



Mechanism of *eve* stripe formation

John Reinitz^a, David H. Sharp^b

^aBrookdale Center for Molecular Biology, Box 1126, Mt. Sinai School of Medicine, One Gustave L. Levy Place, New York, NY 10029, USA

^bTheoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Received 16 May 1994; revision received 19 September 1994; accepted 20 September 1994

Abstract

In this paper we analyze the formation of stripes of expression of the pair-rule gene *eve*. We identify detailed mechanisms which control the formation of stripes 2–5. Each stripe is formed as a result of generalized activation by *bcd* and ubiquitous transcription factors combined with localized repression by gap genes. Each of the eight stripe borders of these four stripes is shown to be under the control of a particular gap gene expression domain. Protein synthesis from *eve* and its controlling gap genes begins at the same time, but localized *eve* expression is substantially delayed relative to localized expression of gap domains. We show that this delay results from a change in the spatial balance between activation and repression due to the intensification and refinement of gap domains during cleavage cycle 14. *eve* stripe formation is ordered in time; stripe 2 appears earlier than stripes 3–5. We show that this happens because the formation of stripe 2 is less dependent on gap domain refinement than is the case for stripes 3–5: Each of stripes 3–5 is controlled by a pair of overlapping gap domains, whereas stripe 2 is controlled by a disjoint pair of gap domains. Finally, we observe that *eve* stripes do not form unless Eve protein has an extremely small diffusivity, and argue that this low diffusivity is a result of the apical localization of pair-rule message. This implies that localization of pair-rule message is required for stripe formation. The essential tool used to obtain these results is the method of gene circuits, which is a new approach to the analysis of gene expression data. Its purpose is to provide a way to use this data to infer how concentrations of products of a given gene change with time and how these changes are influenced by the activating or repressing effects of the products of other genes. The gene circuit method is based on three main ideas, explained in the paper. First is the choice of protein concentrations as state variables for the description of gene regulation. Second is the summary of chemical reaction kinetics by coarse-grained rate equations for protein concentrations. Third is the use of least squares fits to gene expression data to measure phenomenological parameters occurring in the gene circuit.

Keywords: Even-skipped; Gene circuit; Simulated annealing; Computational biology; *Drosophila*; RNA localization; Gene regulation

1. Introduction

The segmented pattern of the insect *Drosophila melanogaster* is established during the blastoderm stage of embryogenesis. At the time of blastoderm formation, the somatic tissues of the embryo form a syncytial ellipsoid of totipotent nuclei. By the time these nuclei become fully enclosed in cellular membranes at the end of the blastoderm stage, the segmental structure of the animal has been stably determined up to a resolution of a half segment. Cell fates have become determined as

finely as possible at this time, since this level of determination corresponds to a spatial resolution in the blastoderm of about one cell. This process is remarkable for its precision and speed: at 25°C, the entire blastoderm stage lasts less than two hours.

Work over the last decade has shown that segmental determination is a result of the action of a coordinated cascade of zygotic and maternal gene products (Akam, 1987; Ingham, 1988). These gene products fall into four classes based on mutant phenotype, regulatory action and the degree of spatial resolution of expression domains. Maternal products present in broad gradients activate gap genes in overlapping patterns which are

* Corresponding author, Tel.: 212 241 1952; Fax.: 212 860 9279;
email: reinitz@kruppel.molbio.mssm.edu.

expressed in domains with spatial resolution on the order of 10 nuclei. Combinations of gap gene products, in concert with other pair-rule gene products and maternal factors, direct the expression of pair rule genes in striped patterns typically containing seven stripes, each of which is about four nuclei in width. The pair rule expression domains are themselves displaced from one another by as little as one nucleus. These overlapping stripes of pair rule expression in turn control the expression of segment polarity genes, such as *engrailed*, in stripes one nucleus wide which directly specify the segments.

The striped pair-rule expression patterns are the first direct manifestation of the periodic pattern of the segments. For this reason, understanding the mechanism of formation of pair-rule stripes is an essential step in understanding the segmentation process itself. In this paper, we analyze the mechanism of stripe formation by the pair-rule gene *even-skipped* (*eve*). *eve* is unique among pair-rule genes in that it shows no cross-regulatory effects from other pair-rule genes until after gastrulation (Frasch and Levine, 1987, Carroll and Vavra, 1989). Since *eve* stripes form during the blastoderm stage of development, this leads to the simplification that for *eve* the mechanism of stripe formation should be understandable on the basis of regulation by gap and maternal gene products alone.

The phenomenon of *eve* stripe formation presents two central questions: How do regulatory interactions control the size and location of each stripe? How do these interactions set the timing of stripe formation? In Section 3 of this paper we report substantial progress towards answering both of these questions. The essential tool used to obtain the results is the method of gene circuits. As explained here and elsewhere (Reinitz et al., 1994), the gene circuit method is a new way to analyze gene expression data which allows one to use this data to infer how concentrations of products of a given gene change with time, and how these changes are influenced by the activating or repressing effects of other gene products.

Previous studies of *eve* expression in gap mutants showed that specific groups of stripes are under the control of specific gap genes. However, the mechanism by which a particular stripe was formed by regulatory input from specific gap gene products remained obscure (Frasch and Levine, 1987). More recent work using P-element mediated transformation of embryos and transfection assays in tissue culture cells has indicated that the anterior and posterior borders of *eve* stripe 2 are under the control of the gap genes *giant* (*gt*) and *Kruppel* (*Kr*), respectively, while the anterior border of stripe 3 is under the control of *hunchback* (*hb*) (Goto et al., 1989; Stanojevic et al., 1991; Small et al., 1991; Small et al., 1992; Small et al., 1993). These results are consistent with observations of *eve* expression in gap mutants. A

further result which implicated *knirps* (*kni*) as the posterior delimiter of *eve* stripe 3 is not consistent with these observations, as *eve* stripe 3 remains intact in *kni*-embryos (Frasch and Levine, 1987, S. Small, personal communication).

In this paper, we analyze the eight borders that delimit *eve* stripes 2–5. Using the gene circuit method, we are able to show that control of each border results from repression of *eve* by a specific gap gene expression domain. In the case of stripe 2 and the anterior border of stripe 3, our results agree with those obtained by other methods. We find that the posterior border of stripe 3 is delimited by *Kr*, a result that is consistent with the observed disruption of stripe 3 in *Kr*–embryos.

Mechanisms unrelated to the regulation of gene expression have been implicated in the formation of pair-rule stripes. *eve*, like other pair-rule genes, produces mRNA that is localized to the apical side of each blastoderm nucleus. It has been demonstrated that this localization results in markedly reduced diffusion of protein synthesized from apical message, but it has not yet been experimentally demonstrated that apical localization is required for stripe formation (Davis and Ish-Horowicz, 1991). An important result of this paper is that correct formation of *eve* stripes, in conjunction with correctly formed gap domains, requires that *eve* have extremely small diffusivity while the gap gene proteins, notably *Kruppel*, have comparatively large diffusivity. On this basis, we conclude that apical localization of pair-rule message is required for stripe formation because it ensures that only a small amount of Eve protein can diffuse to neighboring cells.

The passage from a state of uniform expression of Eve product to fully formed *eve* stripes is characterized by a complex series of intermediate patterns. Understanding the detailed gene circuitry responsible for these time dependent patterns raises a new class of questions about stripe formation. What is the mechanism by which *eve* stripe formation is coordinated in time with the formation of gap domains? Why do some *eve* stripes form before others?

The question of the temporal coordination of *eve* stripe formation with gap domain formation is prompted by the following observations. Transcripts or protein products of *eve* are first detectable at about cleavage cycle 11, which is the time when zygotically produced gap gene products are first found. Although spatially localized gap gene domains are distinguishable by cleavage cycles 12–13, and essentially established save for minor refinement by the beginning of cleavage cycle 14, the distribution of Eve protein remains spatially uniform until the middle of cleavage cycle 14. At this time, the expression of *eve* undergoes a rapid transition, culminating in the formation of stripes. This is somewhat of a paradox, in that the regulatory actions of the gap genes on *eve* appear suspended until the middle of

cycle 14. If gap genes control *eve* stripes, why don't *eve* stripes form slowly in step with the establishment and refinement of gap domains? The delay in the formation of *eve* stripes is not a consequence of the time taken to process gap gene signals through the transcription and translation apparatus, because the same delay is seen at the RNA level even though expression features at the level of transcripts precede those at the level of protein by 10–15 min (Macdonald et al., 1986; Harding et al., 1986).

We show in Section 3 that *eve* expression stays uniform while gap domains are established because of a precise balance of repression among gap domains. As gap gene expression domains refine and intensify during cleavage cycle 14, this balance is broken. The gain in intensity of gap domains provides the extra repression needed to create interstripes. As gap domains refine, nuclei on the margins of the domain lose gap gene product, which decreases the amount of repression in those nuclei, allowing a stripe to form. Stripes cannot form until the refinement process has proceeded to a critical level, which does not happen until the middle of cleavage cycle 14.

Although *eve* stripes form very rapidly, they do not appear simultaneously. *eve*, like the other pair-rule genes, exhibits a distinctive transient expression pattern. Stripe formation is ordered in time: Stripe 1 forms first, then stripe 2, next stripes 3 and 7, then 4 and, finally, stripes 5 and 6 (Frasch et al., 1987). In this paper, we show that the early appearance of stripes 1 and 2 relative to the other stripes can be predicted from a knowledge of the relative locations of stripes and gap domains, together with the fact that stripes do not form until after the beginning of cleavage cycle 14. We show that the early appearance of stripe 2 relative to stripes 3–5 is a consequence of the placement of each stripe relative to the gap gene expression domains that control it. Although gap gene domains form a succession of overlaps along the A-P axis of the embryo, the pair of controlling gap domains for a given stripe may overlap or be disjoint. Thus, *eve* stripe 2 is controlled by a *gt* domain anteriorly and a disjoint *Kr* domain posteriorly, while *eve* stripe 5 is controlled by the posterior *gt* domain posteriorly and an overlapping *kni* domain anteriorly. Stripes that appear early are those controlled by disjoint domains, while late appearing stripes are controlled by overlapping domains.

The basic phenomena connected with the formation of *eve* stripes have been known for several years. Nevertheless, a detailed account of the regulatory actions controlling most of these phenomena has not previously appeared in the literature. It is worth thinking about why this is so. The conventional tools of modern molecular genetics were developed to answer questions about chemical structure (What is the sequence of this gene?) or analytical chemistry (During what developmental

stages can the transcript of this gene be detected?). As is common in science, the extremely successful application of these tools has raised questions whose answers require new tools. The formation of *eve* stripes is the result of a process in which the concentrations of proteins in different blastoderm nuclei change with time. Thus, to ask how *eve* stripes form is really to ask about the rate at which gene products are synthesized in the presence of other gene products, and this question leads beyond structural and analytic chemistry towards chemical kinetics.

An analysis of the control and timing of stripe formation based on fully explicit chemical kinetics would be out of the question today, both because the input data necessary for such a study are lacking and because of the complexity of the system that is to be understood. A practical method for the analysis of gene regulation in stripe formation and similar developmental problems must therefore be based on an approximation to the exact biochemistry. The approximation underlying the gene circuit method used in our work is based on four main ideas. First is the choice of protein concentrations as state variables for the description of gene regulation. Second is the summary of chemical reaction kinetics by coarse-grained rate equations for protein concentrations. Third is the use of a *gene circuit*, which is a simple method to keep explicit track of the activation or repression of one gene by another. Fourth is the use of least squares fits to gene expression data to measure phenomenological parameters occurring in the gene circuit (as resistances occur in electrical circuits), so that the gene circuit is fully determined and usable as an analytic tool. These ideas, and their range of applicability, require very careful explanation. This explanation is given in Section 2, where we also show how the ideas are implemented to produce a concrete method for the study of gene regulation.

The detailed results are given in Section 3. The reader who is interested only in the biological conclusions can go directly to Section 4. The Appendix provides a technical description of the simulated annealing method, which is used to obtain circuit parameters from gene expression data.

2. The gene circuit method

We begin this section by explaining how the interpretation of gene expression data leads to the need for gene circuits (2.1). We next outline the specific reasoning on which the gene circuit method is based (2.2), describe how this method makes direct contact with experiment (2.3) and clarify some issues concerning the prior application of the gene circuit method to gap genes (2.4).

2.1. The need for gene circuits

The development and application over the past

decade of experimental techniques such as single and double labeling using fluorescence tagged antibodies and *in situ* hybridization has produced abundant data on gene expression in model systems such as *Drosophila*. This data has been interpreted to provide convincing evidence for the existence of a network of regulatory interactions among genes which control important developmental processes in the *Drosophila* embryo. For example, the study of expression patterns of segmentation genes in mutant embryos has provided much important information about regulatory circuits. Here it has been found that the expression of a given pair-rule gene is always altered in gap mutants (Frasch and Levine 1987; Carroll and Scott 1986; Carroll et al., 1988) and is often altered in a mutant for another pair-rule gene (Carroll and Scott, 1986; Frasch and Levine, 1987; Carroll et al., 1988), while gap gene expression is altered only in embryos mutant for gap genes but not pair-rule genes (Tautz et al., 1987). Thus gap genes regulate pair-rule genes, while pair-rule genes regulate one another but do not regulate gap genes. Such work represents important progress in understanding gene circuitry.

Nevertheless, many important questions about regulatory interactions remain unanswered. The questions about *eve* stripe formation with which this paper is concerned are examples. To answer these questions, one must fully exploit the information present in gene expression patterns. We believe this requires new methods of analysis based on gene circuit ideas.

The need for gene circuits can be illustrated by several arguments. We first note that the regulatory actions of genes are generally not observed directly. They must be inferred from *in vivo* experimental data and this often cannot be accomplished solely by visual inspection of expression patterns. This can be quite confusing in situations where gene products coarsely distributed in space act in concert to create more spatially refined expression of other genes. For example, it is easy to see that *eve* stripes 4–6 are under the control of *kni*: *kni* expression approximately overlaps with these stripes, and they are disrupted in *kni* mutants. But that does not explain how stripe 5, in particular, is generated. Nor does it explain why stripe 5 is also disrupted in mutants for *Kr* — *Kr* is not expressed between stripe 4 and a region posterior to stripe 7!

More generally, assertions of the form ‘gene *a* activates gene *b*’ are usually based on the observation that expression for gene *b* is reduced in mutants for gene *a*. Of course gene *a* is usually acting as well on genes *c*, *d*, *e* etc, which in turn act back on gene *b*. Thus there is the critical question of separating direct from indirect actions of gene *a*. To do this requires a method of analysis which keeps explicit track of each individual gene’s contribution to the expression of a given gene. This is the purpose of a gene circuit. Note that even if a direct

mechanism of action, such as a segmentation gene product binding to a specific site, is established by biochemical methods *in vitro*, the role of this regulatory mechanism in embryogenesis must still be demonstrated *in vivo* using transgenic embryos. Thus the use of biochemical methods does not remove the necessity of considering how the intact regulatory circuitry acts in an embryo, and the need for gene circuits remains.

A second point is that many questions about gene regulation are essentially quantitative and so cannot be answered by qualitative inferences from data. An example of such a problem concerns the shifts in cell fate markers that occur when the Bicoid gradient is rescaled by changing the number of copies of the gene. Analysis of the quantitative behavior of these shifts using the gene circuit method led to the insight that a second gradient was also involved (Reinitz et al., 1994). Another example of such a question is how the same set of gap gene domains can establish *eve* and *hairy* stripes that are displaced from each other by as little as one nucleus.

The third point, developed in the Introduction, is that understanding the timing of regulatory actions of genes requires a method of analysis which takes account of the rate at which genes are synthesized.

We believe that these points establish the need for a gene circuit model. The form that this model takes must be determined by what is useful and what is feasible as a modeling approach in this scientific context.

