fate map generally coincides with the location of segments that are affected by mutations that lead to a loss of function in that gene. However, it should be emphasized that all of the genes examined are expressed in more than one segment. For example, *Dfd* is expressed at the blastoderm stage in a single stripe, approximately six cells wide, that includes both the maxillary and the mandibulary segment in the posterior head region of the embryo (*3, 26*).

Deletions of the *Antp* locus effect all three thoracic segments. The predominant effect is that the second segment (T2) is transformed toward T1 and head segments (into more anterior segments), suggesting that *Antp* mainly specifies the second thoracic segment. Localization of $Antp^+$ transcripts by in situ hybridization (*57*) and immunolocalization of the *Antp* protein (*16*) are consistent with this idea, since both the RNA and the protein accumulate mainly in T2 and parts of T1 and T3. However, the expression of *Antp* in the nervous system at early embryonic stages extends to the posterior abdominal segments, where it later becomes largely repressed. The first molecular evidence for the interaction between homeotic genes came from studies of *Antp* expression in mutants of the bithorax complex. Removal of the *Ubx* gene, which is expressed posterior to *Antp* (*58*; Fig. 5), results in the ectopic expression of *Antp* in those segments where *Ubx* normally is expressed. This leads to the tentative conclusion that *Ubx* represses *Antp* directly or indirectly in these segments (*48, 50, 59*). Immunolocalization studies with antibodies directed against *Ubx* indicate that *Ubx* in turn is repressed by the next more posteriorly expressed gene *iab-2* (*60*). This suggests a modification of the combinatorial model of Lewis (*54*)—sequential interactions among the homeotic genes may determine the spatial sequence of the body segments. Since the genes involved in this interaction have similar homeo boxes and invariant recognition helices, the interaction may be due to competition for the same DNA-binding sites. This model of competitive sequential interaction can be tested experimentally.

Redesigning the Fruit Fly

Consistent with the above model of sequential interactions between homeotic genes, recessive loss-of-function mutants lead to a

transformation of a given segment into a more anterior one. On the other hand, a considerable number of dominant gain-of-function mutants are known that transform in the opposite (posterior) direction. The nature of these dominant mutations has been an enigma for molecular biologists. In *Antp* the recessive loss of function leads to an anterior transformation mainly of T2 and T3 toward T1 and head segments (*45, 61, 62*), whereas dominant gain-of-function mutants (*61, 63*) lead to a transformation of the antennae into second legs, in other words, in the posterior direction. Most of these dominant mutations at the *Antp* locus are due to chromosomal inversions that separate the 5′ end of the gene from the protein-coding region. Molecular analysis of one of these mutants ($Antp^{73b}$) indicates that the inversion leads to the fusion of the promoter and leader sequences of a foreign gene to the protein-coding region of *Antp* (*64*). Since several inversions that result in similar homeotic phenotypes are known to have different chromosomal breakpoints, it seems unlikely that in all of these cases *Antp* is fused to an antenna-specific promoter. Therefore, we assume that any promoter that is active in the antennal imaginal disk at the appropriate stage of developmental can generate antennal legs. The reason why legs are formed in place of the antennae has to be sought in the control circuits. Previous studies on transdetermination have indicated that in vivo cultured antennal disk cells transdetermine frequently into leg cells without any detectable mutational change (*65*); these studies suggest that the antenna is a weak point in the circuits controlling developmental pathways. If this is the case, it should be possible to induce the formation of antennal legs by a promoter that can be induced in all cells, including the antennal disk.

To test this prediction Schneuwly *et al.* chose the heat shock promoter of the *hsp*70 gene, which is presumably active in all cells and, since the *Antp* gene is too large for transformation experiments, we had to use a cDNA clone. The *Antp* cDNA was inserted into a newly constructed P-vector under the control of the *hsp*70 promoter (*66*). The *hsp*70-*Antp* fusion gene was transferred into the germline of normal flies. The resulting transformants possess two normal $Antp^{+}$ genes and an additional fusion gene under heat-shock control. At normal temperatures these flies do not show any change in phenotype. However, if the transformed larvae are heat-shocked during the early third larval stage, they develop into flies with antennal legs (Fig. 8). Thus, the dominant gain-of-function phenotype appears to be due to overexpression of the *Antp* protein in the antennal disk, an ectopic site where the protein is normally not expressed at detectable levels. Therefore, it is possible to alter the body plan of *Drosophila* by altering the expression of the normal gene (or the normal homeotic protein) in a predictable way, and to "redesign the fruit fly."

