Programming the *Drosophila* Embryo

**JOHN W. BODNAR**

Chemistry Department, U.S. Naval Academy, Annapolis, MD 21402, U.S.A.

(Received on 27 September 1995, Accepted in revised form on 31 October 1996)

A critical step in understanding the mechanisms of development is in defining the steps at the molecular, cellular, and organismal levels in the developmental program for a given organism—so that given the egg one can predict not only how the embryo will develop but also how that embryo evolved from its ancestors. Using methods employed by chemists and engineers in modeling hierarchical systems, I have integrated current theory and experiment into a calculational method that can model early *Drosophila* embryogenesis on a personal computer. This quantitative calculational tool is simple enough to be useful for experimentalists in designing experiments yet detailed enough for theoreticians to derive new insights on the evolution of developmental genetic networks. By integrating the strengths of theoretical and experimental methods into a single engineering model that can compute the cascade of genetic networks in a real organism, I provide a new calculational tool that can apply current theory to current experimental data to study the evolution of developmental programs.

© 1997 Academic Press Limited

1. Introduction

Over the last 20 years, progress in developmental biology has been so dramatic that developmental biologists may be excused for having the view, possibly an illusion, that the basic principles are understood, and that the next 20 years will be spent filling in the details. The most significant advances have come from the application of molecular techniques and a greatly improved understanding of cell biology. So we can begin to ask questions—like whether the egg is computable... How many genes control development—as distinct from providing the housekeeping functions of the cell? ...Will the egg be computable? That is given a total description of the fertilized egg—the total DNA sequence and the location of all proteins and RNA—could one predict how the embryo will develop? This is a formidable task, for it implies that in computing the embryo, it may be necessary to compute the behavior of all constituent cells. It may, however, be feasible if a level of complexity of description of cell behavior can be chosen that is adequate to account for development but that does not require each cell’s detailed behavior to be taken into account... It is not unreasonable to think that enough will eventually be known to program a computer and simulate some aspects of development.

Lewis Wolpert (1994)

There is an enormous literature on development, both theoretical and experimental, yet there is still a gap between the theory and experiment. While overall mechanisms of development are well accepted such as the role of morphogens, gradients, thresholds, and genetic networks, these generic theories cannot yet “compute the egg” of any specific organism well enough to serve as a useful tool for experimentalists. For example, geneticists have isolated about 30 genes that control early embryogenesis in *Drosophila*, and it would be extremely valuable for planning further experiments for the theoreticians to tell them how many they are missing—two, ten, or a hundred? On the other hand, while experimental data is accumulating on many model organisms at an ever-increasing rate, the catalog of genes and genetic interactions appears so complex that theoreticians cannot yet use

*E-mail: Bodnar@nadn.navy.mil*
that detailed data to refine the generic theories into a specific testable computational model for embryogenesis—and adding new data only makes the problem seem more complicated. For example, it would be just as valuable to the theoreticians for the experimentalists to tell them exactly how many genes are needed to model the genetic network for Drosophila early embryogenesis. Thus a workable method to “compute the egg” for a specific organism—however crudely—could provide experimentalists with a very valuable tool to plan experiments and provide theoreticians a way to apply generic theories to specific organisms.

Using methods employed by chemists and engineers in modeling hierarchical systems, I have integrated current theory and experiment into a calculational method that can model early Drosophila embryogenesis on a personal computer. This quantitative calculational tool is simple enough to be useful for experimentalists in designing experiments yet detailed enough for theoreticians to derive new insights on the evolution of developmental genetic networks.

2. Background

MODELING HIERARCHICAL SYSTEMS

“ON THE BACK OF AN ENVELOPE”

The development of an organism is a dynamic process that occurs simultaneously at the molecular, cellular, and organismal levels. Thus, any model for development must be detailed enough to reflect to biology at all three levels yet, to be useful, must still be simple enough to retain an overview of the entire process. While this seems like a daunting task for such a complex process, chemists and engineers routinely address such complexity using semi-empirical calculational methods. Chemists don’t calculate the Shrödinger equation to estimate the effectiveness of a new antibiotic, and engineers don’t calculate the Shrödinger equation to estimate the effectiveness of a new antibiotic. Instead, they use calculational methods that can model early Drosophila embryogenesis “on the back of an envelope” in order to define a single dynamic model that is consistent with both theory and experiment at the molecular, cellular, and organismal levels. The logic behind this method is drawn from a chemical and engineering approach to modeling a hierarchical system and extends the chemical model up to biochemistry, cellular, and developmental biology.

(i) Objects form a hierarchy where any object is composed of individuals that interact to form a unit. Each unit is itself an individual which can form a higher order unit. Protons, neutrons, and electrons, combine to form atoms. Atoms form molecules such as sugars and nucleotides. Molecules form macromolecules such as proteins and DNA. DNA forms genes and defines gene products (proteins). Genes and gene products form cells, and cells form organisms.

(ii) Units can be defined by a characteristic list of individuals that comprise their structure. Atoms can be defined by knowing the number of protons they contain. Sugars, nucleotides, and amino acids can be defined by the arrangement of atoms that comprise them. Genes can be defined by their nucleotide sequence, and gene products can be defined by their amino acid sequence. Cells can be defined by the genes and expressed gene products that comprise them. Organisms can be defined by the combination of cells that comprise them.

(iii) Individuals can be defined by parameters that describe their function. Parameters describe functional aspects of an object that are not present in any of the individuals that comprise it. These include constants that summarize that individual’s potential to interact to form a unit at the next higher level—in space as binding constants and in time as rate constants. Examples are electronegativity for atoms, dipole moment and transition state energy for molecules, or Km and Vmax for enzymes.

(iv) An object is described structurally as a unit by the parameters of the individuals that comprise it but functionally as an individual by its own parameters. The inherent duality of every object, both as an individual and unit, is the key to integrating theory and experiment into semi-empirical methods to describe hierarchical systems.

(v) Theoretical methods can use the parameters of individuals to calculate parameters of units. The Shrödinger equation begins with the number of electrons that comprise a given atom and calculates...
how they interact as they fill orbitals thereby defining the reactivity of a particular type of atom. Similarly, one can use binding and diffusion constants for genes and DNA-binding proteins to calculate the patterns of molecular concentrations in a morphogenic field.

(vi) Parameters can be determined for individuals directly by experiment. Usually the theoretical models give only approximate answers for anything but the most simple units, so parameters are usually measured for those cases, looking at those objects as individuals. For example, one can calculate bond energies from the Schrödinger equation, but it usually more accurate to measure those experimentally using a series of model compounds such as O₂, CO₂, CH₄, CH₃-CH₃, etc.

(vii) A look-up table of parameters is constructed experimentally to provide input to calculate the parameters of the higher level unit that can now be formed. When modeling the interactions among molecules, the parameters of the individual molecules depend on the arrangement of the atoms that comprise them. Calculationally, it is usually faster and more accurate not to begin ab initio with the Schrödinger equation, but rather to select characteristics for each individual bond from a look-up table of bond energies and bond lengths determined experimentally. Any introductory chemistry text is full of tables of thermodynamic data, equilibrium constants, and rate parameters that allow calculation of the hierarchy of subatomic particles/atoms/molecules/solid, liquids, gases.

(viii) Digital parameters rather than analog equations are used in hierarchical models whenever possible. Not only do look-up tables allow more accurate calculations because they are matched against the real system at every level, but digital computers are faster in extracting discrete values from those tables than in calculating continuous functions everywhere in space. Additionally, since each object has dimensions, calculating a parameter as a continuous function at a scale smaller than the object has no physical meaning, and continuous models usually calculate parameter values at a series of points on the scale of the object, e.g. a concentration gradient used to calculate diffusion of molecules means nothing on an atomic scale.

Semi-empirical models in chemistry and engineering are extremely powerful because they draw from the strengths of both theory and experiment: theory provides the tools to calculate a generic multi-level system while experiment provides real parameters that describe a specific system. This makes the models themselves meta-experiments where calculations are continuously refined for a best fit to experiment at all levels in the hierarchy.

THEORETICANS, EXPERIMENTALISTS, ENGINEERS, AND MODELS

To be useful a model must be to be “transparent” (Turkle, 1995), but that means something totally different to theoreticians, experimentalists, and engineers. Theoreticians desire “transparent” models so that they can look inside them to determine how they work, where they came from and how they compare with other models. On the other hand experimentalists desire “transparent” models so that they can look through them to determine how the system works. For example, if one discusses the parameters in a model, theoreticians will want to know how they were calculated while experimentalists will want to know how well the measured values can describe the system and predict new experiments. Engineers need to predict the properties of a system under conditions which not only have not yet occurred but hopefully will never occur, e.g. how much wind shear an aircraft can take and still land safely. Therefore, engineers desire “transparent” models so they can look beyond them to calculate how the system interacts with the environment. Since my modeling efforts are aimed at reconciling theory and experiment, I have divided the exposition of my method into sections which each consider the developmental “model” by a different definition. To begin a dialog among the theoreticians and experimentalists, we need to account for those differences in viewpoint, so I have addressed each outlook in turn. I have then examined how the semi-empirical approaches of the chemists and engineers can unify theory and experiment to something beyond either.

In Section 3, I describe development as a theoretician. I indicate how molecular parameters can be used to calculate cellular parameters and how cellular parameters can then be used to calculate organismal parameters. Here I compile ideas from the theoretical community to form an integrated hierarchical theoretical model. Accordingly, I reference the most cogent references for each idea so that the reader can trace the lineage of the methods, and I suggest Held (1992) for an in depth overview of development from the standpoint of the theoretical community.

In Section 4, I describe development as an experimentalist. I insert experimental molecular and cellular parameters for Drosophila embryogenesis into the model, calculate organismal parameters, and describe the consistency of experimental results and where new experiments would be useful. Here I
compile data from the Drosophila experimental community to form an integrated experimental model. Accordingly, I reference the most cogent references for each parameter so that the reader can trace the sources of the data, and I suggest De Pomerai (1986), Gilbert (1991), or Lawrence (1987) for an in depth overview of Drosophila development from the standpoint of the experimental community.

In Section 5, I describe development as an engineer. I show how the model can be manipulated to calculate how the organism can change in response to changes in molecular parameters. Here I indicate how the model can be used to study mutagenesis and evolution of developmental pathways.

3. Theory of Embryogenesis

The successes of molecular developmental biology over the last ten years have been particularly impressive in those directions favored by its major paradigms . . . Results from several well-studied systems show that morphogenesis of specific structures that are composed of various substructures begins with the regional sets of transcription factors such that each region defines the cells that will ultimately produce a working part of the structure (transient or ultimate). Thus the sole function of this initial process appears to be to install different regulatory states in the territories from which the different parts will develop.

Eric Davidson (1994)

A LOGICAL METHOD TO CALCULATE POSITIONAL VALUES

While the visible mechanisms of development appear to be quite complex and diverse, there is growing evidence that the molecular pathways are both simple and highly conserved. For example, the eyeless gene in Drosophila appears to be a master regulator that can program entire eyes in unusual positions by ectopic expression, and the eyeless gene product is conserved among flies, mice, and nematodes (Halder et al., 1995) suggesting that the basic molecular mechanisms for programming eye development predated the divergence of arthropods and vertebrates. Therefore, a theoretical model for development that can integrate the effects of molecular, cellular, and organismal processes should be able to be applied directly to a wide variety of organisms.

By compiling current ideas in the theoretical community at each level of the hierarchy then applying a chemist’s or engineer’s approach to building hierarchical models, I have formed a hierarchical theoretical model for development. This model considers how molecular parameters can be used to calculate cellular parameters during embryogenesis and how those cellular parameters can be used in turn to calculate the organismal parameters in the growing embryo.

Chromatin switching and the cell cycle

The life of the cell is a cyclic process in which specific genes are expressed to allow the cell to maintain itself and divide into progeny cells. The cell cycle has both genetic and epigenetic components where the identities of the progeny cells are determined both by the genomes and the proteins they contain. In defining the epigenetic components of the cell cycle, I made several assumptions.

(1) Gene regulation can be described as a hierarchy of processes, and each process can be described as a “black box” module in a genetic network (Mittenthal, 1989; Edgar et al., 1989; Reinitz & Sharp, 1995; reviewed in Davidson, 1994). Many steps in gene regulation can be simplified by knowing the inputs and outputs without describing the intermediate process. For example, we can describe the synthesis of a protein in terms of its inputs (the DNA transactivators that bind its gene sequence) and output (the polypeptide itself) without describing the generic process in between (transcription, mRNA processing and transport, and translation). Thus, one can describe networks of interacting genes on the basis of how outputs of one process (transactivator proteins) become the inputs for the next process (transcriptional regulation by binding to DNA sequences of the next gene in the network).

(2) Chromatin structure allows gene switching mechanisms that can activate specific genes only as the cell passes through a cell cycle in the presence of a particular combination of regulatory genes (Boednar & Bradley, 1996). The competition between nucleosomes and transactivators to repress and activate DNA can allow chromatin to act as a switch between inactive and active conformations. One variant of this “chromatin switch” [described in detail in Bodnar & Bradley, (1996)] can be switched only during DNA replication when the appropriate transactivators are available. We can assume that transactivators synthesized in G1 of the cell cycle cannot displace nucleosomes independently but when present can compete for the naked DNA directly behind the DNA replication fork. This means that gene controlled this way can change chromatin states only during S phase of the cell cycle and can, therefore, be “switched” only once per cell cycle.
The “chromatin switch” provides a simple threshold mechanism to convert an “analog” concentration gradient into “digital” gene activation signal (Lewis et al., 1977; Bodnar & Bradley, 1996). For reproducible switching of genes across a morphogen gradient, there must be a way to specify discrete gene states in continuous concentration gradient. The chromatin switch will only be flipped from inactive to active at certain threshold concentrations of transactivator. Therefore, if cells or nuclei are passing through S phase in the presence of a transactivator gradient only those genes in cells where the transactivator exceeds the threshold concentration will switch to active chromatin. Thus, we need not know the exact transactivator concentration to know if it can switch its target genes, only whether it is above or below a threshold.

In a simple epigenetic model of the cell cycle, one must follow two states for each gene: a protein state defined as the concentration of the protein product of a gene at a given time, and a chromatin state defined as the chromatin conformation of a gene at a given time. One can then calculate the values of both the protein and chromatin states for a gene as the cell passes through each cell cycle (Fig. 1).

Given the threshold phenomenon possible through chromatin switching, gene switching during the cell cycle can be treated as a simple digital switching network dependent on the value of the protein state (protein value) and the value of the chromatin state (chromatin value). All genes and the protein products transcribed from those genes were assumed to have four possible states: 0 = gene totally repressed or no protein expressed, 1 = chromatin switch state or expressed protein concentration above a low threshold; 2 = state or concentration above an intermediate threshold, or 3 = state or concentration above a high threshold [Fig. 1(a)]. If the input gene (gene A) is a strong activator of the target gene (gene B), then a gene A protein value of 0 will have no effect on the chromatin value of B, but a gene A protein value of 1, 2, or 3 will return a gene B chromatin value of 3 and will return a gene B protein value of 3 during G1 of the next cell cycle. Similarly if gene A is a strong repressor of gene B, an input gene A protein value of 0 will not affect the gene B chromatin value, but a gene A protein value of 1, 2, or 3 will repress gene B to a return chromatin value of 0 (and a corresponding protein B value of 0 on the next G1). If gene A was

![Fig. 1. Boolean rules for calculating chromatin states. (a) Gene switching depends on gene accessibility as a function of its chromatin state. This can be approximated by a threshold mechanism (Bodnar and Bradley, 1996) where the concentration of a target gene depends on whether input regulatory gene products are above or below specific thresholds. Thus the “chromatin value,” the state of chromatin accessibility for a target gene, is switched each cell cycle by the “protein value,” protein concentration of an input activator or repressor. (b) Threshold activation or repression reduces the input/output rules to a Boolean logic. The Boolean chromatin switching rules are summarized in terms of their input gene type, input protein value, and the return chromatin value for the output gene. These rules and shorthand notation are used for all calculations.](image)
only an intermediate or weak activator or repressor of gene B, the requirement of many more DNA binding sites for gene A on gene B sequences should result in higher concentrations of gene A to affect gene B. This also would result in a sigmoidal activation/repression curve that may not be a single step gradient but which can be calculated with additional thresholds [Fig. 1(a)]. For example, if gene A is an intermediate repressor of gene B, protein A = 0 will have no effect, protein A = 1 will return gene B (chromatin value) = 1, but a value of protein A = 2 or 3 is required for complete repression to gene B = 0. In this case if gene B is already at 0 and if protein A = 1, then the return chromatin value of B must be left unchanged at 0 because a repressor will not increase the value of the target gene. The logic for the other types of activator and repressor is similar, and a complete list of the Boolean gene switching rules for input gene effect on the target gene used throughout all calculations is shown in Fig. 1(b).

Therefore, a simple epigenetic model of the cell cycle program tracks the cell state at any time as a function of the protein states and chromatin states of all the genes in the cell (Fig. 2). The molecular program for the cell cycle then becomes a series of processes which change the cell state [Fig. 2(a)]:

G0 = resting cell; no change in cell state.

New transcription may occur through extracellular signalling but will be neglected (or lumped with G1) in the generic calculation.

G1 = transactivator synthesis; new protein state, but chromatin state is unchanged.

The combination of transactivators synthesized is determined by the chromatin state.

S/G2 = DNA synthesis; new chromatin state but protein state is unchanged.

The combination of genes switched to new chromatin states as the replication fork passes is determined by the protein state.

M = cell division; daughter cells start with protein and chromatin states determined at cell separation.

Identical chromatin states will be passed onto both progeny.

Protein states may be identical or differ by unequal segregation of proteins.
The cell cycle program can be then described by an equivalent logical program [Fig. 2(b)] which can be calculated in a computer programming language as a single DO loop.

- DO for each cell cycle.
  - Display the current values for all gene products.
  - G1: Calculate values for protein state (transactivator concentration) using the current chromatin state for each gene.
  - S/G2: Calculate new values of chromatin state for each gene using current protein state.
  - M: Determine the new position of each daughter nucleus (with same or different protein and chromatin states as the parent).
    (Divide and migrate nuclei by division rules for that particular cycle).
  - [Repeat for each cell cycle].

**Setting positional values in the egg**

The concept of positional information was first formalized by Lewis Wolpert (1969) and has been extensively studied as a mechanism for developmental regulation.

(i) *Positional information allows cells to know where they are via coordinates which they “interpret” as particular states of differentiation* (Wolpert, 1969; reviewed in Held, 1992). Positional information is a concept which proposes that cells are able to recognize their relative positions within a coordinate system. Those cells can recognize discontinuities in the coordinate system and respond by pattern regulation (Malacinski, 1984).

(ii) *Localized determinants stored in gradients in the egg define positional information that is the basis for cell determination* (Kauffman et al., 1978; Kauffman, 1981). Each position in the egg can be assigned a specific “identity” in terms of the concentrations of the regulators. If the gene regulators activate other genes in a combinatorial manner, then each individual combination of concentrations will determine a specific cell fate at that location (Bodnar et al., 1989).

(iii) *Gradients in the egg specify positions by a series of threshold values* (Kauffman, 1981), and morphogen threshold values divide the egg into compartments that provide biochemical boundaries for gene switching (Kauffman et al., 1978; Bodnar et al., 1989). If initial gene activation in the developing embryo depends on chromatin switching and the transactivators necessary for that switching are stored in gradients within the egg, then activation of genes in individual cells will depend on whether the concentration of the required transactivator is above or below the threshold needed to “switch” the chromatin of the gene in that cell.

(iv) *All cells record their positions as a numerical quantity called a positional value* (Held, 1992). The positional value is the information a cell is assumed to possess about its relative position within an organization framework or structure (Malacinski, 1984). The positional value will then be a measure of the state of an individual cell at its current position. Since the state of a given cell can be defined in terms of its current protein state and chromatin state, I assume a positional value can be expressed as a sum of its protein value plus chromatin value. For example, if one describes each of these values as a string variable of the form [ABCDE], then the cell in Fig. 2 at the start of G1 will have a protein value of [11000] and chromatin value of [01100] and its positional value can be described by these two strings.

(v) A positional value is an epigenetic quantity that will change with chromatin switching and protein expression throughout the cell cycle. The example cell (Fig. 2) begins with a protein value of [11000] for [ABCDE] at the beginning of G1; then an extracellular signal will activate protein B to activate expression of gene C during G1; the C protein can then switch on the chromatin of gene D during S; following cell division both progeny cells will then have a protein value of [11110] as they begin G1. As these cells divide and move, their positional values will continue to change. Therefore, the positional value is both a position-dependent and time-dependent parameter that reflects not only the position of the cell within the organism but the time in development as well.

(vi) *Sequential subdivision of the early embryo into compartments encodes a somatic commitment in each region in a combinatorial epigenetic code* (Garcia-Bellido, 1977; Kauffman, 1981). Gierer (1973) proposed a general combinatorial scheme to encode differentiated states in metazoans, and this concept has been developed by many others (reviewed in Held, 1992) to indicate that the positional value can define “compartments” where the numerical values of the positional values are identical, i.e. all cells are localized on the same side of the thresholds for all regulatory genes in the positional value. If a specific positional value can activate a specific “selector gene” which activates the somatic genes that define a particular cell type, then the positional values will form a combinatorial epigenetic code that will define “compartments” of the same cell type by the same positional value.

(vii) *Rapid cell or nuclear divisions in the embryo prior to onset of transcription allow efficient**
interpretation of the preformed gradients before subsequent gene cascades are activated (Bodnar et al., 1989; Davidson, 1994). If a rough gradient of a transactivator protein is found within the egg, the genes that it will activate in a given cell nucleus will depend on whether that nucleus is in a region of the dividing embryo where the transactivator is above a threshold value. If the gradient is not “read” for six nuclear divisions by delaying the activation of transcription, the gradient is sensed when there are $2^6$ or 64 cells, and it is likely that an approximately reproducible number will be found in the correct “compartment” stochastically—merely by the numbers involved.

Initial positional values in an egg are set by the concentration of a morphogen which is subdivided among the cells of the growing embryo (Fig. 3). The egg usually begins growth by three mechanisms—each of which occurs prior to activation of transcription for zygotic genes [Fig. 3(a)]: a blastula where several rapid cell divisions occur; a blastoderm where several rapid nuclear divisions occur within a syncytium; and a growth zone where a single progeny of each cell division divides rapidly. In the blastula and blastoderm the morphogens do not diffuse as the cells or nuclei divide and move to differing regions of the original cytoplasm while in the growth zone the morphogen is diluted as progeny cell grows and divides (or alternatively the morphogen is asymmetrically divided at each division); the net result in each case is a movement of cells or nuclei versus a morphogen gradient. I suggest that all of these embryonic cellular programs correspond to a single logical program [Fig. 3(b)] which rapidly divides the pre-existing pattern of morphogen into multiple compartments with differing positional values that contain at least one cell or nucleus. Then when transcription of zygotic genes begins, the protein value, in terms of morphogen concentration(s), differs across the embryo, and different cells have different positional values. These positional values are the initial conditions used to calculate cell fates in the growing embryo.

**CALCULATING COMMON DEVELOPMENTAL PROGRAMS**

In the central area of gene control in developmental processes, a major issue arising is the nature of gene regulatory networks: their complexity and structure; their use of widely shared subelements; their response to perturbation; their linkage; and their degree of degeneracy . . . As various authors have pointed out, modular regulatory circuit subelements are probably used to control a large number of entirely different downstream processes . . . Soon we will probably have in our hands various semi-complex, modular,

---

**Fig. 3.** Setting initial positional values for the embryo. Initial compartmentalization of morphogens in the fertilized egg sets the positional values for the growing embryo. (a) Three common mechanisms divide the morphogen in a rapid series of divisions into a gradient of concentrations that then serve as initial transactivator states for transcribing zygotic genes. (b) These correspond to a logical program to divide initial positional values that then serve as initial protein values for calculating zygotic protein and chromatin states in embryonic cells.
Fig. 4. Programming the embryo. Once initial conditions are set to determine positional values (Fig. 3) transcription can begin in the embryo and the cell cycle programs a sequential protein value and chromatin value (Fig. 2) according to Boolean switching rules (Fig. 1). Dependent on the initial conditions for protein values several generic patterns can be formed over the course of a few cell cycles. (a) The French flag mechanism divides the embryo into a series of stripes (usually two parasegments or rhombomeres wide) as shown for a syncytial blastoderm. Formation of initial stripes in a blastula can be accomplished in virtually the same way. (b) Using very similar switching rules and morphogen gradients the French flag mechanism can be modified to form stripes in growth zones as seen in other embryos and limbs. (c) The Stripe-doubling mechanism subdivides the first stripes and sets the program for individual parasegments. (d) The Line-drawing mechanism can then define the edges of the segments.

JOHN W. BODNAR (facing p. 398)
‘integrated circuits’ that are to some extent experimentally interchangeable.  

Eric Davidson (1994)

By defining molecular programs for the cell cycle and cellular models for setting positional values as logical models (Figs 2 and 3) one can construct a logical program for defining chromatin and protein states for any cell in an organism as it grows and divides. All one needs to define particular kinds of genetic networks that form specific cell patterns are specific inputs for expression of each gene as Boolean rules and specific initial patterns of morphogens as digital positional values. Then the logical program can calculate the molecular program as the egg divides and the embryo grows.

Three common mechanisms appear to be found in many organisms for setting up particular types of patterns—which I call the “French flag” mechanism, the “stripe-doubling” mechanism, and the “line-drawing” mechanism (Fig. 4). Each of these depends on a simple genetic network (Edgar et al., 1989; Reinitz & Sharp, 1995).

The French flag

A classical problem of stripe formation in development has been called the “French flag” problem in theory of positional values [reviewed in Held, (1992)]. As Wolpert (1992) has stated it, “Imagine a line of cells each of which can turn blue, white, or red. The problem is to invent a mechanism which generates a pattern that looks like a French flag, that is with the first third of the line blue, the middle third white, and the last third red.” A mechanism which can reproducibly program a three stripe pattern like the French flag can then be used with minor modifications to produce much more complex patterns (Meinhardt, 1978, 1986; Meinhardt & Gierer, 1980).