2.2. Formulating the gene circuit method

The gene circuit method is formulated by successive approximations to the exact biology. The approximations made proceed from general assumptions which determine the overall approach to very specific approximations made in applying these ideas to the *Drosophila* blastoderm.

We intend to make these approximations plausible. When an approximation is made, we shall argue that it is in accord with some specific experimental evidence. Nevertheless, the specific form of the gene circuit model is not dictated by experiment. Scientific judgement is required in deciding both what to include and what to leave out. The important thing is that when something that turns out to be necessary has been left out, say, this will show up in comparing to experiment and the model can be corrected.

The fundamental hypothesis of our work is that gene regulation can be usefully studied with concentrations of gene products taken as fundamental state variables, and with gene interactions represented by their effect on the synthesis rate of gene products. We believe that gene expression patterns furnish the correct experimental data base in which to seek answers to many (not all) important questions about gene regulation.

This hypothesis reflects a coarse-grained approach to

gene regulation, compared to one based directly on detailed biochemistry. This is a central point and the reason for adopting this approach requires careful explanation.

We first explain why we have not attempted to base our method on a fine-grained, biochemical level of description. The short answer is that such a method would be too complicated to use in the context of development and that the experimental input data which would be needed is not available. For example, to use such a method, one would need extensive knowledge of the relationship between the binding state of a transcription complex and concentration history. However, the relationship of binding state to transcription rate cannot be assayed until there is a faithful in vitro assay for RNA polymerase II regulation using only purified factors.

A coarse-grained method is always incomplete relative to the total scientific picture. This is reflected in two important ways. One is that the coarse-grained description will contain parameters, not fixed by the model, which must be determined either by experiment or by a more fundamental theory. Second, the scope of the method is limited. Application of the method in new experimental regimes will generally require changes in the model and certain important questions will fall outside its scope.

Both of these features are present in the gene circuit method. The method does not determine the coefficients defining the coupling strength of genes in the regulatory circuit; these coefficients must be found by fits to expression data. The method would have to be extended if one were to describe the effect of promoter structure on expression patterns, or the influence of cell-cell interactions. Likewise, the method must be modified when it is applied at a time or to a region of the embryo where new genes become active.

We emphasize that these features do not prevent a coarse-grained method from having both predictive power and explanatory value. Our previous work on gap genes and the work on *eve* reported in Section 3 of this paper are examples. Extremely useful analysis can be carried out at a coarse level of description even when a finer level of description is available.

State variables carry the information necessary to provide a complete description of a system at a given level of approximation. A consequence of our fundamental hypothesis is that the state variables defining a gene circuit are observable, since most genes active during blastoderm have been cloned and the expression patterns of cloned genes can be monitored with specific reagents such as antibodies.

The fact that the state variables used in formulating the gene circuit method are directly observable is of course of critical importance in bringing this model into contact with experiment. We believe that the obser-

vability of state variables in the blastoderm provides an opportunity to analyze the dynamics of important developmental processes at an unprecedented level of detail.

The full set of state variables used to describe an embryo in the gene circuit method consists of the concentration of regulatory molecules and the number and type of biological entities which are present at a given time. The next idea is that it is possible to write a fairly simple set of equations which determine how these quantities change with time. This is done as follows.

The regulatory molecules represented in the current implementation of the gene circuit method are proteins synthesized by certain gap and pair rule genes. RNA is not included because there is no evidence, that we are aware of, for a direct role of RNA or small ligands in the regulation of zygotic segmentation genes, although the formation of maternal gradients involves the selective localization (*bcd*, *nanos*) and translation (*hb^{mat}*) of RNA. This is the first of many assumptions we make which are system specific. Such an assumption is evidently not fundamental and could be relaxed if necessary.

The change in time of concentrations of proteins is governed by three basic processes: Direct regulation of protein synthesis from a given gene by the protein products of other genes (including auto-regulation as a special case); transport of molecules between cell nuclei; and decay of protein concentrations. Each of these processes must be represented in any method which has a chance of being correct and the equations we shall write down are the simplest set of equations which do this.

The form of the term accounting for the direct regulation of protein synthesis is suggested by general experience with regulated enzymatic reactions. Such reactions are characterized by a maximum reaction rate and the property that they can vary from zero up to this maximum rate. We thus introduce a scale factor setting the maximum rate of synthesis for each gene, as well as a term which describes the regulatory effect of uniformly distributed transcription factors. We next introduce a 'regulation-expression function' which makes the synthesis rate of a given gene a monotonic, saturating function of the regulating gene products (Fig. 1). This choice of regulation-expression function is the simplest way to interpolate between the state in which a gene is completely turned off and its state of maximum activity. It is an important, and necessary, aspect of our approach that its main results do not depend sensitively on the exact choice of this function.

We next discuss the dependence of the regulation-expression function on protein concentrations. This dependence must allow for the regulatory effects of various genes on the concentration of a particular gene. We believe that the precise binding states of regulator

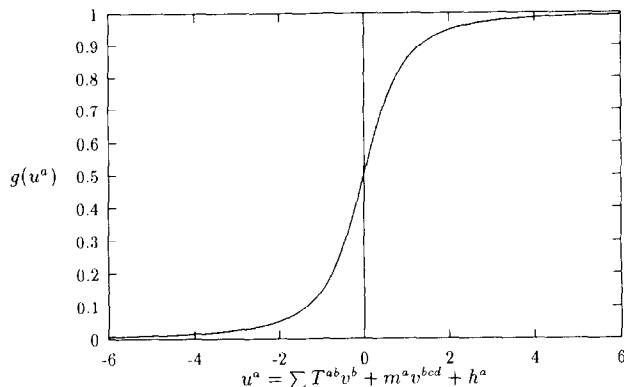


Fig. 1. The regulation-expression function $g(u^a)$. The total regulatory input u^a to the promoter of gene a is shown on the x-axis. The elements of T^{ab} and m^a have dimensions of inverse concentration so that u^a is dimensionless. The y-axis shows $g(u^a)$, a dimensionless quantity that describes the relative activation level of gene a . When $g(u^a) = 1$, the gene is transcribed at the maximum possible rate. Note the rapid approach to saturation at values of u^a above 2 and the rapid approach to zero for values of u^a below -2.

proteins are not required for a description of the regulative state of a gene. Although the regulative state is in a general sense a result of ligand binding, these binding configurations are highly redundant and thus do not provide a useful way of specifying the regulative state of a gene. This, of course, is a restatement of our coarse-graining approximation and it leads to the important consequence that the direct regulatory action of one gene on another can be described simply in terms of its effect on concentrations. We further suppose that these interactions can be characterized by a single real number (for each pair of genes). In a specific problem, we will generally need to allow for an interaction between any pair of genes represented in the gene circuit, so that we must consider a collection of regulatory coefficients T^{ab} . This collection of numbers is conveniently represented as a matrix, T . The matrix of regulatory coefficients defines a gene circuit. It is thus the fundamental theoretical object in the gene circuit method. In the future it may be possible to relate the coefficients T^{ab} to quantities having direct biochemical significance. We emphasize that at present these coefficients have the status of phenomenological parameters, which must be determined by experimental data. The way in which this is done is described in Section 2.3.

We observe that in assuming that gene interactions can be characterized by a single real number per pair of genes, we are excluding certain kinds of interactions which are consistent with the general idea that gene regulation can be modeled in terms of concentrations. For example, a more complicated approach would be required if gene a activated/repressed gene c , depending on whether the product of gene b was present/absent. This could actually happen. At the moment, we proceed

assuming it does not, pending evidence to the contrary. Another possibility, that gene interactions depend on the space derivatives of concentrations, is excluded for a different reason. At the early stage of embryonic development which we model, it seems clear that cell nuclei must be represented as discrete objects. The spatial dependence of gene concentrations is therefore represented as a discrete variable, and a gradient is represented by a finite difference operator, determined by neighboring cell nuclei. The regulatory coefficients would thus not be a property of a single cell nucleus, contradicting a basic biological fact.

The gene circuit method also includes transport of gene products between cell nuclei and decay of gene products. Exchange of gene products is modeled as a classical diffusion process (with discretized space derivatives). This way of modeling molecular transport is not likely to be correct at a fine level of description, but it is our judgement that classical diffusion is sufficiently accurate for our present purposes. Decay of gene products is modeled as a simple exponential, with a rate constant that must be supplied from experiment.

To summarize the discussion so far, our model for the change in concentration of a gene product leads to an ordinary differential equation having the schematic form

$$\left(\begin{array}{l} \text{time rate of change} \\ \text{of protein conc.} \end{array} \right) = \text{Regulation} + \text{Diffusion} + \text{Decay} \quad (1)$$

A differential equation of this kind holds during interphase, while gene products are synthesized. It is also essential to represent mitosis in the model. We do not model the process of mitosis, which is itself quite complicated, in detail. Instead, we introduce a further coarse-graining assumption and model mitosis as an elementary event in which just three things occur: The synthesis of gene products is suspended; the number of cell nuclei is doubled; and the gene products present in a cell nucleus at the onset of mitosis are distributed among the progeny. At the mathematical level, this is accomplished by adjoining a rule to the differential equation which suspends the equation and re-initializes the state variables according to the schedule of mitosis. In general, it is possible to arrange for this rule to be triggered when the dynamics brings a given cell to a particular state (so that, for example, a specific gene product exceeds a concentration threshold). In applying the equation to the *Drosophila* blastoderm, however, we will take advantage of the simplifying fact that the timing of cell division is under maternal control to implement a pre-determined schedule of mitosis. In either case, the resulting mathematical framework is an example of a hybrid dynamical system, consisting of linked continuous and discrete time evolution. This scheme can

also be understood, more generally, as a dynamical grammar in the sense of Lindenmayer (1968) and Mjolsness et al. (1991).

The foregoing discussion has focussed specifically on the gene circuitry. In a typical embryo, many other processes are coupled to gene regulation. These include cell-cell interactions, hormonal regulation and morphogenetic movements. Elsewhere (Mjolsness et al., 1991) we have sketched out how the gene circuit approach can be enlarged to include such processes.

These additional complications can be avoided in applying the gene circuit method to the *Drosophila* blastoderm. During the blastoderm stage of development, the genetic regulatory network is effectively isolated from other developmental processes. This is a consequence of the specifics of *Drosophila* development. The mechanical milieu is set by substances provided by the mother so that, although a complex series of nuclear movements and divisions occur, they can be regarded as a prescribed arena in which gene regulation dynamics takes place. Prior to gastrulation, the zygotic genes required for segmentation regulate only one another; they do not directly affect morphological processes like cell movements, cell shape or cell division. This de-coupling between mechanical variables and the action of zygotic genes prior to gastrulation is demonstrated by the fact that the earliest visible alterations in embryonic morphology as a result of mutating a zygotic segmentation gene are seen in the gastrula or later (Lehmann and Nusslein-Volhard, 1987; Wieschaus et al., 1984; Petschek et al., 1987; Strecker et al., 1988; Merrill et al., 1988). Mutations in early acting zygotic genes do alter the patterns of expression of other early acting zygotic genes in the blastoderm (Jäckle et al., 1986; Reinitz and Levine, 1990; Mohler et al., 1989; Pankratz et al., 1989; Kraut and Levine, 1991a; Kraut and Levine, 1991b; Eldon and Pirrotta, 1991), but these altered patterns of expression do not result in morphological changes until after the onset of gastrulation. Cell-cell interactions are not important because cells have not yet formed; any spatial interactions between cell nuclei can be treated by exchange of gene products. Lineage is known not to restrict cell fate until the blastoderm has cellularized (Simcox and Sang, 1983). Thus the *Drosophila* blastoderm has many properties which simplify the analysis of its gene circuitry.

To make the ideas we have discussed fully explicit, we must make further approximations which are specific to the *Drosophila* blastoderm.

In applying the gene circuit method to study the role of gap genes in the formation of *eve* stripes, we analyze the blastoderm during cleavage cycles 11–14 and focus on the region of the blastoderm which generates the thoracic and anterior abdominal segments. Here, the four gap genes *Kr*, *kni*, *gt* and *hb*, all under the control of *bcd*, and the pair-rule gene *eve* form an approximately isolat-

ted set of mutually regulating genes, during the interval of time under consideration.

This approximation is based on several facts. First, genes affecting the anterior-posterior (A-P) and dorsal-ventral (D-V) axes are uncoupled in the central part of the blastoderm, as is known from the fact that mutations in D-V genes do not affect the expression of A-P genes and vice versa. In the middle region of the embryo, the main gap genes active during cycles 11–14 are the four listed above, with genes such as *tailless* exerting an effect (in wild type embryos) only at the anterior and posterior ends of the region modeled.

Although cross-regulation and auto-regulation of pair-rule genes are critical for their correct function, we need not consider pair-rule regulation of *eve* here. Each pair-rule gene has a characteristic time at which altered expression becomes visible in pair-rule mutants; for *eve*, this time occurs after the onset of gastrulation (Frasch and Levine, 1987). Moreover, the onset of cross-regulatory effects on *eve* occurs later than for any other pair-rule gene (Carroll and Vavra, 1989).

We further note that in the middle region of the embryo, the level of expression of gap and pair-rule genes is, approximately, a function only of position along the A-P axis. This implies that it is sufficient to consider a one-dimensional system consisting of a line of nuclei running along this axis.

Let the position of a cell nucleus along the A-P axis be indexed by *i*, such that nucleus *i*+1 is immediately posterior to nucleus *i*. Each cell nucleus contains a copy of a regulatory circuit composed of *N* genes, and which is determined by an *N* × *N* matrix *T*. The concentration of the *a*th gene product in nucleus *i* is a function of time, denoted by *v_i^a(t)*. In this notation, the schematic equation (1) can be written in the explicit form

$$\begin{aligned} \frac{dv_i^a}{dt} = & R_a g_a \left(\sum_{b=1}^N T^{ab} v_i^b + m^a v_i^{bcd} + h^a \right) \\ & + D^a(n) \left[(v_{i-1}^a - v_i^a) + (v_{i+1}^a - v_i^a) \right] \\ & - \lambda_a v_i^a \end{aligned} \quad (2)$$

where *N* is the number of zygotic genes included in the circuit (= 5 in the present application). The first term on the right hand side of the equation describes gene regulation and protein synthesis, the second describes exchange of gene products between neighboring cell nuclei and the third represents the decay of gene products.

In (2), *T^{ab}* is the previously discussed matrix of genetic regulatory coefficients, whose elements characterize the regulatory effect of gene *b* on gene *a*. This matrix does not depend on *i*, a reflection of the fundamental fact that the cell nuclei of a multicellular organism con-

tain identical genetic material. The *bcd* input is given by $m^a v_i^{bcd}$, where v_i^{bcd} is the concentration of Bicoid protein in nucleus i and m^a is the regulatory coefficient of *bcd* acting on zygotic gene a . g_a is the regulation-expression function (Fig. 1), which we assume takes the form $g_a(u^a) = (1/2)[(u/\sqrt{u^2 + 1}) + 1]$ for all a , where $u^a = \sum_{b=1}^N T^{ab} v_i^b + m^a v_i^{bcd} + h^a$. R_a is the maximum rate of synthesis from gene a and h^a summarizes the effect of general transcription factors on gene a . The diffusion parameter $D^a(n)$ depends on the number n of cell divisions that have taken place and varies inversely with the square of the distance between nuclei. We assume that the distance between adjacent nuclei is halved after a nuclear division. λ_a is the decay rate of the product of gene a .

hb^{mat} protein concentration is incorporated into the circuit as an initial value (at cleavage cycle 11) for the concentration of *hb* product. This is a mathematical expression of the fact that the observed concentration of *hb* protein consists of both maternal and zygotic components. Initial values of the other, purely zygotic, gene products are taken to be zero at cleavage cycle 11.