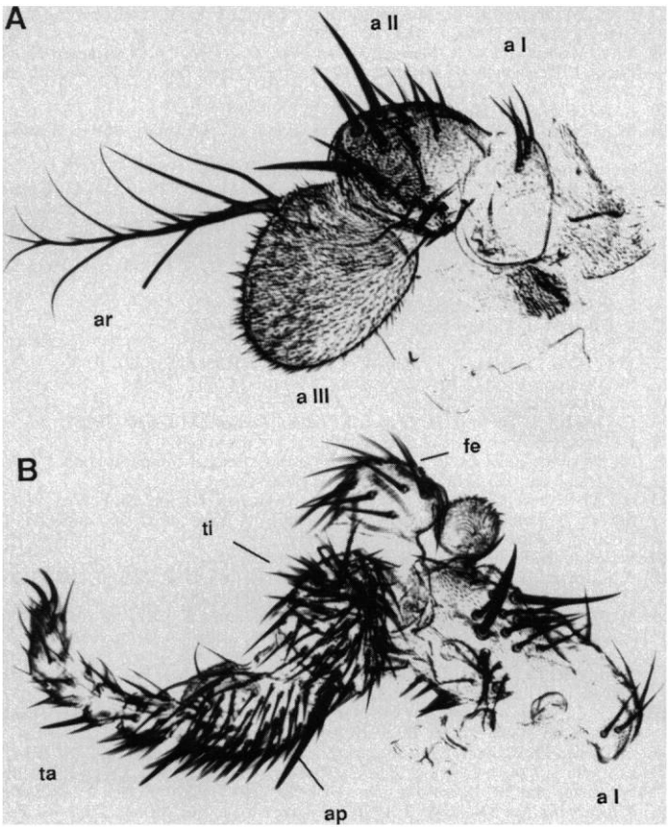

Fig. 8. Transformation of the antenna into a leg in a transformed fly carrying an additional Antennapedia protein coding sequence under the control of a heat-inducible promoter. [Taken from (*66*) courtesy of *Nature*.] (**A**) Normal antenna (control); (**B**) transformant after heat induction during the early third larval stage. The antenna is almost completely transformed into a middle leg. Abbreviations: aI, aII, and aIII are first, second, and third antennal segment; ar, arista; ta, tarsus; ti, tibia; fe, femur; ap, apical bristle.

Conclusions

The discovery of the homeo box in *Drosophila* has allowed the rapid isolation of a family of genes specifying the body plan. It has been possible to alter the body plan by isolating the *Antp* gene, fusing its normal protein-coding region to a heat-inducible promoter, and reintroducing this fusion gene into the germline of recipient flies. The results of this experiment, which were predicted on the basis of genetic data, provide definitive evidence for the idea that *Antp* specifies mainly the second thoracic segment. A series of other homeotic genes have been cloned that specify other parts of the body plan. The model of sequential competitive interactions among the homeotic genes can be tested experimentally. However, the network of genetic interactions may be rather complex. The recent results on *bcd* suggest that homeotic genes may also specify the positional information in the egg. Genetic data, DNA sequence comparison, immunolocalization, and in vitro binding studies suggest that homeotic genes encode gene regulatory proteins that bind to specific DNA sequences by means of the homeo domain. Transformation studies with *ftz*–β-galactosidase fusion genes suggest that *ftz* encodes an enhancer-binding protein that interacts autocatalytically with its own enhancer. However, it is not known which genes on the next lower level of the hierarchy are in turn controlled by homeotic genes.

The homeo box has provided an entry point for cloning homologous genes in higher organisms including humans. So far, none of the cloned genes has been assigned to a known mutation, but in the mouse this should be possible in the future. Alternatively, transgenic animals may reveal the function of these genes. Sequence comparison of homeotic genes from insects and vertebrates suggests that the basic mechanisms of genetic control of development may be similar despite the different modes of development.