As shown in Fig. 4 forming a reproducible three stripe “French flag” pattern requires a single input gene (gene A) in a gradient and three target genes (genes B, C, and D) that will ultimately form the three stripes. The series of genes forms a activation-repression network with the following characteristics: (1) there is a sequential activation starting with the gradient input gene (A→B→C)*; (2) there is feedback inhibition of the target genes (D→C→B) which defines the downstream border of each target gene stripe, i.e. D→C defines the C border on the side with D; and (3) there is feedforward inhibition of the target genes (A→C→D) which defines the upstream border of each target gene, i.e. A→C defines the C border on the side with B.

One can step through the formation of the French flag pattern manually by this mechanism as shown in Fig. 4(a) for a blastoderm or blastula and Fig. 4(b) for a growth zone. In an egg that will form a blastoderm [Fig. 4(a)] the concentration of input protein for gene A is a step gradient that divides the egg in an anterior/posterior direction at three thresholds while protein for target gene D is found homogeneously throughout the egg. This divides the egg into four compartments with different initial protein values: [ABCD] = [3003],[2003],[1003],[0003]. During S phase of the first cell cycle proteins A and D are available to compete with chromatin and each other on the DNA of genes B, C, and D. Any nuclei in a given compartment sense the local protein concentrations and switch the chromatin of all target genes in accordance with the Boolean gene switching rules as the DNA replication forks pass. Protein A switches on gene B (A→B) in the first three compartments while protein A inhibits gene D (A→D) in a gradient and totally abolishes expression of D in the first anterior compartment. The positional values in each compartment also change because they are epigenetic quantities dependent on the chromatin configurations of the nuclei in those compartments.

During S phase of the second cell cycle interactions begin among the target genes. Target gene B is still activated by A. Gene C is activated as a target of protein B (B→C) in the two middle compartments, but activation of C is inhibited by feedback repression by A (A→C) in the first compartment, and the levels of C are limited by A in the second compartment and D (D→C) in the third compartment. The protein D gradient is maintained by autoregulatory activation of D itself and continued feedforward repression by A. Thus at the beginning of cell cycle three target gene interactions form three overlapping stripes of proteins B, C, and D.

During the third cell cycle the stripes resolve to a sharp three stripe pattern. The accumulation of protein C during the previous cycle now can sharpen the boundaries of the stripes during S phase of cell cycle three. In the middle two compartments the strong repressor protein C abolishes any expression of genes B and D by feedback and feedforward inhibition respectively. In the first and fourth compartments the stripes of B and D are maintained as in the previous cell cycle.

*Abbreviations for gene interactions are given in Fig. 1(b), e.g. A→B means A is a strong activator of B and D→C means D is a weak repressor of C.
After the third cell cycle a “French flag” with three sharp stripes of proteins B, C, and D is formed and will be maintained in any subsequent cell cycles because the combination of sequential activation, feedback, and feedforward forms a stable network of gene interactions. In the first compartment gene B is maintained by autoregulatory activation but expression of protein A inhibits any new transcription of C and D; in the second and third compartments gene C expression is maintained by autoregulatory activation but protein C represses both genes B and D; finally in the fourth compartment protein D is maintained by autoregulatory activation and D itself inhibits any new transcription of genes B or C. A French flag can also develop in an embryo with a growth zone by minor changes in gene inputs and rules for cell division [Fig. 4(b)].

In summary we can use chromatin switching theory coupled with the concept of positional information to devise a simple molecular mechanism to determine a French flag three-striped pattern. This mechanism can easily be expanded to make four or more stripes by addition of target genes regulated in the same way (e.g. a gene B’ activated by B with feedforward repression by a and feedback repression by C).

The stripe-doubling mechanism

The second common mechanism for pattern formation in embryos of segmented organisms is a stripe-doubling mechanism where stripes that are two segments or rhombomeres wide are subdivided into two stripes each determining an individual segment or rhombomere.

Doubling the number of stripes in an embryo [Fig. 4(c)] is possible with only minor modifications to the processes just described. In a field of overlapping stripes the total number of stripes can be doubled if one new stripe is formed at each overlap of the old stripes then one new stripe is formed in the center of each old stripe (Meinhardt, 1986; Nagorcka, 1988; Reinitz & Sharp, 1995). The simplest stripe-doubling combination [Fig. 4(c) as an example for a blastoderm] requires four genes that form a sequential inhibition network: all four genes are activated by upstream input genes which define the first level in the hierarchy; two genes are the second level of hierarchy (genes B and C) and form adjacent stripes by mutual inhibition where the inhibition is sufficiently weak to allow an overlap of the stripes; the third gene (gene E) is downstream (third level) such that it is inhibited directly by both striped genes; the fourth gene (gene F) is still further downstream (fourth level) such that it is inhibited directly by the third gene. The rules for these interactions are surprisingly simple once the first set of overlapping stripes is formed.

The stripe-doubling mechanism begins at the junction between two overlapping stripes [genes B and C in Fig. 4(c)]. Both downstream genes (genes E and F) are activated on the first cell cycle by the input genes. Repression of gene E by proteins B and C causes a striped pattern where gene E is expressed in the junctions between B and C (where their concentrations are low) but repressed by high concentrations of B and C in the centers of the stripes. Since proteins B and C have no direct effect on the furthest downstream gene (gene F), that gene is expressed throughout the entire region of both B and C stripes.

During the second cell cycle alternate bands of the downstream genes form with tightly defined borders. Going into S phase of that cell cycle gene E is found in stripes at the junctions of B and C; if gene E now maintains its own expression by autoregulation, those stripes remain stable as long as B and C are expressed. However, gene F—which began expression throughout the entire region—is strongly inhibited by protein E so that F is repressed in the stripes just formed by E. In essence, the final gene in this pathway (F) is expressed everywhere before the upstream gene (E) is present, and the upstream gene “punches holes” in the homogeneous expression pattern of the final gene. This results in a stable pattern of alternating stripes of the downstream genes (E and F) with the total number of stripes (E + F) equal to twice the number of stripes of either the upstream genes (B or C). Also, in genetic studies the upstream genes (B and C) will appear to activate the last gene (F), but such activation is indirect—through lifting inhibition of the last gene (F) by the gene (E) in the middle of the hierarchy.

The line-drawing mechanism

The third common mechanism for pattern formation in embryos is a line-drawing mechanism where a single line a cells is determined at the juncture of two compartments. This can be accomplished if cells reciprocally activate or exclude gene activation along a boundary (Meinhardt & Gierer, 1980).

Drawing a line of cells [Fig. 4(d)] requires the addition of cell-cell interactions to the mechanisms described above. The overall process begins with upstream input genes that activate the second level gene [gene F in Fig. 4(d)] and one third level gene (gene G). The second level gene (F) is regulated to form an asymmetric stripe which has a very high level at one edge trailing off to no expression in an interstripe region. This second level gene is a strong
repressor of one third level gene (G) and very weak activator of the other third level gene (H). When the asymmetric stripes of gene F form, gene G can be expressed in narrow stripes in the interstripe regions (as a stripe-doubling mechanism). However, protein G is a secreted autocrine or paracrine factor that can bind receptors on the same or adjacent cells, and receptor binding allows those cells to continue through the cell cycle (or allows activation of pre-existing transactivator for gene H). The final gene in the pathway (H) is weakly activated by protein F but only if the protein G factor is bound to receptors on the same cell. This only occurs at the asymmetric stripe boundary where gene G expression is highest; there protein F concentration is high enough to activate gene H and in the adjacent cell protein F concentration is low enough to lift repression of gene G. On the other boundary gene G may be expressed in the interstripe region, but no gene H expression is possible because protein F levels are too low. Thus, only in the cells “next” to one boundary of the protein F stripe can gene H be expressed by gene F expression within the cell and protein G secretion from the adjacent cell. The result is a single line of cells expressing gene H directly adjacent to a narrow band of cells expressing gene G, and the diagnostic pattern is a G/H double stripe with a defined polarity. Further feedback of gene H expression can “lock in” gene G expression by activation of another paracrine factor by gene H to bind to the receptors of the cells secreting gene G to allow continued expression of gene G. This allows the stable expression of the G/H line even after fading of upstream signals; this feedback is neglected in my simple genetic networks.

4. Parameters of Drosophila Embryogenesis

A compilation of ideas from the theoretical community has been used to assemble a hierarchical theoretical model for development. This provides a tool to calculate the molecular, cellular, and organismal parameters for common cellular patterns found in a variety of growing embryos. However, modeling a specific organism requires the integration of experiment. If the theoretical model is constructed well, the semi-empirical approach will only require inserting measured parameters in the experimental embryo for the calculated parameters in the theoretical embryo. In the hierarchical model for embryogenesis this will require constructing a lookup table for the regulatory genes in the organism and the parameters for their interactions and lookup table for the distribution of morphogens in the egg. With those in place one should then be able to compute the egg.

Calculating the Drosophila Program from Egg to Blastoderm

The development of the Drosophila blastoderm is one of the most intensely studied experimental systems for understanding pattern formation. Current data suggest that pattern formation in the Drosophila embryo depends mainly on transcriptional control of a few master regulatory genes that are used combinatorially to define cell fate (Garcia-Bellido, 1977; Kauffman et al., 1978; Martinez-Arias & Lawrence, 1985; Carroll et al., 1988). Therefore, I have chosen to use Drosophila blastoderm development as a specific model to test the new calculational method on a real organism—with the realization that most experimental tests of these models were likely already complete for Drosophila and need only to be retrieved from the literature, compiled, and compared with the theory. Accordingly, a Drosophila embryogenesis pilot project was completed using a MacIntosh Plus personal computer and Turbo Pascal which provided all the computing power necessary to integrate theory and experiment by this simple calculational tool.

Given the enormous amount of theory, data, collation, and analysis required in this project, I present results only as they relate to the problem at hand—how can the Drosophila egg be computed? Therefore, I discuss the general aspects of the program and confine the details of the individual Drosophila genes and mutants to Appendices, and I reference the most comprehensive papers and reviews rather than attempting to give credit to the thousands of investigators upon whose work I base my analysis. This not only keeps manuscript length and the reference list manageable but also provides two ways to read the manuscript: those interested in an overview of the method to study the epigenetics of development need only read a general description and use the most informative sources for further reading while those interested in the ability of the method to direct further experimentation on Drosophila can use the detailed descriptions in the Appendices for further reference—with detailed references for the gene interactions in Appendix A, detailed descriptions of the calculations for each individual gene (along with hypotheses on new experiments or on the evolution of the individual genes) in Appendix B, and the results of calculations for mutants in Appendix C.

Data on Drosophila embryogenesis indicate that it is a sequential stepwise process [reviewed in De Pomerai (1986), Gilbert (1991) and Nüsslein-Volhard (1991)]. The fertilized nucleus undergoes nine rapid synchronous nuclear divisions within the ooplasm.
whereupon the nuclei migrate to the periphery of the egg and divide five more times; at this point cytoplasmic membranes form around the individual nuclei to form cells. The embryo containing dividing nuclei for the first 13 nuclear divisions is termed the syncytial blastoderm while the embryo formed by cellularization at mitotic cycle 13 is termed the cellular blastoderm [reviewed in De Pomerai (1986) and Schweisguth et al. (1991)]. The cellular blastoderm undergoes several more cycles of cell division prior to gastrulation and germ band extension, but here I will only address the processes involved in the formation of the cellular blastoderm. Deducing the rules for genetic interactions in blastoderm formation seemed possible because, as Christiane Nüsslein-Volhard and co-workers have noted, “first, the number of genes that are specifically involved in the establishment of positional information in the egg is quite small. About 30 genes have been identified so far, and the total number is unlikely to be much higher than this. Second, the two body axes are established independently, as mutations either affect the anterior-posterior pattern or the dorsal-ventral pattern, but never both. Third, the number of embryonic phenotypes observed is much smaller than the number of genes (St. Johnston & Nüsslein-Volhard, 1992)”. The cellular blastoderm is an excellent waypoint in the study of Drosophila embryogenesis for several reasons: (1) it is defined by the last synchronous nuclear division; (2) it is the last stage at which the fate map can be projected directly as a plane; (3) it is the last step when cell-cell interactions can be neglected; and (4) it is the step for cell determination. The synchronicity of nuclear division means that all nuclei have the same cell cycle duration so simulating nuclear division in the syncytial blastoderm can assume all nuclei have the same division program. At the cellular blastoderm a single layer of cells is found on the surface of a roughly cylindrical surface so that a Mercator projection (Dunlap & Shufeldt, 1972) of this surface made by cutting the cylinder down opposite sides (Fig. 5) will result in a plane of cells representing one side of the embryo; this can be easily modeled as a rectangular grid of cells. This planar representation can later be mapped against the mitotic and phenotypic fate maps of the cellular blastoderm to integrate the genotype and phenotype (Bodnar, manuscript in preparation). Cell membranes do not form around the syncytial nuclei until the 13th nuclear division; therefore, up to that point the role of cell-cell interactions can be neglected because there are no distinct cells. The cellular blastoderm is also the time of cell determination for the majority of cell types in the embryo; if a nucleus from the syncytial blastoderm is transplanted into another embryo at a new location, it assumes the identity of the cells at the location of transplant, but if a cell from the cellular blastoderm is transplanted to another embryo in a new location, it retains its original identity (Illmensee, 1978). Therefore, programming the steps leading to the cellular blastoderm provides a well defined system for analysis, and defining that program will both allow better understanding of the processes of pattern formation and cell determination in the Drosophila embryo and serve as a preliminary example for a logical method to compute the egg.

Assumptions of the Drosophila calculation

In light of the massive amount of theoretical and experimental literature available on Drosophila development, it was necessary to use the appropriate theory (as summarized above) to provide a simplifying framework for the calculation. Therefore, the following assumptions were made.

(i) The genes in the regulatory network are “black boxes” with the following characteristics:

—chromatin switching allows continuous gradients to be approximated by step gradients where the steps correspond to threshold values for gene activation or repression [Fig. 1(a)];
—genes respond to input concentrations of regulatory proteins to return output concentrations of gene product proteins according to the Boolean gene switching rules in Fig. 1(b);
—switching of the regulatory genes can only occur once per cell cycle during S phase (Fig. 2).

(ii) Step gradients of maternal effect gene products in the egg set initial positional values (Fig. 3).

(iii) All steps starting with a fertilized egg through the growth of the cellular blastoderm occur in a syncytial blastoderm so cell-cell interactions can be neglected. After cellularization (cycle 14) new rules for cell-cell interactions were added for programming segment polarity.

(iv) Nuclei respond to concentrations of regulatory proteins at the position in the embryo (i.e. positional value) where they are during S phase of a particular cell cycle.

(v) The programming of the formation of the 14 parasegments plus acron and telson was calculated. The change in “register” from parasegments to segments was indicated in the formation of the engrailed stripes which form the anterior edge of each
5. Growth model for Drosophila nuclei in syncytial blastoderm. In the Drosophila embryo the single fertilized nucleus goes through 13 synchronous nuclear divisions within the embryo. After nuclear cycle 9 the nuclei migrate to the surface of the embryo forming a single layer along the outer membrane. This process can be modeled by projecting the position of the nuclei onto the surface of the embryo and cutting a plane that represents a “Mercator” map of one side of the embryo (left). Divisions of the projected nuclei are calculated as alternate vertical and horizontal divisions of the plane (center) where each block represents a region (positional value) that contains on average a single nucleus. After 13 nuclear divisions the plane contains 64 × 32 blocks corresponding to individual regions that contain a single nucleus when membranes form around the nuclei to define the cellular blastoderm. The anterior/posterior and dorsal/ventral computer programs follow the fate of a single line of blocks through embryogenesis.

Fig. 5. Growth model for Drosophila nuclei in syncytial blastoderm. In the Drosophila embryo the single fertilized nucleus goes through 13 synchronous nuclear divisions within the embryo. After nuclear cycle 9 the nuclei migrate to the surface of the embryo forming a single layer along the outer membrane. This process can be modeled by projecting the position of the nuclei onto the surface of the embryo and cutting a plane that represents a “Mercator” map of one side of the embryo (left). Divisions of the projected nuclei are calculated as alternate vertical and horizontal divisions of the plane (center) where each block represents a region (positional value) that contains on average a single nucleus. After 13 nuclear divisions the plane contains 64 × 32 blocks corresponding to individual regions that contain a single nucleus when membranes form around the nuclei to define the cellular blastoderm. The anterior/posterior and dorsal/ventral computer programs follow the fate of a single line of blocks through embryogenesis.

parasegment and posterior edge of each segment (De Pomerai, 1986).

(vi) Patterns within the acron (parasegment 0) and telson (parasegment 15) were not considered.

(vii) Changes in gene product patterns are calculated and discussed, but phenotypes regulated by combinations of genes or phenotypic effects of mutations are left for later discussion (Bodnar, manuscript in preparation).

(viii) The anterior/posterior and dorsal/ventral determinations of the blastoderm were considered independently because there is a little interaction between the determination of the two body axes prior to gastrulation (St. Johnston & Nüsslein-Volhard, 1992).

(ix) In cases where several genes act directly in a series, the effect of the series is calculated by using the critical gene or most commonly studied gene in the pathway. For example, dorsal encodes a transactivator protein that is initially found homogeneously throughout the egg, but the activity of the dorsal protein is controlled by a changes in nuclear compartmentalization regulated in a gradient by an extracellular ligand (Rushlow et al., 1989; Steward, 1989). However, for the purpose of a calculation of the positional information in gradients, the “dorsal”
gene can be used as a marker for the activity of the entire pathway (i.e. dorsal = snake− → Toll− → pelle/tube− → cactus− → dorsal; Ray et al., 1991).

(x) Some regulatory genes shift patterns or are regulated differently during germ band extension or later embryogenesis. Only the patterns and regulatory interactions up to blastoderm formation are considered or discussed.

**Overview of the Drosophila calculation**

The developmental program for formation of the *Drosophila* blastoderm requires a cascade of gene activations both in time [Fig. 6(a)] and in space [Fig. 6(b)]. Maternal effect gene products or mRNAs deposited in the egg in gradients set the initial positional values for the regulatory networks. The maternal effect genes begin a program involving interactions among the gap genes and terminal genes that set a seven stripe pattern where each stripe is two parasegments wide. The gap and terminal genes then program an initial pair-rule gene pattern consisting of two alternating broad seven stripe patterns where each stripe is now one parasegment wide. Interactions among the pair-rule genes themselves then refine the pair-rule pattern to sharply demarcated stripes. The pair-rule genes also set polarity within the

---

**Fig. 6.** (a) Overview of the temporal program for *Drosophila* early embryogenesis. *Drosophila* development requires a cascade of gene activations beginning with the fertilized egg until the commitment of developmental compartments at the formation of the cellular blastoderm [adapted from Levine and Harding, (1989)]. The approximate timeline (top to bottom) correlates the timing of the major processes and the approximate nuclear division cycles when they occur (italics) with the classes of regulatory genes involved (bold face) and the particular genes in those classes used in the calculations. The pathways modeled are shown by arrows (—→) and those not modeled are shown in dashed arrows (----). (b) see facing page. Overview of spatial programming in *Drosophila* early embryogenesis. The cascade of transcriptional regulation in *Drosophila* pattern formation has components both in time and space. The overall A/P spatial patterns are summarized in a flow corresponding to the temporal flow in (a). Calculated spatial patterns are indicated starting with positional values set by morphogen gradients (top) through the setting of homeotic selector (Program 1) and segment polarity (Program 2) codes.
MATERNAL EFFECT GENES
Initial gradients set positional values

anterior TERMINAL GENES posterior

Ant P Ubx Abd A Abd B HOMEOTIC GENES

Setting homeotic selector code

Homeotic selector code

Segment polarity code

John W. Bodnar (facing p. 404)
one-segment-wide stripes by activation of the *engrailed* and *wingless* genes, and intrasegmental polarity is further defined by cell-cell interactions among the segment polarity genes after cellularization of the blastoderm. During this process the homeotic genes are also activated first in broadly defined stripes programmed by the gap and terminal genes and then in refined stripes in the cellular blastoderm through interactions with the pair-rule genes and the homeotic genes themselves. The calculation was set up as two Pascal computer programs which considered the process of pattern formation as a step-by-step cascade from the maternal effect genes through the formation of the 14 individual segments with defined polarity. The computation also considered the initial programming of the homeotic genes. Finally, the dorsal/ventral pattern is set by a single gene cascade under control of the *dorsal* gene, and this process was considered separately as a manual calculation.

The overall process of *Drosophila* blastoderm formation is shown in Fig. 5. I assumed that the fertilized nucleus undergoes a series of 13 synchronous nuclear divisions within the syncytium of the egg. After each nuclear division the two daughter nuclei separate in random directions, but this can be approximated by assuming that the separation is in the anterior/posterior (A/P) direction on one division and the dorsal/ventral (D/V) direction on the next division (Fig. 5). Therefore, any square region projected on the surface of the syncytium that starts with a single nucleus at its center will have four nuclei evenly distributed within it two nuclear cycles later. I assumed there are $2^7 = 128$ nuclei at nuclear cycle 7 when transcription begins; if these were projected on the surface of the syncytium about half (64) would be found on each side thus giving a planar projection of $16 \times 4$ boxes on each side of the embryo each containing a single nucleus. I then assumed that each nucleus then divided in the D/V direction during cycle 7 and alternate A/P and D/V directions through cycle 13. During G1 of each nuclear cycle, each nucleus expressed those genes regulated by the combination of transactivator proteins expressed in the box. During mitosis each daughter nucleus would carry the new chromatin conformations as the two nuclei divided in the appropriate direction at which point the nuclear cycle would be repeated as before. Therefore, after cycle 13 when the segments begin to form there are approximately $2^{13} = 8192$ nuclei of along the surface of the embryo of which about half would be found on one side, and about half of those $= 2048$ would be found in a $64 \times 32$ projection on each side of the embryo that approximates the region which will ultimately form segments.

This simple projection and division scheme follows the experimental data well. While nuclei can migrate in any direction following each division, the fate of each nucleus depends on the particular region to which it migrates at cellularization (Minden et al., 1990); in the calculation I assumed that on average for each nucleus migrating out of a region there would be one nucleus migrating into that region. At the cellular blastoderm stage the embryonic cross-section contains about 72 cells in circumference (Rushlow & Levine, 1990); therefore, each side of the embryo contains about 36 nuclei in the D/V direction in excellent agreement with the calculated value of 32. At the onset of cellularization there are 14 segments and each segment is about four cells wide; thus the segmented region is about 56 cells wide in excellent agreement with the calculation which ends up with 56 cells in segments and the remainder of the 64 A/P cells in the acron and telson.

The actual computer programs are available from the author*, but the overall characteristics of the programs can be described without reference to the programming steps themselves. The first program describes the fate of a single row of cells in the A/P direction from the fertilized egg through formation of the cellular blastoderm. The second program describes the fate of two segments from the formation of overlapping *hairy/even-skipped* stripes through the setting of segment polarity of the cells within each segment. Growth of a D/V row of cells was calculated manually as described in the text.

The aim of this pilot project was to assess the feasibility of synthesizing theory and experiment into an epigenetic program of a real organism, and the programs were, therefore, kept as simple as possible. Integration of all the A/P and D/V programs would require: (1) changing the Boolean transcription rules [Fig. 1(b)] to include gradients with greater than three steps, and (2) adding to the first computer program to allow for two dimensional “growth” of the embryo. The first program has the arrays defined that can allow two-dimensional calculation, but the

---

*The simplest method to follow the text and discussion, especially for calculation of mutants, is to step through the program cycle by cycle in conjunction with reading the text and figures. Therefore, the programs (in both text and compiled versions) are available on 3.5 inch disc from the author along with reprints or can be downloaded directly from the author’s web site (http://www.nadn.mil/ChekDep/). The programs have been tested to run on MacIntosh Plus, SE, Mac II, and Power Book personal computers.
program in that form runs very slowly on a personal computer. These simple additions could be added later, but the current version is both simpler to follow in terms of how the theory directly affects the data and in terms of making the method available to scientists who only have access to or desire to use a personal computer. In outline form the computer programs are set up as follows.

- Define all generic procedures needed. This includes the logical steps for:
  - Boolean gene switching rules [Fig. 1(b)].
  - Setting initial values for maternal effect genes (Fig. 3).
  - Rules for nuclear division, i.e. how to divide nuclei A/P or D/V on alternate cycles (Fig. 5).
  - Procedures to allow mutants (null or ectotopic values, or aberrant maternal effect gene gradients).

- Define all Drosophila-specific procedures (using lookup tables of experimental data—Appendix A).
  - Initial pattern values for Drosophila maternal effect genes
  - Input gene switching rules for each Drosophila gene (i.e. which genes activate/repress each gene).

- DO loop for each nuclear cycle (Fig. 2).

As written the two calculations are straightforward Pascal programs differing only in the specific initial patterns or gradients of protein values and specific input gene switching rules. As such they may potentially be modified to describe other organisms or later times in Drosophila development.

Specific methods of the Drosophila calculation
The development of the Drosophila blastoderm involves two different French flag genetic networks, four individual stripe-doubling genetic networks, a line-drawing genetic network, and a final selector genetic network. While these genetic networks interact combinatorially to define cell fates, description and calculation of the developmental program is straightforward because their effects are distinctly divided in time [Fig. 6(a)] or space [Fig. 6(b)]. Therefore, I describe the individual genetic networks in their temporal sequence.

In long germ band insects such as Drosophila all the abdominal segments are determined simultaneously in the embryo, a mechanism which appears to be recent in evolution; in the short germ band insects from which Drosophila has evolved, the head and thorax are determined in the blastoderm, and the abdominal segments grow sequentially later in development [reviewed in Klingler (1994)]. In the analysis of the Drosophila embryonic program, it became apparent that several pathways have evolved from short germ band insects to allow the simultaneous determination of all segments in the long germ band Drosophila. This appears to be mainly due to the posterior terminal system picking up functions in determination of the abdomen with the posterior system becoming vestigial in Drosophila embryogenesis. The posterior system was, therefore, omitted from the program (simulating a double mutant in maternal hunchback and maternal nanos—see Appendix B for detailed discussion).