In implementing the rule for cell division, we incorporate the facts that in the *Drosophila* blastoderm, mitosis lasts about four minutes and that gene products appear to be equally distributed between the two daughter nuclei.

2.3. How experimental data determines a gene circuit

Equation (2) forms the basis of the gene circuit method for analyzing time dependent gene expression data. The function of the method is two-fold: It takes gene expression data as input and produces a gene circuit, T^{ab} , as output. Conversely, given T^{ab} , the method can be used to achieve biological insight into the ways in which genes, acting collectively, form expression patterns in the blastoderm.

In this section, we will explain how gene expression data is used to determine the coefficients T^{ab} . There are two basic ideas. The first is to formulate the problem of finding T^{ab} as that of obtaining a least squares fit to gene expression data; the second is to solve this problem using the method of simulated annealing.

As applied to gene circuits, the least squares fitting procedure works as follows. The input data consists of information about the concentration of a number of proteins in a large number of cells, at a number of different times and possibly in several different genotypes as well. These data are fit with protein concentrations obtained as solutions to Equation (2). For fixed initial conditions, a solution of Equation (2) can be thought of as a function of the parameters T^{ab} , R_a etc. The values of these parameters are varied so as to minimize the summed squared deviations of the computed protein concentrations from the experimentally observed concentrations.

This defines a least squares problem for gene circuits. The key to solving this problem is in the method used to systematically vary the parameters so as to achieve the desired minimization. We use a minimization algorithm known as ‘simulated annealing’ (Kirkpatrick et al., 1983). Although it is sometimes possible to solve minimization problems by fairly straightforward gradient descent methods (i.e., by going ‘downhill’), such methods fail on this problem. The function we are trying to minimize is very ‘bumpy’, i.e., it has many local minima that will trap a local downhill search. Simulated annealing has the advantage that it can find the true global minimum in such circumstances, although this is accomplished at the price of intensive computation. (Each fit reported in Section 3 required about 1 week of CPU time on a Sparc2 workstation.)

Simulated annealing works as follows. Imagine that we have just calculated the protein concentrations by integrating Equation (2) (once for every genotype for which there is data) and used this solution to compute the summed squared deviations from the data. We call this quantity the ‘cost function’. Next we change a parameter in Equation (2) and recalculate this cost function. If the new cost function has a smaller value than the old (we have gone ‘downhill’), we keep the new value of the cost function. If the new value is larger (an ‘uphill move’), we may or may not keep it: the decision is made according to chance. The bigger the uphill step, the smaller the probability that it will be accepted. More importantly, the chance of making an uphill move decreases slowly throughout the entire procedure, so that when the fit is over, the chance of an uphill move is zero. The ability to make uphill moves is why simulated annealing does not get trapped in local minima; the gradual reduction of the probability of an uphill move to zero is why it eventually finds the global minimum. Because simulated annealing is a random process, one must obtain the same results in multiple annealing runs in order to gain confidence that the global minimum has been found. A more detailed description of how simulated annealing is applied to the genetic circuitry problem is given in the Appendix.

2.4. Comments on the use of the data

It is clear that the quality of the conclusions that come out of the gene circuit method depend critically on the quality of the data that goes in. In this section we discuss that data, together with its limitations. We consider what types of information can and cannot be reliably extracted from the data with the methods we are currently using.

The data used in this paper were obtained by visual inspection of photomicrographs of embryos simultaneously stained with two fluorescently tagged antibodies. Estimates of the staining intensity were made visually and given numerical intensities ranging from

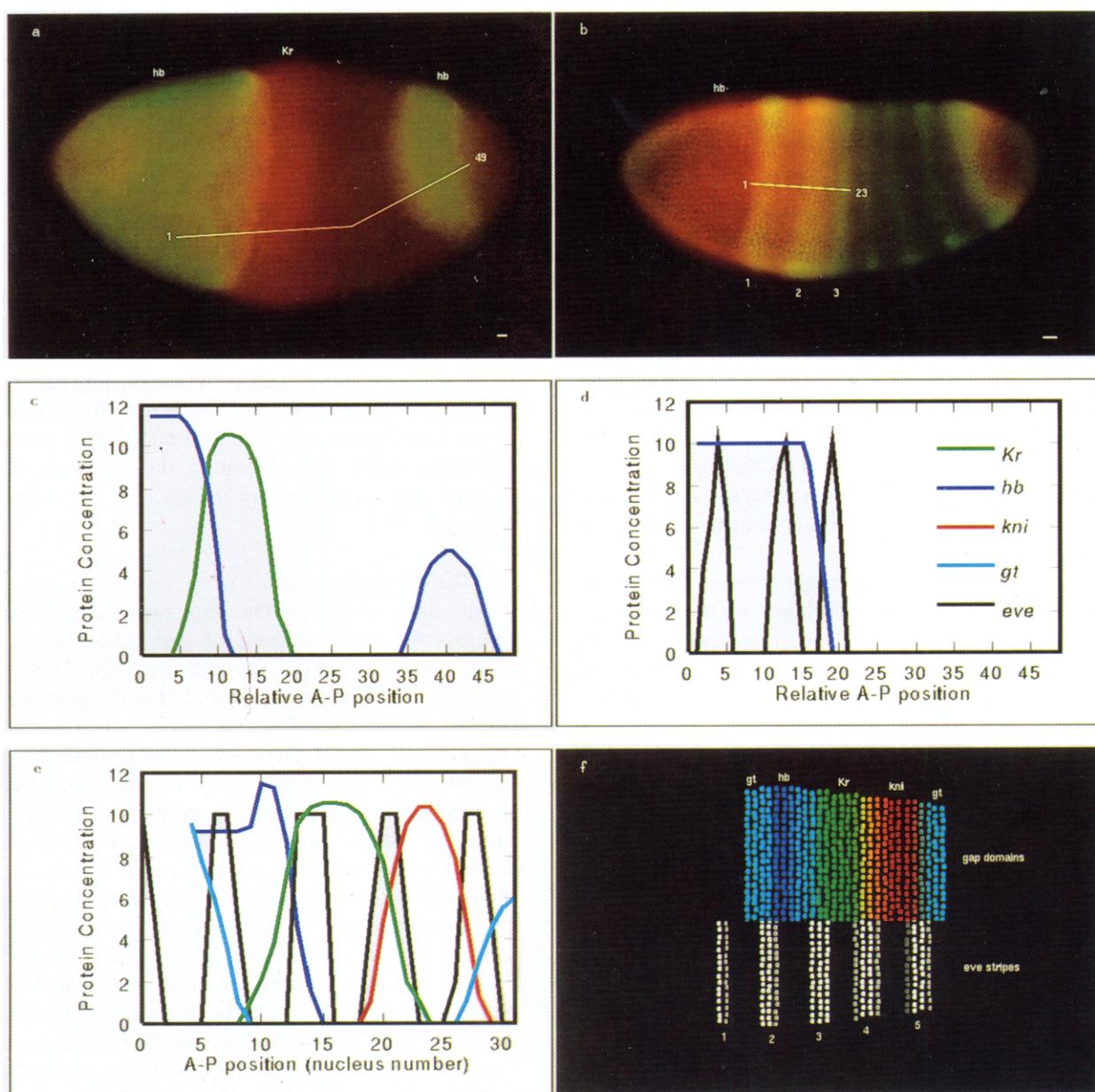


Fig. 2. Construction of a dataset for analyzing *eve* stripes. Panels (a) and (b) show photomicrographs of late cleavage cycle 14 embryos oriented with anterior to the left and dorsal up. The dorso-laterally oriented embryo in (a) has been stained with FITC tagged antibody to *hb* (green) and rhodamine tagged anti-*Kr* (red) (Frasch and Levine, 1987; Stanojevic et al., 1989; Kraut and Levine, 1991a), while the laterally oriented embryo in (b) has been stained for *hb* (red) and *eve* (green) (M. Levine, unpublished data). In (a) and (b), the white scale bar on the lower right corresponds to 10 microns. These two photomicrographs and about 30 additional ones comprised our raw data. (c) and (d) show the estimated protein levels derived from (a) and (b), respectively. In (c) and (d), the x-axis gives the relative anterior-posterior location in terms of number of nuclei along the embryo, while the y-axis gives estimated protein concentrations in terms of perceived relative fluorescence. In graphs (c)–(e), different protein species are indicated by color as shown in (d). The protein levels were obtained along an anterior-posterior strip near the lateral equator. The data in (c), comprising 49 nuclei, were obtained along the white line over embryo (a), with nucleus 1 and 49 located as shown. This line is in two segments to approximate the path of the lateral equator. The data in (d), comprising 23 nuclei, were obtained from the embryo in (b). This data was taken along the white line shown on the embryo in (b), with nucleus 1 and 23 as shown. (e) shows the composite late cleavage cycle 14 dataset derived from photomicrographs of 17 embryos, including (a) and (b). The y-axis shows relative concentration as in (c) and (d). The horizontal axis shows position along the anterior-posterior axis, with anterior to the left, in terms of the number of nuclei posterior to the middle of *eve* stripe 1. One nucleus is about 1% egg length. Note that there is no gap expression data in the anterior four nuclei. (f) is a representation of the late cleavage cycle 14 dataset shown in (e) as it would appear in a stained embryo, viewed from a tangential plane of focus. The protein levels are color coded as in (c)–(e), except that *eve* is white. More intense color indicates higher protein concentration. *Eve* levels are shown on the bottom half and gap domains on the top. The cyan band between the *Kr* and *hb* domains is a visual artifact resulting from the mixture of Kruppel and Hunchback; the two cyan bands at the edge are *gt* domains.

zero to ten. These photomicrographs were made for every pairwise combination of gene product and then grouped by developmental stage. The level of temporal resolution permitted a clear division into early middle and late cleavage cycle 14; earlier times were grouped by cleavage cycle. A full dataset was bootstrapped by aligning common expression domains. Because the critical spatial information for each expression domain is its position relative to every other expression domain considered, we obtained critical overlap data ourselves as part of an earlier study (Reinitz et al., 1994). This procedure is illustrated in Fig. 2.

Datasets obtained in this way have high quality information about the overlap of domains. They have poor quality information about the absolute levels of gene product concentrations. All of the data contain an unknown scale factor relating the observed staining intensity to the actual concentration. This means that certain types of information about the circuitry can be extracted from the data reliably, while others cannot, as we now explain.

Any information about gene expression that is not critically dependent on the absolute scale of concentrations can be extracted using the gene circuit method. This includes information about the specification of positional information by gradients (Reinitz et al., 1994), as well as the qualitative circuitry features discussed in Section 3. It is important to bear in mind that most important facts about development are fundamentally qualitative in nature. With current data, we are essentially inputting qualitative data and drawing qualitative conclusions. The fact that the intermediate workings of the method are carried out in quantitative terms should not obscure this fact. The key objective is to find features of the answer that are insensitive to this scaling uncertainty: we briefly indicate how this is done below and return to the topic in Section 3.

The uncertainty in scale factor for the concentration v^a of a gene product a has the following effect. If v^a is multiplied by a scale factor β , the equations remain the same if R^a , D^a , λ^a and the column of the T matrix T^{ab} are multiplied by $1/\beta$. That is, an uncertainty of scale in v leads to an uncertainty of scale of those parameters attached to that v . Thus, it is significant if a parameter turns out to zero or nearly so; similarly, for parameters like T^{ab} that can take on positive or negative values, the sign is significant.

We emphasize that it is false to think that the construction of a quantitative model is an all- or-nothing process in which all data must be perfectly quantified before any scientific conclusions can be drawn. It is a fact that nearly all currently available information about gene regulation in the blastoderm is derived from visual inspection of photographs of embryos. We are carrying this mode of analysis further by mapping the expression domains as carefully as possible and systematically

analyzing this data with the gene circuit method. This analysis will be further improved by the acquisition of quantitative data using CCD cameras or confocal microscopes. But even with current data, the gene circuit analysis of blastoderm circuitry is a major advance compared to what can be achieved by visual inspection of patterns.

3. The formation of *eve* stripes

In this Section, we present a detailed analysis of the formation of *eve* stripes 2–5. We restrict our attention to a region of the blastoderm which extends from the middle of *eve* stripe 1 to the middle of the interstripe between stripes 5 and 6. This is approximately the region where the four gap genes we consider constitute a full set of gap-type regulators. Extension of the analysis to the remaining stripes would require the inclusion of additional gap genes such as *tailless* (*tll*) and *huckbien* (*hkb*).

3.1. The regulatory circuit

The data used for the gap expression domains specifies the concentrations of gap gene products in early, middle and late cleavage cycle 14. This data, together with data specifying maternal gradients of *bcd* and *hb*, is extracted from a dataset given elsewhere (Reinitz et al., 1994). Here we use the central 32 nuclei (i.e. nucleus 17 to nucleus 48) of that dataset. The late cleavage cycle 14 gap gene expression data used in this study are shown in Fig. 2. Published observations (Frasch et al., 1987) were used to generate *eve* data for early and late cycle 13 as well as early cycle 14 (See Fig. 3). At each of these times the *eve* data is spatially uniform in the region considered. We also included *eve* data from late cycle 14 (Stanojevic et al., 1989, M. Levine, unpublished data), at which time the stripes have clearly formed (Figs. 2 and 3). Because we did not use *eve* data for mid cycle 14, we included *eve* data for two different times during cycle 13. This ensured that the same total number of data points was available for each gene, so that the fit was not biased in favor of any particular gene's expression pattern.

An important remark is in order about this dataset. We have input no information which directly specifies the transient pattern of *eve* expression. The *eve* data essentially tells the model that *eve* expression should be spatially uniform through the beginning of cycle 14 and fully striped by the end of cycle 14. It says nothing about the route taken from uniform expression to stripes.