REFERENCES AND NOTES

1. C. B. Bridges and T. H. Morgan, *Carnegie Inst. Washington Publ.* **327**, 137 (1923).
2. W. Bender *et al.*, *Science* **221**, 23 (1983); R. L. Garber, A. Kuroiwa, W. J. Gehring, *EMBO J.* **2**, 2027 (1983); M. Scott *et al.*, *Cell* **35**, 763 (1983).
3. W. McGinnis, M. S. Levine, E. Hafen, A. Kuroiwa, W. J. Gehring, *Nature (London)* **308**, 428 (1984); W. McGinnis, R. L. Garber, J. Wirz, A. Kuroiwa, W. J. Gehring, *Cell* **37**, 403 (1984).
4. M. P. Scott and A. J. Weiner *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4115 (1984).
5. A. E. Carrasco *et al.*, *Cell* **37**, 409 (1984).
6. W. McGinnis, C. P. Hart, W. J. Gehring, F. H. Ruddle, *ibid.* **38**, 675; M. Levine, G. M. Rubin, R. Tjian, *ibid.*, p. 667.

7. J. C. W. Shepherd, W. McGinnis, A. E. Carrasco, E. M. DeRobertis, W. J. Gehring, *Nature (London)* **310**, 70 (1984).
8. V. L. Mackay and T. R. Manney, *Genetics* **76**, 272 (1974); J. N. Strathern, J. B. Hicks, I. Herskowitz, *J. Mol. Biol.* **147**, 357 (1981); K. Tatchell, K. Nasmith, B. D. Hall, C. Astell, M. Smith, *Cell* **27**, 25 (1981).
9. A. Johnson and I. Herskowitz, *Cell* **42**, 237 (1985).
10. M. Hall and A. Johnson, *ibid.*, in press; S. Porter and M. Smith, *Nature (London)* **320**, 766 (1986).
11. A. Laughon and M. P. Scott, *Nature (London)* **310**, 25 (1984).
12. C. O. Pabo and R. T. Sauer, *Annu. Rev. Biochem.* **53**, 293 (1984). A lucid discussion of gene regulation in bacteriophage lambda is found in M. Ptashne, *A Genetic Switch: Gene Control and Phage Lambda* (Blackwell Scientific, Oxford, 1986).
13. R. A. H. White and M. Wilcox, *Cell* **39**, 163 (1984); P. A. Beachy, S. L. Helfand, D. S. Hogness, *Nature (London)* **313**, 545 (1985).
14. S. B. Carroll and M. P. Scott, *Cell* **43**, 47 (1985).
15. S. DiNardo, J. M. Kuner, J. Theis, P. H. O'Farrell, *ibid.*, p. 59.
16. S. B. Carroll, R. A. Laymon, M. A. McCutcheon, P. D. Riley, M. P. Scott, *ibid.* **47**, 113 (1986); J. Wirz, L. Fessler, W. J. Gehring, *EMBO J.* **5**, 3327 (1986).
17. M. Macdonald and G. Struhl, *Nature (London)* **324**, 537 (1986).
18. M. Mlodzik and W. J. Gehring, *Cell* **48**, 465 (1987).
19. C. Desplan, J. Theis, P. H. O'Farrell, *Nature (London)* **318**, 630 (1985).
20. U. Weber, J. Wirz, W. Gehring, unpublished observations.
21. M. Krasnow, E. Gavis, P. Beachy, D. Hogness, personal communication.
22. H. Krause and W. Gehring, unpublished observations.
23. S. J. Poole *et al.*, *Cell* **40**, 37 (1985); K. G. Coleman, S. J. Poole, M. P. Weir, W. C. Soeller, T. Kornberg, *Genes Dev.* **1**, 19 (1987); A. Fjose, W. J. McGinnis, W. J. Gehring, *Nature (London)* **313**, 284 (1985).
24. U. Walldorf, R. Fleig, W. J. Gehring, in preparation.
25. A. L. Joyner, T. Kornberg, K. G. Coleman, D. R. Cox, G. R. Martin, *Cell* **43**, 29 (1985); A. Joyner and G. Martin, *Genes Dev.* **1**, 29 (1987).
26. M. Regulski, N. McGinnis, R. Chadwick, W. McGinnis, *EMBO J.* **6**, 767 (1987); R. Chadwick and W. McGinnis, *ibid.*, p. 779; A. Laughon, S. B. Carroll, F. A. Storfer, P. D. Liley, M. Scott, *Cold Spring Harbor Symp. Quant. Biol.* **50**, 253 (1985); R. P. Harvey, C. J. Tobin, D. A. Melton, *EMBO J.* **5**, 1237 (1986); F. Mavilio *et al.*, *Nature (London)* **324**, 664 (1986).