Program I

Programming Positional Values by Maternal Effect Genes

Programming the Gap Gene Pattern by Positional Values

Programming Pair-Rule Stripes by Gap Genes

Initial Programming of Homeotic Genes by Gap Genes

Setting the Homeotic Selector Code by the Homeotic Genes

The first calculation was a Turbo Pascal program that begins with positional information stored as morphogen gradients and traces how that positional information is converted by a transactivator cascade to a homeotic selector code to define the identities of all the segments.

Programming positional values by maternal effect genes. Information for pattern formation is stored in the egg as gradients of specific morphogens. For the anterior/posterior (A/P) pattern formation in Drosophila, all the information necessary to simulate programming of the 14 segments and the pattern of homeotic genes within those segments is contained in the bicoid gradient and the terminal system gradients under the control of the torso gene [Fig. 6(a,b)].

The A/P pattern is modulated by the bicoid gene product, a transactivator whose mRNA is deposited in a gradient from the anterior end of the egg, and the torso gene product, a signal transducer that is regulated in a gradient from both ends of the egg. The actual transcriptional effects of torso are mediated by transactivators—tailless, huckebein, forkhead, and gene “Y”—which are activated by torso then regulate the zygotic gap genes. In the program I assumed an initial step gradient pattern of deposition of bicoid or...
activation by \textit{torso} in 16 compartments (Fig. 7). While \textit{tailless} is not a maternal effect gene, its expression begins early in embryogenesis in gradients from either end of the egg mediated by high activity of \textit{torso}, and it is not further regulated by the gap genes; therefore, I have included the \textit{tailless} gradients as part of the original gradient pattern with \textit{bicoid} and \textit{torso}. (In essence, this simulates a six step \textit{torso} gradient where \textit{tailless} is activated in a gradient by \textit{torso} at the highest three concentrations). Since \textit{huckebein} and \textit{forkhead} affect terminal regions in the acron and telson outside the segmented region, they were neglected in the calculation.

The initial combination of \textit{bicoid}, \textit{torso}, and \textit{tailless} gradients can form 16 compartments that will ultimately correspond to parasegments (PS) 0 through 15, and the localization of the initial gradients is consistent with the experimental distribution of the three gene products (see Gilbert, 1991). Note that the segment pattern is offset one compartment posterior from the center of the embryo in the calculation with one additional acron compartment defined (but not included in the calculation) by the anterior edge of the morphogen gradients [see Figs 6(b) and 7]; this asymmetric location of the compartments is consistent with the actual formation of the segmented region of the embryo being displaced slightly toward the posterior end of the egg. Calculated \textit{bicoid} concentrations form a gradient from the anterior end through Parasegment 7 (about 50\% of egg length) consistent with experimental evidence for \textit{bicoid} activity over the anterior two thirds of the egg (St. Johnston & Nüsslein-Volhard, 1992). Calculated \textit{tailless} concentrations form gradients from both ends of the egg forming stripes of about 20\% of egg length consistent with experimental evidence for terminal \textit{tailless} stripes detectable over 15\% from the termini (Mahoney & Lengyel, 1987). (As discussed in detail in Appendix B, the posterior system—in which \textit{nanos} mediates a zygotic gradient of the maternal \textit{hunchback} gene—appears to be a vestige of evolution and not necessary at this stage. Therefore, the posterior system was not included in the \textit{Drosophila} calculation).

The step gradients assumed in the calculation form the initial conditions for zygotic gene activation; these gradients are then “read” according to the Boolean gene switching rule as described above for the French flag and stripe-doubling mechanisms. Using only a three step gradient, 16 compartments can be formed to specify “positional values” for 15 of the 16 compartments. A calculated transcriptional cascade (see below) can turn that positional code into a homeotic gene selector code for 15 distinct parasegments (Fig. 7). The homeobox selector code then defines the identity of each parasegment.

The combinatorial positional values, corresponding to protein values deposited in the egg, for (\textit{bcd}, \textit{tor}, \textit{tll}) as shown in Fig. 7 define 15 different parasegments (PS). One more threshold in the \textit{torso} gradient would be necessary to resolve PS14 from the telson (PS15). Rather than overly complicate the computer program I chose to stay with a simple three-step gradient model and accept that PS14 is fused with the telson.

In summary gradients of morphogens are deposited in the egg, and threshold values of these morphogens forms a digital positional code as follows (where, for
example, bcd3 indicates a bicaudal concentration of 3, i.e. above a high threshold).

<table>
<thead>
<tr>
<th>Acron</th>
<th>PS1</th>
<th>PS2</th>
<th>PS3</th>
<th>PS4</th>
<th>PS5</th>
<th>PS6</th>
<th>PS7</th>
</tr>
</thead>
<tbody>
<tr>
<td>bcd3</td>
<td>bcd3</td>
<td>bcd3</td>
<td>bcd3</td>
<td>bcd3</td>
<td>bcd2</td>
<td>bcd1</td>
<td></td>
</tr>
<tr>
<td>tor3</td>
<td>tor3</td>
<td>tor3</td>
<td>tor3</td>
<td>tor2</td>
<td>tor1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tll3</td>
<td>tll2</td>
<td>tll1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This positional code provides the input information necessary to program Drosophila embryogenesis beginning with the programming of the gap gene pattern.

*Programming the gap gene pattern by positional values.* The first step in the A/P programming of the Drosophila blastoderm is the formation of a two-segment-wide stripe pattern (usually called the gap gene pattern) programmed directly by maternal effect gene gradients [reviewed in Gilbert, (1991), Ingham (1988) and Gaul & Jäckle (1990)]. This pattern appears to be programmed by non-periodic cues in the overlap of activation and repression of both maternal effect genes and gap genes (Pankratz & Jäckle, 1990). In any analysis the gap genes and terminal genes are considered together because it became apparent during the development of the calculation that both systems are required to form the double-segment stripe pattern and that at least one gene—Krüppel—acts both as a gap gene and a terminal gene. The results of the calculation for sequential activation and interactions of the terminal and gap genes and comparison with experimental effects of the terminal and gap genes in the cellular blastoderm (Gilbert, 1991) are shown in Fig. 8.

Mutations in the zygotic gap genes delete broad regions of the embryonic A/P pattern, in each case spanning several contiguous segments; the term gap gene thus refers to the “gap” of sequential segments lost in each gap gene mutant (De Pomerai, 1986). Historically, the genes in this class have been defined as hunchback (hb), Krüppel (Kr), giant (gt), and tailless (tll) (De Pomerai, 1986). From my analysis of blastoderm pattern formation there appears to be a distinct difference in the “gap” functions regulated by the terminal system and those regulated by the A/P system; therefore, in my discussion below I will discuss “gap” gene functions in terms of “terminal” regulation and “A/P” regulation. Terminal regulation is a stripe-doubling mechanism under the control of the torso gene and includes tailless, Krüppel, and giant while A/P regulation is a French flag mechanism under control of bicaudal and includes Krüppel, knirps, and hunchback [Fig. 6(b)]. The interactions of these two systems (including direct effects on Krüppel from both systems) allows the formation of the two-segment-wide stripe pattern of the “gap” genes.

<table>
<thead>
<tr>
<th>PS8</th>
<th>PS9</th>
<th>PS10</th>
<th>PS11</th>
<th>PS12</th>
<th>PS13</th>
<th>PS14</th>
<th>Telson</th>
</tr>
</thead>
<tbody>
<tr>
<td>tor1</td>
<td>tor2</td>
<td>tor3</td>
<td>tor3</td>
<td>tor3</td>
<td>tor3</td>
<td>tor3</td>
<td>tll1</td>
</tr>
<tr>
<td>tll2</td>
<td>tll3</td>
<td>tll3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The terminal system is a stripe-doubling system where the gradients of the terminal (ter) genes at each end of the egg form the initial two stripes (ter--------ter). A wide Krüppel (Kr) stripe then forms between the two terminal stripes (ter---Kr---ter) followed by another stripe-doubling of giant (gt) stripes forming between the terminal and Krüppel stripes (ter-gt-Kr-gt-ter). Simultaneously the A/P system forms a French flag system with bicaudal as the input gene [gene A in Fig. 4(a)] in a gradient from the anterior end of the egg and with hunchback, Krüppel, knirps, and hunchback as the target genes [genes B and C in Fig. 4(a)]. The posterior end of the A/P system (gene D) is defined by repression from the terminal system. Feedforward and feedback interactions as described for the French flag pattern set up a stable stripe pattern (hb-Kr-kni-hb).

The calculation assumes that transcription starts at nuclear cycle 6 and all the gap genes are activated at that time (Fig. 8). Over the course of five cell cycles the terminal pattern (ter-gt-Kr-gt-ter) and the A/P pattern (hb-Kr-kni-hb) form simultaneously. However, Krüppel is a member of both classes and additional mutual repressions (e.g. hb and gt) have evolved between the systems to form a reproducible pattern where the two giant stripes form adjacent to the hunchback stripes for a final seven stripe pattern (gt-hb-Kr-kni-gt-hb-tll) in the region of the syncytial blastoderm that will later give rise to the segmented region of the Drosophila embryo. Therefore, over the course of five cell cycles the (bcd, tor, tll) positional value code is converted into a new positional value code of the seven gap gene stripes (Fig. 8) as indicated by the computed values for the gap genes at cycle 11. The complex interactions among these genes stabilize at cycle 11 into a steady state pattern. The interactions, however, are totally dependent on the original (bcd, tor, tll) positional values.

In summary the terminal genes and the A/P genes act together to convert the positional values stored in the maternal effect gene gradients into a parasegment positional code (Fig. 8). This pattern begins to form as transcription starts in the embryo (cycle 6), and over the next five cell cycles the regulation network of terminal and gap genes forms a stable pattern where
Fig. 8. Programming the gap gene pattern. Calculation of the gap gene pattern assumed the interactions (top) with Boolean transcription rules as in Fig. 1(b). Starting with morphogen gradients at cycle 6 (Fig. 7), values for each gene were calculated each nuclear cycle on the basis of current values of all gene products present. The final calculated values at cycle 11 are compared with experimental results defining segments affected by mutants in each of the gap genes (Ingham et al., 1986; Petschek et al., 1987; Lehmann, 1988; Gilbert, 1989; Gaul & Jäckle, 1990).
Fig. 9. Programming hairy pair-rule stripes by gap genes. The pattern (top) of maternal effect and gap genes at cycle 11 (Fig. 8) serves as the input for programming the pattern of the primary pair-rule gene *hairy* by a stripe-doubling mechanism. The hairy stripes are programmed by several enhancers in the 5' region of the *hairy* gene (middle) where each enhancer determines a stripe by binding an activator, a repressor to define the anterior edge, and a repressor to define the posterior edge. The calculated patterns for *hairy* (green), *giant* (blue), and Krüppel (red) are compared with experimental data [bottom; reproduced with permission from Paddock et al. (1993). *Biotechniques* **14**, 42–46].
the regions that will become the parasegments express a unique combination of the gap genes in overlapping stripes. The positional values in the calculation at cycle 11 are as follows:

\[
\begin{array}{cccccccc}
\text{Acron} & PS1 & PS2 & PS3 & PS4 & PS5 & PS6 & PS7 \\
bcd3 & bcd3 & bcd3 & bcd3 & bcd3 & bcd3 & bcd2 & bcd1 \\
tll3 & tll2 & tll1 & hb1 & hb2 & hb2 & kr1 & kr2 \\
hb1 & hb2 & hb2 & kr2 & kr3 & kr3 & kni1 & kni2 \\
tll2 & hb1 & kr1 & kr3 & kr3 & kni3 & kni3 & gt1 \\
tll1 & hb2 & kr2 & kni3 & kni3 & kni3 & kni1 & gt1 \\
tll3 & hb1 & kr1 & kni1 & kni1 & gt1 & gt3 & gt1 \\
\end{array}
\]

(Again with the simple three step gradient PS14/Telson are fused because they have the same positional code). This gap gene positional code (which is also plotted in detail in Fig. 9) next serves as the baseline code to program both the pair-rule gene stripes and the initial homeotic gene patterns.

Programming pair-rule stripes by gap genes. In *Drosophila* the broad overlapping stripes of the gap genes lead to expression of two sets of seven stripe patterns of the pair-rule genes in alternating register. This step in the cascade appears also to be a stripe-doubling mechanism cued by non-periodic signals in the overlap of activation and repression of the gap genes that program two independent stripe patterns of *hairy* (*h*) and *even-skipped* (*eve*) which in turn program stripes of *fushi tarazu* (*ftz*) ([Meinhardt, 1986; Pankratz & Jäckle, 1990; Carroll, 1990; Pankratz et al., 1990]). The parasegments are defined initially by two complementary pair-rule functions: odd-numbered parasegments correspond to the seven stripes of *eve* (*E*) and *h* (*H*) expression, while even-numbered parasegments correspond to the seven stripes of *ftz* expression ([De Pomerai, 1986]). In each case the parasegment border (as defined by *engrailed* expression) coincides cell by cell with the anterior margin of the zone expressing *eve* or *ftz* products ([De Pomerai, 1986]). Expression of the pair-rule genes begins about cycle 10 when the nuclei migrate to the embryo surface suggesting that during the migration pair-rule expression is activated or derepressed likely by factors regulated at the embryo membrane ([Davis & Ish-Horowicz, 1991]). Independent genetic networks for *eve* activation ([Reinitz & Sharp, 1995]) and *h* activation operate at the same time both programmed by the gap genes.

For the stripe-doubling mechanism ([Fig. 4(c)]) four levels of regulation are necessary. In the *Drosophila* blastoderm the first level [input genes in Fig. 4(c)] is the gap gene pattern. The second level [genes B and C in Fig. 4(c)] is also the gap gene pattern. The third level [gene E in Fig. 4(c)] has two separate genes—*hairy* (*h*) and *even-skipped* (*eve*)—that are regulated independently into seven stripe patterns by the gap genes. The fourth level [gene F in Fig. 4(c)] is the *fushi tarazu* (*ftz*) gene that forms stripes by repression from both *h* and *eve*. The computer calculation only addresses the programming of the initial *h* and *ftz* stripe patterns both because the *eve* and *h* patterns are partially redundant and because the data for *eve* are incomplete.

As shown in Fig. 4(c) the stripe-doubling mechanism starts with activation of the third level gene (*E*) in regions of overlap between two upstream genes (*B* and *C*). In the calculation the third level gene is *h*; it is repressed by *tramtrack* (*ttk*) until *ttk* is degraded at cycle 9 when the *h* is activated in a seven stripe pattern. The *h* promoter has several enhancer regions that are independently regulated by different combinations of genes (Fig. 9). Each *h* stripe (gene *E*) is formed in a region of overlap between two gap genes where each individual enhancer can program a narrow stripe with three inputs: an activation signal from a particular gap gene (input gene), repression on the anterior border by another gap gene (gene *B*), and repression on the posterior border by a third gap gene (gene *C*). As shown in Fig. 9 for each enhancer there is only in a narrow region of the blastoderm in which the activator is present but not the anterior or posterior repressors. Therefore, in the program the Boolean transcription rules were determined for each enhancer (Fig. 9), and *h* was considered activated if any one of the enhancers was activated by its required gap gene combination. Overall the seven *h* stripes form in the overlaps of the gap gene stripes (Fig. 9): stripe 1 between *tll* and *hb*, stripe 2 between *tll* and *Kr*, stripe 3 between *hb* and *kni*, stripe 4 between *bed* and *kni*, stripe 5 between *Kr* and *gt*, stripe 6 between *Kr* and *tll*, and stripe 7 between *kni* and *tll*.

The fourth level gene (*F*) in *Drosophila* the pair-rule pattern is *ftz*. The *ftz* gene is also repressed by *tramtrack* until *ttk* is degraded at cycle 9 at which point it is expressed throughout the embryo. The *ftz* stripe pattern is then formed by repression from *h* in the calculation (and from both *h* and *eve* in the organism) to leave *ftz* expression only in the regions devoid of *h* repression.
In summary the pattern of alternating h/eve and ftz seven stripe patterns is formed as a stripe-doubling mechanism dependent on inputs from the gap gene pattern. These genes provide an odd/even parasegment code where eve expression (which overlaps h expression) is the odd selector gene, and ftz is the even selector gene. The calculated patterns at cycle 14 are consistent with those mechanisms:

<table>
<thead>
<tr>
<th>Acron</th>
<th>PS1</th>
<th>PS2</th>
<th>PS3</th>
<th>PS4</th>
<th>PS5</th>
<th>PS6</th>
<th>PS7</th>
<th>PS8</th>
<th>PS9</th>
<th>PS10</th>
<th>PS11</th>
<th>PS12</th>
<th>PS13</th>
<th>PS14</th>
<th>Telson</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>h</td>
<td>ftz</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>ftz</td>
<td>h</td>
<td>ftz</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
</tr>
</tbody>
</table>

The h/eve stripes and the ftz stripes provide the odd/even parasegment input for the homeotic selector code. The initial h/eve and ftz stripe patterns also provide the inputs for the refinement of the pair-rule genes and for setting segment polarity.

**Initial programming of homeotic genes by the gap genes.** The identities of the entire parasegments are programmed through activation of the homeotic genes. The homeotic genes appear to be the “selector genes” that ultimately program the A/P cell fate in the Drosophila embryo (Garcia-Bellido, 1977). These genes contain a consensus DNA-binding domain known as the homeobox, and the homeobox genes that encode the Antennapedia (Antp) class of homeobox genes all map to the Antennapedia and Bithorax complexes collectively known as the Homeotic complex or HOM-C (McGinnis & Krumlauf, 1992). The homeobox genes are aligned in the two loci in the order of their effects anterior to posterior in the embryo: labial (lab), proboscipedia (pb), Deformed (Dfd), Sex combs reduced (Scr), and Antennapedia (Antp) in the Antennapedia locus and Ultrabithorax (Ubx), abdominal-A (abd-A), and Abdominal-B (abd-B) in the Bithorax locus (McGinnis & Krumlauf, 1992). Of these only proboscipedia has no effect on the embryonic phenotype, and the labial gene is important only in the acron. Therefore, pb and lab were not used in the computer calculation.

The current model for homeobox function is based on a model by E. B. Lewis where all segments posterior to T2 are controlled by subfunctions of the bithorax of genes, with successively more posterior segments requiring the activity of more and more bithorax genes [reviewed in De Pomerai (1986) and Gilbert (1991)]. The initial expression patterns for the homeotic genes appears to be defined by patterns of gap gene expression but later refined by the pair-rule genes and homeotic genes themselves (Harding & Levine, 1988; Gaul & Jäckle, 1990).

Temporal control of the homeotic genes in vivo may be regulated by the length of the gene sequences themselves. For example, the genes knirps (kni) and knirps-related (knrl) encode proteins that contain almost identical DNA-binding domains, complement the other’s phenotype, and their spatial expression patterns coincide; however, the kni gene is expressed early in blastoderm formation but knrl is not expressed until the 13th nuclear cycle (Rothe et al., 1992). Due to a difference in intron size the kni transcription unit is 3 kb long while the knrl transcription unit is 23 kb, and it appears that the knrl gene may be transcribed but that unfinished nascent transcripts are aborted during mitosis in the rapid early nuclear cycles. Thus transcript length may be a limiting factor for gene expression early in blastoderm formation, and long genes may not be expressed until the nuclear cycle time increases just prior to the formation of the cellular blastoderm (Rothe et al., 1992). Since the Drosophila homeotic gene loci are hundreds of thousands of nucleotides long, the expression of the homeobox loci is likely inhibited early in embryogenesis by the same mechanism as the knrl gene. Therefore, in the calculation I assumed that homeobox gene expression was inhibited prior to nuclear cycle 12—even if the appropriate combination of transactivators was present.

The gap gene pattern appears to have all the positional information necessary to program the initial homeotic gene pattern while the homeotic genes themselves have some indication of a French flag interactions in which individual homeobox gene stripes are defined by feedback repression from genes more posterior in the pattern (Fig. 10). Therefore, the calculation depended almost exclusively on known gap gene activations or repressions of the genes in the HOM-C loci along with interactions among the HOM-C genes themselves. These few interactions did an excellent job of modeling the initial rough HOM-C loci expression patterns at the cellular blastoderm stage.

Many cross regulations among the homeotic genes lead to fine control in which levels of an individual homeobox protein vary segmentally throughout the regions in which it is expressed. These may involve the production of multiple alternately spliced mRNAs that produce slightly different protein products (such
Fig. 10. Initial programming the homeotic gene pattern. Calculation of the homeotic gene pattern assumed the gap gene interactions (top) with Boolean gene switching rules as in Fig. 1(b). Starting with the steady state gap gene pattern at cycle 11 (Fig. 8), values were calculated each nuclear cycle on the basis of current values of all gene products present. The final calculated values are compared with the localization of the homeotic gene proteins detected in situ in the epidermis of the cellular blastoderm (Ingham et al., 1986; Petschek et al., 1987; Lehmann, 1988; Gilbert, 1989; Gaul & Jäckle, 1990).
Fig. 11. Refinement of pair-rule stripes. Overlapping $h$ and $eve$ stripes provide the input to begin refinement of the pair-rule stripes. Following derepression of $run$, $ftz$, and $prd$ (through degradation of $ttk$) the pair-rule genes interact among themselves to sharpen the stripe patterns. The calculated patterns are compared with the early (cycle 12) and refined (cycle 14) experimental pair-rule patterns (Howard, 1990). The register of the stripes is compared with the anterior edge of the parasegments (vertical gray stripes) which will become stripes of *engrailed* expression.
as for Ubx; see Lopez & Hogness, 1991) or several mRNAs that are produced at different levels in different segments (such as for Abd-B; see Crosby et al., 1993). Since the primary purpose of the model was to test whether current knowledge on gap gene activation of the homeobox genes was sufficient to explain the initial expression of those genes, these fine tuning effects were neglected in the calculation. Similarly, the levels of Scr, Antp, and Ubx are elevated in PS2, PS4, and PS6 respectively by the activity of ftz (Scott & Carroll, 1987; Ingham, 1988); this fine tuning mechanism was also neglected.

The homeobox loci are believed to have evolved sequentially by gene duplications in a scenario that controlled the evolving morphological organization of the arthropod body plan in the insect-myriapod lineage (Akam et al., 1988). Indications for evolution of the homeobox loci also became evident in the analysis of the regulatory patterns for those genes which paralleled the suggested changes in body plan. The primordial homeobox locus was likely \[\text{lab-}pb-Dfd\] under control of an ancient terminal system—defining an arthropod with a head (Acron-PS1) followed by identical trunk segments. The next locus was likely \[\text{lab-}pb-Dfd-Scr-Antp\] with the new genes under control of the A/P system (\text{bed, hb, Kr})—defining a myriapod with a head, additional defined segments (PS2-PS5), followed by trunk. The \text{Scr-Antp}\ genes then could have evolved to define thorax segments as well. The final gene duplication then added \[\text{lab-}pb-Dfd-Scr-Antp[Ubx-}

The homeotic gene pattern is then refined by interactions among the homeotic genes and inputs from the pair-rule genes and provides the input for a homeotic gene selector code.

**Setting the homeotic selector code.** The homeotic genes are believed to be “selector” genes in that they program final gene expression rather than more regulatory genes (Garcia-Bellido, 1977). It is likely that the programming of the homeotic gene pattern is the step at which “selection” of the cell phenotype is mediated by transcriptional activation of the appropriate combinations of somatic genes that will be expressed in the differentiated cells of the embryo (Bodnar et al., 1989).

In the calculation of *Drosophila* embryogenesis the final output combinations of homeotic genes in combination with the \text{ftz}\ gene in the even parasegments (or with \text{eve} in the odd parasegments) can be used to define a parasegment homeotic selector code [Figs 6(b) and 7]. In the head and thorax region these codes are defined by simple combinatorial codes: PS1 = Dfd, PS2 = \text{Scr + ftz}, PS3 = Scr, PS4 = Antp + ftz, and PS5 = Antp + Ubx. In the abdomen the homeotic selector code likely depends both on the identity of the homeobox genes expressed as well as the levels of homeobox proteins available to switch chromatin in individual genes (Crosby et al., 1993):

\[
\begin{align*}
\text{PS6} &= \text{Ubx3 + ftz}, \\
\text{PS7} &= \text{Ubx2 + abdA1}, \\
\text{PS8} &= \text{Ubx1 + abdA2 + ftz}, \\
\text{PS9} &= \text{Ubx1 + abdA2}, \\
\text{PS10} &= \text{Ubx1 + abdA2 + abdB1 + ftz}, \\
\text{PS11} &= \text{Ubx1 + abdA1 + abdB2}, \\
\text{PS12} &= \text{Ubx1 + abdB2 + ftz}, \\
\text{PS13} &= \text{abdB2}, \\
\text{PS14} &= \text{abdB2 + ftz}.
\end{align*}
\]

While these selector codes are based on only the initial programming of the homeotic genes, I suggest that only minor modifications of this code will be necessary to integrate the fine tuning of HOM-C expression patterns by further interactions among themselves and the pair-rule genes.

As suggested by simple theory on combinatorial gene regulation [reviewed in Bodnar et al., (1989)], each parasegment has a unique set of one to four evolutionary possibilities are presented in Appendix B.

In summary the homeotic genes in the HOM-C loci are the final genes in the transcriptional cascade to determine the identity of the embryonic parasegments. Their initial pattern is programmed by the gap genes and later refined by interactions with themselves and the pair-rule genes. This initial pattern can be calculated quite well using only the gap gene interactions and interactions among the homeotic genes themselves.
transactivators that can define its identity by “selection” of genes to be expressed in that cell by combinatorial DNA-protein interactions to activate DNA domains (Bodnar, 1988). The calculation indicates that the “selector” code for programming Drosophila segmentation requires only seven home-box genes used combinatorially. The calculation is also consistent with the data that the segment polarity code works in a segmental register while the homeotic selector code works in a parasegmental register.

Program 2
Refinement of Pair-Rule Stripes
Setting Segment Polarity

The second calculation is a Turbo Pascal program that starts with positional values stored in the h and eve pair-rule stripes and computes how the pair-rule stripes are refined by interactions among themselves then how the refined pair-rule stripes can set segment polarity by programming the wingless (wg) and engrailed (en) stripes.