Our dataset also includes expression data for *eve*⁻ for the same times at which we have wild type data. Although, in general, the mutant expression patterns would be quite different, they are the same in this particular case. This is because gap expression is unaltered in *eve* mutants and *eve* expression is unaltered in *eve*⁻

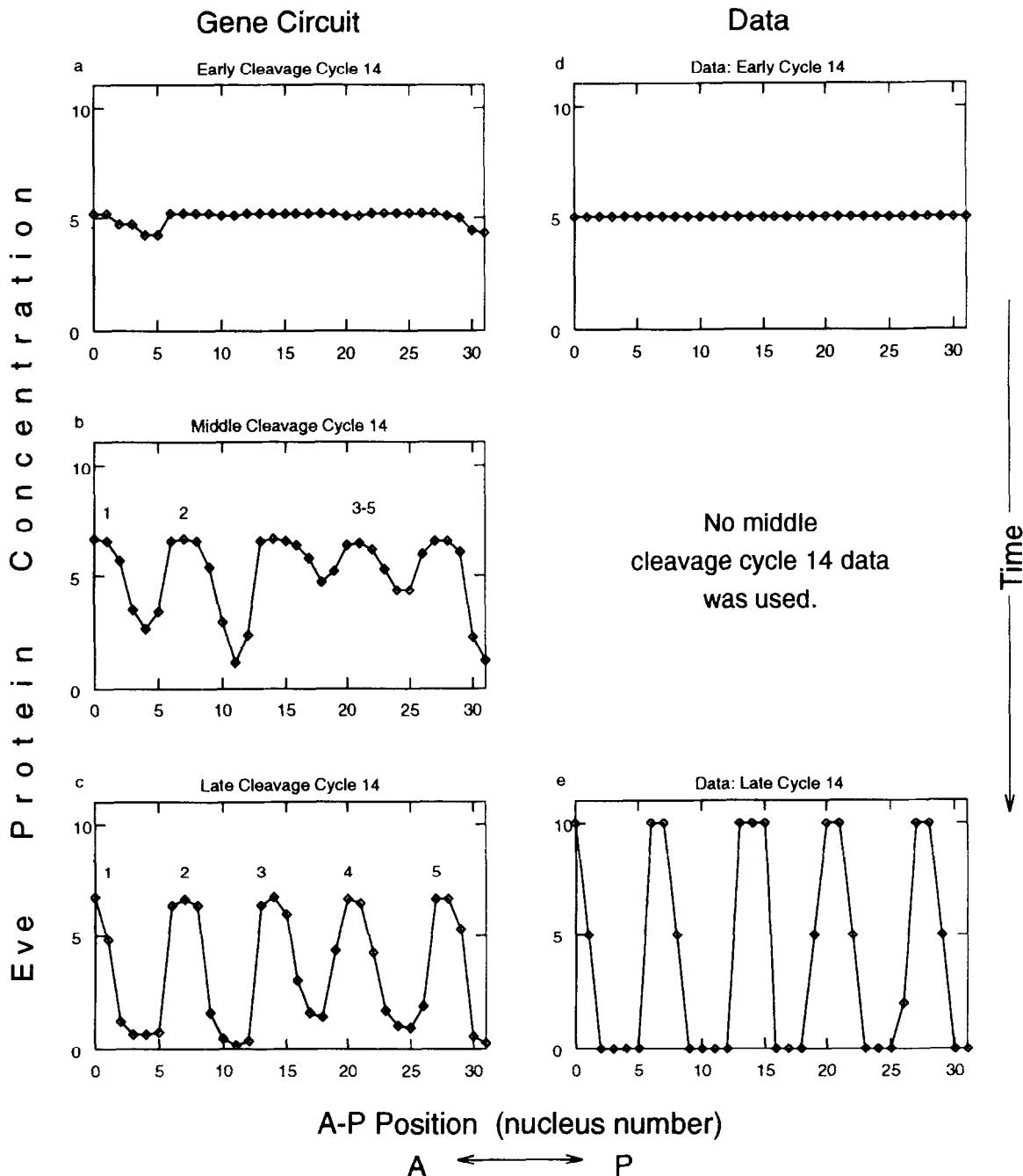


Fig. 3. Comparison of *eve* stripes formed by the gene circuit to input data. (a)–(c) show the behavior of the circuit at early, middle and late cleavage cycle 14 respectively. In (b), stripes 1 and 2 are distinct and labeled, while stripes 3–5 have just begun to form. In (c), all five stripes have formed and are labeled. (d) and (e) show the *eve* data from early and late cycle 14 that was used for the fit; no data for middle cycle 14 was used. The axes for the graphs are as described for Fig. 2e. Each small diamond indicates the *eve* concentration in a single nucleus. The left hand column shows the behavior of the circuit and the right hand column shows the *eve* expression data used for the fit. Additional spatially uniform *eve* expression data for early and late cleavage cycle 13 are not shown. The *eve* expression level at early cycle 13 was set to 4.0 and at late cycle 13, to 4.5. Time is measured from the completion of the tenth nuclear division. As measured from this marker, 'early cycle 13' occurs at 30.66 min, 'late cycle 13' at 48 min, 'early cycle 14' at 57.33 min, 'middle cycle 14' at 73.66 min and 'late cycle 14' (the onset of gastrulation) at 89.66 min (Foe and Alberts, 1983). Time increases downward, as indicated.

until after gastrulation. We incorporate this information into the method as follows. For each evaluation of the cost function, the equations are integrated and compared with data twice; first to the wild type data and then to the mutant data. When integrating the equation for the purpose of comparing with mutant data, the T-matrix elements that describe the regulatory effect of *eve* on all the other genes are set to zero. This describes a mutant which is functionally null but nevertheless synthesizes inactive protein. This procedure incorporates the fact that gap gene expression patterns are unchanged in mutants for *eve*.

Application of the simulated annealing procedure described in Section 2.3 determined a set of parameters for Equation (2) that gave the closest possible fit to the data. Several simulated annealing runs gave scores which agreed to within 1% and nearly identical values for the parameters, so that the results are reproducible and can be reliably taken to represent the global minimum.

Solving the equation using these parameters gave the results shown in Fig. 3. This figure shows the *eve* expression patterns at early, middle and late cycle 14, as well as the experimental data used. Fig. 4 provides a

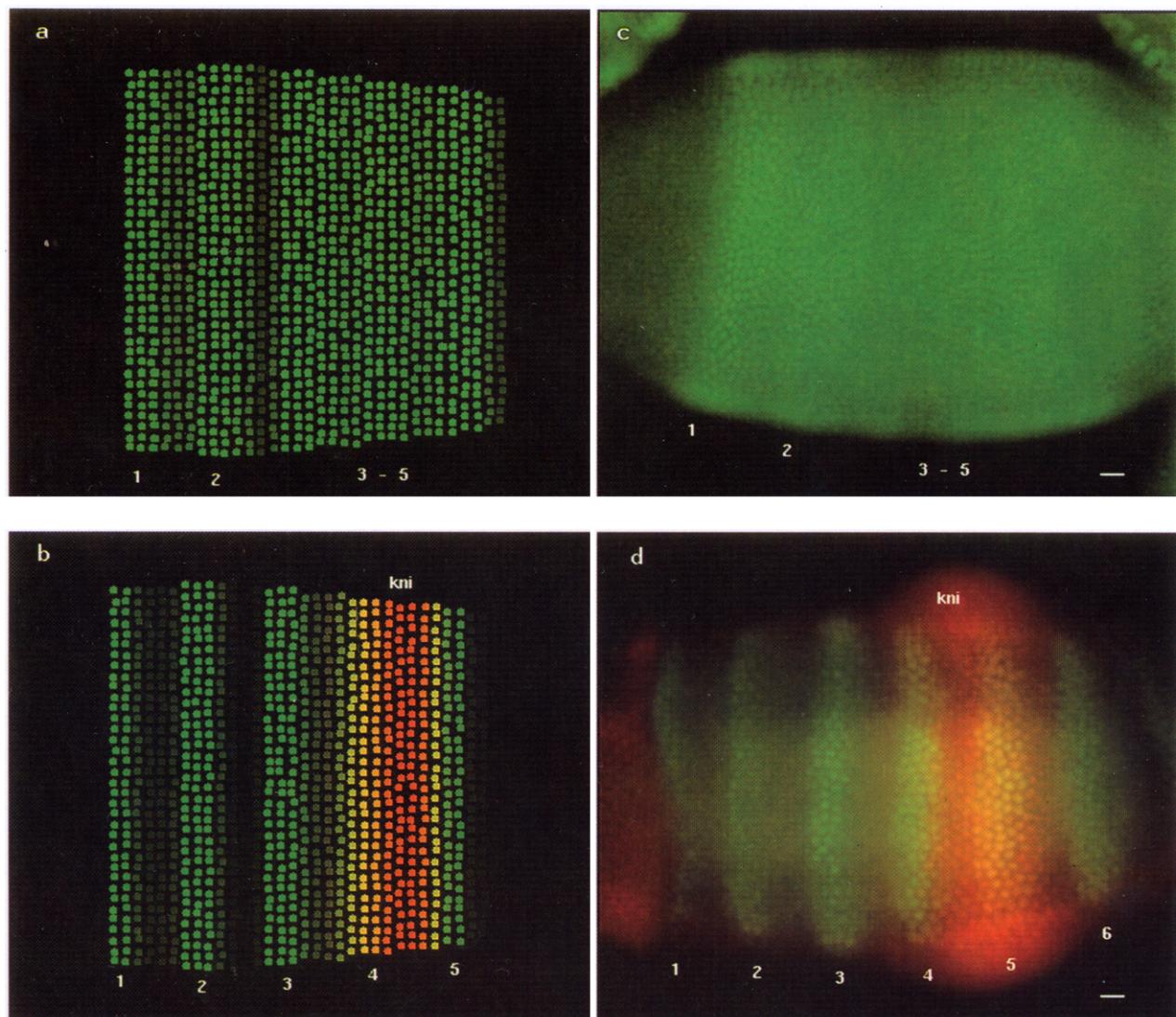


Fig. 4. A comparison of *eve* stripes formed in the gene circuit with those formed by embryos. (a) and (b) show the behavior of the gene circuit at middle and late cleavage cycle 14, respectively. (c) and (d) show stained embryos at middle and late cleavage cycle 14. (a) and (b) represent the protein concentrations as they would appear in a stained embryo (see Fig. 2f); (a) shows *Eve* alone (green), while (b) shows *Eve* concentration (green) and *Knirps* (red). (c) shows an embryo stained for *eve* alone, while (d) (M. Levine, unpublished data) shows a ventral view of a late cleavage cycle 14 embryo stained for *eve* (green) and *kni* (red). In (c) and (d), the scale bar at the lower right corresponds to 10 microns; each embryo is oriented with anterior to the left; the lateral embryo in (c) is oriented dorsal up. In (a) and (c), stripes 1 and 2 have formed and are labeled; in (b) and (d) all five *eve* stripes have formed and are labeled, as is the *kni* domain.

Table 1
The T-matrix determined by the experimental data shown in Figs. 2 and 3

Target	Regulator					
	Kr	hb	gt	kni	eve	bcd
Kr	+0.34	-0.86	-0.44	-1.6	+0.005	+1.8
hb	+0.013	+0.11	-0.90	-0.23	+0.002	+7.6
gt	-2.5	-0.28	+0.15	-0.24	+0.001	+1.1
kni	-0.33	-4.31	-1.6	-0.076	+0.003	-1.8
eve	-2.3	-3.0	-1.9	-1.4	+0.039	+14

The $a b^T$ element of T specifies the regulatory effect of b on a . The last column, with bcd as regulator, corresponds to m^a in the equations. These terms may be thought of as an extra column of T^{ab} . Each entry in the table has dimensions of concentration⁻¹.

representation of the results in Fig. 3 as they would appear in a stained embryo, together with photomicrographs of actual stained embryos for comparison.

We call attention to the principal features of the solution. Five distinct *eve* stripes are formed by the end of cycle 14. They are in the proper location and have the correct width, although there is a small amount of residual expression in some of the interstripes. The peaks are of the same height, as is the case in the dataset, although they are a little too low. Because of uncertainties in the absolute scale of intensities, discussed in Section 2.4., deviations from the data that amount to a uniform scale factor are not significant.

The time dependence of the *eve* expression pattern is also given correctly (Figs. 3 and 4). As seen in these figures, *eve* expression is essentially uniform through early cycle 14. A phased process of stripe formation is initiated in middle cycle 14, beginning with the formation of stripes 1 and 2, followed by the formation of stripes 3–5. Stripe formation is completed by the end of cycle 14.

It is to be emphasized that key features of the pattern predicted by the gene circuit for mid cycle 14 are in substantial agreement with the observations here (Fig. 4; Reinitz and Sharp, unpublished observations) and elsewhere (Frasch et al., 1987). These observations show that as uniform expression breaks up into stripes in mid cycle 14, stripes 1 and 2 appear first, followed quickly by stripes 3 and 7. This gives four stripes and a broad belt of expression containing the presumptive stripes 4–6. Stripe 4 resolves next, followed by the splitting of stripes 5 and 6. In the circuit, stripes 1 and 2 appear first, followed by the synchronous appearance of stripes 3–5. The early appearance of stripes 1 and 2 is an important feature of the transition pattern that is correctly predicted by the gene circuit, although no direct information about this pattern was included as part of the input data. This is a striking fact, which raises the question of how the gene circuit gets this right. We show that

this can be understood by looking in more detail at this gene circuit.

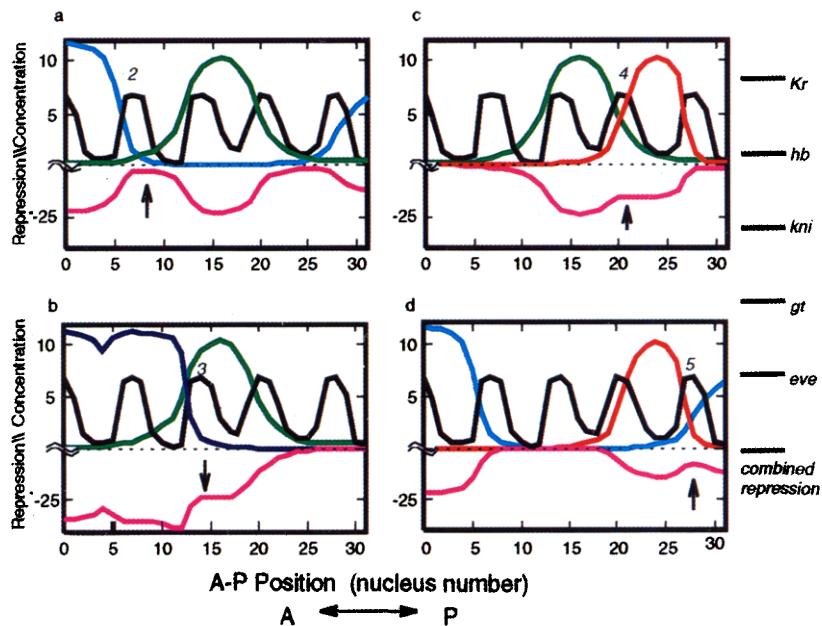
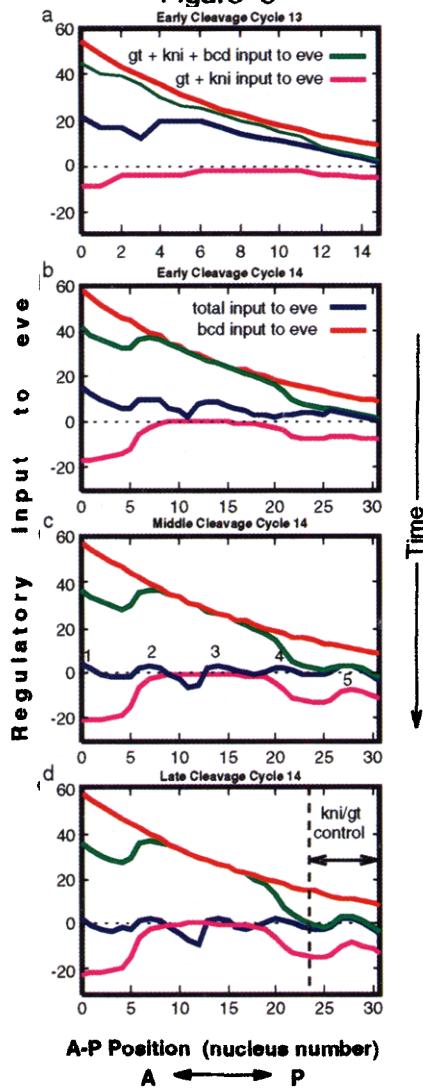
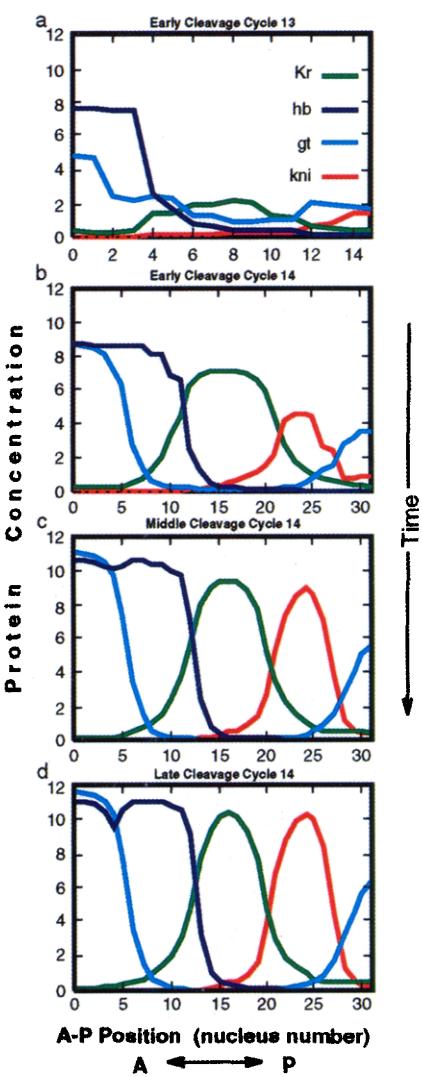
The T-matrix which determines this circuit has five notable features (See Table 1). (1) The diagonal (auto-regulatory) terms of the matrix are positive. (2) The off-diagonal (cross-regulatory) terms are negative. Specifically, all the gap genes repress *eve*. (3) The input from *bcd* is positive for all genes with the exception of *kni*. (4) All *eve* outputs are zero — *eve* does not regulate gap genes. (5) An exception to (1) and (2) is that a few of the T-matrix elements are zero. We can infer from this a picture in which there is generalized activation of expression domains by *bcd* and general transcription factors. This activation is modulated by local repression to form stripes.