27. W. J. Gehring, *Cell* **40**, 3 (1985); B. Hogan, P. Holland, P. Schofield, *Trends Genet.* **1**, 65 (1985); J. M. W. Slack, *J. Theor. Biol.* **114**, 463 (1985).
28. A. Theiler, D. Varnum, L. C. Stevens, *Z. Anat. Entwicklungsgesch.* **145**, 75 (1974).
29. L. Chan and W. J. Gehring, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 2271 (1971); E. Wieschaus and W. J. Gehring, *Dev. Biol.* **50**, 249 (1976); E. Steiner, *Wilhelm Roux's Arch. Dev. Biol.* **180**, 9 (1976); P. Lawrence and G. Morata, *Dev. Biol.* **54**, 40 (1977); A. Simcox and J. Sang, *ibid.* **97**, 212 (1983).
30. J. A. Campos-Ortega and V. Hartenstein, *The Embryonic Development of* Drosophila melanogaster (Springer-Verlag, Heidelberg, 1985).
31. C. Nüsslein-Volhard, in *Determinants of Spatial Organization*, S. Subtelny and I. R. Königsberg, Eds. (Academic Press, New York, 1979), p. 185.
32. C. Nüsslein-Volhard and E. Wieschaus, *Nature (London)* **287**, 795 (1980).
33. See for review, W. Gehring and Y. Hiromi, *Annu. Rev. Genet.* **20**, 147 (1986).
34. H. G. Frohnhöfer *et al.*, *J. Embryol. Exp. Morphol.* 97 (suppl.), 169 (1986).
35. H. G. Frohnhöfer and C. Nüsslein-Volhard, *Nature (London)* **324**, 120 (1986).
36. G. Frigerio *et al.*, *Cell* **47**, 735 (1986); D. Bopp *et al.*, *ibid.*, p. 1033.
37. H. Meinhardt, *J. Cell Sci.* **23**, 117 (1977).
38. M. Mlodzik, A. Fjose, W. J. Gehring, *EMBO J.* **4**, 2961 (1985).
39. J. Mohler and E. Wieschaus, *Genetics* **112**, 803 (1986).
40. A. Kuroiwa, E. Hafen, W. J. Gehring, *Cell* **37**, 825 (1984); A. Weiner, M. Scott, T. Kaufman, *ibid.*, p. 843.
41. P. Macdonald, P. Ingham, G. Struhl, *ibid.* **47**, 721 (1986).
42. U. B. Rosenberg *et al.*, *Nature (London)* **319**, 336 (1986); R. Schuh *et al.*, *Cell* **47**, 1025 (1986).
43. I. Duncan, *Cell* **47**, 298 (1986).
44. B. T. Wakimoto, F. R. Turner, T. C. Kaufman, *Dev. Biol.* **102**, 147 (1984).
45. B. T. Wakimoto and T. C. Kaufman, *ibid.* **81**, 51 (1981).
46. E. Hafen, A. Kuroiwa, W. J. Gehring, *Cell* **37**, 833 (1984); M. P. Weir and T. Kornberg, *Nature (London)* **318**, 433 (1985). The fate map reflects the fate of a cell in the course of normal development at a given location in the embryo. It indicates where the precursors of the mature cells are located.
47. D. Ish-Horowicz, K. R. Howard, S. M. Inchin, P. W. Ingham, *Cold Spring Harbor Symp. Quant. Biol.* **50**, 135 (1985); P. Gergen and E. Wieschaus, *Cell* **45**, 289 (1986); personal communication; K. Kilchherr *et al.*, *Nature (London)* **321**, 493 (1986).
48. K. Harding, C. Wedeen, W. McGinnis, M. Levine, *Science* **229**, 1236 (1985).
49. J. Mohler and E. Wieschaus, *Cold Spring Harbor Symp. Quant. Biol.* **50**, 105 (1985).
50. W. J. Gehring, in *Molecular Biology of Development* (Liss, New York, 1984), pp. 3–22.
51. Y. Hiromi, A. Kuroiwa, W. J. Gehring, *Cell* **43**, 603 (1985).
52. S. B. Carroll and M. P. Scott, *ibid.* **45**, 113 (1986).
53. Y. Hiromi and W. J. Gehring, in preparation.
54. E. B. Lewis, *Nature (London)* **276**, 565 (1978).
55. T. C. Kaufman, in *Time, Space and Pattern in Embryonic Development* (Liss, New York, 1983), pp. 365–383.
56. G. Struhl, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7380 (1982).