Refinement of pair-rule stripes. After the primary pair-rule pattern of h, eve, and ftz forms with broad overlapping stripes, the individual stripes narrow with more highly defined edges. The refinement of pair-rule stripe patterns comes from complex interactions of those genes themselves along with runt (run) and paired (prd) [reviewed in Howard (1990) and Kania et al., (1990)].

The positional information for refinement of the broad overlapping pair-rule stripes into sharply defined stripes is contained in the initial pair-rule stripe pattern itself. While the computer program to simulate embryogenesis could calculate the pattern of h and ftz stripe patterns based on the original positional code stored in the maternal effect gene gradients, I found that to continue the calculation to refine the pair-rule gene stripes and set segment polarity required a second computer program. This second program followed conceptually from the output pair-rule stripe patterns in the first program but required two corrections: (1) the calculation up to this point considered only the h and ftz primary pair-rule genes, and the second program needed to consider the eve stripes as well; (2) the simplicity of the calculation in the three-step gradients resulted in one-segment-wide h and ftz stripes with sharp edges, and to reflect reality the second program needed “sloppy” broad h stripes to begin the refinement process (Note that this is the only place in the calculations that assumes any diffusion of a morphogen).

Therefore, the second computer program (Fig. 11) calculated refinement of the pair-rule stripes and setting of segment polarity beginning with broad overlapping stripes of h and eve. The program models pattern formation of four adjacent parasegments assuming: (1) eve stripes are included which are posterior to the h stripes but with considerable overlap (see above and Howard & Struhl, 1990); and (2) the h and eve stripes which are stepped the first calculation (e.g. values of 00111100) are rounded on their edges (e.g. 01232100) as they are in vivo at this point. The ftz pattern was begun in a continuous band throughout the embryo as before. The refinement of the pair-rule stripes requires three more levels of stripe-doubling based on the initial h and eve patterns (Fig. 11).

The first stripe-doubling in refinement of the pair-rule stripes depends on interactions among h, eve, and the runt (run) gene where run appears in stripes in the gaps between the h/eve stripes and modulates the eve pattern (Edgar et al., 1989). The h and run proteins repress each other, and the run pattern forms directly in the fissures between the h stripes; the h stripes do not change in response to run activation because existing h expression limits run expression. However, build-up of run between the h stripes is enough to decrease eve levels at the posterior edge of the eve stripes. The eve gene also is autoregulatory (Goto et al., 1989; Jiang et al., 1991) and will, without further input, increase its own expression to high levels. This combination of h/run/eve interactions means that the eve stripes become asymmetric with high levels on the anterior edge (due to eve autoregulation in the center of the h stripes) dropping to low levels of the posterior edge (due to run repression). These asymmetric eve stripes are inputs for setting segment polarity in the odd parasegments (see below).

The second stripe-doubling mechanism in refinement of the pair-rule stripes depends on repression of ftz by h and eve to modulate the ftz pattern. The eve and ftz genes repress each other so that the ftz stripes form directly in the fissures between the eve stripes; as above with h and run, the eve stripes do not change in response to ftz expression because eve is upstream of ftz in the cascade. The ftz gene like eve is autoregulatory and will increase to high levels in the absence of further input (Kaufman et al., 1990; Pick et al., 1990), h is also a repressor of ftz (but weaker than eve), and h is expressed in the posterior regions of the ftz stripes. Therefore, this combination of h/eve-ftz interactions means that ftz stripes become asymmetric with high levels on the anterior edge (due to ftz autoregulation in the gap between h/eve stripes) dropping to low levels at the posterior edge (due to h repression). These asymmetric ftz stripes are inputs for setting segment polarity in the even parasegments (see below).
Fig. 12. Setting segment polarity. The refined eve and ftz stripes provide input for setting segment polarity. Expression of paired is doubled into a 14 stripe pattern through repression by ftz. At cycle 14 cells form, beginning cell-cell interactions. Expression of wg and en is controlled by a “line-drawing” mechanism [Fig. 4(d)] controlled by eve and ftz. The register changes from parasegment to segment as the en stripes define the anterior edge of the parasegments and posterior edge of the segments. The calculation (top) is compared with experimental results [bottom; adapted with permission from Gilbert (1991). Developmental Biology. Sunderland, MA: Sinauer Associates Inc. p. 669].
Fig. 13. Programming dorsal/ventral genes. Calculation of the D/V gene pattern assumed the interactions (left) with Boolean gene switching rules as in Fig. 1(b) (plus one additional $dl$ threshold as a very strong repressor of $zen$). Starting with a dorsal activity gradient, values for each gene were calculated each nuclear cycle on the basis of the current values of all genes present. The final calculated values at cycle 14 are compared with experimental values (right) for each of the D/V genes (Anderson, 1987; Ray et al., 1991).
The third stripe doubling mechanism in the refinement of pair-rule stripes depends on repression of the paired (prd) gene by eve and ftz. It appears the prd is activated by gap gene products, but its expression is limited by eve and ftz; this gives a prd pattern where there is a prd stripe at each eve/ftz or ftz/eve border leading to the expression of 14 prd stripes at the cellular blastoderm. While other pair-rule genes show 14 stripe patterns such as odd-paired (opa) and odd-skipped (odd), there was little data available on these genes, and I suggest that the mechanism of their regulation is similar to prd. Therefore, prd was the only 14 stripe pair-rule gene considered in the pilot project.

In summary the refinement of the pair-rule genes depends on stripe-doubling interactions among the pair-rule genes themselves. The output of this refinement is an odd/even parasegment polarity code dependent on the asymmetric eve and ftz stripes.

Lines of cells in parasegments

<table>
<thead>
<tr>
<th>Even</th>
<th>Even</th>
<th>Even</th>
<th>Odd</th>
<th>Odd</th>
<th>Odd</th>
<th>Odd</th>
</tr>
</thead>
<tbody>
<tr>
<td>ftz3</td>
<td>ftz3</td>
<td>ftz1</td>
<td>prd</td>
<td>prd</td>
<td>prd</td>
<td>prd</td>
</tr>
</tbody>
</table>

This parasegment polarity code next serves as the input for setting segment polarity.

Setting segment polarity. The two seven stripe-pair-rule patterns of alternating asymmetric eve and ftz stripes are next converted into a 14 stripe pattern where determination of the individual 14 segments depends on a repeating pattern of the segment polarity genes—engrailed (en), wingless (wg), patched (ptc), and hedgehog (hh). The single-cell-wide 14 stripe pattern of engrailed is set up by the pair-rule genes independent of wingless activity although the en stripes fade soon after cellularization without expression of wg. It is at this point in Drosophila development that cell-cell interactions become important when wg/en interactions in adjacent cells are necessary for the full expression of the segment polarity genes within each segment [reviewed in DiNardo & Heemskerk (1990), Ingham & Arias (1992) and Perrimon (1994)].

The parasegments appear to be the fundamental units of metameric organization in Drosophila, but this pattern is converted to a segmental organization. At its inception, each parasegment is about four cells wide; the posteriormost line of cells expresses wg while the anteriormost expresses en (De Pomerai, 1986; Lawrence et al., 1987). The shift in parasegment to segment registers occurs when a series of shallow transverse grooves appear in the embryo during germ band extension. This change likely involves the contraction of a “posterior” (p) compartment anterior to an “anterior” (a) compartment in each parasegment thus shifting the frame of reference from [p]ap[p]ap[p] in the parasegments (in the blastoderm) to [p]ap[p]ap[p] in the segments later in development [reviewed in De Pomerai (1986)]. I suggest that contraction of the engrailed stripes (which form on the anterior edge of the parasegments and later define the posterior edge of the segments; De Pomerai, 1986) defines this shift in reference frames.

The programming of segment polarity is a line-drawing mechanism dependent on inputs from the refined pair-rule genes (eve, ftz, and prd) and on cell-cell interactions to refine the segment polarity genes into stripes that are a single cell wide. Full determination of all four rows of cells in each segment requires all the wg, en, hh, and ptc genes and several pathways for cell-cell interactions. However, it appears that the initial determination of the segment anterior and posterior borders depends on the interactions between the wg and en genes (Meinhardt & Gierer, 1980, Meinhardt, 1994). This was modeled (Fig. 12) as a line-drawing mechanism [Fig. 4(d)] using en and wg—including the first cascade in Drosophila development that requires cell-cell interactions.

The initial definition of segment polarity appears to depend on the expression of wg at the posterior edge of each parasegment. The refinement of the eve and ftz to asymmetric stripes leads to high eve or ftz expression at the anterior edge of the parasegments fading to no eve or ftz expression at the posterior edge of the parasegments; because both eve and ftz repress wg, the repression of wg will be lifted at the posterior edge of each parasegment as the eve and ftz stripes refine. At that time wg can be activated by prd, or possibly opa in the even parasegments (Ingham & Arias, 1992), in a narrow stripe at the posterior edge of each parasegment. This wg stripe likely sets the register for subsequent definition of polarity within the segments.

The eve and ftz genes were considered weak activators of en. Therefore, about the same time as wg is first expressed the eve and ftz stripes refine by
autoactivation at their anterior edges, and the very high levels of *eve* and *ftz* will activate narrow stripes of *en*.

It is here that the first cell-cell interactions are required in the A/P patterning. Starting at cycle 14 when cell membranes form around the syncytial nuclei, *wg* protein is secreted by cells that express it to act on receptor on adjacent cells. Data indicate that *wg* does not diffuse more than a few cell widths (Ingham & Arias, 1992). I assumed that *wg* expression is permissive for *en* expression “next” to the *wg*-expressing cell, i.e. *en* is activated by *eve* and *ftz* but cell cycle progression is permitted by *wg* so that cells expressing *en* will continue to do so only if they are adjacent to a cell expressing *wg*. Thereafter, *wg* “locks in” *en* expression in adjacent cells, and later on (not calculated) *en* activation of *hh* can “lock in” *wg* in adjacent cells, thus “drawing a line” of *wg/en*-expressing cells at the parasegment borders. Further refinement of the *wg/en* stripes to *hh/wg/en/ptc* stripes in each parasegment depends on multiple cell-cell interaction pathways during gastrulation and was considered beyond the scope of the calculation.

In summary, drawing the lines at segment boundaries in the programming of the *Drosophila* blastoderm depends on inputs from the refined asymmetric *eve* and *ftz* stripes. Expression of *wg* begins in those gaps, allows initial expression of *en* in cells adjacent mediated by *eve* and *ftz* activation. These initial *wg/en* double stripes then provide the positional information for setting segment polarity and shifting the register of embryo organization from parasegments to segments (Fig. 11). The *en* stripe is the anterior row of cells in each parasegment but contracts to become the posterior row of cells in each segment.

The final segment register in the 14 *engrailed* stripes is set in the initial *bcd/tor/tll* maternal effect gradients. The positional code stored in the morphogen gradients initially programs the overlapping gap gene stripes; overlapping *h/eve* stripes then form in the overlaps between the gap genes; the offset *h* and *eve* stripes provide the cue for the *eve/ftz* stripes; and the *en* stripes form at the anterior edge of the alternating *eve* and *ftz* stripes. Ultimately the *en* stripes form at the point of maximum *h* expression and in the middle of the gap between *h* stripes (see Figs 11 and 12). Therefore, the theory and data both support the notion that in programming the *Drosophila* blastoderm the initial positional values (stored in morphogen gradients) can program segment formation down to the level of individual stripes of cells in determining segment polarity.

**Program 3**

*Programming Dorsal/Ventral Patterns*

The third program is a manual calculation that starts with positional information stored in the dorsal/ventral (D/V) direction as a gradient of nuclear *dorsal* protein and computes the determination of the horizontal stripes of cells in the blastoderm (Fig. 13).

Mutations in some 20 genes (both maternal and zygotic) alter the embryonic D/V pattern (reviewed in Anderson (1987) and Ray et al. (1991)). These fall into several groups which interact to provide a single function in D/V patterning, i.e. *Toll*, *easter*, *pelle*, are all required for nuclear localization of the *dorsal (dl)* protein so that activation of *dorsal* function by nuclear transport is the regulatory step in pattern formation; therefore, I will consider these groups in terms of the critical transactivator or extracellular signal protein. Genetic studies have indicated that the *dl* protein is the only maternal effect gene required for determination of the D/V pattern (Arora & Nüsslein-Volhard, 1992). The *dl* gradient appears to be interpreted on three levels: ventrally, *dl* activates

<table>
<thead>
<tr>
<th>Change of register from parasegment (PS) to segment (Seg)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PS</strong> Even Even Even Even Odd Odd Odd Odd Even Even Even Even Odd Odd Odd Odd</td>
</tr>
<tr>
<td><strong>Seg</strong> Odd Even Even Even Even Odd Odd Odd Odd Even Even Even Even Odd Odd Odd Odd</td>
</tr>
<tr>
<td>• = location of <em>en</em> stripes = anterior edge of parasegments = posterior edge of segments.</td>
</tr>
</tbody>
</table>

The *en* stripes are set in the initial *bcd/tor/tll* maternal effect gradients. The positional code stored in the morphogen gradients initially programs the overlapping gap gene stripes; overlapping *h/eve* stripes then form in the overlaps between the gap genes; the offset *h* and *eve*...
At the formation of the cellular blastoderm the embryo is approximately 72 cells around the dorsal-ventral circumference (Fig. 13): 15–16 cells at the dorsal midline will become mesoderm and muscle; 13 ventrolateral cells on each side will give rise to the ventral nerve cord and ventral epidermis; 10 dorso-lateral cells on each side will become the dorsal epidermis; and 5–6 cells at the dorsal midline will form the amnioserosa (Anderson, 1987). Individual cells at different dorsal-ventral positions become committed to particular cell fates at different times starting at gastrulation (Anderson, 1987).

The D/V pattern is determined by a single French-flag mechanism mediated by a gradient of nuclear localization of the dl protein (Fig. 13). The dl protein transactivator activates twist and snail and represses dpp and zen—all by direct DNA binding. The threshold concentrations of dl required to bind each of these gene promoters determines the D/V pattern. In the calculation one additional threshold for dl needed to be defined—a very strong repressor where dl can inhibit zen at very low protein concentrations. The complete D/V pattern appears to require additional genes such as toloid, shrew, and screw, but these are likely involved in posttranscriptional regulation of dpp levels (Arora & Nüsslein-Volhard, 1992) and were, therefore, not included in the pilot project. Additionally, a line-drawing mechanism appears to activate singleminded (sim) under the control of sna and rho. Note that like the sim stripe at the rho/sna border there is another single row of cells at the zen/dpp border that forms the dorsal line and gives rise to neural cells; while not specifically calculated I suggest that the dorsal line is also produced by cell–cell interactions at the zen/dpp border by a line-drawing mechanism similar to en and sim activation.

In the formation of the Drosophila embryo cellular compartments can also be defined as groups of cells called mitotic domains that divide synchronously during the cell divisions immediately following cellularization (Foe, 1989). The D/V selector codes can be classified both by the phenotypes and the mitotic domains they produce. Also, it appears that the positional codes for D/V specification in the region between twist and dpp expression defaults to dorsal neuroectoderm without additional input other than the dl gradient.

<table>
<thead>
<tr>
<th>Localization</th>
<th>dl value</th>
<th>Positional Code</th>
<th>Phenotype</th>
<th>Mitotic Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dorsal</td>
<td>0</td>
<td>dpp + tld → zen</td>
<td>amnioserosa</td>
<td>dA</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>zen/next/dpp</td>
<td>dorsal line</td>
<td>d19</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>dpp</td>
<td>ectoderm</td>
<td>d11</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>twi → rho</td>
<td>dorsal</td>
<td>dN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>neuroectoderm</td>
<td>dM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sim</td>
<td>ventral</td>
<td>d14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>twi + sna</td>
<td>mesoectoderm</td>
<td>d10</td>
</tr>
<tr>
<td>Ventral</td>
<td>3</td>
<td></td>
<td>mesoderm</td>
<td></td>
</tr>
</tbody>
</table>

This selector code forms independently of the A/P homeotic selector code, so that when the cellular blastoderm is formed, the intersection of the two selector codes can then program a complex mosaic of cell fates throughout the embryo (see Foe, 1989; Bodnar, manuscript in preparation).

5. Engineering the Embryo

The semi-empirical model for embryogenesis allows implementing the theoretical model to a specific organism by inserting experimentally measured parameters. Insertion of data collected by the Drosophila community into the model provides a specific tool for experimentalists to study the developmental program for a real organism that can address the consistency and completeness of the collected data at the molecular, cellular, and organismal levels.

An engineer or chemist uses a semi-empirical model for a hierarchical system as a tool to look beyond theory and experiment to calculate experiments yet undone in which the system is changed or optimized or in which the system must interact with a changing environment. The embryogenesis model can also look beyond theory and experiment to manipulate the embryo by changing the input parameters. Thus one
A powerful tool for deducing the role of specific genes in a developmental program is to produce genetic mutants and study their effects on the expression of other genes or on the phenotype of the organism. If a computer program for development accurately reflects the genetic networks required for organismal growth, it should be able to be used in a similar manner. One can “mutate” the program by causing the chromatin value to be positive in every cell to model an ectopic mutant. One can then deduce the effect of that mutant by examining the changes in expression patterns calculated for the other genes or in the calculated homeotic selector codes [Fig. 7(c)].

To test the power of the calculational method, I compiled a comprehensive list of experimental data on Drosophila mutagenesis up to the blastoderm then compared the experimental results with the calculated results.

For example, in Program 1:

(i) In bcd− mutants the acron is transformed into a second telson (De Pomerai, 1986). In calculated bcd− mutants the anterior end of the embryo is replaced with inverted versions of PS8 to telson as judged by homeobox selector code.

(ii) In a dominant (ectopic) mutation of the tor gene, the entire anterior half of the embryo is converted into acron and the entire posterior half into telson (De Pomerai, 1986; Gilbert, 1991). In calculated tor/tll ectopic mutants the embryo is filled with acron and PS14/telson.

(iii) In Kr− mutants only four broad h stripes are seen roughly in the positions of Stripe 1, a fused Stripe 2/3/4, a fused Stripe 5/6, and Stripe 7 (Carroll & Vavra, 1989; Hooper et al., 1989). The calculated h pattern has a normal stripe 1, stripes 2 to 6 fused into a wide stripe, and a normal stripe 7.

(iv) In hb− mutants h Stripe 3 is missing; Kr expression spreads in an anterior direction compressing h Stripes 1 and 2 toward the anterior end of the embryo and spreading out the posterior stripes 4 through 7 (Carroll & Vavra, 1989; Hooper et al., 1989). The calculated h pattern is normal except for a missing stripe 3.

(v) Ectopic bcd expression results in a second Dfd stripe symmetrically placed at the posterior end of the embryo (Jack & McGinnis, 1990). This second Dfd stripe is also apparent with calculated ectopic bcd expression.

(vi) In hb− mutants Ubx expression shifts anteriorly (Steward, 1989). In calculated hb− mutants Ubx expression spreads anteriorly into PS4.

In Program 2:

(vii) Expression of the odd-numbered en stripes is lost in eve mutants while the expression of the even-numbered stripes is lost in ftz− mutants (De Pomerai, 1986; Ingham & Arias, 1992). Only the even stripes are expressed in calculated evo− mutants and odd stripes in calculated ftz− mutants.

(viii) Prolonged ectopic expression of run prior to activation of en repressed the odd-numbered en stripes. However, short pulses of ectopic run expression result in a wg/en pattern where the odd-numbered en stripes are anterior to the adjacent wg stripes; these patterns correspond with ftz expression in a single wide band (Manoukian & Krause, 1993). Calculations of ectopic run shows the pattern seen in vivo for short pulses—alternate reversals of wg/en polarity (wg/en—en/wg—wg/en—en/wg—etc.).

A complete list of calculated mutants compared with experimental results is found in Appendix C. In all, 84 mutants were found that affected the 28 genes of interest, and 66 of these were modeled adequately by the calculations. Considering the simplicity of the method, having over three quarters of the calculated mutants correspond to experiments on the real organism indicates the power of such a calculational tool in the study of genetic networks and developmental programs.

**THE EVOLUTION OF DEVELOPMENTAL PROGRAMS**

A speculation that is now common in our field is that the evolution of metazoan development may someday be understood in terms of assembly and reassembly of such intermediate-level regulatory network subelements . . . An important consequence is that the molecular biology of development is starting to converge with the molecular biology of evolution. This is particularly well illustrated in comparative studies underlying metamerization in long- and short-germ-band insects.

_Eric Davidson (1994)_

Since calculated mutant Drosophila embryos compared so well with the real mutant Drosophila embryos I next wondered, “If we can postulate what mutations to the developmental program led to the divergence of fruit flies from their short germ band ancestors, can we mutate the calculated Drosophila to
Fig. 14. Programming the beetle embryo. Calculation of gap gene and homebox gene patterns for a hypothetical short germ band insect assumed the interactions (top) modified from calculated *Drosophila* pathways (Fig. 8). Starting with morphogen gradients at cycle 6, values for each gene were calculated each nuclear cycle on the basis of current values of all gene products present. Cycles 6 through 9 were calculated as for *Drosophila* (Fig. 8). Cellularization was assumed starting with cycle 9; on subsequent cycles only cells in contact with the posterior end were assumed to cycle and divide forming a growth zone through cycle 17 when the full gap gene pattern was generated. Homebox genes were programmed as for *Drosophila* (Fig. 10) with the same interactions as before (except with hb-> AbdB).
‘de-evolve’ into that ancestor?’ Genetic studies on Drosophila embryogenesis have indicated that the posterior gradient system (nos plus hh) is totally dispensable for the developmental program (Hülskamp et al., 1989; St. Johnston & Nüsslein-Volhard, 1992—described in detail in Appendix B). In the evolution of long germ band insects from short germ band insects, it is believed that the system which causes the abdomen segments to grow sequentially in short germ band insects was mutated to allow all those segments to be determined simultaneously in the long germ band insects [reviewed in Klingler (1994)]. The analysis and computer programming of embryogenesis suggested that short germ band insects had a viable posterior system which defined the abdomen but that long germ band insects diverged by a single mutation—causing a nurse cell to inject the tor control factor (torsolike) from both ends of the egg rather than just the anterior end therefore allowing the posterior terminal system to supplant the posterior system in programming the abdomen. This mutation would allow more rapid embryogenesis in long germ band insects because all segments would be defined at once rather than sequentially. I tested this hypothesis by “de-evolving” the calculated fruit fly into its short germ band ancestor.

The program for beetle embryogenesis

Starting with the Drosophila Program 1 for parasegment and homeotic gene patterning, I made a few minor changes to simulate the embryogenesis of a short germ band insect such as a beetle or grasshopper (Fig. 14). The program was the Drosophila Program 1 modified as follows:

(i) The posterior terminal (tor/tll) gradient was deleted so that the egg had an anterior tor/tll gradient and bcd distributed evenly throughout.

(ii) The nanos (nos) protease was added in the posteriormost compartment, and a single nos pathway added where nos−→hh (i.e. nos inhibits hh by proteolytic degradation). This program is in place (and redundant) in the real Drosophila but neglected in the Drosophila program.

(iii) The Kr/kni interactions neglected in the Drosophila calculation (since Kr was both a terminal and A/P gene) were added back so that: Kr −→ kni and kni −→ Kr.

(iv) The h enhancers under control of the terminal system (Stripes 1, 2/6, and 7) were deleted.

(v) Cells divided as for Drosophila for nuclear cycles 6 through 9 when “cellularization” occurred; at that point a “growth zone” was initiated [see Fig. 4(b)] so only the posteriormost cells could divide—simulating a paracrine growth factor secreted by the cells at the posterior edge of the embryo.

(vi) The maternal effect gradients did not change as nuclei moved around in the syncytial blastoderm, but following “cellularization” maternal effect factors (bcd and nos) in the growth zone were diluted out as the original concentration was divided between daughter cells.

With these few minor changes the program could simulate beetle embryogenesis quite well!

The computer program for beetle embryogenesis (Fig. 14) was set to begin like the Drosophila program with the same projection of nuclear patterns (as in Fig. 5), with transcription starting at nuclear cycle 6, and with the same nuclear division program up to cycle 9. However, since there was only one tor/tll gradient from the anterior end of the embryo, the positional codes in the egg corresponded to those for PS0 to PS5, and by cycle 9 there was only single tor/tll gradient from the anterior end of the embryo, the positions in the egg corresponded to those for PS0 to PS5 (Fig. 14). At cycle 9 cellularization was assumed and now the only new cell growth was extension of the germ band. I assumed that regulation of germ band extension was by a growth factor that allows only the cells at the posterior edge to progress through the cell cycle, switch chromatin fates, and divide. This was modeled by changing the rules for division such that only the cell in the growth zone (the rightmost cell) would cycle and divide following cycle 9.

Gradient in space vs. gradients in time

This sequential cell division in the growth zone of the extending germ band had the effect of setting up the same gap gene pattern for the abdomen as seen for Drosophila but regulated by gradients in time rather than gradients in space! The determination of the abdomen began by diluting out the bcd and nos stored in the posterior cells; as each cell grew and divided it was assumed that factors stored in that cell (which had only a given quantity since they were not still being expressed) would be diluted each time the cell divided into two new cells. The gap gene A/P cascade then began in a temporal sequence in the growth zone. The expression of Kr increased, and since hb was destroyed by nos, kni expression could begin. Increasing kni expression then repressed Kr and activated gt. Increasing gt expression repressed kni, and with no further kni repression, hb was then activated again. Finally, increasing hb expression repressed gt and activated Kr. Comparison of the calculated gap gene patterns for Drosophila (Fig. 8) and beetle (Fig. 14) showed that both patterns had
the same progression and overlap of gap genes—including the second hb stripe near the posterior end of both embryos. Additionally, the beetle calculation had a recurrence of Kr expression at the posterior end. The posterior hb stripe has been seen in both Drosophila an flour beetle embryos (Wolff et al., 1995), and the posterior Kr stripe is also seen experimentally in Drosophila (see, for example, the experimental data in Fig. 9). Thus, the feedforward activation coupled with feedback repression of the gap genes seen spatially in Drosophila occurred temporally within the growth zone in the calculated beetle. The changes between the calculated fruitfly and beetle were consistent with prior indications that several Drosophila pathways may be redundant or evolutionary relics.