To understand how this happens, one must look in detail at the spatial variation of gap gene expression and its regulatory effect on *eve*. It is in the analysis of this question that the gene circuit method shows its real power. Each gene active in the region is explicitly represented and we can isolate its individual effect on the expression pattern of *eve*. The effect of spatially varying regulation of gene a on *eve* is given by $T^{eve} - {}^a v_i^a$ (or by $m^{eve} v_i^{bcd}$ in the case of *bcd*). In particular, at each spatial location i , the various contributions are summed. In analyzing stripe formation, we need consider only the genes that are expressed in the region where the stripe is formed.

3.2. The formation of stripes

We first consider the repressive effects of pairs of gap genes on *eve* expression. The four panels in Fig. 5 show the results obtained using the gene circuit method for the expression of *eve*, together with the expression patterns of four different pairs of gap genes at late cleavage cycle 14. The existence of a minimum of pairwise repression coinciding with stripes 2 (Fig. 5a; repression by *Kr* and *gt*) and 5 (Fig. 5d; repression by *gt* and *kni*) suggests that pairwise repression is sufficient to form a stripe. For stripes 3 (Fig. 5b; repression by *hb* and *Kr*) and 4 (Fig. 5c; repression by *Kr* and *kni*), the situation is more ambiguous: A clear feature in the pairwise repression curve corresponds to the peak of each stripe but, in this case, the feature is a shoulder in the curve, rather than a minimum. Because each side of these minima or shoulders of repression are formed by a different gap domain, it is reasonable to suppose that each stripe border is under the control of a particular gap domain. Fig. 5 thus suggests several interesting hypotheses about the role of repression by pairs of gap genes in stripe formation. We turn to more detailed analyses which sharpen and confirm these general observations.

The main thing that this analysis must do is separate out the gene contributions that actually make a given stripe. We will carry out this analysis for each stripe, beginning with stripe 5. The four panels in Fig. 6

Figure 5**Figure 6****Figure 7**

correspond to early cycle 13 and early, middle and late cycle 14. Each panel shows four curves: *bcd* activation, $m^{eve}v_i^{bcd}$, the net repression of *eve* by *gt* and *kni*, $T^{eve-gt}v_i^{gt} + T^{eve-kni}v_i^{kni}$, the sum of the foregoing contributions and the total input to *eve*, u^{eve} (see Equation (2)). These curves carry the information which allows us to determine whether regulation by a given pair of gap genes is in fact sufficient to form a particular stripe. To do this we need to have the curve for the combined repression of the two gap genes under consideration. Since the only activating gene product is *bcd*, we must also have the curve of *bcd* activation. To determine the balance between activation and repression, we must also have the sum of these two curves. Finally, we display u^a to demonstrate that the combined effects of *bcd* and the gap genes accounts accurately for the total input to *eve*. The behavior of u^a allows us to track the formation of stripes.

Now we can understand how the two gap genes *gt* and *kni* control the formation of stripe 5. The joint repressive effect of these two genes combines with the *bcd* gradient to produce a local maximum of *eve* activation at the location of the peak of stripe 5, surrounded by local minima of *eve* activation corresponding to the two interstripes. These maxima and minima are amplified by the regulation-expression function $g(u)$ to produce *eve* stripe 5. In particular, *kni* sets the anterior border of stripe 5 and *gt* sets its posterior border. Thus the virtually exact coincidence of the net contribution of *bcd*, *gt* and *kni* with the total *eve* input in the stripe 5 region clearly accounts for the formation of this stripe. The patterns of maxima and minima of *eve* activation first becomes apparent at the beginning of cycle 14 and strengthens thereafter. This is a consequence of the

increase in peak gap gene expression during cycle 14, shown in Fig. 7. The increase in peak levels increases repression in the interstripes, which forms a gap in repression leading to the formation of a stripe. Thus the rate of growth of gap domains combined with the exact values of connection strengths control the timing of stripe formation. Although *gt* and *kni* domains are present in cycle 13 and early cycle 14, their combined effect is spatially uniform (compare Fig. 7a and 7b to Fig. 6a and 6b).

We next discuss the formation of stripes 4 and 3. Referring to Fig. 8a–d, we draw attention to the time development of the combined repression of *Kr* and *kni* (magenta curve). The shoulder appearing in Fig. 5 also appears here, as early as middle cycle 14. In cycle 13 and early cycle 14, this feature is absent. The appearance of the shoulder correlates well with the appearance of three peaks of activation of u^{eve} in middle cycle 14 that are the first indicators of stripes 3–5. The coincidence of the green curve summing the effects of *bcd*, *Kr* and *kni* with u^{eve} (blue) throughout the region where stripe 4 and its surrounding interstripes will form indicates that these three genes are solely responsible for the formation of stripe 4. It is thus evident from the figure that the posterior margin of stripe 4 forms as a consequence of *kni* repression together with the small differences in Bicoid level across the stripe. The anterior margin of stripe 4 is delimited by *Kr*. The formation of the shoulder in repression, and thus the stripe, is a result not only of the growth of *Kr* and *kni* repression to make interstripes but also of the spatial refinement of these domains (Fig. 7). Gap protein concentration declines in a few nuclei at the edge of the domains during the refinement process, which lowers repression and contributes

Fig. 5. Placement of *eve* stripes relative to pairs of repressing gap domains as given by the gene circuit. The x-axis of each graph shows anterior-posterior position as described for Fig. 3e. Positive values along the y-axis represent relative concentrations as in Fig. 3e, while the negative values indicate the net repressive effects of the pair of gap genes shown. (a) shows the placement of *eve* stripe 2 (labeled) relative to *gt* (cyan) and *Kr* (green) expression domains. The magenta curve shows $T^{eve-Kr}v_{Kr} + T^{eve-gt}v_{gt}$, the combined repressive effect of *gt* and *Kr* on *eve*. The arrow indicates a gap in repression corresponding to stripe 2. (b) shows the placement of *eve* stripe 3 (labeled) relative to domains of *hb* (blue) and *Kr* (green) expression. The magenta curve shows $T^{eve-Kr}v_{Kr} + T^{eve-hb}v_{hb}$, the combined repressive effect of *Kr* and *hb* on *eve*. The arrow indicates a shoulder in repression corresponding to stripe 3. (c) shows the placement of *eve* stripe 4 (labeled) relative to domains of *kni* (red) and *Kr* (green) expression. The magenta curve shows $T^{eve-Kr}v_{Kr} + T^{eve-kni}v_{kni}$, the combined repressive effect of *kni* and *Kr* on *eve*. The arrow indicates a shoulder in repression corresponding to stripe 4. (d) shows the placement of *eve* stripe 5 (labeled) relative to domains of *gt* (cyan) and *kni* (red) expression. The magenta curve shows $T^{eve-kni}v_{kni} + T^{eve-gt}v_{gt}$, the combined repressive effect of *gt* and *kni* on *eve*. The arrow indicates a gap in repression corresponding to stripe 5.

Fig. 6. Regulatory input to *eve* stripe 5. The x-axis of each graph shows anterior-posterior position as described for Fig. 3e. In (a), the axis extends from 0 to 15 since it represents cleavage cycle 13 data in which half as many nuclei are present compared to cleavage cycle 14. The y-axis shows regulatory input to (dimensionless units). The panels display the same quantities at each of four times. The red line represents $m^{eve}v^{bcd}$, the magenta line $T^{eve-kni}v_{kni} + T^{eve-gt}v_{gt}$, the green line $T^{eve-kni}v_{kni} + T^{eve-gt}v_{gt} + m^{eve}v^{bcd}$ and the blue line, u^{eve} . Where the blue line overlaps the green line *kni*, *gt*, and *bcd* provide all input to *eve*, leaving *eve* under the effective control of *kni* and *gt*. The region of this overlap is marked ‘*kni/gt* control’ in (d). The prepattern of the five stripes in u^{eve} is visible by middle cleavage cycle 14 and the stripes are labeled in (c).

Fig. 7. Evolution of gap domains as determined by the gene circuit. The x-axis of each graph shows anterior-posterior position as described for Fig. 3e. In (a), the axis extends from 0 to 15 since it represents cleavage cycle 13 data in which half as many nuclei are present compared to cleavage cycle 14. The y-axis of each graph shows the protein concentration in relative units. The gap gene expression domains are color coded as shown. (a)–(d) show the gap domains as given by the circuit for early cycle 13, early cycle 14, middle cycle 14 and late cycle 14 respectively.

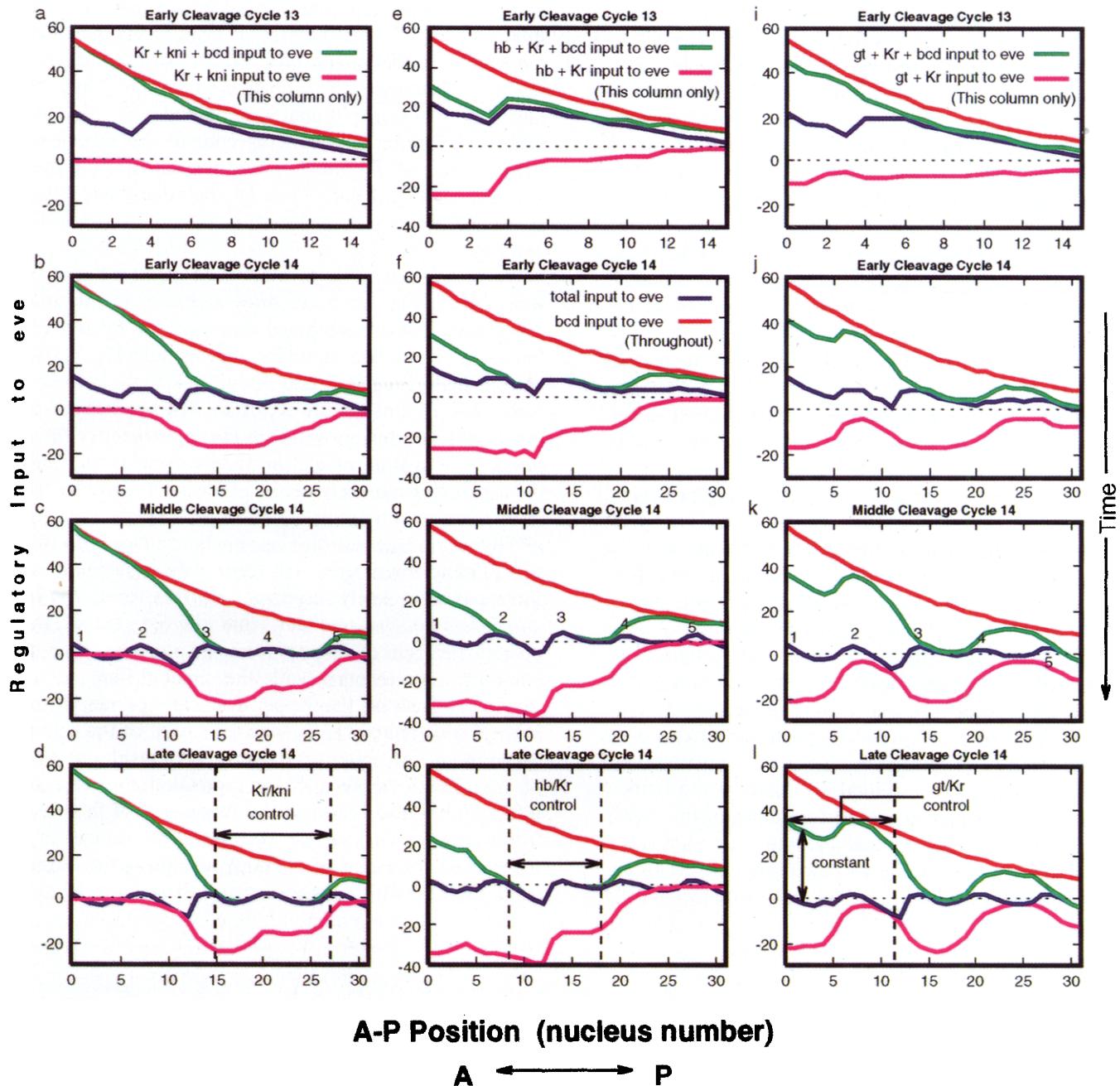


Fig. 8. Regulatory input to eve stripes 2–4. The axes are the same as in Fig. 6. Each row displays a given time. In each panel, the red line represents $m_{eve,bcd}$, and the blue line u^{eve} . In (a)–(d), the magenta line shows $T^{eve} - kni_y kni + T^{eve} - Kr_y Kr$ and the green line shows $T^{eve} - kni_y kni + T^{eve} - Kr_y Kr + m_{eve,bcd}$. In (e)–(h), the magenta line shows $T^{eve} - hb_y hb + T^{eve} - Kr_y Kr$, and the green line shows $T^{eve} - hb_y hb + T^{eve} - Kr_y Kr + m_{eve,bcd}$. In (i)–(l), the magenta line shows $T^{eve} - gt_y gt + T^{eve} - Kr_y Kr$ and the green line $T^{eve} - gt_y gt + T^{eve} - Kr_y Kr + m_{eve,bcd}$. The prepattern of the five stripes in u^{eve} is visible by middle cleavage cycle 14 and the stripes are labeled in (c), (g) and (k). For further information, see the text.

to the formation of the stripe. Similar arguments account for the formation of stripe 3 (See Fig. 8e–h). In this case, the gap genes *Kr* and *hb* combine with *bcd* to form the pattern of repression and activation leading to stripe formation. Fig. 8h also shows that the anterior border of stripe 3 is set by *hb* and its posterior border is set by *Kr*.

We now turn to stripe 2. Fig. 8i–l shows the combined repressive effects of *Kr* and *gt* on *eve*, *bcd* activation of *eve*, the sum of *bcd* activation with repression from *Kr* and *gt*, and the total activation u_i^{eve} of *eve*. Here the green curve does not coincide with the blue u_i^{eve} over the region where stripe 2 will form, since *hb* expression is constant in space in that region. Rather, the green

Table 2
The additional parameters of Equation (2) as determined from experimental data

Parameter	Symbol	Gene	Kr	hb	gt	kni	eve
Threshold (effect of general transcription factors)	h^a		-1.5	-18.	+0.59	-4.5	+4.4
Maximum synthesis rate of promoter (minutes ⁻¹)	R^a		0.49	0.96	1.1	1.1	0.80
Protein half life (minutes)	$\ln 2/\lambda^a$		19.	8.2	7.6	8.7	6.0
Diffusion operator (minutes ⁻¹)	D^a		0.18	3.7×10^{-4}	0.070	0.030	7.6×10^{-6}

curve is displaced by a constant value from u_i^{eve} from the anterior margin of the region under consideration back to the middle of the interstripe separating stripes 2 and 3. Since the contribution of *hb* is constant throughout that region, it cannot have an essential role in the formation of stripe 2 and its associated interstripes. In contrast with stripes 3–5, which are controlled by pairs of gap genes having overlapping domains, stripe 2 is under the control of *Kr* and *gt* which have cleanly separated domains. In this case, *gt* sets the anterior border and *Kr* sets the posterior border. Note that the minimum of repression and corresponding peak of u_i^{eve} that will eventually lead to the formation of stripe 2 is detectable at early cycle 14. This is because disjoint domains of gap gene expression form a minimum of repression without the need for refinement.