57. E. Hafen, M. Levine, R. L. Garber, W. J. Gehring, *EMBO J.* **2**, 617 (1983).
58. M. E. Akam, *ibid.*, p. 2075; ——— and A. Martinez-Arias, *ibid.* **4**, 1689 (1985); R. A. H. White and M. Wilcox, *ibid.*, p. 2035.
59. E. Hafen, M. Levine, W. J. Gehring, *Nature (London)* **307**, 287 (1984).
60. G. Struhl and R. A. H. White, *Cell* **43**, 507 (1985).
61. G. Struhl, *Nature (London)* **292**, 635 (1981).
62. S. Schneuwly and W. J. Gehring, *Dev. Biol.* **108**, 377 (1985).
63. R. E. Denell, *Genetics* **75**, 279 (1973); T. Hazelrigg and T. Kaufman, *ibid.* **105**, 581 (1984).
64. S. Schneuwly, A. Kuroiwa, P. Baumgartner, W. J. Gehring, *EMBO J.* **5**, 733 (1986); S. Schneuwly, A. Kuroiwa, W. J. Gehring, *ibid.* **6**, 201 (1987); L. E. Frischer, F. S. Hagen, R. L. Garber, *Cell* **47**, 1017 (1986).
65. W. Gehring, *J. Embryol. Exp. Morphol.* **15**, 77 (1966); *Dev. Biol.* **16**, 438 (1967); E. Hadorn, in *The Genetics and Biology of* Drosophila, M. Ashburner and T. R. F. Wright, Eds. (Academic Press, New York, 1978), vol. 2C, pp. 555–617.
66. S. Schneuwly, R. Klemenz, W. J. Gehring, *Nature (London)* **325**, 816 (1987). In a similar experiment, Struhl [*Nature (London)* **318**, 677 (1985)] has fused *ftz* to a heat-shock promoter. The ubiquitous expression of *ftz* leads to the deletion of pattern elements in alternate body segments, largely anterior to those deleted in *ftz*$^-$ embryos. This corresponds to a nearly reciprocal phenotype.
67. I would like to thank all the members of my groups who have enthusiastically contributed to this work, in particular, S. Schneuwly, Y. Hiromi, J. Wirz, U. Weber, M. Mlodzik, and H. Krause for allowing me to include their unpublished data, J. Shepherd for the computer analysis and stimulating discussions, C. O'Kane and G. Gibson for critical reading of the manuscript, and E. Wenger-Marquardt for the typing. Supported by a grant from the Swiss National Science Foundation and by the Kantons of Basel.

Tinkering with Enzymes: What Are We Learning?

JEREMY R. KNOWLES

It is now possible, by site-directed mutagenesis of the gene, to change any amino acid residue in a protein to any other. In enzymology, application of this technique is leading to exciting new insights both into the mechanism of catalysis by particular enzymes, and into the basis of catalysis itself. The precise and often delicate changes that are being made in and near the active sites of enzymes are illuminating the interdependent roles of catalytic groups, and are allowing the first steps to be taken toward the rational alteration of enzyme specificity and reactivity.

THE DEVELOPMENTS IN MOLECULAR BIOLOGY OF THE PAST few years have created the opportunity to change essentially any amino acid in any protein (*1*). To mechanistic enzymologists, interested in the origins of the formidable catalytic efficiency of enzymes, this opportunity is proving irresistible. Yet where is the resulting flood of new work leading us? Are we being starry-eyed to allow that "the ultimate goal is to design tailor-made enzymes for every reaction. . ." (*2*)? Are we illuminating existing problems, or merely creating a large number of new ones? There are perhaps 10^6

The author is in the Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, MA 02138.