An engineering model for embryogenesis can provide a method to study the evolution of the genetic networks of development using a computer. I suggest that in the future the power of this kind of computer calculation will not be that this simple method can program any individual Drosophila pattern better than any previous models but rather that a single integrated model and calculation can account for the entire transcriptional cascade of Drosophila blastoderm formation and then be applied directly to study insect evolution as well.

6. Discussion

Much like the blind men and the elephant, theoreticians and experimentalists have been analysing development with a cacophony of voices. All of the voices seem to agree that development is a process which likely will prove to be very similar in virtually all organisms. Therefore, describing development can probably be a simpler matter than one might have thought a few years ago—if only we all can find a common language. The chemists and engineers have found such a common language to integrate theory and experiment in semi-empirical hierarchical models. However, the highest level of hierarchy and complexity the chemists need to consider is molecules; in studying development biologists need to consider additional levels—multiplying the complexity of theory and data needed. Now, with the significant advances over the last few years from both the theoretical and experimental communities studying development, all the pieces are available at all levels of the developmental hierarchy to look for a common language. I have presented a method for synthesizing theory and experiment into an integrated model by calculating the Drosophila egg on the “back-of-the-envelope” As in any “back-of-the-envelope” calculation, its primary aim is to identify the specific interactions and parameters at each level required to describe the overall system and to point toward both refined theoretical and experimental methods to describe not only the individual parts but the whole system as well.

Any integrated model depends on the strengths of the individual fields it comprises. In biology theoreticians can describe processes very well by looking inside a model, while experimentalists can describe structures very well by looking through a model, but both usually still work at a single level of hierarchy. Engineers and chemists can order hierarchies and describe the interplay of function and structure by looking beyond a model. In constructing the “back-of-the-envelope” calculation of Drosophila embryogenesis, I have taken the viewpoints in turn of a theoretician, an experimentalist, and an engineer and have made the most significant progress in constructing the model when I have looked to how the strengths of each field point toward refinements in the other fields.

EXPERIMENTS POINT TOWARD REFINED THEORY

Theoretical methods calculate how processes and interactions among individuals cause unit parameters to emerge. However, at every increasing level of hierarchy the number of possible types of individuals increases in theory, but not in what actually occurs. For example, chemists can calculate reactivities of over a hundred different atoms, but only one—carbon—is the basis for all living systems. On the next level organic chemists can calculate the properties of millions of different kinds of polymers, but only a few—proteins, nucleic acids, and carbohydrates—are found in living systems. Thus, theoreticians can calculate many different mechanisms for development at higher hierarchical levels but must depend on experimental results to indicate which of those myriad possibilities describe existing organisms. As a theoretician I have asked how well established hypotheses derived from experimental results point to particular theoretical constructs out of the many current models.

Monotonic gradients can program a striped pattern which occurs by rather precise local reading of concentration levels in the gradient by individual cells or nuclei, and the formation of multiple stripes is regulated by simple subdivision of the existing stripes. Experimental results indicate that Drosophila embryogenesis follows a multiple step transcriptional cascade starting with gradients of only three morphogens, and the experiments point toward
mechanisms which read morphogen thresholds individually rather than those which form patterns by reaction-diffusion mechanisms. Stripes in *Drosophila* appear to require multiple levels of subdivision where individual stripes are programmed at the junction of two previous stripes. In particular, the seven striped *hairy* (*h*) pattern (Fig. 9) is programmed by multiple enhancers which react to gradients of different gap genes, and deletion of a single gap gene (*hb* for example) can sometimes delete only a single stripe in the overall pattern. Therefore, experimental results pointed toward a theoretical model which accounts for a cascade of transcriptional activation with multiple temporal levels of genes each of which is programmed by threshold mechanisms using a small subset of the previous genes.

*Gene switching in genetic networks is coupled to intracellular molecular events through the cell cycle.* A cell is an active chemical system that controls intracellular concentrations of regulatory molecules. Second messenger and transactivator concentration or activity is modulated directly in various compartments within a single cell by mechanisms such as transcription, nuclear transport, or phosphorylation and can vary widely within a single cell through a single cell cycle [reviewed in Bodnar & Bradley (1996)]. Therefore, a concentration of gradient is “read” independently by the individual cells or nuclei—and ultimately by the chromatin structure of individual genes—as they progress through their individual cell cycles (Bodnar & Bradley, 1996). Each individual cell senses a gradient to switch chromatin, protein, and cell state both in space and time. Consequently, experiments indicate that an integrated model must account for the gene switching events in each individual cell during each cell cycle.

*Gene switching logic in genetic networks is stored as chromatin states, and cells contain a memory which allows sequential genetic networks to be coupled together into developmental programs.* Experiments suggest that each individual cell “remembers” its state in the combination of equilibrium chromatin configurations and transactivator concentrations as it progresses through the cell cycle. All the rules for activation of any individual gene are contained in all cells but only recalled as the input gene products become available during a developmental program. Patterns take a fixed number of cell cycles to develop—fourteen from the egg to the *Drosophila* blastoderm and eleven from the egg to the mature *C. elegans*. Therefore, experiments point toward models in which genes are switched between defined states at the cellular level—based on a memory stored in the nuclear structure.

**THEORY POINTS TOWARD NEW ENGINEERING MODELS**

Engineers can compute the dynamic characteristics of hierarchical systems, but need a refined theory at each level of the hierarchy to use as a basis for their calculations. The theories are then simplified if possible and combined to link the various hierarchical levels. As an engineer I have examined theory of gene regulation at all levels of the hierarchy and asked what characteristics of the theoretical models are well established and how can they be simplified into a hierarchical engineering model.

*Gene switching in response to morphogen gradients can be approximated by Boolean logic.* Theoreticians have shown by a variety of methods, including the most established reaction-diffusion and genetic network models (Meinhardt, 1986; Nagoreka, 1988; Edgar et al., 1989; Lacalli, 1990; Lyons & Harrison, 1992; Reinitz & Sharp, 1995), that genes activated or inhibited by a morphogenic gradient form a steep sigmoidal activation curve. Analysis of molecular and cellular biological models for gene activation supports a threshold model for gene activation (Bodnar & Bradley, 1996). Regardless of the exact details, all these theoretical models can be approximated into an engineering model as a step function at a threshold of input transactivator concentration.

**Digital or continuous models for gene switching apply at different scales.** A chromatin switching model assumes a digital threshold while reaction-diffusion and genetic network models assume continuous morphogen concentration gradients for gene activation. Both the continuous and digital approaches use the same chemical principles—merely on a different scale. Just as quantum mechanical approaches to physics correspond to traditional approaches as one goes from individual atoms to many atoms (see Ehrenfest, 1927), the digital and continuous approaches correspond as one goes from the few individual cells seen early in development to the continuous sheets of many cells seen late in development. Thus the digital approach is useful for modeling early embryogenesis while continuous models are more useful for modeling processes later in development.

**ENGINEERING MODELS POINT TOWARD NEW EXPERIMENTS**

Experimentalists can measure parameters of the structures on any level of the hierarchy, but cannot define the processes that link them. For example, chemists can measure electronegativities of atoms and organic chemists can measure dipole molecules of molecules, but both must depend on theoretical calculations to explain the processes by which they are related. Similarly, experimentalists can determine the
primary structure of a protein and grow crystals of that protein but must depend on the complex calculational tools of X-ray crystallography to determine its secondary, and tertiary structures.

In studying development experimentalists can measure the interactions of genes, gene products, or cells that form an organism but need a calculational tool to link the molecular, cellular, and organismal events in embryogenesis. The “back-of-the-envelope” calculation provides detailed estimates to many cross-disciplinary questions which can be valuable in linking experimental results and planning new experiments (see Calculations, Appendices, and computer programs themselves). Should I expect any more input genes to affect my target gene directly? Which effects on the target gene are direct and which are indirect through the transcriptional cascade, i.e. is the regulator a direct activator or is it an inhibitor of an inhibitor? Are any genes missing from the currently known genetic network? If there are any missing genes, where in the network do they work, what do they program, and when in embryogenesis can I expect them to be expressed? What effects on development can I expect if I mutate a particular gene? How can I catalog all the interactions among the many developmental genes in a manner where I have an overview of the entire process of embryogenesis yet can easily trace the effects and interrelationships of any individual gene?

ENGINEERING MODELS POINT TO ALL LEVELS OF THE PROCESS

Engineers usually begin with the simplest model possible—a “back-of-the-envelope” calculation. This strategy is based on the notion that if one understands the underlying theory of a process, a reasonably good approximation to that process can be obtained by a very simple model. The model obtained also keeps an overall perspective on the interplay of process and structure—studied as theory and experiment. The interplay between theory and experiment is maintained in the semi-empirical approach by using theoretical calculations to define the required parameters while using experimental results to set their values. Here I have taken complicated theory and made only a few new assumptions which approximate cell cycle-dependent chromatin switching networks as simple digital computing networks and have written two very simple computer programs (containing about 50 k of code in their compiled version) which insert experimentally defined parameters and describe Drosophila embryogenesis very well. The digital nature of the calculation allows an overview of the entire process to be maintained at any time yet one can easily focus on any individual part of the process—on an individual gene, genetic network, or cell—and trace how theory and experiment interact to explain Drosophila embryogenesis at any level. For engineering “back-of-the-envelope” calculations, if the theory is appropriate and quantitated properly through experiment, simple calculations can give tremendous insight into complex processes.

Engineering models integrate theory and experiment into a coherent whole that can provide an overview of system dynamics yet focus directly on the role individual system components. Now that we can calculate the egg, both experimental and theoretical biologists can draw on the integrated model to address several more general questions about development. “How many genes control development—as distinct from providing the housekeeping functions of the cell (Wolpert, 1994)” How do morphogen gradients install different regulatory states in the territories from which the different parts of the organism will develop? What kinds of gene regulatory networks are required to control development? What is their complexity and structure; their use of widely shared subelements; their response to perturbation; their linkage; and their degree of degeneracy? As a community we now can start to catalog the “various semi-complex, modular, ‘integrated circuits’ [of development] that are to some extent experimentally interchangeable” (Davidson, 1994).

Engineering models also allow one to ask very easily “what would happen if I changed the system?” In the simplest case here, that allowed testing and refining the model by comparing calculated mutations with actual mutations. The calculation could also be modified to study evolution by asking “what changes are there between the developmental programs of short germ band and long germ band insects?” Additionally, the computer program as written is modular so that by changing the rules for cell division (with certain genes segregating asymmetrically each division), the exact genes in the organism and logical rules for their calculation, and the distribution of maternal effect genes in the egg, one could program the developmental pathway for C. elegans by changing only the organism-specific procedures in the computer program.

PROGRAMMING BIOLOGICAL COMPUTERS

In the midst of writing the computer program for a developmental program it became apparent that the double-entendre of “programming” reflects a fundamental characteristic of information systems (Bodnar,
that biological information systems are, in essence, biochemical computers. Much is known about the “hardware” of biochemical computers, but currently little is known about the “software”. As I wrote the computer program in Pascal, the logic and syntax of the Pascal programming language closely paralleled the logic and syntax of the developmental program being modeled; DNA domains corresponded directly to string variables, gene products to procedures, cell cycles to DO loops, patterns of nuclei to arrays, and growth of the organism to running the compiled computer program. It is also clear that the correspondence of syntax would be even closer if written in an Object-Oriented Programming (OOP) language (Waldrop, 1993). I suggest that deciphering the programs for development written on the non-coding portions of the genome will not correspond to reading the “book of life” like a novel but rather to deciphering the compiled version of an Objects Pascal computer program. The methods outlined here suggest the type of syntax used in the logic statements in the “software” of developmental gene regulation.

PROSPECTS

During the analysis, modeling, and calculation of Drosophila embryogenesis, I often found the theory suggested a few specific possible interactions that could allow the program to run better. By searching the literature for experimental evidence of those potential interactions, I usually found that most of the experimental tests were already done. The enormity of the data accumulated to date on Drosophila and other model organisms suggests that by applying engineering models as outlined here an enormous amount of current theory and data can be integrated without doing any new laboratory experiments. One straightforward followup example is obvious from the results of this Drosophila calculation; by comparing (1) the stripe patterns formed at the Drosophila blastoderm stage to define a transcriptional fate map, (2) the mitotic domains of synchronously dividing cells in the blastoderm to define a cell cycle fate map, an (4) mutational studies of defects in the resulting embryo to define a phenotypic fate map, one should be able to synthesize an integrated transcriptional, cell cycle, and phenotypic fate map that can account for genetic pathways in the phenotypic programming of the Drosophila embryo (Bodnar, manuscript in preparation). A short catalog of current literature suggests that enough data is available to do the same for C. elegans embryogenesis as well as limb development and initial organ specification in vertebrates.

I thank Charles Rowell, Margaret Bradley and Philip Moen for helpful discussion and critical reading of the manuscript. I thank Erik Chaum, Will Henrickson, and Robert Siddon for insights into applying engineering models to biological systems. I thank Aileen Buchanan for cataloging the many literature references and Amy Breitig and Barbara Yoakum for help in using the library databases. This project was funded in part by the Naval Academy Research Council.

REFERENCES


selective nuclear transport in neurogenesis.


knowing your neighbors: Cell interactions determine intrasegmental patterning in Drosophila. TIGS 8, 243–249.


the genetic basis of patterned baldness in Drosophila. Cell 76, 781–784.

history and failure. Amer. Sci. 80, 523–526.


The control of cell fate along the dorsal-ventral axis of Tribolium (Coleoptera). Nature 361, 448–450.


Spacing ensures autonomous expression of different stripe enhancers in the even-skipped promoter. Development 119, 767–772.

Regulation of even-skipped stripe 2 in the Drosophila embryo. EMBO J. 11, 4047–4057.


Individual stripe elements in the Drosophila hairy promoter respond to maternal, gap, and pair-rule genes. Genes and Dev. 5, 840–854.


The expression and regulation of Sex combs reduced protein in Drosophila embryos. Genes and Dev. 1, 716–730.


The graded distribution of the dorsal morphogen is initiated by selective nuclear transport in Drosophila. Cell 59, 1165–1177.

The Drosophila spitzen gene encodes a putative EGF-like growth factor involved in dorsal-ventral axis formation and neurogenesis. Genes and Dev. 6, 1503–1517.


Spacing ensures autonomous expression of different stripe enhancers in the even-skipped promoter. Development 119, 767–772.

Regulation of even-skipped stripe 2 in the Drosophila embryo. EMBO J. 11, 4047–4057.


Involvement of an ortholog of the Drosophila pair-rule gene hairy in segment formation of the short germ-band embryo of Tribolium (Coleoptera). Nature 361, 448–450.


The polarity of the dorsoventral axis in the Drosophila embryo. Dev. Biol. 113, 64–76.

Relocalization of the dorsal protein from the cytoplasm to the nucleus correlates with its function. Cell 59, 1179–1188.


APPENDIX A

Boolean gene switching rules for Drosophila Embryogenesis

The initial conditions and rules used in calculating the Drosophila and beetle development programs were written as Boolean logic statements as summarized below. Three separate calculations were done (two computer programs and one manual calculation) for Drosophila embryogenesis and one for beetle embryogenesis as described in the text.

Experimental interactions are referenced, and postulated interactions required to make the calculation run properly are listed as “presumptive”. Most experiments did not discriminate between weak, intermediate, or strong activators and repressors; the exact rule listed is the fit from the experimental data that allows the calculation to reflect the overall development process the best.

Initial conditions—indicate experimentally determined localization in the egg (with supporting references) for the maternal effect genes.

Boolean gene switching logic—indicates the upstream regulators of each gene and the switching rules for its regulation used in the calculations. The logic statements for each individual gene using the shorthand notation from Fig. 1(b) are listed in the order used to step through the computer program.

Program I

Programming Positional Values by Maternal Effect Genes

Programming the Gap Gene Pattern by Positional Values

Programming Pair-Rule Stripes by Gap Genes

Initial conditions

Maternal effect genes

**bicoid (bcd)**

mRNA stored in A/P gradient over anterior 2/3 of egg

(St. Johnston & Nüsslein-Volhard, 1992)

**torso (tor)**

mRNA homogeneous in egg

(Casanova & Struhl, 1989)

Protein activated in gradients from both ends of egg (likely by *torsolike*)

(Casanova & Struhl, 1989)

(St. Johnston & Nüsslein-Volhard, 1991)

**nanos (nos)**

mRNA in posterior-anterior gradient (neglected)

(St. Johnston & Nüsslein-Volhard, 1992)

**maternal hunchback (hb)**

3.2 kb mRNA homogeneous throughout egg (neglected)

(Schröder et al., 1988)

**tramtrack (ttk)**

Homogeneous throughout egg

(Brown & Wu, 1993)

Terminal genes

**tailless (tll)**

\(tor \rightarrow tll\)

(Mahoney & Lengyel, 1987; Brönnner & Jäckle, 1991; Liaw & Lengyel, 1992; Pignoni et al., 1992)

**bcd**\(\rightarrow tll\) (neglected)

(Mahoney & Lengyel, 1987; Brönnner & Jäckle, 1991; Liaw & Lengyel, 1992; Pignoni et al., 1992)

**dl**\(\rightarrow tll\) (neglected)

(Mahoney & Lengyel, 1987; Brönnner & Jäckle, 1991; Liaw & Lengyel, 1992; Pignoni et al., 1992)
Posterior genes—neglected in calculation

**nanos (nos)**
maternal effect only (neglected in calculation)
(St. Johnston & Nüsslein-Volhard, 1992)

Gap genes and terminal genes

**Krüppel (Kr)**

constitutive promoter  (Gaul & Jäckle, 1990)

\( hb \rightarrow Kr \)
(Meinhardt, 1986; Jacob et al., 1991; Hoch et al., 1992)

\( bcd \rightarrow Kr \)
(Meinhardt, 1986; Jacob et al., 1991; Hoch et al., 1992)

\( tor \rightarrow Kr \)
(Weigel et al., 1990)

\( gr \rightarrow Kr \)
(Capovilla et al., 1992)

\( tll \rightarrow Kr \)
(Gaul & Jäckle, 1990; Hoch et al., 1992)

\( kni \rightarrow Kr \) (neglected)
(Hoch et al., 1992)

**giant (gt)**

Presumptive

\( tor \rightarrow gt \)

\( hb \rightarrow gt \)
(Struhl et al., 1992)

\( tll \rightarrow gt \)
Presumptive

\( Kr \rightarrow gt \)
(Capovilla et al., 1992)

**kniips (kni)**

constitutive promoter

Presumptive

\( Kr \rightarrow kni \)
(Pankratz et al., 1989)

\( bcd \rightarrow kni \)
(Hulskaump & Tautz, 1991)

\( tll \rightarrow kni \)
(Pankratz et al., 1989)

\( hb \rightarrow kni \)
(Treisman & Desplan, 1989)

\( Kr \rightarrow kni \) (neglected)
(Pankratz et al., 1989)

**zygotic hunchback (hb\{z\})**

constitutive promoter

Presumptive

\( bcd \rightarrow hb \)
(Schröder et al., 1988)

\( hb \rightarrow hb \)
(Treisman & Desplan, 1989)

\( Kr \rightarrow hb \)
(Treisman & Desplan, 1989)

\( tll \rightarrow hb \)
(Schröder et al., 1988)

\( kni \rightarrow hb \)
Presumptive

Pair rule genes

**tramtrack (tkk)**

Degraded starting at Cycle 9
(Brown & Wu, 1993)

**hairy (h)**—(Individual enhancers upstream of transcription start site)

- **Stripe 1 enhancer** (−4.9 to −4 kilobase pairs upstream [kb])

\( bcd \rightarrow h1 \)
(Howard & Struhl, 1990; Riddihough & Ish-Horowicz, 1991)

\( tll \# \rightarrow h1 \)
Presumptive (# = tll activation only if \( bcd \) also present)

\( h \rightarrow h1 \)
(Howard & Struhl, 1990; Riddihough & Ish-Horowicz, 1991)

- **Stripe 2 and 6 enhancer** (stripe 2 = −9.4 to −4 kb and stripe 6 = −9.1 to −5.2 kb)

\( bcd \rightarrow h2/6 \)
Presumptive

\( kni \rightarrow h2/6 \)
(Pankratz et al., 1990; Riddihough & Ish-Horowicz, 1991)

\( Kr \rightarrow h2/6 \)
(Pankratz et al., 1990; Riddihough & Ish-Horowicz, 1991)

\( tll \rightarrow h2/6 \)
(Pankratz et al., 1990; Riddihough & Ish-Horowicz, 1991)

\( gt \rightarrow h2/6 \) (neglected)
(Riddihough & Ish-Horowicz, 1991)

- **Stripe 3 and 4 enhancer** (−12.4 to −11 kb)

**Stripe 3**

\( hb \rightarrow h3 \)
(Hartmann et al., 1994)

\( h \rightarrow h3 \)
(Howard & Struhl, 1990; Riddihough & Ish-Horowicz, 1991)

\( kni \rightarrow h3 \)
(Pankratz et al., 1990; Hartmann et al., 1994)

**Stripe 4**

\( Kr \rightarrow h4 \)
(Howard & Struhl, 1990; Riddihough & Ish-Horowicz, 1991)

\( kni \rightarrow h4 \)
(Pankratz et al., 1990; Riddihough & Ish-Horowicz, 1991)

\( bcd \rightarrow h4 \)
(Howard & Struhl, 1990; Riddihough & Ish-Horowicz, 1991)

- **Stripe 5 enhancer** (−6.8 to −4 kb)

\( kni \rightarrow h5 \)
(Pankratz et al., 1990; Riddihough & Ish-Horowicz, 1991; Langeland & Carroll, 1993)

\( Kr \rightarrow h5 \)
(Pankratz et al., 1990; Langeland & Carroll, 1993)

\( gt \rightarrow h5 \)
(Riddihough & Ish-Horowicz, 1991; Langeland & Carroll, 1993)

\( Kr \rightarrow h5 \) (neglected)
(Riddihough & Ish-Horowicz, 1991)

- **Stripe 7 enhancer** (−11 to −9.4 kb)

\( tor \rightarrow h7 \)
(Howard & Struhl, 1990; Pankratz et al., 1990; Riddihough & Ish-Horowicz, 1991)

\( tll \rightarrow h7 \)
Presumptive

\( kni \rightarrow h7 \)
(Howard & Struhl, 1990; Pankratz et al., 1990; Riddihough & Ish-Horowicz, 1991)

\( bcd \rightarrow h7 \)
Presumptive

\( ttk \rightarrow h \)
(Brown & Wu, 1993)

**even-skipped (eve)**—neglected in calculation

- **Stripe 2 enhancer** (−1.65 to −1.15 kb)

\( bcd, hb \rightarrow eve2 \)
(Driever & Nüsslein-Volhard, 1988b; Small et al., 1991, 1992)

\( gt, Kr \rightarrow eve2 \)
(Stanojevic et al., 1989, 1991; Small et al., 1993)

- **Stripe 3 enhancer** (−3.8 to −2.9 kb)

\( bcd, hb \rightarrow eve2 \)
(Goto et al., 1989)
"fushi tarazu" Deformed Homeotic genes (Sex combs reduced)

**Ultrabithorax (Ubx)**

- **Antp → Ubx** (Presumptive) (González-Reyes et al., 1990)
- **Abd-B → Ubx** (Lamka et al., 1992)
- **ill → Ubx** (Reinitz & Levine, 1990; Qian et al., 1991)
- **hb → Ubx** (Qian et al., 1991; Zhang et al., 1991)

**abdominal-A (abd-A)**

- **kni → abd-A** (Presumptive) (Lamka et al., 1992)
- **Abd-B → abd-A** (Lamka et al., 1992)

**Abdominal-B (Abd-B)**

- **tor → Abd-B** (Reinitz & Levine, 1990)
- **hb → Abd-B** (Busturia & Bienz, 1993)
- **Kr → Abd-B** (Busturia & Bienz, 1993)
- **bcd → Abd-B** (Presumptive)
- **kni → Abd-B** (neglected) (Harding & Levine, 1988; Busturia & Bienz, 1993)

### Program 2

**Refinement of Pair-Rule Stripes**

**Setting Segment Polarity**

**Initial conditions**

**hairy (h)**

The output of the first program gave h stripes that were 4 nuclei wide at cycle 12 but with square edges (00111100). The input for second program (Fig. 9) looked at two h stripes that were 5 cells wide and broad on the edges (0123100).

**even-skipped (eve)**

The other input for the second program was an eve pattern of four nuclei-wide stripes (0122100) assumed to be programmed by the gap genes. Once the second calculation began interactions of the pair-rule genes modulated the eve pattern.