We emphasize that the early forming stripe 2 is controlled by gap genes with non-overlapping expression patterns, while the later appearing stripes 3–5 are controlled by pairs of genes having overlapping domains. These stripes cannot form until the expression domains have refined their spatial structure; it is this fact which accounts for the delay in formation of stripes 3–5 as compared to stripe 2.

We do not analyze stripe 1 because our data pertains only to the posterior side of this stripe and there is no reason to think that we have represented all the genes that control the formation of stripe 1.

Table 2 shows that *eve* product has very low diffusivity. This feature of the fits occurs whenever each gene product is allowed to have its own diffusivity. We have also fit to data in such a way as to constrain all proteins to have the same diffusivity. In this case, the fits exhibit one of two *eve* pathologies. In one class, one or more interstripes are missing — that is, *eve* stripes are fused. In another class, *eve* stripes are formed correctly, but the *Kr* domain is incorrectly formed, in that it appears as a rectangular rather than a bell-shaped expression domain. When *eve* stripes do form, the transient pattern is as shown above. These results indicate that *Kr* and *eve*, at least, have quite different diffusivities in the embryo.

3.3. Sensitivity studies

We have carried out many studies to explore the sensitivity of our conclusions to variation in the details of the simulated annealing procedure used to obtain the fits. The following features of the analysis are insensitive to these details. (1) The sign and relative magnitudes of T^{ab} and associated circuit parameters. (2) The observation that the circuit gives the correct transient pattern without that pattern being used in the input data. (3) The conclusions reached about the role of pairs of gap genes in forming stripes. (4) The fact that the overall timing of stripe formation depends on the increase in gap gene expression levels over cycle 14, together with their refinement. (5) The fact that the relative timing of the appearance of stripe 2 compared with that of 3–5 is a consequence of the fact that stripes 3–5 are controlled by two overlapping gap domains while stripe 2 is controlled by a pair of disjoint domains.

Limits on the search space must be imposed in order for the simulated annealing procedure to be feasible. These limits were imposed in different ways on different parameters. It turned out to be unnecessary to impose limits on the search space for D . The search space for R and λ was controlled by setting minimum and maximum values for these parameters. The search space for T^{ab} , h^a and m^a , the parameters entering u^a , was controlled by limiting the maximum value of $g(u^a)$. Use of different search spaces for R and λ , with data and other annealing conditions unchanged, did not lead to any significant changes in the results discussed above. However, the effect of *bcd* on the formation of stripes 3–5 showed a dependence on the maximum allowed value of $g(u^a)$. This dependence was as follows. As the maximum value of $g(u^a)$ was varied from 0.999 to 0.99999, the size of m^{eve} increased. This increase in m^{eve} correlated with the involvement of *bcd* in setting posterior margins of stripes at more and more posterior positions. When the maximum value of $g(u^a)$ was set to 0.999, *bcd* helped set the posterior boundary of stripe 3 but not stripe 4. When the maximum value was set to 0.9999 (in the results discussed above), there was *bcd* involvement in stripe 4. Further increase in the maximum value of

$g(u^a)$ to 0.99999 showed even more *bcd* involvement; these runs were not characterized further because the search space was so large the annealer was no longer effective. The effect of *bcd* on stripe 2 did not show this sensitivity. For these reasons, our results on the involvement of *bcd* in setting the posterior margins of stripes 3 and 4 should be regarded with some reservation. Further details are given in the Appendix.

Fits performed using only wild type data gave the same transient pattern of *eve* expression described above and had qualitatively similar circuit elements for gap auto-regulation, cross regulation and regulation of *eve* by gap genes. However, these fits gave incorrect results in the sense that *eve* itself had a prominent role in the formation of gap domains, which were profoundly altered in *eve*⁻. Moreover, there was considerably more variation in the T matrix than was the case for the results presented in the Tables. These points notwithstanding, the same features which appeared in the timing analysis above were present.

Fits were also done with *eve*⁻ data that had Eve protein levels set to zero, i.e. a deficiency for *eve*. These gave incorrect results in that they contained a substantial auto-regulatory term and stripes were completely disrupted in *eve*⁻, although gap gene expression was normal. Although *eve* auto-regulation is very important after gastrulation, no decrease of Eve protein levels in *eve*⁻ relative to wild type is seen before gastrulation, the time for which our data applies. With this exception, these results were identical to those described above.

To summarize, whenever the gene circuit forms five *eve* stripes and correct gap domains, it gives the same transient pattern, has the same controlling gap domains for each stripe and gives very low diffusivity for Eve protein. This is an inescapable consequence of three things: Spatially uniform expression of *eve* at the beginning of cycle 14; fully formed *eve* stripes at the end of cycle 14; and the relative location of *eve* stripes and gap domains.

4. Discussion

Our results bear on several topics of broad interest. In this section, we compare our conclusions to those based on studies of *eve* and other pair-rule genes using more conventional methods. This comparison shows that a variety of methods lead to a common picture of the basic regulatory actions required for stripe formation. We draw implications of our findings on the low diffusivity of Eve protein for understanding the apical localization of pair-rule message. We also relate our results to previous theoretical work on pattern formation in *Drosophila*.

4.1. General activation-local inhibition circuits

A substantial body of work suggests that the striped

pattern of *eve* forms by generalized activation combined with local repression.

The first analyses of *eve* regulation by gap genes were performed by monitoring *eve* expression in gap mutants (Frasch and Levine, 1987). There were two difficulties in interpreting these experiments. Since a gap gene expression domain is larger than a single stripe, abolishing that gap domain will affect multiple stripes. This complicates the assignment of a particular regulatory action from a given gap domain to a single stripe. Also, gap genes cross regulate (Jäckle et al., 1986; Reinitz and Levine, 1990) so that, in effect, multiple gap domains are sometimes altered in mutants for a single gap gene. Nevertheless, these experiments provide a useful guide to which stripes are *not* regulated by a given gap gene. Thus, in stripes 2–5, these results indicate that *hb* does not regulate stripe 5, *kni* does not regulate stripes 2 and 3 and *gt* does not regulate stripes 3 and 4. Our conclusions are consistent with all of these facts, which were in no way incorporated into the gene circuit beforehand.

One way to circumvent the complexity of interpreting gene expression data from mutants is to perturb the system by pharmacological means. α -amanitin can be used to halt RNA polymerase II dependent transcription, affording a means to estimate RNA lifetimes. Similarly, cycloheximide injections will halt translation. These techniques can indicate which aspects of the evolution of an expression pattern are dependent on the synthesis of new proteins. Studies of this kind were performed on *fz* (Edgar et al., 1986) and subsequent studies with cycloheximide only were performed on *eve*, *h* and *run* (Edgar et al., 1989). For *fz*, turnover of RNA was very rapid with a half-life of 7–15 min, in good agreement with the 6 minute value for the half-life of Eve protein obtained here. Cycloheximide injections up to the middle of cycle 14 caused *eve* transcripts to accumulate in interstripes, indicating that stripe formation is dependent on the synthesis of repressors at least through that time. The results reported here are consistent with this picture and extend it by identifying the repressing factors with specific gap genes.

Another approach to untangling circuitry is to use a promoter-reporter construct. This approach has two essential features. First, since the reporter gene typically codes for a bacterial enzyme (usually β -galactosidase) that does not regulate other genes, the promoter does not feed back into the regulation network as it would for an intact gene synthesizing regulatory product. Second, small fragments of the promoter can be assayed for regulatory action and these sometimes have expression patterns that are readily interpretable. The *eve* promoter has been extensively studied using this methodology. Early work demonstrated that the late autoregulation (Harding et al., 1989) and pair-rule cross-regulation (Goto et al., 1989) of *eve* was controlled by a different

region of the promoter than that which controls the establishment of stripes 2, 3 and 7. This work showed that stripes 2 and 3 were controlled by small discrete regions of the promoter, while stripe 7 was under the control of a large region that contained the discrete elements responsible for establishing stripes 2 and 3.

A particularly thorough set of investigations of this kind has established that stripe 2 is formed as a result of general activation by *bcd*, followed by repressive refinement by *gt* on the anterior edge and *Kr* on the posterior edge (Goto et al., 1989; Stanojevic et al., 1991; Small et al., 1991; Small et al., 1992). This result was established using three types of construct. The proximal 1.7 kb of the *eve* promoter together with its native TATA region direct expression that is coextensive with the native stripe 2 in late cleavage cycle 14. This stripe extends posteriorly in *Kr*⁻ mutants and anteriorly in *gt*⁻. Removal of an internal segment of about 480 bp in length abolishes stripe 2 expression (Goto et al., 1989), and an almost identical segment fused to the *eve* TATA element will direct a faithful stripe 2 by the onset of gastrulation (Small et al., 1992); this segment is referred to as the stripe 2 MSE (Minimal Stripe Element). The expression of the stripe 2 MSE differs from the proximal 1.7-kb construct only in that early expression is always posteriorly delimited at the eventual posterior limit of stripe 2, while the proximal 1.7 kb of promoter directs considerably more posterior expression, which is then restricted to form the stripe.

Within the stripe 2 MSE are binding sites for *bcd*, *hb*, *Kr* and *gt* proteins, with most sites located in clusters on either end of the element. Transcription of a construct containing the two clusters fused to the *Drosophila* hsp70 TATA region and the *E. coli* CAT gene and transfected into Schneider cells with suitable effector plasmids was repressed by Kruppel and Giant proteins, activated by Bicoid, and activated by Hunchback only in the presence of Bicoid protein (Small et al., 1991). These regulatory actions were shown to be dependent on DNA binding. Furthermore, site-directed mutagenesis of all of the *bcd* and *gt* binding sites gave essentially the

same expression pattern seen in *bcd*⁻ and *gt*⁻ embryos, respectively (Small et al., 1992). These results on gene fragments are consistent with studies of the intact gene, in which stripe 2 was altered for mutants in *gt*, *Kr* and *hb*, but not *kni* or *tll* (Frasch and Levine, 1987).

We cite these results in detail because they constitute a very important biochemical control on our circuitry results for stripe 2 (Fig. 9). The single substantive discrepancy between the results reported here and those discussed above concerns the role of *hb*, for which those authors observed a weak activating role. We note that both the biochemical and the gene circuit methods indicate that *hb* plays an auxiliary role in forming stripe 2: Biochemically, one out of 12 sites in the stripe 2 MSE bind Hunchback; moreover, in Schneider cells, Hunchback can only activate MSE 2 sequences in the presence of Bicoid, while the converse is not true. Using the gene circuit method, we find that Hunchback balances the Bicoid activation over the presumptive stripe 2 so that it will form at the proper time. In both cases, the *hb* input is not significant without *bcd*.

The biochemical analysis of the stripe 2 element is particularly significant because it assigns specific physiological roles to binding sites. The gene circuit method as formulated here cannot do that, because as yet it contains no explicit representation of promoter substructure. At the same time, biochemical dissections of promoters are hampered by their inability to reliably assay the regulation of an intact gene from a fragment of its promoter.

The difficulty is the following. The labor involved in analyzing transformant lines dictates that promoter analyses will focus on the study of contiguous fragments of DNA. Since the stripe 2 regulating sites are tightly clustered, this method of analysis leads to definitive results. Stripe 7, on the other hand, is controlled by sequences distributed over about 4.5 kb that overlap localized elements for stripes 2 and 3. One can imagine that there exist, say, 10 isolated regions of 50 bp each that mediate stripe 7 function in a concerted manner. Perhaps these sequences could be put together so as to

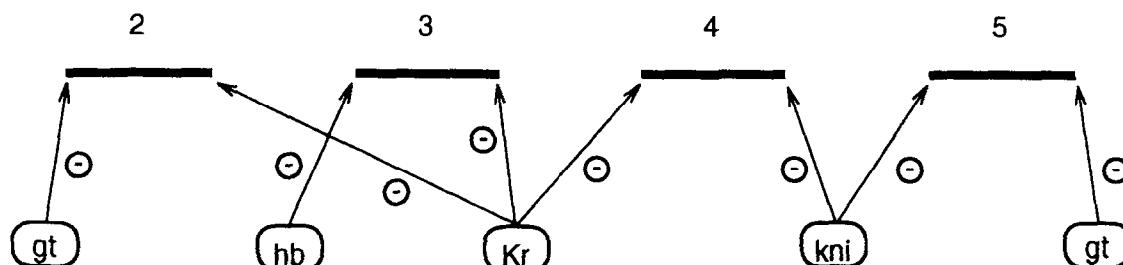


Fig. 9. Summary of gap gene control of the eight borders of *eve* stripes 2 through 5; ⊖ indicates repression.

form a stripe 7 MSE that is as well characterized as the stripe 2 MSE, but finding these dispersed sequences will be difficult and laborious using current methods.

Even with a well defined functional element in hand, it is not clear how faithful the common DNA binding assays are to *in vivo* physiology. The problem here is that *in vivo* binding to DNA occurs in the context of intact chromatin. Some proteins known to bind specifically to intact chromatin do not bind in footprinting or gel shift assays (Zink and Paro, 1989; Paro, 1990), while some sites that bind in footprinting assays have no apparent physiological activity (Jiang et al., 1991).

These considerations show the importance of characterizing the genetic circuitry at the level of the intact gene as an aid to more detailed biochemical investigations of promoter function. In essence, understanding the regulatory function of the intact gene acts as an experimental check on promoter dissections: Is the same answer obtained by both methods? For the stripe 2 MSE, the answer is yes.

In the case of stripe 3, the situation is more complex. Our results indicate that stripe 3 is delimited on the anterior side by *hb* and on the posterior side by *Kr*. There is a minimal stripe element for stripe 3 (stripe 3 MSE) which gives *lacZ* expression that is coextensive with the native expression of stripe 3 in wild type embryos (Small et al., 1993, S. Small, personal communication). The stripe 3 MSE has multiple Hunchback binding sites, which suggests that the anterior boundary of this stripe is set by *hb*. The same result was obtained in this work. Our results differ from those obtained by P-transformation with respect to the posterior delimiter of stripe 3. We found that the posterior boundary of stripe 3 was set by *Kr*. By contrast, expression of the stripe 3 MSE is unaltered in *Kr*⁻ but extends very far posteriorly in *kni*⁻ (S. Small, personal communication). Since

the expression of stripe 3 in an intact *eve* gene is unaltered in *kni*⁻ and disrupted in *Kr*⁻ (Frasch and Levine, 1987), it is likely that sequences outside the stripe 3 MSE contribute to the regulation of that stripe. The observed absence of strong *Kr* binding sites in the stripe 3 MSE (Stanojevic et al., 1991) is consistent with this picture, since the concentration of Kruppel is much higher at the posterior border of stripe 3 compared to the posterior border of stripe 2.

Our finding that stripe 4 is delimited anteriorly by *Kr* and posteriorly by *kni* and that the anterior and posterior borders of stripe 5 are delimited by *kni* and *gt*, respectively, is consistent with the observation that stripe 4 is narrower than normal in *Kr* heterozygotes, stripe 5 is narrower than normal in *kni* heterozygotes and that stripe 5 forms more slowly in *gt* heterozygotes (Frasch and Levine, 1987). It will be of great interest to see how these regulatory actions are reflected in more distal regions of the *eve* promoter.

It is remarkable that we find that *Kr* can set three separate stripe borders (two posterior, one anterior), even though the regulatory action of *Kr* on *eve* is described by a single number. This is surprising, because inspection of expression patterns in mutant embryos often seems to require different regulatory actions in different parts of the embryo. For example, in *tll*⁻, *hb* expression is increased anteriorly and decreased posteriorly, which was interpreted as separate activating and repressing actions in these two regions (Reinitz and Levine, 1990). Our work suggests that complex behavior of this type can result from combining simple interactions of several genes, rather than from complex interactions of a single gene. This can be the case despite the undeniable complexity of the molecular apparatus underlying transcriptional control. Indeed, a situation of exactly this kind is found in translation: The detailed molecular architec-

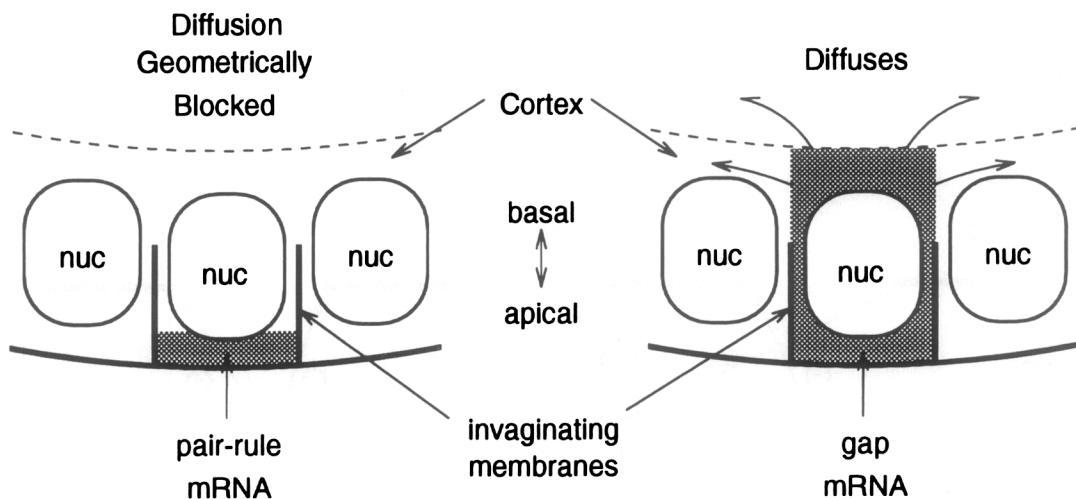


Fig. 10. Illustration of proposed mechanism linking apical localization of pair-rule message to the observed low diffusivity of Eve protein.

ture of the ribosome is highly complex but a functional description satisfactory for most purposes is given by the genetic code.

4.2. The role of diffusion: Why pair-rule message is apically localized

A striking result of our analysis is that correct formation of *eve* stripes in the context of correctly formed gap domains requires Eve protein to have extremely small diffusivity while the gap gene proteins, notably Kruppel, must have comparatively large diffusivities. This result supports a previously proposed mechanism (Edgar et al., 1987, Davis and Ish-Horowicz, 1991) to account for the observed apical localization of pair-rule message.

The mRNA of pair-rule genes, including *eve*, *ftz* and *h* is confined to cytoplasm on the apical side of each blastoderm nucleus that expresses it, while gap gene message is distributed throughout the cortex of the egg. Hence the sources of gap and pair-rule proteins are distributed differently in each energid (cytoskeletal unit in the syncytium that will become a cell). Since cell membranes invaginate in an apical → basal direction, this growing geometric obstacle will affect apically synthesized pair-rule proteins more than gap proteins synthesized both apically and basally (see Fig. 10). This maps to a sharp difference in diffusivity of gap and pair-rule proteins in our one dimensional circuit.

There is experimental evidence for this mechanism. It has been shown that β -galactosidase synthesized from apically localized transcripts diffuses much less than from unlocalized transcripts (Davis and Ish-Horowicz, 1991). It was conjectured that apical localization was required for stripe formation because it prevented pair-rule protein from diffusing. The causal link could not be determined experimentally, however, because the sequences that control localization also control message lifetime. Our result that very low diffusivity of Eve protein is required for proper stripe formation provides that causal link. This is an important example of a problem in developmental biology that can best be solved by numerical methods in conjunction with experiment.

4.3. Dynamics of expression is functionally important

The finding that progressive refinement of gap domains during cycle 14 is essential for proper function highlights an important feature of the hierarchical relation between gap and pair-rule genes. Maternal genes act earlier than other segmentation genes simply because they are transcribed from the maternal genome; segment polarity genes are most downstream because their transcripts are not detected until the onset of gastrulation. In these cases, the hierarchical relationship is set from outside the segmentation gene system. In contrast, the hierarchical relationship between the gap and pair-rule genes is set entirely by their regulatory relationships to one another.

Our work suggests that in order to understand the specification of segments in the blastoderm, it is not sufficient to consider a 'snapshot' of the blastoderm at one particular time. The gap gene patterns are dynamic, changing throughout the blastoderm period and after the onset of gastrulation. Changes in the gap expression pattern control the sequence of formation of *eve* stripes. This shows the importance of careful mapping of expression patterns not only in space but also in time. Furthermore, each pair-rule gene has a characteristic transient pattern, which may give important information about how the expression of that gene is controlled. For example, while *eve* stripes 5 and 6 are the last stripes to form, *ftz* stripe 5 is the second stripe seen (after 1) (Karr and Kornberg, 1989).

4.4. Implications for pattern formation theories

We next discuss the relationship of our work to previous studies of pattern formation in *Drosophila*. Such studies have their origin in the seminal paper of Turing (Turing, 1952).

Turing proposed a state description approach to the analysis of development. He considered 'masses of tissues which are not growing, but within which certain substances are reacting chemically and through which they are diffusing' (ibid, p. 38). This system was modeled by a set of reaction-diffusion equations. Our interphase equations (2) are similar to Turing's equations, generalized to include N reacting species, nuclear divisions, degradation of proteins and with explicit expressions substituted for Turing's general reaction functions f and g .

Turing was chiefly concerned with how chemical patterns could arise from a spatially uniform initial state. He was able to treat this question using analytic methods, since a linear approximation to his equations can be used close to a homogeneous state. Turing himself was well aware that the important biological problems to be solved did not involve the formation of pattern from a homogeneous initial state: 'Most of the organism, most of the time, is developing from one pattern into another, rather than from homogeneity into a pattern.' (ibid, pp. 71–72).

In the absence of critical information regarding the identity of the actual morphogens, a tradition arose of treating the simplest possible system that could give rise to a particular pattern. For example, in a reaction-diffusion system composed of two substances, one substance (the 'activator') is assumed to diffuse slowly and to exert an autocatalytic effect on its own synthesis. A second substance (the 'inhibitor') diffuses faster than the activator and inhibits production of the activator. Such systems can give rise to spots or stripes (Meinhardt, 1982, for review). The stripes produced are often irregular in shape, but they can be stabilized in somewhat more complex models by various methods (Lacalli et al.,

1988; Meinhardt et al., 1982; Hunding et al., 1990; Lacalli et al., 1990; Meinhardt et al., 1986; Meinhardt et al., 1988; Hunding et al., 1990; Lacalli et al., 1990; Nagorecka et al., 1988). Other workers have used mathematical results from dynamical systems theory to show that idealized gene expression patterns can be regarded as a generic series of bifurcations of standing waves (Goodwin and Kauffman, 1990).

This work has been useful in illustrating different ways in which patterns can be generated. The fact that generic processes can produce observed patterns in a stable and precise manner is of particular importance. However, the same features of the model that lead to generic results are a drawback in relating them to specific experiments. The dynamical equations for these models have been selected to be mathematically tractable; they were not chosen to represent the behavior of specific gene products. Moreover, all of these models must assume something about the regulatory interactions between gene products; this information is put in by hand. However, these regulatory interactions are in fact not known experimentally; finding them is a fundamental objective of much work on gene regulation.

Previous work using an approach similar to our own considered a very specific model of pair-rule gene interactions using both the concentrations of protein products and mRNA as state variables (Edgar et al., 1989). It was shown in this model that schematic pair-rule genes could amplify a periodic positional cue to form a striped pattern that becomes stable against short pulses of gene product. The periodic gap gene cue postulated by these authors is analogous to the periodic pattern of *eve* input (u^{eve} ; see Fig. 8) that we observed. We have extended their results by showing in detail how and when this periodic cue is generated by specific gap genes.

An important conclusion that emerges from our results is that a local activation/long range inhibition model is far too restrictive to describe real developmental systems. In the circuit reported in this work, *Eve* protein does not diffuse but is inhibited by diffusible gap gene products. This much is similar to the local activation/long range inhibition model. However, prior to gastrulation, *eve* is activated by products distributed even more widely than its inhibitors, rather than autocatalytically. Similarly, it is a fundamental postulate of local activation/long range inhibition models that differential diffusion is required to form pattern. We have shown elsewhere that differential diffusion is not required to form gap gene patterns (Reinitz et al., 1994). It is not that the mechanisms postulated in local activation/long range inhibition models are wrong. In fact autocatalysis, differential diffusion and long range inhibition have all been observed in various circuits which we have constructed from experimental data. One does not, however, necessarily see all of these mechanisms occurring together, nor do we find that

morphogens always come in autocatalytic/inhibitory pairs. Which mechanisms operate in a specific case must be determined experimentally, rather than postulated ab initio.

4.5. Limitations and future prospects

There are presently three main limitations of the gene circuit method. The first is that this method depends critically on the quality and scope of the experimental input data, which is currently not entirely adequate for quantitative modeling purposes. The second limitation is that not everything is represented in the model that should be. A third limitation is that computational expense restricts the scope of feasible numerical investigations. These limitations can be addressed by suitable improvements in data quality, computational power and theoretical analysis.

For example, limits on the search space for simulated annealing (Section 3.3 and Appendix) are due to computational expense. Limitations from this source and from inadequate data are reflected in the fact that the gene circuit did not correctly describe all the features of the transients in the *eve* expression pattern. Stripes 1 and 2 formed first in the circuit, as is observed in embryos, but we did not see the full transition pattern in which stripe 3 forms next, followed in a short time by stripe 4. This transition pattern changes quite rapidly and presumably depends on fine details of the refinement kinetics of gap domains. These details may well not be present in the gap gene expression data we used, which refers to just three distinct times. Additional errors may have made in assigning a developmental stage to particular embryos. These problems can be overcome with better data. With regard to computational resolution, we note that Equation (2) was integrated using 80-s time steps. This time resolution may be insufficient to capture all the significant features of the rapidly changing transient pattern.

Difficulties stemming from all three sources are indicated by the fact that the gene circuit fails to give correct results for expression patterns of zygotic gap mutants if wild type data alone is used to determine the circuit parameters (Reinitz and Sharp, unpublished observations). This problem could result from the uncertainty in absolute scale of the visually estimated expression data, which leads to a corresponding uncertainty in R^a (Section 2.4). For a system of two genes that repress one another, so that one is always active and the other repressed, which gene ends up being active is sensitive to the ratio of the maximum synthesis rates R of each gene (Reinitz and Vaisnys, 1990). The gap genes repress one another. When one gap domain is removed by mutation, others will compete to be expressed in that region and the final expression pattern will depend on the precise ratio of synthesis rate constants R for the respective genes. It could also be that expression pat-

terns from some gap mutants must be included in the dataset in order to predict the expression patterns of others which are not included. Recall that in this study it was necessary to use experimental data on *eve*⁻ in order to obtain a well defined circuit. Finally, it is possible that there is an error in the way that a mutant gene is represented in the circuit. These questions are under investigation.

Key projects for the future are to obtain fully quantitative data by digital methods, to include more genes in the model and to find a way to represent promoter substructure in the model. If the latter can be done, it will provide a way to incorporate the extremely important experimental results linking promoter sub-structure to the expression patterns of intact genes and thereby to address some critical questions concerning them. Our approach analyses a gene circuit from the top down, while studies of enhancers and binding sites analyze a circuit from the bottom up. An exciting task for the future is to bridge this gap. The correspondence previously noted between the regulatory relations we have found and those found by biochemical methods give hope that this will be possible.

5. Appendix

Here we give a more technical description of the simulated annealing method. We first discuss the annealing algorithm itself, which is independent of the particular problem to be solved. We next discuss two problem-specific features of the implementation: the construction of a cost function and the formulation of a move generation strategy. Finally, we discuss the performance of the algorithm as implemented for the gene circuitry problem.

5.1. The annealing schedule

Simulated annealing is derived from statistical mechanics (Metropolis et al., 1953), where it models the slow cooling (annealing) of a physical system to its lowest energy state; later it was generalized by Kirkpatrick (Kirkpatrick et al., 1983). In the following, the function to be minimized (the ‘cost function’) is $E = f(x_1, \dots, x_i, \dots, x_n)$ and T is a parameter (the ‘temperature’) that starts off large and slowly gets smaller.

The simulated annealing procedure is as follows. Let x_i , $i = 1, \dots, n$ be a set of random numbers. Then

1. Compute $E = E_{\text{old}}$ from the variables x_i .
2. Make a change in one (or more) of the x_i . (This is referred to as a ‘move’).
3. Compute $E = E_{\text{new}}$ from the newly generated set of x_i .
4. Compute $\exp\left(\frac{E_{\text{old}} - E_{\text{new}}}{T}\right)$

5. If the above quantity is bigger than a random number between zero and one, keep the new x_i ’s (‘accept the move’). Otherwise, restore the old x_i ’s (‘reject the move’).
6. Repeat while allowing T to decrease slowly from a large value to zero.

This sequence of steps will converge to the global minimum of E as long as T decreases sufficiently slowly. In practice, it is desirable to lower T as fast as possible while still obtaining a sufficiently accurate answer. Good performance of the algorithm is also dependent on the strategy of move generation. The formula for decreasing T is called an ‘annealing schedule’; we use the Lam schedule (Lam and Delosme, 1988a; Lam and Delosme, 1988b). The Lam schedule is accurate and rapid, as it decreases the temperature after every move. It is given by:

$$\begin{aligned} s_{n+1} &= s_n + \lambda \left(\frac{1}{\sigma(s_n)} \right) \left(\frac{1}{s_n^2 \sigma^2(s_n)} \right) \\ &\times \left(\frac{4\rho_0(s_n)(1 - \rho_0(s_n))^2}{(2 - \rho_0(s_n))^2} \right) \end{aligned} \quad (3)$$

where $s_n = 1/T_n$ and T_n is the temperature at the n th evaluation of E . $\sigma(s_n)$ is the variance of E at this step and $\rho_0(s_n)$ is the acceptance ratio; that is, the ratio of accepted to proposed moves. The four factors play the following roles.

λ is a quality factor. Making λ smaller increases the quality of the answer but also increases the computation time. We used $\lambda = 0.0001$ in the studies reported here. [Note that the quality factor λ should not be confused with the decay rate parameter λ^a which occurs in equation (2).]

$(1/\sigma(s_n))$ measures how close the system is to thermal equilibrium. One of the features of the Lam algorithm is a set of statistical estimators, derived from control theory, for $\sigma(s_n)$ and the average energy $E(s_n)$. These estimators require two ‘memory length’ parameters known as lambdamemv (for $E(s_n)$) and lambdamemv (for $\sigma(s_n)$) that set the time scale for taking quasi-equilibrium statistics while the temperature is changing. Lambdamemu was set to 0.2; lambdamemv was set to 100.

$\left(\frac{1}{s_n^2 \sigma^2(s_n)}\right)$ is the inverse of the specific heat.

$\left(\frac{4\rho_0(s_n)(1 - \rho_0(s_n))^2}{(2 - \rho_0(s_n))^2}\right)$ is equal to $\frac{\rho_2}{2}$,

where ρ_2 is the variance of the average energy change during a move. It is a measure of how effectively the state space is sampled and is at a maximum value when $\rho_0 \approx 0.44$. We maintain ρ_0 close to this value by adaptively changing the size of moves (see below).