**Boolean gene switching logic**

**hairy (h)**

run → h (Carroll & Vavra, 1989; Warrior & Levine, 1990; Manoukian & Krause, 1993)

**run** (run)

- **ttk → h** (Brown & Wu, 1993)
- **run → eve** (Carroll & Vavra, 1989; Warrior & Levine, 1990; Manoukian & Krause, 1993)
- **ttk → eve** (Read & Manley, 1992)

**even-skipped (eve)**

- **eve → eve (late)** (Goto et al., 1989; Jiang et al., 1991)
- **prd → eve (neglected)** (Treisman et al., 1991)
- **h → eve** (Carroll & Vavra, 1989; Warrior & Levine, 1990; Manoukian & Krause, 1993)
- **ttk → eve** (Read & Manley, 1992)

**fushi-tarazu (ftz)**

- **ftz → ftz** (Kaufman et al., 1990; Pick et al., 1990)
- **h → ftz** (Carroll & Vavra, 1989)
- **eve → ftz** (Manoukian & Krause, 1992)
- **ttk → ftz** (Read & Manley, 1992)

**paired (prd)**

Gap genes → prd (Presumptive)
\[ \text{eve} \rightarrow \text{prd} \quad \text{(Manoukian & Krause, 1992)} \]
\[ \text{fitz} \rightarrow \text{prd} \quad \text{Presumptive} \]
\[ \text{fitz} \rightarrow \text{prd} \quad \text{(neglected)} \]
\[ \text{(Baumgartner & Noll, 1991)} \]
\[ \text{h, run} \rightarrow \text{prd} \quad \text{(neglected)} \quad \text{(Ingham, 1988)} \]
\[ \text{wingless (wg)} \]
\[ \text{prd} \rightarrow \text{wg} \quad \text{Presumptive} \]
\[ \text{eve} \rightarrow \text{wg} \quad \text{(Manoukian & Krause, 1992)} \]
\[ \text{fitz} ightarrow \text{wg} \quad \text{(Ish-Horowicz et al., 1989)} \]
\[ \text{wg, en, hh} ightarrow \text{wg} \quad \text{(neglected)} \quad \text{(Peifer & Bejsovic, 1992)} \]
\[ \text{engrained (en)} \]
\[ \text{ftz} ightarrow \text{en} \quad \text{(Ish-Horowicz et al., 1989; Morrissey et al., 1991)} \]
\[ \text{eve} ightarrow \text{en} \quad \text{(Morrissey et al., 1991)} \]
\[ \text{wg} ightarrow \text{next} ightarrow \text{en} \quad \text{(Heemskerk et al., 1991; Peifer & Bejsovic, 1992)} \]
\[ \text{run} ightarrow \text{en} \quad \text{(neglected)} \quad \text{(Manoukian & Krause, 1993)} \]

**APPENDIX B**

### Regulation of Individual Genes

The discussion of each individual gene includes the following: a brief description of the gene, additional important literature data needed in the analysis and previous hypotheses on regulation of the gene.

**Calculation**—description of the computer calculation, the best estimate of mechanisms for regulation of that gene’s localization, and any discrepancies between the calculations and experimental data.

**Prospects**—predictions and speculation from the theory and calculation on the current state of knowledge of the gene. Is the regulation of this gene well established experimentally? What experiments are needed to fill in the gaps? What does the specific analysis suggest about importance, mechanisms of activation, or evolution of this gene? Where possible specific experiments are suggested to test the hypotheses stated.

**Program 3**

**Programming Dorsal/Ventral Patterns**

**Initial conditions**

**dorsal (dl)**
Homogeneous throughout the egg (Steward, 1989; (Rushlow et al., 1989)

**Boolean gene switching logic**

**dorsal (dl)**
Activated in a gradient by nuclear translocation by Toll (Roth et al., 1989; (Stein et al., 1991)

**twist (twi)**
\[ dl ightarrow twi \] (Thisse et al., 1991)
\[ twi ightarrow twi \] (Leptin, 1991)

**snail (sna)**
\[ dl ightarrow sna \] (Ip et al., 1992b)
\[ twi ightarrow sna \] (Leptin, 1991; Kosmann et al., 1991; Ip et al., 1992b)

**decapentaplegic (dpp)**
\[ dl ightarrow dpp \] (Huang et al., 1993)
\[ tld ightarrow dpp \quad \text{(neglected)} \] (Ferguson & Anderson, 1992)

**zerknuilt (zen)**
\[ dpp \rightarrow (tld) ightarrow zen \] (Rushlow & Levine, 1990)
\[ dl ightarrow zen \] (Rushlow & Levine, 1990; Thissse et al., 1991)

**rhomboid (rho)**
\[ dl ightarrow rho \] (Ip et al., 1992a)
\[ twi ightarrow rho \] (Ip et al., 1992a)
\[ sna ightarrow rho \] (Ip et al., 1992a)

**singleminded (sim)**
\[ twi ightarrow sim \] (Presumptive)

\[ \text{rho} \rightarrow \text{next} \rightarrow \text{sim} \quad \text{Presumptive} \]
\[ \text{sna} ightarrow \text{sim} \quad \text{(Kasai et al., 1992)} \]
Prospects. The identity of bcd as the anterior morphogen has been solidly established experimentally. The calculation indicated that bcd programs the fate of the head, thorax, and anterior end of the abdomen.

Torso (tor)

Torso (tor) appears to be the crucial gene for organizing the terminal pattern in the embryo (Gaul & Jäckle, 1990). The mRNA for tor is distributed homogeneously throughout the fertilized egg, but translation of the tor protein can be detected in the fourth or fifth nuclear division (Casanova & Struhl, 1989). The tor protein is a cell surface receptor found ubiquitously throughout the embryo but apparently only activated by a spatially restricted ligand, possibly the torsolike gene product (Casanova & Struhl, 1989; Nüsslein-Volhard, 1991). The activity of tor appears to induce an unknown transactivator (“gene Y”) that determines the domains of expression of tailless and huckebein in a concentration-dependent manner, defining two thresholds (Nüsslein-Volhard, 1991).

Calculation. Torso was modeled as a maternal effect gene where gradients of tor activity are set in the egg and unchanged thereafter [Fig. 6(a)]. Activation of downstream genes assumed a transactivator intermediate (“gene Y”) that mediates the tor gradient.

Prospects. The torso gene has been solidly established as a morphogen, but since it is a cell surface receptor, it must act through a transactivator gene—gene “Y” (Nüsslein-Volhard, 1991) which is currently not known. The calculation is consistent with the hypothesis that another transactivator besides tailless and huckebein is necessary to mediate the torso gradient to the nucleus. The need for the torso signal throughout most of the abdomen was a surprise, but this localization is consistent with torso taking over the nanos/hunchback maternal function in the patterning of the abdomen in long germ band insects (see below).

Tailless (tll)

The tailless (tll) protein is a putative transcriptional activator of the steroid receptor superfamily (Pignoni et al., 1992). The tll gene product is not present in the egg but is activated by the terminal system in anterior and posterior caps covering 15% of the embryonic termini; the anterior cap becomes a stripe at the cellular blastoderm stage mediated by bicoid and dorsal repression (Mahoney & Lengyel, 1987; Liaw & Lengyel, 1992; Brönner & Jäckle, 1991). Thus, tll appears to be one of several transactivators that decode the torso activity gradient in terms of a transcriptional signal.

Calculation. As indicated above tll was modeled as a maternal effect gene due to its role far upstream in the regulatory pathway; therefore, the gradient of tailless was set initially [Fig. 6(a)] and remained the same throughout the calculation. The bcd and dl interactions with tll were neglected because they allow regulation within the acron outside the segmented region and also integration of A/P and D/V signals within the head region.

Prospects. Tailless, huckebein, and forkhead (not modeled) appear to mediate the torso gradient at both termini of the embryo. This allows the torso gradient to act through multiple thresholds where at low concentrations torso regulates Krüppel and giant in the center of the embryo, at higher concentrations regulates tailless in PS1, PS2, PS12, and PS13, and at even higher concentrations regulates huckebein and forkhead in the acron and telson (Weigel et al., 1989, 1990; Casanova, 1990). Therefore, a more refined version of the calculation could address pattern formation within the acron and telson by assuming that activation thresholds of huckebein and forkhead are at even higher torso levels than tailless and modeling their regulation and that of downstream genes (such as orthodenticle, buttonhead, empty spiracles, and spalt) (Cohen & Jurgens, 1990) as was done for tailless.

Nanos (nos)

Nanos (nos) and hunchback (hb) appear to be the critical genes in a posterior determining system in which nos prevents repression of gap genes in the abdominal region by hb; this effect is likely mediated by nos-induced degradation of maternal hb mRNA in the posterior regions of the syncytial blastoderm (St. Johnston & Nüsslein-Volhard, 1992; Wang & Lehmann, 1991). However, those hb/nos system appears to be redundant or an evolutionary remnant, since hb−/nos− double mutant eggs develop normal abdomens and give rise to fertile adults (Hülskamp et al., 1989; St. Johnston & Nüsslein-Volhard, 1992).

Calculation. The nanos/hunchback maternal effect system was not used in the calculation. This makes the model the equivalent of a maternal nos−/hb− mutant.

Prospects. The calculation of Drosophila embryogenesis without the nos/hb maternal posterior system is consistent with the notion that this system is an evolutionary relic (St. Johnston & Nüsslein-Volhard, 1992) which was used in embryogenesis of short germ band insects but displaced in function by the terminal system to define the abdomen in long germ band insects.
Maternal hunchback (hb)

There are two hunchback (hb) transcripts which both express the same protein product: the maternal product in the egg is a 3.2 kb transcript deposited homogeneously throughout the egg and a 2.9 kb transcript is expressed zygotically (Schröder et al., 1988). The maternal hb product is lost from the posterior end of the embryo through the action of the nanos gene product.

Calculation. The maternal hb was not used in the calculation as described above.

Gap genes and terminal genes

The individual terminal (Krüppel, giant) and gap (hunchback, knirps) genes are considered separately. Regulatory programs used in the calculation for each gene are defined in terms of the input regulatory genes where the Boolean transcription rules [Fig. 1(b)] are listed in the order used in the computer program.

Krüppel (Kr)

The Krüppel (Kr) gene encodes a zinc-finger protein expressed in a broad stripes in the center of the embryo. The Kr protein appears to activate transcription as a monomer but homodimerizes at high concentrations to become a repressor acting through the same target DNA sequence (Sauer & Jäckle, 1990). It appears that Kr receives only negative inputs from the maternal genes (Gaul & Jäckle, 1990).

Calculation. Krüppel appears to be both a terminal gene and A/P gene. In the calculation initial expression of Kr was constitutive but then defined as a broad band in the center of the embryo by repression from both ends through torso. In the wild type (wt) calculation no others inputs were necessary for the localization of the Kr band. However, analysis of mutants indicated that Kr is part of the A/P French flag mechanism with bcd as an input gene, hb marking the anterior edge of the Kr stripe by feedforward regulation, and kni marking the posterior edge by feedback regulation.

In the simple 3-step gradient the kni interaction was neglected because no stable solutions were possible if it was incorporated along with the terminal regulation. This is likely due to very weak repression of kni by Kr homodimers at very high Kr concentrations which affect but cannot totally abolish kni expression.

Prospects. The calculation indicated that Kr is regulated both as a terminal and A/P gene, but that the major inputs appear to be through the terminal system. This suggests that Kr was originally an A/P gene in short germ band insects and some of those controls remain after its regulation was supplanted by the terminal system in the long germ band insects. This is consistent with the very weak Kr/kni interactions which needed to be neglected for pattern stability in the three-step gradient model. A more refined computer model could account for the interactions between the terminal and A/P systems as well as the neglected posterior system by allowing weak (vestigial) interactions at this point.

Giant (gt)

The giant (gt) gene is a transactivator required for formation of PS3 and PS10-12 with weak effects on PS4 and PS13 (Petschek et al., 1987).

Calculation. The gt gene was modeled as a terminal gene. Therefore, the localization of the two gt stripes was determined by a stripe-doubling mechanism where gt is activated by tor between the central Kr stripe and the terminal till stripes. This accounts well for its distribution in two symmetric stripes about equidistant from either end of the embryo. An additional repression from the hb gene adds crosstalk from the A/P system and modulates the anterior gt stripe.

Prospects. The giant gene appears to be a very recent addition in the evolution of insect patterning. While its location in the transcriptional cascade is very early, it looks to be a recent terminal gene that helps define the abdomen in the long germ band and to provide a new subdivision within the head.

Knirps (kni)

The knirps (kni) gene is a transactivator necessary for correct determination of the abdomen (Nauber et al., 1988).

Calculation. The knirps gene was modeled as an A/P gene under feedback regulation by bcd and feedback regulation by hb. Therefore, upon activation its anterior border was determined mainly by bcd; additional regulation of its anterior edge comes from Kr, but as noted above in the simple 3-step gradient calculation kni/Kr interactions gave no stable results and were neglected. The posterior edge of the kni stripe was determined by hb.

Prospects. The kni gene (possibly in conjunction with the related knrl gene) likely determines the abdominal segments in a cascade mediated by the Ultrabithorax gene (see below). Again, I found that Kr/kni interactions needed to be neglected for stability in my simple model. It is likely that Kr activates kni at low protein concentrations (Pankratz et al., 1989) but weakly represses kni at very high concentrations (Sauer & Jäckle, 1993); this suggests
that Kr can reduce but not totally repress kni at high Kr concentrations. However, the simple 3-step gradient model does not allow the refinement to account for the crosstalk between the terminal and A/P classes of genes that comes from having Kr as a member of both classes.

**Zygotic Hunchback** (hb)

A single **hunchback** (hb) protein is expressed from two separate transcripts, a maternal 3.2 kb mRNA deposited homogeneously in the egg (see above), and a 2.9 k mRNA transcribed zygotically (Schröder et al., 1988). Zygotic hb transcripts first appear in the anterior half of the syncytial blastoderm, but soon become confined to an anterior region which abuts on the Kr transcript zone (Schröder et al., 1988; Gaul & Jäckle, 1990). Slightly later, a narrow band of hb transcripts appears in a posterior position corresponding to the primordia of A7/8 (De Pomerai, 1986).

**Calculation.** The zygotic hb gene has two regions of activation. The wide anterior stripe was modeled as the first target gene in the A/P French flag mechanism where it is activated by the bicoid input gene, its posterior edge regulated by feedback from Kr, and its anterior edge defined by the terminal tll product. The narrow posterior stripe was defined by the terminal system; there is a gap in the abdomen pattern between kni and tll in which hb is not repressed leading to an additional “terminal” hb stripe.

**Prospects.** The zygotic hb gene as an A/P gene under the control of bicoid appears to be a major determinant of the head and thorax, a role it likely has in short germ band insects. I suggest that the border between hb and kni at the PS5/PS6 boundary was the anterior border of the germ band in the short germ band insects. However, there is crosstalk with the terminal system in the regulation of hb where the second hb stripe is determined in the posterior terminal region.

**Pair rule genes**

**tramtrack (ttk)**

**Tramtrack** (ttk) encodes two related proteins which appear to inhibit transcription by sequence-specific DNA-binding through alternately spliced zinc-finger domains (Brown & Wu, 1993). Staining for the ttk protein is observed during growth of the syncytial blastoderm but fades and is completely absent in the cellular blastoderm (Brown & Wu, 1993).

**Calculation.** Since ttk represses all the primary pair-rule genes (see h, eve, ftz, and runt below), the degradation of ttk at cycle 9 was used as a timing signal in the calculation to begin expression of the pair-rule pattern after the gap gene pattern was established. However, the result of the degradation was not seen until cycle 12 because of the dynamics of cell cycle-dependent chromatin switch: if the protease to degrade ttk is switched on and expressed in G2 of cycle 9 any free ttk will disappear during cycle 9 but DNA-bound ttk will be unaffected; DNA-bound ttk is displaced during S phase of cycle 10 and degraded while activators of the pair-rule genes bind their DNAs at that time; during G1 of cycle 11 the pair-rule gene products are finally expressed in the calculation; and chromatin switched by those products is evident in G1 of cycle 12.

**Prospects.** The ttk gene appears to be a developmental timing signal that allows smooth transition from expression of the gap genes to stripe-doubling in expression of the pair-rule gene patterns. Since the gap gene pattern is complete by cycle 11, repressing the pair-rule genes until that time allows a stable gap gene pattern to be established prior to activation of the pair-rule genes. The degradation of ttk occurs about the same time as migration of the nuclei to the membrane of the syncytial blastoderm, and it is likely that the signal for nuclear migration and ttk degradation are related.

**Hairy (h)**

The hairy (h) promoter includes extensive upstream regions covering 14 kb necessary for correct striping (Howard & Struhl, 1990; Pankratz et al., 1990; Riddihough & Ish-Horowicz, 1991). As shown above and in Fig. 9, mutational analysis of upstream regions indicates that individual h stripes are independently regulated by different enhancers where deletions of the appropriate DNA sequences only causes the loss of the given stripe(s) [reviewed in Gilbert (1991)]. However, the expression of the h promoter is not autoregulated (Hooper et al., 1989). Sequences necessary for expression of Stripes 3 and 4 extend outside the 14 kb upstream region, but a construct containing −12.4 to −11 kb drives expression in a broad central stripe (3/4), roughly from Stripe 3 to interstripe 4/5 (Riddihough & Ish-Horowicz, 1991). Constructs containing the Stripe 6 defining sequences (−9.1 to −5.2 kb) also drive weak expression of Stripe 2 (Riddihough & Ish-Horowicz, 1991). The Stripe 2 and 6 response appears to overlap since a construct that gives only Stripe 6 triggers a Stripe 2 response in the reverse orientation (Howard & Struhl, 1990). Expression of h in an anterior domain (Stripe 0) requires sequences 3′ to the h transcription unit (Riddihough & Ish-Horowicz, 1991).
Calculation. Each h stripe enhancer was considered in turn. Boolean gene switching rules were written for each enhancer, and transcription of h was considered active if any one of the enhancers was activated by the gap gene positional code in that region. In general, as described above each stripe had a gap gene input, a gap gene anterior repressor, and a gap gene posterior repressor (Fig. 9).

Stripe 1 was driven by its own enhancer which depends mainly on the terminal genes for regulation. The calculation used tll as the input gene which must act synergistically with bicoid to activate the stripe 1 promoter; the anterior repressor is tll; and the posterior edge is defined by the tll activation threshold (Fig. 9).

Stripes 2 and 6 were driven by overlapping sequences in a common enhancer. Both are repressed by tll and Kr where tll regulates the anterior edge of stripe 2 and the posterior edge of stripe 6 while Kr regulates the posterior edge of stripe 2 and the anterior edge of stripe 6. (This would account for the overlapping sequences in their enhancers). However, activation of stripe 2 appears to be by bcd while activation of stripe 6 is by kni.

Stripes 3 and 4 share a common enhancer. This enhancer appears to work in a two-step mode for stripe formation where a single wide stripe is first set up that is later separated by repression in the forthcoming interstripe region (Hartmann et al., 1994). The stripe 3/4 enhancer contains multiple hb and kni binding sites required for the initial formation of the wide stripe 3/4 while the runt (run) gene product (which is a downstream pair-rule gene that can repress h—see below) is required for correct separation of stripes 3 and 4 (Hartmann et al., 1994). The initial programming of h stripes 3 and 4 was modeled where the activators were hb (in the anterior end) and Kr (in the posterior end), the anterior repressor was higher levels of hb or bcd, and the posterior repressor was kni. I suggest that in vivo this combination along with an additional hb threshold would be enough to program the separation of stripes 3 and 4 by only hb, kni, and Kr (without the bcd input needed in the calculation); the stripe 3 enhancer would use hb as activator, hb as an intermediate repressor, and kni as a strong repressor confining h expression to PS5; stripe 4 would use Kr as a weak activator, hb as a very strong repressor (mediated by the bcd gradient), and kni as an intermediate repressor so that very low levels of hb would repress h in PS6 and confine h expression to PS7. Note that since separation of these stripes would depend on fine gradations in hb levels (which is regulated by both bcd and Kr) that one might, therefore, expect that separation of these stripes would occur late in the programming of the h stripe pattern—when run expression (see below) could insure repression of h in the 3/4 interstripe.

Stripe 5 was driven by its own enhancer where the input is kni, the anterior repressor is Kr, and the posterior repressor is gt.

Stripe 7 was driven by its own enhancer where the input is through tor activation, the anterior repressor is kni, and the posterior repressor is tll. An additional repression (by bcd or hb) was also required to prevent stripe 7 expression in the symmetrical location of the tor gradient at the anterior tip of the embryo.

Expression of the anterior domain (Stripe 0) was not modeled.

Prospects. The formation of the seven h stripes involves the stripe-doubling mechanism, but with each stripe determined by an individual set of gap gene inputs. While this appears to be a very sloppy way to form the h stripes, it is an effective way to program a pattern of stripes for the same gene at multiple junctions of genes that are different. The mechanism suggests a slow evolution of h regulation where individual h function was added slowly throughout evolution by addition of new enhancers for the h gene. The calculation indicates that h (along with eve) is the primary pair-rule gene that reads the gap gene positional code and transforms it into a parasegment code to define the odd parasegments.

Even skipped (eve) The seven stripes of even-skipped (eve) expression appear about the same time as the seven h stripes; the equivalent h and eve stripes overlap but with the each eve stripe lying posterior to the overlapping h stripe (Hooper et al., 1989; Howard & Struhl, 1990). The expression of eve stripe 2 has been extensively studied (Driever & Nüsslein-Volhard, 1988b; Stanojevic et al., 1989; 1991; Small et al., 1991, 1992, 1993), and the data for stripe 2 regulation are consistent with the stripe-doubling mechanism for eve expression. The activator for eve stripe 2 (like h stripe 2) is bcd, while the repressors are gi and Kr. This would place eve stripe 2 and h stripe 2 at a similar location at the anterior edge of the Kr stripe but likely with eve stripe slightly offset posteriorly from h stripe 2. Similarly, the posterior edge of eve stripe 3 and h stripe 3 appear to be determined by Kr repression again leading to overlapping but likely slightly offset eve and h stripes. Unfortunately, there was not enough experimental data available on the remaining eve stripes to include eve in the calculation. However, modeling of genetic networks by Reinitz & Sharp (1995) has indicated
that repression to form the eve stripes follows a very similar pattern as that of hgt–[eve2]–Kr; hb–[eve3]–Kr; Kr–[eve4]–kni; kni–[eve5]–gt.

Calculation. The initial eve stripe pattern was not included in the calculation due to lack of data.

Prospects. It appears from the available data that the h and eve seven stripe patterns are set up in parallel by independent combinations of gap gene inputs. By comparing the inputs for h and eve stripes 2 and 3, the equivalent h and eve stripes are in a similar location but with eve displaced posteriorly by slightly different thresholds of inputs or by differing input genes. Therefore, further study of eve expression should indicate that the eve enhancers are likely driven by combinations of inputs similar to those seen for h expression.

Fushi tarazu (ftz)

The expression of the fushi tarazu (ftz) gene is observed just before formation of the cellular blastoderm. Initially low levels of of ftz expression are detected throughout the blastoderm, but this distribution becomes rapidly modulated to seven ftz stripes in alternating parasegments (De Pomerai, 1986). The pattern of expression for the fushi tarazu (ftz) gene is generated through general activation throughout the germband and localized repression in the ftz interstripe regions (Dearolf et al., 1989; Read & Manley, 1992) where the anterior margin of each ftz stripe appears to be defined by the posterior eve boundary (Ish-Horowicz et al., 1989).

Calculation. The ftz gene was modeled as the gene at the last level of the stripe-doubling hierarchy [gene F in Fig. 4(c)]. As such ftz expression was allowed by degradation of its repressor ttk at cycle 9. Expression of ftz then was seen throughout the embryo, but after activation of h, ftz was repressed in the regions of h expression and found in the even parasegments as in vivo. Note that ftz is repressed by both h and eve so that the ftz pattern in vivo is likely formed by repression in the overlapping h and eve stripes. Repression through the tor gradient appears to be involved in setting the posterior edge of ftz stripe 7, but with the 3-step tor gradient this effect was neglected leaving ftz stripe 7 fused with the telson.

Prospects. The ftz gene appears to be the last gene in the stripe-doubling mechanism for parasegment determination under the control of both h and eve. Since both h and eve mechanisms appear to operate independently, this suggests that originally only eve (or h) worked with ftz in determining parasegment odd/even identity, but gene duplication (or enhancer duplication) allowed the redundant pathway for both h and eve to work along with ftz. This type of mechanism allows the h/eve/ftz initial pattern to develop very rapidly in all regions of the embryo simultaneously and may be yet another case where gene and pathway duplication allows rapid determination of all segments in a long germ band insect.

Homeotic genes

Deformed (Dfd)

The expression of Deformed (Dfd) transcripts is first detectable at cycle 13 in a wide circumferential stripe; as the cellular blastoderm forms, levels of Dfd increase and become limited to a circumferential stripe about 6 cells in width. (McGinnis et al., 1990).

Calculation. As shown in Fig. 10 the Dfd gene appears to be regulated by the terminal system where activation and repression is through terminal activators—here presumed to be tll—but only in conjunction with high levels of bcd. (Note that this is one of the few places in the calculation where the synergy of two separate activators is required for target gene activation). The posterior edge of the Dfd stripe was then defined by the tll activation threshold, and the anterior edge was defined by the higher tll repression threshold.

Prospects. The Dfd gene appears to be the primary selector gene for PS1. The labial, proboscipedia, Dfd genes in the S’ end of the Antp locus appear to be the most ancient homeobox genes. (Akam et al., 1988). Dfd, lab, and pb (data not shown) appear to be all under regulation directly by the terminal system (with A/P input to limit their expression to the anterior end of the embryo). This suggests that the arthropod ancestor, which had only a head plus other identical segments had its head segments (corresponding to the current Acrorn-PS1) defined by the terminal system plus a lab-pb-Dfd homebox locus.

Sex combs reduced (Scr)

The Sex combs reduced (Scr) gene product is first detectable just as the parasegmental boundaries become visible, and stained Scr protein is visible in the nuclei of all ectodermal cells of PS2 with some staining of cells in PS3 (Riley et al., 1987). There appears to be no extensive overlap between cells expressing Dfd and Scr (Mahaffey et al., 1989).

Calculation. Each of the homeobox genes appears to have a major gap gene activator to set its initial expression which is then modified by other homeobox genes. For Scr the major gap gene activator was modeled as hb which sets the Scr stripe in the PS1-PS5 region; the edges of the stripe are simultaneously narrowed by anterior inhibition from tll and posterior inhibition by Kr and Ubx.
Antennapedia (Antp)

The Antennapedia (Antp) locus has two promoters, termed P1 and P2, but the P1 promoter appears to be dispensable for embryonic development (Kaufman et al., 1990). Antp expression is first detected early in cycle 14 in a broad band between 15–65% of egg length but rapidly becomes localized predominantly to PS4 (Harding & Levine, 1988). At the cellular blastoderm stage, transcripts arising from the P1 promoter span a broad region of the embryo, while those derived from P2 are restricted to a narrow stripe of about three to four cells wide approximating PS4 (Irish et al., 1989b). Early transcription from the P1 promoter closely mirrors the domain of Kr expression (Irish et al., 1989b).

Calculation. The calculation accounts for the Antp P1 promoter. Initial activation of the Antp P1 promoter was modeled to be by the gap gene Kr. This resulted in a broad Antp stripe at cycle 12. The anterior edge of this stripe was defined by the anterior edge of Kr expression which remained through to the cellular blastoderm. The posterior edge of the broad Antp stripe was narrowed by feedback inhibition through new expression of the abdominal regulating genes—kni and Ubx.

Prospects. The Antp gene appears to be the primary selector gene for PS4 (with ftz) and PS5. As described above, Antp and Scr appear to have been produced by gene duplications in myriapod-like ancestral arthropods (Akam et al., 1988). Their regulatory pathways are consistent with this notion where: (1) they differ from the previous class, [lab-pd-Dfd], in their primary activation through the A/P system rather than the terminal system; (2) they are expressed in two-segment-wide stripes that are later modulated by ftz expression rather than single-segment-wide stripes of [lab-pd-Dfd]; and (3) they define new segments (posterior head and thorax) that appeared in myriapods and the first insects. I suggest that the P2 promoter may be activated by hb, where [Scr-Antp] expression would follow localization of hb in PS2–PS5—paralleling the regulation of new segments by Scr and Antp (PS2–PS5).

Ultrabithorax (Ubx)

The BX-C locus can be divided into three independent genetic units: Ubx, abd-A and Abd-B (Sanchez-Herrero et al., 1985). The Ubx protein is expressed in PS7–PS13 where the anterior border is defined by repression from the hb gene product (Steward, 1989). While there are only three genes in the Ubx locus, there are iab mutations for each of the abdominal segments that correspond to multiple enhancers for regulation of the three genes (Karch et al., 1985).

Calculation. Activation of the Ubx gene was presumed to be by a gap gene as before—in this case by kni. This led to an initial wide band of expression at cycle 12 followed by slight expansion forward by feedforward activation by Antp and a downregulation from the posterior edge by feedback by abd-A, Abd-B and ttl. The final anterior edge of Ubx expression was defined through repression by hb.

Prospects. The Ubx gene appears to be involved as a selector gene for PS5 through PS12 and to be the primary selector gene (with ftz) for PS6. The Ubx gene appears to another homeobox gene duplication that added abdomen determination to ancestral insects. Its localization on a different chromosome from the Antp locus supports this notion, but also there several regulatory differences between the [Scr-Antp] region and Ubx: (1) Scr and Antp likely are activated by hb while Ubx is repressed by hb; (2) Ubx is likely activated by a gap gene, kni (possibly in conjunction with the related knr genes), that would be unnecessary in myriapods but “new” in insects; and (3) the demarcation line between hb and kni regulation in PS6 also marks the region of germ band extension in short germ band insects (see below).

Abdominal-A (abd-A)

The abd-A locus is involved in specification of A1p to A4 (PS7 to PS9) (Sanchez-Herrero et al., 1985).

Calculation. While little data is available for abd-A regulation, it was presumed to be activated by kni since its expression closely follows the region of kni expression, and its posterior edge was then limited by feedback inhibition through Abd-B.

Prospects. Abd-A likely appeared by another gene duplication of Ubx in the HOM-C locus which would be consistent with an activation by kni (the gap gene that controls abdomen determination).

Abdominal-B (abd-B)

The Abd-B locus is involved in specification of segments A5 to A8 (Sanchez-Herrero et al., 1985). Two gene functions called m and r control
morphogenesis in different posterior regions of the embryo, \( m \) in PS10-13 and \( r \) in PS14 (Busturia & Bienz, 1993). A control region \( IAB5 \) found over 40 kb 3' to the \( Abd-B \) gene regulates \( Abd-B \) expression in PS10-14 apparently through strong repression by \( Kr \) and weak repression by \( kni \) (Busturia & Bienz, 1993).

**Calculation.** The \( Abd-B \) gene has complex regulation from two separate promoters and several enhancers. The calculation, therefore, used the experimental data to indicate that \( Abd-B \) is first activated through the terminal system but inhibited on its anterior edge by the A/P system (\( bcd, hb, Kr, kni \)) directly or indirectly. As the last gene in the homeobox cascade, \( Abd-B \) has controls that were too complex to fit any more closely in the pilot project.

**Prospects.** A final gene duplication in the HOM-C locus would give [\( Ubx-abdA-AbdB \)] to regulate the identities of the abdominal segments. The regulation of this locus differs from the previous [\( Scr-Antp \)] region in that: (1) it is much more complex with multiple enhancers for only three genes (Karch et al., 1985); (2) it is activated by \( kni \) and repressed by \( hb \) rather than activated by \( hb \) and repressed by \( kni \); (3) the homeotic selector code defined by \( Scr \) and \( Antp \) is simple with little overlap of the genes, but the homeotic selector code in the abdomen requires combinations of \( Ubx, abd-A, \) and \( Abd-B \) with considerable overlap of those genes.

**Program 2**

**Refinement of Pair-Rule Stripes**

**Setting Segment Polarity**

**Pair rule genes**

**Hairy (h)**

After formation of the rough \( h \) stripes as described above additional input comes from the runt (run) gene which is a weak repressor of \( h \) (Carroll & Vavra, 1989; Warrior & Levine, 1990; Manoukian & Krause, 1993) and which also is needed to repress \( h \) to separate \( h \) stripes 3 and 4 (Hartmann et al., 1994; see above).

**Calculation.** The output of the first program gave \( h \) stripes that were four nuclei wide at cycle 12 but with square edges (00111100). The input for second program (Fig. 9) looked at two \( h \) stripes that were five cells wide and broad on the edges (01232100). In calculation of the wt embryo, activation of \( h \) before run maintained \( h \) expression and excluded run from the \( h \) stripes.

**Prospects.** In Drosophila it appears that the seven stripe \( h \) pattern sets the register for the parasegment code and \( eve \) expression then shifts the register for the programming of the segments. The baseline \( h \) stripe pattern stays fixed until the \( eve \) and \( fitz \) expression can be modulated into asymmetric stripes to set further patterning.

**Runt (run)**

**Calculation.** The run gene was assumed to be expressed following degradation of its repressor \( ttk \) at cycle 9. It was also assumed that expression of \( run \) required at least one cell cycle longer than activation of the initial \( h \) and \( eve \) patterns. This allowed \( run \) expression in the gaps between the overlapping \( h/eve \) stripes. Note that since \( run \) represses both \( h \) and \( eve \), if \( run \) expression preceded setting up of the \( h \) and \( eve \) stripes, the \( run \) expression would repress both the \( h \) and \( eve \) patterns. In this instance the expression of mutually excluding repressors is very sensitive to which is expressed first (Edgar et al., 1989).

**Prospects.** The expression of \( run \) sets the graded posterior edge of the \( eve \) stripes which are required for programming segment polarity. Additionally, the separation of \( h \) stripes 3 and 4 requires expression of \( run \) (see above); this may require slow enough buildup of \( h \) protein so that high level \( run \) expression in the center of the wide stripe 3/4 precedes high level \( h \) expression allowing \( run \) to predominate and repress \( h \).

**Even skipped (eve)**

After initial programming of the seven stripe even-skipped (eve) pattern, a new \( eve \) enhancer is activated by the expression of \( eve \) itself which mediates high levels of \( eve \) protein (Goto et al., 1989; Jiang et al., 1991). The \( eve \) protein appears to repress transcription by direct DNA binding and represses several different classes of promoters that contain \( eve \) binding sites (Han & Manley, 1993). In wt embryos the anterior margin of the \( eve \) stripe is derived from nuclei that showed maximal levels of both \( h \) and \( eve \) at the cellular blastoderm (Warrior & Levine, 1990).

**Calculation.** The other input for the second program was an \( eve \) pattern of four nuclei-wide stripes (00122100) assumed to be programmed by the gap genes. Once the second program began interactions of the pair-rule genes modulated the \( eve \) pattern. Expression of \( eve \) at the anterior edge of the stripes increased to high levels through autoregulation while expression at the posterior edge was limited through repression by \( run \).

**Prospects.** Setting the \( eve \) stripes slightly posterior to the \( h \) stripes results in a slight shift in register of stripe patterns; the \( h \) pattern appears to set the parasegment code while the \( eve \) pattern appears to line up the segments. This is consistent with the interesting experimental result that the homeotic selector code (programmed by gap genes) and the
segment polarity code (programmed by eve and ftz) end up out of register where several homeobox genes work in the posterior compartment of one segment and the anterior compartment of the next segment. The asymmetry in the refined eve stripes at this point provides the input to program the polarity of the odd numbered segments.

**Fushi tarazu (ftz)**

The pattern of expression for the fushi tarazu (ftz) gene is generated through general activation throughout the germ band and localized repression in the ftz interstripe regions (Dearolf et al., 1989; Read & Manley, 1992) where the anterior margin of each ftz stripe appears to be defined by the posterior eve boundary (Ish-Horowicz et al., 1989). A “zebra” element within 740 bp of the ftz gene cap site interacts with the products of h, run, and the gap genes to form the seven stripe pattern, and an upstream enhancer binds the ftz protein for autoregulation (De Pomerai, 1986). The ftz protein bands soon become asymmetric, their anterior borders remaining stable and sharply demarcated while their posterior borders shift such that each stripe becomes narrower and more intense (De Pomerai, 1986).

**Calculation.** The initial ftz distribution was assumed to be homogeneous following derepression by degradation of ttk. The ftz pattern was then modulated by repression by eve in the odd parasegments, by h to repress the posterior edge, and by autoactivation to cause high level expression of ftz on the anterior edge of the ftz stripes.

**Prospects.** The refinement of the eve and ftz stripes follows a parallel pattern, and the ftz stripes on the anterior side of each h stripe set the segment register for the even segments just as the eve stripes on the posterior side of each h stripe set the segment register for the odd segments. The asymmetry in the refined ftz stripes at this point provides the input to program the polarity of the even numbered segments.

**Paired (prd)**

The expression of paired (prd) begins at nuclear cycle 12 or 13 in a broad anterior band; in the next nuclear cycle the band splits into a seven band pattern (each band about 6 cells wide) and subsequently as cellularization proceeds prd is repressed in the two cells in the middle of each band 2 to 7 followed by splitting of band 1, leading to a final 14 stripe pattern at the cellular blastoderm (Ingham, 1988; Baumgartner & Noll, 1991). None of the pair-rule genes are required for the initial activation of prd (Baumgartner & Noll, 1991). All available data indicate that prd is located at the bottom of the hierarchy of pair-rule genes (Baumgartner & Noll, 1991) and appears to mediate the transition to segment polarity genes directly. Although prd is expressed at every parasegment boundary, it appears to act only in alternate parasegments; therefore, it has been suggested that apa, which has yet to be analyzed at the molecular level, may be expressed in a similar pattern and function in a complementary manner to prd (Ingham & Arias, 1992).

**Calculation.** The prd gene was assumed to be activated by gap genes (or activated by relief of ttk repression) at cycle 12 (Fig. 11). Therefore, initial prd expression is homogeneous throughout the embryo. As the eve and ftz stripes refine and eve or ftz levels increase at the anterior edges of the stripes, it appears that eve or ftz repression “punches holes” sequentially in the prd pattern of two stripes each in sequence (Fig. 10). The prd pattern therefore goes from homogeneous to six cell-wide stripes spaced by two cell gaps, then to prd expression disappearing in the middle two cells of the wide stripes, going finally to 14 two cell stripes. In the calculation the ftz pattern comes up first so the prd six cell stripes are first in register with the eve stripes; however, it appears that in vivo the eve stripes refine first so the prd six cell stripes are first in register with the ftz stripes (Kilcherr et al., 1986).

**Prospects.** The prd gene lies at the end of the pair-rule cascade and appears to be involved in the setup of the wg/en segment polarity gene patterns (see below). In the calculation I considered only the prd gene at this level of the pair-rule cascade, but it appears that this prd function is shared with odd-paired (opa) and odd-skipped (odd) (Ingham & Arias, 1992; Benedyk et al., 1994) which all are regulated by pair-rule or gap genes and provide inputs into the wg/en pattern.

**Segment polarity genes**

**Wingless (wg)**

The wingless (wg) gene encodes a secreted protein and is the homolog of the vertebrate Wnt-1 gene (Peifer & Bejsovic, 1992). It acts as a transcriptional activator but only for cells that are touching (or possibly within a few cell lengths from) the wg-expressing cell (Ingham & Arias, 1992).

**Calculation.** Expression of wg was assumed to be activated by relief of ttk gene repression at which point expression was seen only in a narrow stripe at the posterior edge of each parasegment in the fissures between eve and ftz expression.

**Prospects.** The expression of wg as the last stripe-doubling in programming the blastoderm likely
provides the first cue for setting segment polarity. Refinement of segment polarity then depends on cell-cell interactions.

Engrailed (en)

The engrailed (en) gene encodes a variant homeobox sequence which is split by an intron (De Pomerai, 1986). The expression of en is first seen at the cellular blastoderm state in a 14 stripe pattern where each stripe is but a single cell wide (Ingham, 1988). The en function is required specifically to determine the posterior cells in each segment where each stripe is but a single cell wide (Ingham, 1988). The anterior boundaries of stripes expressing en coincide, cell by cell, with the anterior boundaries of the stripes expressing ftz (for the even-numbered parasegments) and eve (for the odd-numbered parasegments) (Lawrence, 1987). Early in development, wg-expressing cells act to maintain the fate of their posterior neighbors, the en-expressing cells (Peifer & Bejsovic, 1992).

Calculation. As described above, the calculation assumed a line-drawing mechanism for en activation by wg with eve or ftz.

Prospects. The expression of en sets the segment register by determining which rows of cells will contract to become the posterior row of cells in each segment. Setting the en pattern is the last step of the transcriptional cascade in the syncytial blastoderm, and after cellularization of the blastoderm the en stripes set the input register for continued definition of segment polarity but by cell-cell interactions after cellularization.

Program 3

Programming Dorsal/Ventral Patterns

Dorsal (dl)

The dorsal (dl) protein is a sequence-specific DNA-binding protein related to mammalian NF-KB (Thissen et al., 1991). The maternal dorsal protein is found in the cytoplasm of the entire egg complexed with the cactus gene product. Ligand activation of the Toll membrane receptor in a dorsal-ventral gradient then mediates nuclear translocation of the dl protein through release of dll by phosphorylation of cactus (Steward, 1989; Rushlow et al., 1989; Roth et al., 1989; Stein et al., 1991).

Calculation. The calculation assumed that dl was the only D/V morphogen, and its activity was found in a step gradient from the ventral to dorsal side of the embryo (Fig. 13). Therefore, like bcd and tor the dl concentrations were set in the egg and remained unchanged during the calculation. The activation of the D/V genes was assumed to begin after cycle 9 when the nuclei migrate to the cortex. If Toll begins a phosphorylation cascade for dl during cycle 10, then dl will be in the nucleus at G1 of cycle 11 switching the chromatin of the genes it binds during G2 of cycle 11, leading to first expression of the D/V gene during nuclear cycle 12.

The dl gradient required an additional threshold over the Boolean switching rules as in Fig. 1(b) (as a very strong repressor of zerknult) because it appears that the dl gradient can be read at four individual thresholds for programming the D/V pattern. (Note that this is still consistent with the step gradient model—it merely means that dl is a strong repressor for both dpp and zen but that the threshold concentration for each differs slightly).

Prospects. The dl gradient has been well established as the major morphogen for Drosophila D/V pattern formation. This calculation reaffirms that notion and indicates that the dl gradient may be “read” at multiple thresholds by two mechanisms—an activation gradient from the ventral side (programming twist and snail) and a repression gradient toward the dorsal side (programming decapentaplegic and zerknult).

Twist (twi)

Twist (twi) is a helix-loop-helix protein involved in the determination of the Drosophila mesoderm (Leptin, 1991). Expression of twist is graded with peak levels in ventral regions and progressively lower levels in ventrolateral regions (Ip et al., 1992b). There are at least five dl sites in the twi promoter (Ip et al., 1992b). Transcripts of twi and snail are first detected during nuclear cycle 11–12 in a single continuous stripe comprising the ventralmost 20% of the embryo (Ray et al., 1991).

Calculation. The twi gene was modeled to be the first gene activated in the French flag mechanism [Fig. 4(a)] mediated by the dl gradient. It was activated directly by dl forming a gradient of activation from the ventral edge of the embryo.

Prospects. The twi gene appears to be the primary selector gene for the activation of mesoderm genes and genes in the lateral neuroectoderm.

Snail (sna)

Snail (sna) is a zinc-finger protein involved in determination of the Drosophila mesoderm; twi and snail act together so that only the absence of both gene products results in the complete loss of all mesodermal characteristics. Both proteins first appear
during nuclear cycle 12 in a diffuse band along the ventral surface of the embryo (Leptin, 1991). The sna gene product is likely not necessary to activate mesodermal genes, but to repress ventrolateral genes in the mesoderm territory (Leptin, 1991). The expression of sna abruptly ends at the lateral limits of the presumptive mesoderm, and the sharp sna borders coincide with the boundary between the mesoderm and neuroectoderm (Kosman et al., 1991; Ip et al., 1992b). There are 10 dl and two twi binding sites in the sna promoter (Ip et al., 1992b). It is possible that twi responds to dorsal activity in a more or less linear way, whereas sna responds to the combination of dorsal and twi proteins in a cooperative way, thus creating a sharper boundary (Leptin, 1991).

Calculation. The sna gene was modeled to be activated by both dl and twi. As a weak activator dl would begin sna expression on cycle 12, and sna expression would then be further enhanced by twi when its expression was activated by dl. Thus the combination of dl and twi would be able to form a sharp boundary near the threshold for sna activation by dl.

Prospects. It appears that twi and sna act together to determine the mesoderm and ventrolateral neuroectoderm (Fig. 13). The twi transactivator would activate mesoderm genes (directly or through an undetermined selector gene) and ventrolateral genes by activating the selector gene rhomboid. Then sna repression of rhomboid (see below) would limit the expression of ventrolateral genes in the mesoderm.

Decapentaplegic (dpp)

High levels of dpp activity are necessary to specify the amnioserosa, and progressively lower dpp levels are needed to specify dorsal and lateral ectoderm with dpp as the central element of this system (Wharton et al., 1993). The dl protein inhibits dpp transcription through low affinity binding at multiple sites in intron 2 of the dpp gene (Huang et al., 1993). Five transcripts that all encode the same or similar polypeptides are produced from the dpp gene controlled by cis-regulating elements spaced over more than 55 kB of DNA (St. Johnston et al., 1990). Prior to cycle 14 dpp is expressed in the dorsalmost 40% of the blastoderm with transcripts extending around both anterior and posterior poles (Ray et al., 1991). The tolloid (tld) gene product appears to increase dpp activity; since the mammalian analog of tld (BMP-1) complexes with the analog of dpp (TGF-beta), this increase in dpp activity by tld is likely through post-translational interaction (Ferguson & Anderson, 1992; Wharton et al., 1993).

Calculation. The expression of dpp was modeled by be constitutive following migration of the nuclei to the cortex but strongly inhibited by dl. Therefore, dpp formed a stripe in the dorsalmost 40% of the blastoderm.

Prospects. The dpp gene is likely the major positive determinant of cell fates in the dorsal portions of the embryo. The simple calculation indicated that dpp expression can be correlated to the ectoderm and amnioserosa, but additional modulation of the dpp transcripts from multiple enhancers and post-translational modification of dpp by tolloid suggests that the dl gradient can be read in fine detail in that region to allow determination of several different cell types within the ectoderm and amnioserosa by differing concentrations of dpp and differing activities of the expressed dpp.

Zerknult (zen)

Transcripts for the zerknult (zen) gene are among the first zygotic genes to appear, being detected at nuclear cycle 10–11 (Rushlow & Levine, 1990). By the cellular blastoderm zen expression is limited to a middorsal stripe 7 cells wide and 70 cells long (Kaufman et al., 1990).

Calculation. Activation of zen was modeled to be under control of dpp as modified by tld. The dl gene was assumed to be a very strong repressor of zen (although an equivalent possibility is that dpp concentration plus modification by tld—see above—is only high enough to activate zen in the dorsalmost region of the embryo). This combination gave zen expression only in the very top of the blastoderm.

Prospects. The zen gene appears to be the selector gene for the amnioserosa under regulation by dpp and tld.

Rhomboid (rho)

Rhomboid (rho) encodes a putative transmembrane protein that may function as a receptor in a cell-signaling pathway and is required for the differentiation of a subset of the ventral epidermal cells that arise from the neuroectoderm (Ip et al., 1992a). The 5' upstream region of the rho gene contains a total of 10 high affinity binding sites for the dl, twi, and sna transactivators (Ip et al., 1992a). Expression of rho is first seen in cycle 13 and is restricted to the presumptive neuroectoderm by cycle 14 (Ip et al., 1992a).

Calculation. The calculation assumed that rho was activated by dl throughout the ventral region of the blastoderm but increased by strong twi binding in the ventral neuroectoderm region and totally repressed by sna in the mesoderm region. The stable expression
of rho was then in a stripe corresponding to the ventral neuroectoderm.

Prospects. The data suggest that rho is the selector gene for the neuroectoderm. This is consistent with the hypothesis that snail works in defining the mesoderm by suppressing neuroectoderm fate (by repressing rho) in the mesoderm region.

Singleminded (sim)

The singleminded (sim) gene encodes a basic helix-loop-helix transcription factor that is required for proper development of the CNS midline lineage. It is transcribed near the end of cellular blastoderm stage in the two lateral stripes of mesectodermal precursor cells. The snail gene product binds to the sim gene promoter and appears to act as a repressor of sim expression (Kasai et al., 1992). Expression of snail does not overlap the sim stripe, but the sim stripe lies within the region of graded expression of the twi protein (Leptin, 1991). Loss of sim function eliminates both neuronal and non-neuronal derivatives of the ventralmost ectoderm (Ingham, 1988).

Calculation. The expression of sim is found in a single row of cells at the border of the mesoderm and ventral neuroectoderm. This suggested a line-drawing mechanism for sim activation. I assumed that sim expression could be activated by a high vicinity concentration but only in cells presenting the rho membrane receptor for another activation signal secreted by snail-expressing cells (possibly the spitz EGF-like growth factor which has a phenotype similar to rho−) (Rutledge et al., 1992). However, since snail would repress sim within the snail-expressing cells themselves, sim expression would be limited to a single line of cells adjacent to the region of snail expression. This mechanism, therefore, defines a feedback loop for sim expression similar to that defined for wg/en stripes (as above).

Prospects. The single line of sim-expressing cells appears to be fated to become the ventral nerve cord and other mesoectoderm cells. The formation of the single-cell-wide stripe is another example (like wg/en) of cell-cell interactions at the junction of two transcriptional compartments “drawing a line” for an additional cell type.

APPENDIX C

Calculation of Developmental Mutants

The most stringent test of any model for development is its ability to simulate correctly the effects of mutations in regulatory mutants. Therefore, a comprehensive collection from the literature of the effects of regulatory gene mutants was compiled and compared with the calculated pattern for each mutant. In the calculation the maternal positional values calculate directly to a homebox selector code that define the individual parasegments (Fig. 7); the effects of any mutants on parasegment determination could then be assessed by their effect on the pattern of homebox genes. The individual mutants are marked with a plus (+) if the calculated pattern was consistent with the experimental data or a minus (−) if it was not. In all, 84 mutants were found that affected the 28 genes of interest, and 66 of these were modeled adequately by the calculations.

Program 1

Maternal effect genes

bicoid (bcd)

+ In bcd− mutants the acron is transformed into a second telson (De Pomerai 1986). In calculated bcd− mutants the anterior end of the embryo is replaced with inverted versions of PS8 to telson as judged by homeobox selector code [Fig. 6(c)].

torso (tor)

Deletion of torso would result in no expression of tailless while dominant mutants of torso would result in ectopic expression of tailless. Therefore, tor− mutants were modeled as tor−/tll− double mutants, and dominant tor mutants were modeled as tor/tll ectopic mutants.

+ In tor− mutants both the acron and telson are deleted and the segmented portion of the egg takes up all the embryo (De Pomerai, 1986; Gilbert, 1991). In calculated tor−/tll− mutants the embryo is filled with PS6, 7, and 8.

+ In a dominant mutation of the tor gene, the entire anterior half of the embryo is converted into acron and the entire posterior half into telson (De Pomerai, 1986; Gilbert, 1991). In calculated tor/tll ectopic mutants the embryo is filled with acron and PS14/telson.

+ Superabundance of the tor protein throughout the embryo does not alter body patterning (Casanova & Struhl, 1989)—likely because the activated torso protein is still restricted in its spatial pattern.

tailless (tll)

+ Mutants lacking tll delete acron and telson structures which are a subset of those lost in tor mutant embryos (De Pomerai, 1986) and have defects both posterior in the tail region and anterior in the acron (Strecker et al., 1986; Mahoney & Lengyel,
1987; Reinitz & Levine, 1990). In calculated \textit{tll}− mutants PS0 and PS1 are missing at the anterior end of the embryo and PS12 through PS15 are missing at the posterior end. Also, in the calculation high concentrations of \textit{tll} (i.e. \textit{tll} = 3) correspond to the gaps in the segmentation pattern seen in \textit{tll}− mutants (Fig. 8) suggesting that the “gap” function of \textit{tll} requires high concentrations of the \textit{tll} gene product.

+ The expression of \textit{tll} is unaffected in the absence of \textit{hb}, \textit{Kr}, \textit{kni}, or \textit{gt} (Brönner & Jäckle, 1991). The \textit{tll} pattern is unaffected in the four calculated mutants because \textit{tll} is upstream of the gap genes.

\textit{nanos (nos)}

+ Mutations of \textit{nos} prevent abdominal segmentation, but maternal double mutants in \textit{hb} and \textit{nos} form viable offspring (Hülskamp et al., 1989).

\textit{maternal hunchback (hb)}

+ Embryos lacking the maternal \textit{hb} gene product can be rescued if a wild-type copy of the gene is provided paternally, indicating that the zygotic expression alone is sufficient for normal development (Gaul & Jäckle, 1990).

+ Maternal double mutants in \textit{hb} and \textit{nos} form viable offspring (Irish et al., 1989a; Hülskamp et al., 1989). The presence of \textit{hb} as both a maternal effect gene and zygotic gene is likely a remnant of the evolution of the long germ band insects.

\textit{Gap genes and terminal genes}

\textit{Krüppel (Kr)}

+ In \textit{Kr}− mutants PS4 to PS10 are affected and segments T1 to A5 are lost (Ingham et al., 1986; De Pomerai, 1986; Gilbert, 1991). In the calculation the final stable \textit{Kr} stripe correctly encompasses PS4 to PS10, but in calculated \textit{Kr}− mutants only PS4 to PS6 are lost according to the homeobox selector code. This suggests that the \textit{Kr}/\textit{kni} interaction in PS7 to PS10 which was neglected is involved in the final determination of that region.

+ Ectopic \textit{gt} expression from a heat shock promoter causes suppression of the \textit{Kr} stripe (Capovilla et al., 1992). In calculated ectopic \textit{gt} mutants the \textit{Kr} stripe is reduced to a low intensity.

+ In \textit{hb}− mutant the \textit{Kr} domain extends anteriorly and is also somewhat shifted toward the anterior, that is, the posterior border of the \textit{Kr} domain lies at a more anterior position (Gaul & Jäckle, 1990). The simple step gradient calculation does not account well for band broadening, and there is no effect on the \textit{Kr} band in calculated \textit{hb}− mutants. Alternatively, the calculation neglects maternal \textit{hb} and \textit{nos} which may also have effects on \textit{Kr} in the \textit{hb}− mutants.

+ In \textit{kni}− mutants the \textit{Kr} domain extends slightly toward the posterior, and the width of the \textit{Kr} stripe is about twice as wide in \textit{kni}− mutants as wt by the onset of gastrulation (Gaul & Jäckle, 199). These effects were not seen in the calculation because \textit{kni} repression of \textit{Kr} was neglected.

\textit{giant (gt)}

+ Null mutants in \textit{gt} affect two regions—PS2 to PS4 and PS10 to PS12 (Gilbert, 1991)—causing fusion of labial and prothoracic segments and the fusion of abdominal segments A5 to A7 and sometimes to A8 (De Pomerai, 1986; Capovilla et al., 1992). The calculated localization in two stable stripes at PS2 to PS4 and PS10 to PS12 correctly reflect the experiments, but no parasegments are lost according to the homeobox selector code; this is likely due to only weak calculated effects between \textit{gt} and \textit{Antennapedia} and between \textit{gt} and the UBX-C genes (discussed in detail below).

+ In \textit{tll}− mutants the posterior stripe is broader and extends posteriorly as compared with wt while in \textit{tor}− mutants the posterior \textit{gt} stripe extends to the posterior tip (Brönner & Jäckle, 1991). In calculated \textit{tll}− mutants the posterior \textit{gt} stripe extends to the posterior tip.

+ In \textit{Kr}− mutants the anterior edge of the posterior \textit{gt} stripe is expanded (Mohler et al., 1989). In calculated \textit{Kr}− mutants the anterior stripe is correctly unchanged, but the posterior stripe is also unchanged (likely due to neglecting \textit{Kr}/\textit{kni} interactions).

+ In \textit{hb}− and \textit{kni}− mutants the posterior edge of the posterior \textit{gt} stripe is expanded (Mohler et al., 1989). Stripe broadening is not seen in the calculation since the \textit{gt} stripes are totally defined by the terminal system.

\textit{knirps (kni)}

+ The \textit{kni} gene affects PS6 to PS12 (Gilbert, 1991), and in \textit{kni} mutants segments A1–A7 are fused and replaced by a single segment (Nauber et al., 1988). The calculated localization of \textit{kni} was in PS6 to PS12, and in calculated \textit{kni}− mutants PS7 to PS12 (corresponding to segments A2 to A7) have abnormal homebox selector codes while the rest of the embryo is normal.

\textit{zygotic hunchback (zhb)}

+ In zygotic \textit{hb}− mutants segments Labial to T3 and A7/A8 are affected corresponding to PS1 to PS5 and PS13 (Gilbert, 1991). The calculated \textit{hb} pattern was two stripes in PS2 to PS5 and in PS13.
+ Ectopic $hb$ expression causes embryos lacking abdominal segments mimicking the effects of $nos^-$ mutants (Struhl, 1989). In calculated ectopic $hb$ mutants the embryo has normal PS0 to PS4 but the selector codes for all the abdominal segments are severely disrupted.

− In $tor^-$ mutants $hb$ is detected in the posterior half of the cellular blastoderm, but the posterior $hb$ stripe is missing (Tautz, 1988). In the calculated $tor^-/tll^-$ mutant no $hb$ expression was seen. Maternal $hb$, which was neglected in the calculation, may account for this result.

**Pair rule genes**

$hairy$ ($h$)

+ In $Kr^-$ mutants only four broad $h$ stripes are seen roughly in the positions of stripe 1, a fused stripe 2/3/4, a fused stripe 5/6, and stripe 7 (Carroll & Vavra, 1989; Hooper et al., 1989). The calculated $h$ pattern is $[1][2][3][4][5][6][7]$ (i.e. normal stripe 1, fused stripes 2 to 6, and normal stripe 7).

− In $hb^-$ mutants $h$ stripe 3 is missing; $Kr$ expression spreads in an anterior direction compressing $h$ stripes 1 and 2 toward the anterior end of the embryo and spreading out the posterior stripes 4 through 7 (Carroll & Vavra, 1989; Hooper et al., 1989). The calculated $h$ pattern is $[1][2][3][4][5][6][7]$ (i.e. missing stripe 3). The loss of stripe 3 in $hb^-$ mutants indicated that $hb$ is the activator for stripe 3 but not for stripe 4.

+ In $kni^-$ mutants $h$ stripes 1, 2, and 3 appear normal but stripe 4 is missing and stripes 5–7 are fused in a wide band (Carroll & Vavra, 1989; Hooper et al., 1989). The calculated $h$ pattern is $[1][2][3][4][7]$. In $tll^-$ mutants $h$ stripe 7 is missing (Mahoney & Lengyel, 1987; Hooper et al., 1989). The calculated $h$ pattern is $[1][2][3][4][5][6][7]$.

+ In $eve^-$ mutants $h$ expression is virtually normal with $eve$ needed only for the proper maintenance of $h$ stripe 2 which is usually seen weakly in the proper position (Hooper et al., 1989). The calculation assumed $h$ and $eve$ expression are independent.

**even-skipped (eve)**

Studies with a lacZ− stripe 2, 3, 7 construct indicated: in $gt^-$ embryos early stripe 2 is abnormally broad; in $Kr^-$ mutants early stripes 2 and 3 are fused into one strong and broad band; in $hb^-$ mutants stripes 2 and 3 are fused and stripe 7 is reduced; and in $tll^-$ mutants stripe 7 is missing while stripes 2 and 3 are unaffected (Goto et al., 1989). These mutants were not modeled.

$fushi tarazu$ ($ftz$)

+ In $h^-$ deletion mutants $ftz$ protein is found in nearly all of the nuclei where it is normally absent, except that no $ftz$ protein is seen in the anterior 30% or posterior tip of the embryo (Howard & Ingham, 1986; Carroll & Scott, 1986). In calculated $h^-$ mutants $ftz$ is continuous as a single band throughout the embryo.

The $ftz$ gene is downstream of both $h$ and $eve$, and its stripe pattern depends on both $h$ and $eve$ expression. Since the $eve$ gene was not considered in the calculation, effects of other gap gene mutations on the early $ftz$ pattern were not considered although the current data have been catalogued for completeness: $tor^-$ (Casanova & Struhl, 1989), $tll^-$ (Mahoney & Lengyel, 1987), $gt^-$ (Petschek & Mahowald, 1990), ectopic $gt$ (Carroll & Scott, 1986; Capovilla et al., 1992), $kni^-$ (Carroll & Scott, 1986).

**Homeotic genes**

*Deformed* ($Dfd$)

+ In $Dfd^-$ mutants the mouth hooks and cirri are missing, both of which are derived from the ventral and ventral posterior regions of the maxillary segment (McGinnis et al., 1990), and the $Dfd$ stripe is localized in the maxillary segment. The calculated $Dfd$ stripe is localized to PS1 (Mx segment).

− In $bcd^-$ mutants no $Dfd$ can be detected at any stage in development (McGinnis et al., 1990; Jack & McGinnis, 1990). In calculated $bcd^-$ mutants no $Dfd$ is expressed.

+ Ectopic $bcd$ expression results in a second $Dfd$ stripe symmetrically placed at the posterior end of the embryo (Jack & McGinnis, 1990). This second $Dfd$ stripe is also apparent with calculated ectopic $bcd$ expression.

− In $hb^+$ mutants no blastoderm expression of $Dfd$ is seen (Jack & McGinnis, 1990). The $Dfd$ pattern is unchanged in $hb^-$ calculations suggesting that $Dfd$ activation may require A/P inputs from $bcd$ indirectly through activation of zygotic $hb$ (Jack & McGinnis, 1990).

+ Ectopic $hb$ expression leaves the $Dfd$ pattern unchanged (Jack & McGinnis, 1990). The $Dfd$ pattern is unchanged in calculations of ectopic $hb$.

+ In $Antp^-$, $Scr^-$, $Ubx^-$, or double and triple mutants in those loci $Dfd$ expression appears normal (McGinnis et al., 1990). This is true in all three calculations because $Dfd$ is under control of the terminal system.

**Sex combs reduced ($Scr$)**

− In $gt^-$ mutants $Scr$ is expressed in PS3 but not PS2 (Riley et al., 1987). No change in $Scr$ expression
is seen in calculated \(gt^-\) mutants suggesting that there may also be refinement of the Scr stripe directly by \(gt\) or indirectly by changes in Antp expression through loss of \(gt\).

+ In \(hb^-\) mutants the Scr stripe narrows and no expression is seen in PS3 (Riley et al., 1987). In calculated \(hb^-\) mutants no Scr expression is seen suggesting that there is a second activator for Scr expression.

− The Scr stripe expands posteriorly in Antp− mutants to the anterior compartment of T2 (Riley et al., 1987). No change in Scr expression is calculated for Antp− mutants, but in Kr− mutants Scr expression spreads posteriorly into PS5. This suggests that the documented Kr inhibition of Scr (Riley et al., 1987) may be indirect through the pathway Kr− \(\rightarrow\) Antp−\(\rightarrow\)Scr.

+ No change in Scr distribution is seen in Ubx− or Dfd− mutants (Riley et al., 1987). Normal Scr expression is calculated in both Ubx− and Dfd− mutants.

+ In the absence of the Ubx locus, Scr is expressed in an ectopic domain spanning the posterior compartments of the thorax and abdominal segments (Pelaz et al., 1993). In the absence of a Ubx locus inhibitor there is a second Scr stripe calculated in the abdominal region.

Antennapedia (Antp)

− In \(gt^-\) embryos the Antp stripe extends more anteriorly than in the wt (Reinitz & Levine, 1990). No change in Antp expression is calculated in \(gt^-\) mutants; this likely reflects a low \(gt\) concentration in the calculation of the gap genes that has no inhibitory effect on Antp.

+ Kr− mutants P1 is not active (Irish et al., 1989b).

− The Antp stripe is slightly broader than in wt and is shifted to a slightly more posterior position (Harding & Levine, 1988). No Antp is expressed in calculations of Kr− mutants since the program only accounts for the P1 promoter.

+ In kni− mutants the Antp transcripts are detected in a broad band with the normal anterior edge but with a posterior edge corresponding to the more posterior edge of Kr expression seen in these mutants (Harding & Levine, 1988). The Antp band broadens in calculated kni− mutants but later splits where Ubx expression is highest.

Ultrabithorax (Ubx)

+ Deletion of Ubx results in the region from the posterior compartment of T1 through the anterior compartment of A1 being transformed into repetitions of T1a/T2p or PS4 (Sanchez-Herrero et al., 1985). In calculated Ubx− mutants Antp expression spreads into PS6 making PS4 through PS6 totally defined by Antp.

+ In \(hb^-\) mutants Ubx expression shifts anteriorly (Steward, 1989). In calculated \(hb^-\) mutants Ubx expression spreads anteriorly into PS4.

+− There is no effect on Ubx patterns in Ser−, Antp− or Ser−/Antp− mutants (Struhl & White, 1985). There is no effect on Ubx in calculated Scr− mutants but Ubx is lost in PS5 in Antp− mutants.

+ In \(abd-A^-\) mutants Ubx expression extends from PS5 to PS12 while in \(abd-A^-/abd-B^-\) double mutants Ubx expression is detected all the way to PS13 (Struhl & White, 1985). Low levels of Ubx are calculated in PS7 to PS12 where the Ubx concentration is inhibited by \(abd-A\) and \(abd-B\) repression.

Abdominal-A (\(abd-A\))

+ Deletion of \(abd-A\) results in the region from A1p to A4 (PS7–PS9) being transformed into repetitions of T3p/A1a (PS6) (Sanchez-Herrero et al., 1985). In the calculation \(abd-A\) appears in combination with Ubx in PS7 to PS9, suggesting that the Ubx/abd-A combination is the selector code for those parasegments. Loss of \(abd-A\) would then give only Ubx in those parasegments reverting them to PS6.

Abdominal-B (\(abd-B\))

+ Deletion of \(abd-B\) results in the segments A5 through A8 (PS10–PS13) developing as repetitions of A4 (PS9) (Sanchez-Herrero et al., 1985). In the calculation the combination of Ubx/abd-A/abd-B appears to be the selector code for PS10 to PS12; therefore loss of \(abd-B\) would revert those parasegments to the Ubx/abd-A code which is the selector for PS7–PS9.

− In Kr− mutants \(abd-B\) mRNA expression is found in a second stripe in the region of PS4 in addition to its posterior expression domain (Harding & Levine, 1988). \(abd-B\) expression is unchanged in calculated Kr− mutants.

− In kni− mutants the \(abd-B\) stripe is expanded in the anterior direction with its normal posterior boundary (Harding & Levine, 1988). \(abd-B\) expression is unchanged in calculated kni− mutants.

Program 2

**Pair rule genes**

**hairy (h)**

− Although both \(h\) and eve patterns are affected in run mutants, the \(h\) pattern is more severely disrupted (Warrior & Levine, 1990).
Ectopic expression of run does not affect h patterns unless run genes under heat shock promoters were heat shocked for 30–45 min at which point run can repress all h expression (Manoukian & Krause, 1993). Ectopic run expression represses all h expression in the calculations.

**run (run)**

+ In run− mutants stripes 2, 3, 4, 6 and 7 bifurcate into sets of two sharply defined stripes while stripes 1 and 5 are incompletely split (Goto et al., 1989). The eve stripes are sharp on both sides in calculated run− mutants.
+ Ectopic expression of run causes loss of eve expression (Manoukian & Krause, 1993). Calculated ectopic run expression represses eve to low levels.
+ Ectopic expression of prd has no effect on the distribution of eve protein (Ingham & Arias, 1992). Since prd is downstream of eve in the calculation, no effect of prd on eve is seen in the calculation.

**fushi tarazu (ftz)**

+ Ectopic expression of run does not affect the ftz pattern unless run is activated prior to ftz stripe resolution in which case ftz is seen as a single wide stripe filling the entire trunk of the embryo (Manoukian & Krause, 1993). Calculation of ectopic run expression gives ftz stripes in their proper locations (but with sharp edges on both sides).
+ In h− mutants prd is ectopically expressed between bands 2 and 7 but bands 1 and 2 appear normal; at the cellular blastoderm this splits into a normal band 1 plus seven bands with double-segment repeat (Baumgartner & Noll, 1991). In calculated h− mutants prd is seen in a seven stripe pattern.
+ In eve− mutants the gaps in the early 7 stripe pattern are not repressed and prd appears ectopically as a single long band; during cellularization the band does split into a 7 stripe pattern (Baumgartner & Noll, 1991). In calculated eve− mutants prd is seen in a seven stripe pattern.
+ In ftz− mutants the 7 stripe pattern appears normally but splits into alternate 1 cell odd stripes and 3 cell even stripes (Baumgartner & Noll, 1991). Calculations of ftz− mutants show a seven stripe pattern that does not split further.

**Segment polarity genes**

**wingless (wg)**

+ In cells lacking eve or ftz, wg becomes transcribed throughout the normal regions of expression of either pair-rule gene (Ingham and Arias, 1992). Normal wg is calculated for both eve− and ftz− mutants.
+ Ectopic expression of run prior to activation of wg did not affect the wg pattern unless continued for greater than 30 min in which case the 14 narrow stripes changed to 7 wide stripes positioned between the broadened ftz stripes (Manoukian & Krause, 1993). In the calculation 7 wide wg stripes are expressed between the ftz stripes.

**engrailed (en)**

+ Expression of the even-numbered en stripes is lost in ftz mutants (De Pomerai, 1986; Ingham & Arias, 1992). Only the odd stripes are expressed in calculated ftz− mutants.
+ Expression of the odd-numbered en stripes is lost in eve mutants (De Pomerai, 1986; Ingham & Arias, 1992). Only the even stripes are expressed in calculated eve− mutants.
+ In h− mutants only seven wide en stripes can be detected (Howard & Ingham, 1986). In calculated h− mutants 14 en stripes are expressed in mirror image polarity (wg/en−/en/wg/en−/en/wg−/etc.).
+ Null prd mutants delete alternate en stripes, specifically those belonging to the odd-numbered segments (De Pomerai, 1986). No en expression is calculated in prd− mutants.
+ With heat shock ectopic ftz expression all 14 en stripes are still expressed initially with the even stripes wider than normal, but expression soon decays in the even stripes (Ish-Horowicz et al., 1989). No en is expressed in calculations of ectopic ftz.
+ Ectopic expression of the prd gene causes the odd-numbered en stripes to be expanded posteriorly to span half the parasegment (Morrissey et al., 1991; Ingham & Arias, 1992). There is no effect on en in calculations of ectopic prd.
+ Prolonged ectopic expression of run prior to activation of en repressed the odd-numbered en stripes. However, short pulses of ectopic run expression result in a wg/en pattern where the odd-numbered en stripes are anterior to the adjacent wg stripes; these patterns correspond with ftz expression in a single wide band (Manoukian & Krause, 1993). Calculations of ectopic run shows the pattern seen in vivo for short pulses—alternate reversals of wg/en polarity (wg/en−/en/wg/en−/en/wg−/etc.).

**Program 3**

**twist (twi)**

+ In twi− mutants the ventral furrow and the mesoderm anlage are narrower than in the wild type
(Leptin, 1991), and cells in the position of the mesoderm do not divide following cellularization (Arora & Nüsslein-Volhard, 1992). Calculation of a
\( \text{twi}^- \) mutant results in loss of high level \( \text{sna} \) expression and all \( \text{rho} \) expression indicating that mesoderm proliferation requires both \( \text{twi} \) and \( \text{sna} \).

\( \text{snail} \) (\( \text{sna} \))

+ In \( \text{sna}^- \) mutants the mesoderm is lost and cells in the region of the normal mesoderm divide but do not invaginate. In calculated \( \text{sna}^- \) mutants the positional code for the mesoderm (\( \text{twi} + \text{sna} \)) is lost.

\( \text{decapentaplegic} \) (\( \text{dpp} \))

+ In \( \text{dpp}^- \) mutants all cells dorsal to the dorsal neuroectoderm disappear; double mutants of \( \text{dpp} \) with other D/V genes do not change this pattern indicating that \( \text{dpp} \) plays a key role in forming the D/V pattern (Arora & Nüsslein-Volhard, 1992). In the calculation \( \text{dpp} \) is expressed in all cells dorsal to the dorsal neuroectoderm.

+ In mutants lacking maternal \( \text{dl} \) protein \( \text{dpp} \) is expressed uniformly throughout the blastoderm (Huang et al., 1993). Calculation of a \( \text{dl}^- \) embryo gave no repression of \( \text{dpp} \) anywhere by \( \text{dl} \) and expression of \( \text{dpp} \) throughout the entire embryo.

+ In \( \text{tolloid} \) mutants the width of the dorsal epidermis from the dorsal edge of the embryo is decreased and the amnioserosa is missing (Ferguson & Anderson, 1992). While \( \text{tld} \) was not included in the calculation, these data indicate that \( \text{dpp} \) with \( \text{tld} \) modification is required to program the amnioserosa (likely through activation of \( \text{zerknult} \)).

\( \text{zerknult} \) (\( \text{zen} \))

+ In \( \text{zen}^- \) mutants the amnioserosa does not form; in double mutants of \( \text{zen} \) and other D/V genes loss of \( \text{zen} \) expression does not change phenotypes of the other genes suggesting that \( \text{zen} \) is downstream of other D/V genes (Arora & Nüsslein-Volhard, 1992). In the calculation \( \text{zen} \) is the last downstream gene in the D/V pathway and is expressed in the region that becomes the amnioserosa.

+ Ectopic expression of \( \text{dl} \) causes a complete repression of \( \text{zen} \) (Rushlow & Levine, 1990). As a very strong repressor of \( \text{zen} \), \( \text{dl} \) would completely repress \( \text{zen} \) when expressed ectopically.

\( \text{rhomboid} \) (\( \text{rho} \))

+ Deletion of the \( \text{dl} \) binding sites virtually abolishes expression of \( \text{rho} \) while deletion of the two \( \text{twi} \) binding sites shows marked reduction in \( \text{rho} \) expression levels and limits in the promoter/lacZ constructs (Ip et al., 1992a). In the calculation loss of \( \text{dl} \) or \( \text{twi} \) activity reduces or abolishes \( \text{rho} \) expression.

+ Deletion of the \( \text{sna} \) binding sites in the promoter/lacZ constructs caused a dramatic increase in \( \text{rho} \) levels in both the lateral and ventral regions (Ip et al., 1992a). In calculated \( \text{sna}^- \) mutants \( \text{rho} \) was expressed throughout both the mesoderm and neuroectoderm.

\( \text{single-minded} \) (\( \text{sim} \))

+ In \( \text{twi}^- \) mutants the two \( \text{sim} \) stripes lie closer together than in the wild type but are irregular and usually two cells wide (Leptin, 1991; Arora & Nüsslein-Volhard, 1992). In \( \text{twi}^- \) mutants the border of \( \text{sna} \) expression which would abut or overlap \( \text{rho} \) expression would be less clearly defined because the \( \text{dl}/\text{twi} \) combination forms a sharp edge to the \( \text{sna} \) stripe. Therefore, \( \text{sim} \) expression at the \( \text{sna}/\text{rho} \) border would also be less clearly defined and closer to the ventral edge of the embryo.

− In the absence of both \( \text{twi} \) and \( \text{sna} \), \( \text{sim} \) is not expressed, but the presence of either \( \text{twi} \) or \( \text{sna} \) alone allows the activation of \( \text{sim} \). In the calculation \( \text{twi} \) was the only activator of \( \text{sim} \) so that in \( \text{twi}^- \) mutants no \( \text{sim} \) was seen; this suggests that there is additional activation pathway for \( \text{sim} \) which is not repressed by \( \text{sna} \) in the same way.

**Beetle Embryogenesis Program**

Again, to validate to the calculational method I compared the calculated patterns caused by “de-evolved” mutations from *Drosophila* to a short germ band insect with the developmental program of short germ band insects using data from the literature.

+ In the red flour beetle *Tribolium*, \( \text{Kr} \) expression is first seen at the posterior end of the blastoderm and persists in the same cells as the abdomen is added resulting in a single broad central stripe (Sommer & Tautz, 1993). In the calculation the \( \text{Kr} \) stripe begins at the posterior edge of the blastoderm and expands into a single central broad stripe.

+ In *Tribolium* \( \text{hb} \) expression is first seen at the anterior end of the embryo, but as the germ band extends, a second \( \text{hb} \) stripe is seen at the posterior end of the embryo (Wolff et al., 1995). A similar \( \text{hb} \) pattern is seen in the calculated beetle embryo.

+ *Tribolium* has a single homeotic gene locus representing the homologs of the *Antennapedia* and bithorax loci in juxtaposition but with homologs of all the homeotic genes modeled above present in the combined locus in the same order—
Even with little change in the computer program for the beetle homeotic genes (only the addition of \textit{hb} to activate \textit{Abd-B} at posterior end of the embryo was varied from the \textit{Drosophila} rules in Fig. 10) the calculated homeotic gene patterns were very similar between \textit{Drosophila} (Fig. 10) and beetle (Fig. 14) and the homeotic selector codes they would program were virtually the same from PS0 through PS12.

Additionally, there is a band of \textit{Scr} expression at the posterior end of the calculated beetle embryo, and it is interesting to note that \textit{Scr} is expressed after \textit{Drosophila} gastrulation in the malphigian tubes and other posterior structures (Riley \textit{et al.}, 1987)—leading one to speculate that a mechanism of \textit{Scr} activation in temporal order in beetles remained in spatial order in \textit{Drosophila} after change of gap gene regulation.

Expression of the \textit{h} homolog in \textit{Tribolium} begins after \textit{Kr} expression in the region of \textit{Kr} expression similar to the location of \textit{Drosophila} \textit{h} stripes 3 and 4, and an additional \textit{h} band is seen later at the posterior end; as growth continues \textit{h} stripes are seen in the developing germ band ectoderm (Sommer & Tault, 1993). In the beetle calculation only the \textit{h} enhancers under gap gene control (stripes 3, 4, and 5) were used and gave a pattern for \textit{h} expression similar to the bettle pattern (data not shown).

In the \textit{Drosophila} calculation it appeared that the form of pair-rule regulation involving stripe-doubling may have evolved to allow all segments to be defined at the same time in the blastoderm (see above). Therefore, one would expect differing regulation for the pair-rule and downstream genes in short germ band insects, i.e. the only activation of the pair-rule and segment polarity genes would likely be in the growth zone. This is seen experimentally; the grasshopper homolog of \textit{eve} does not serve a pair-rule function in early development but rather forms a band at the posterior end of the growing germ band that does not overlap any \textit{en} expression (Patel \textit{et al.}, 1992); there is no evidence that beetles have a homolog of the \textit{ftz} gene with a similar location and function (Patel \textit{et al.}, 1989); and in the development of the grasshopper the \textit{engrailed} stripes appear one at a time at the posterior edge of each abdominal segments as the individual segments are added (Patel \textit{et al.}, 1989, 1992).