Annealing runs were started with random parameter values. To erase any dependence on the initial state, the cost function was evaluated 100 000 times at a starting temperature of 1000. Then all statistics were discarded and the function was evaluated another 100 000 times to collect initial statistics, whereupon cooling began. During cooling, the temperature was lowered every 10 steps, and the statistics which determine the cooling rate were recalculated every 100 iterations. When the system changed its energy by a factor of less than 10^{-5} it was deemed frozen; after being found in the frozen state, five times the final values of the parameters were printed, terminating the run. Final temperatures were about 10^{-3} . A single simulated annealing run entailed 10^6 – 10^7 evaluations of the cost function.

5.2. The cost function

Our cost function is given by

$$E = \sum_{\text{all } a, i, t \text{ and geno-} \atop \text{types for which data} \atop \text{exists}} (v_i^a(t)_{\text{model}} - v_i^a(t)_{\text{data}})^2 + (\text{penalty terms}) \quad (4)$$

The first term is the sum of the squared deviation of the solutions to (2) for each time, gene product, genotype and nucleus for which data exists. The penalty terms Π are

$$E_{\text{penalty}} = \Pi_{R^a} + \Pi_{\lambda^a} + \Pi_{u^a},$$

where Π_{R^a} and Π_{λ^a} represent search space limits for R^a and λ^a , respectively. When these search space limits were exceeded, the move was rejected without integrating the equations, so this term may be thought of as being zero within the search space and infinity outside it. The third penalty term is given by

$$\Pi_{u^a} = \begin{cases} \exp(\Lambda (\Sigma_{ab} (T^{ab} v_{max}^b)^2 + (m^a v_{max}^{bcd})^2 \\ + (h^a)^2)) - 1 & \text{iff } \Lambda (\Sigma_{ab} (T^{ab} v_{max}^b)^2 + (m^a v_{max}^{bcd})^2 \\ + (h^a)^2) > 0 \\ 0 \text{ otherwise} \end{cases} \quad (5)$$

Here v_{max}^b and v_{max}^{bcd} are the largest values of v for gene a and v^{bcd} found in the dataset. Λ controls the size of

the search space for terms that contribute to u^a , since a very large penalty will arise whenever the argument of the exponential is greater than $1/\Lambda$. Thus the effect of the penalty on u^a is to limit the maximum saturation of u to $(1 - \Lambda)$.

5.3. Move generation strategy

The move generation strategy is based on that used (Lam and Delosme, 1986b) for the Traveling Salesman Problem. Let x_i be a parameter that is being adjusted and let ξ be a random number between zero and one. Then a move consists of changing x_i to x_i^{new} by setting

$$x_i^{new} = x_i \pm \bar{\theta}^i \ln \xi, \quad (6)$$

where the sign is chosen randomly and $\bar{\theta}^i$ is a scaling parameter that is dynamically adjusted to keep the acceptance ratio close to 0.44. This move generation procedure amounts to selecting the move from an exponential distribution with mean $\bar{\theta}^i$. In general, increasing the size of moves lowers the acceptance ratio while decreasing the size of moves increases it. As annealing progresses, moves are made on each variable in turn. A set of moves that attempt to change each variable once is called a ‘sweep’. Every 100 sweeps, the full set of $\bar{\theta}^i$ is updated using

$$\ln \bar{\theta}_{new}^i = \ln \bar{\theta}^i + 3 (\rho_0 - 0.44), \quad (7)$$

where ρ_0 is the observed acceptance ratio and the constant 3 was found by numerical experiment. Independent scaling of move size for each parameter was essential for proper performance of the algorithm; these often differed by two orders of magnitude during the course of the run.

4.4. Performance and reliability of the algorithm

The reliability of the simulated annealing procedure was assessed in three ways. First, the annealing procedure as described above was tested on a small system consisting of two genes. Parameters were randomly assigned, the equations were solved and the solutions at a few timesteps were used as synthetic ‘data’ for the annealing procedure. In this system, the root mean square (rms) deviation of the solution was $< 10^{-3}$ and the original parameters were recovered to an accuracy of 0.5–3%. For annealing runs on biological data, accuracy in finding the global minimum was assessed by checking that multiple runs resulted in nearly the same minimum scores and closely similar values for the parameters. (See Table 3). The accuracy was also checked using the parameters determined by a fit to generate synthetic data and then fitting to this dataset. These controls gave rms deviations of about 0.1 and parameters

Table 3
Degree of uniqueness in parameter determination

Parameter	Run number		
	1	2	10
rms deviation (cost)	1.348	1.352	1.347
$T^{Kr} - Kr$	0.4491	0.2665	0.3377
$T^{Kr} - hb$	-1.345	-1.485	-0.8580
$T^{Kr} - gt$	-0.3577	-0.5250	-0.4425
$T^{Kr} - kni$	-2.052	-1.745	-1.585
$T^{Kr} - eve$	0.00645	0.00633	0.00550
$T^{hb} - Kr$	0.06908	0.07881	0.01269
$T^{hb} - hb$	0.1708	0.1916	0.1116
$T^{hb} - gt$	-0.7278	-0.7175	-0.8970
$T^{hb} - kni$	0.09466	0.01679	-0.2270
$T^{hb} - eve$	0.00101	0.00178	0.00225
$T^{gt} - Kr$	-2.404	-2.596	-2.550
$T^{gt} - hb$	-0.3075	-0.2749	-0.2771
$T^{gt} - gt$	0.1742	0.1191	0.1525
$T^{gt} - kni$	-0.2710	-0.2485	-0.2440
$T^{gt} - eve$	0.00059	0.00193	0.00122
$T^{kni} - Kr$	-0.2653	-0.2651	0.3308
$T^{kni} - hb$	-5.788	-4.771	-4.307
$T^{kni} - gt$	-1.339	-1.511	-1.580
$T^{kni} - kni$	-0.0333	-0.0426	-0.0756
$T^{kni} - eve$	0.00170	0.00425	0.00320
$T^{eve} - Kr$	-2.177	-2.222	-2.259
$T^{eve} - hb$	-2.969	-3.007	-2.998
$T^{eve} - gt$	-1.780	-1.813	-1.871
$T^{eve} - kni$	-1.300	-1.335	-1.369
$T^{eve} - eve$	0.03495	0.03735	0.03869
m^{Kr}	2.720	3.203	1.814
m^{hb}	6.292	6.156	7.561
m^{gt}	1.226	1.177	1.075
m^{kni}	-1.117	-1.837	-1.768
m^{eve}	14.36	14.50	14.58
h^{Kr}	-2.401	-2.707	-1.456
h^{hb}	-15.53	-15.38	-17.71
h^{gt}	-0.9440	-0.7101	-0.589
h^{kni}	3.361	4.361	4.526
h^{eve}	-4.752	-4.667	-4.422
D^{Kr}	0.1819	0.1939	0.1755
D^{hb}	0.00831	0.01817	3.73×10^{-4}
D^{gt}	0.08234	0.06181	0.06918
D^{kni}	0.01970	0.02974	0.0340
D^{eve}	5.43×10^{-6}	2.44×10^{-5}	7.59×10^{-6}
R^{Kr}	0.4803	0.4847	0.4886
R^{hb}	1.122	1.123	0.9612
R^{gt}	1.087	1.093	1.125
R^{kni}	1.123	1.065	1.108
R^{eve}	0.8101	0.8128	0.7963
$\ln 2/\lambda^{Kr}$	17.93	17.99	17.98
$\ln 2/\lambda^{hb}$	6.992	6.962	8.215
$\ln 2/\lambda^{gt}$	7.939	7.964	7.607
$\ln 2/\lambda^{kni}$	8.520	8.793	8.724
$\ln 2/\lambda^{eve}$	6.007	6.001	6.001

The three runs summarized in this Table had total scores that agreed to within 1% (see line 2, rms deviation). The scatter in the parameters listed allow one to judge the degree to which various parameters are constrained by the data. Note that parameters connected with *eve* (e.g. D^{eve} and m^{eve}) are very tightly constrained, while parameters associated with the gap gene circuit (e.g. D^{hb} and $T^{kni} - Kr$) are less constrained.

were recovered with an accuracy of about 10%. The equations were integrated using the Euler method with a stepsize of 80 seconds, which led to integration errors of less than 10%, based on comparison with integration at smaller step sizes.

In order for simulated annealing to be effective, the search space must be sampled in a sufficiently dense manner. The sampling of the search space is controlled by the annealing schedule so that sampling can always be improved by reducing the quality factor λ . In practice, this is limited by available computing power. If the search space is enlarged while keeping the annealing schedule fixed, the algorithm eventually becomes unreliable. Unreliability is indicated by a failure to obtain repeatable minimum scores with similar parameters and often by poorer scores.

The procedure was most sensitive to limits on the search space for R and the protein half-life $t_{1/2}$ ($t_{1/2} = \ln 2/\lambda^a$). Search spaces were uniform across protein species. It appears that annealing reliability is dependent on the product of the ranges that R and $t_{1/2}$ are allowed to assume during annealing. This can be understood by considering a system with one gene and one nucleus where $g(u) = 1$. In that case, the protein concentration will tend toward a steady state v_0 given by

$$v_0 = R t_{1/2} / \ln 2.$$

It is likely that when the search space for R and $t_{1/2}$ is large, there are many combinations of synthesis and decay rates that can give rise to a given concentration. The work described here used a search space for R between 0.38 and 1.12 min $^{-1}$ and for $t_{1/2}$, between 6 and 18 min. A search space where R ranged from 0.25 to 0.75 min $^{-1}$ and $t_{1/2}$, from 8 to 24 minutes gave scores about 10% worse. Larger search spaces rendered the annealer ineffective.

The search space on u was set to three different ranges, equivalent to limiting $g(u)$ to 99.9%, 99.99% and 99.999% of its saturation value. The annealer operated reliably at the first two of these values, but not the last. Scores were improved by about 20% by enlarging the saturation limit from 99.9% to 99.99%. They improved somewhat when it was increased to 99.999%, although we are not confident that we obtained a true minimum in that case. In all cases, there was a nonzero value for Π_u^a contributing to the final score. This was apparently due to m^{eve} , which consistently grew larger as the search space was enlarged. We believe this is due to absence of a posterior cutoff for the Bicoid gradient in our dataset, equivalent to a minimum concentration below which Bicoid is ineffective. Larger values of m^{eve} , since they multiply v^{bed} , have the effect of rescaling the Bicoid gradient to make it larger in posterior regions of the blastoderm.

Acknowledgments

We thank M. Levine and S. Small for access to unpublished data, for encouragement and for stimulating conversations. S. Small made many helpful comments about the manuscript. We thank C. Alonso for help with embryos and color images, and M. Frasch for discussions and eve antiserum. JR thanks P. Miller for encouragement and valuable discussions. We thank the authors of the GNU tools, ImageMagick, xv and other shareware/public domain software used in this project. We thank Hewlett-Packard for a grant of equipment to the Yale Center for Computational Ecology, and Sun Microsystems for a grant in equipment to the Yale Center for Medical Informatics. This work was supported by grants LM 07056 and RR 07801 from the National Institutes of Health. This is publication #164 from the Brookdale Center For Molecular Biology.

References

- Akam, M. (1987) *Development* 101, 1–22.
- Carroll, S.B., Laughon, A. and S. Thalley, B. (1988) *Genes Dev.* 2, 883–890.
- Carroll, S.B. and Scott, M.P. (1986) *Cell* 45, 113–126.
- Carroll, S.B. and Vavra, S.H. (1989) *Development* 107, 673–683.
- Davis, I. and Ish-Horowicz, D. (1991) *Cell* 67, 927–940.
- Edgar, B.A., Odell, G.M. and Schubiger, G. (1987) *Genes Dev.* 1, 1226–1237.
- Edgar, B.A., Odell, G.M. and Schubiger, G. (1989) *Dev. Genet.* 10, 124–142.
- Edgar, B.A., Wier, M.P., Schubiger, G. and Kornberg, T. (1986) *Cell* 47, 747–754.
- Eldon, E.D. and Pirrotta, V. (1991) *Development* 111, 367–378.
- Foe, V.A. and Alberts, B.M. (1983) *J. Cell Sci.* 61, 31–70.
- Frasch, M., Hoey, T., Rushlow, C., Doyle, H.J. and Levine, M. (1987) *EMBO J.* 6, 749–759.
- Frasch, M. and Levine, M. (1987) *Genes Dev.* 1, 981–995.
- Goodwin, B. and Kauffman, S. (1990) *J. theor. Biol.* 144, 303–319.
- Goto, T., MacDonald, P. and Maniatis, T. (1989) *Cell* 57, 413–422.
- Harding, K., Hoey, T., Warrior, R. and Levine, M. (1989) *EMBO J.* 8, 1205–1212.
- Harding, K., Rushlow, C., Doyle, H.J., Hoey, T. and Levine, M. (1986) *Science* 233, 953–959.
- Hunding, A., Kauffman, S.A. and Goodwin, B.C. (1990) *J. theor. Biol.* 145, 369–384.
- Ingham, P.W. (1988) *Nature* 335, 25–34.
- Jackle, H., Tautz, D., Schuh, R., Seifert, E. and Lehmann, R. (1986) *Nature* 324, 668–670.
- Jiang, J., Hoey, T. and Levine, M. (1991) *Genes Dev.* 5, 265–272.
- Karr, T.L. and Kornberg, T.B. (1989) *Development* 106, 95–103.
- Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. (1983) *Science* 220, 671–680.
- Kraut, R. and Levine, M. (1991a) *Development* 111, 601–609.
- Kraut, R. and Levine, M. (1991b) *Development* 111, 611–621.
- Lacalli, T.C. (1990) *J. theor. Biol.* 144, 171–194.
- Lacalli, T.C., Wilkinson, D.A. and Harrison, L.G. (1988) *Development* 104, 105–113.
- Lam, J. and Delosme, J.M. (1988a) Technical Report 8816. Yale Electrical Engineering Department, New Haven, CT.
- Lam, J. and Delosme, J.M. (1988b) Technical Report 8817. Yale Electrical Engineering Department, New Haven, CT.
- Lehmann, R. and Nusslein-Volhard, C. (1987) *Dev. Biol.* 119, 402–417.
- Lindenmayer, A. (1968) *J. theor. Biol.* 18, 280–315.
- Macdonald, P.M., Ingham, P. and Struhl, G. (1986) *Cell* 47, 721–734.
- Meinhardt, H. (1982) *Models of Biological Pattern Formation*. Academic Press, New York.
- Meinhardt, H. (1986) *J. Cell Sci. Suppl.* 4, 357–381.
- Meinhardt, H. (1988) *Development* 104 Suppl., 95–110.
- Merrill, P.T., Sweeton, D. and Wieschaus, E. (1988) *Development* 104, 495–509.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M.N., Teller, A. and Teller, E. (1953) *J. Chem. Phys.* 21, 1087–1092.
- Mjolsness, E., Sharp, D.H. and Reinitz, J. (1991) *J. theor. Biol.* 152, 429–453.
- Mohler, J., Eldon, E.D. and Pirotta, V. (1989) *EMBO J.* 8, 1539–1548.
- Nagorcka, B.N. (1988) *J. theor. Biol.* 132, 277–306.
- Pankratz, M.J., Hoch, M., Seifert, E. and Jackle, H. (1989) *Nature* 341, 337–340.
- Paro, R. (1990) *Trends Genet.* 6, 416–420.
- Petschek, J.P., Perrimon, N. and Mahowald, A.P. (1987) *Dev. Biol.* 119, 175–189.
- Reinitz, J. and Levine, M. (1990) *Dev. Biol.* 140, 57–72.
- Reinitz, J., Mjolsness, E. and Sharp, D.H. (1994) *J. Exp. Zool.*, in press.
- Reinitz, J. and Vaisnys, J.R. (1990) *J. theor. Biol.* 140, 57–72.
- Simcox, A.A. and Sang, J.H. (1983) *Devl. Bio.* 97, 212–221.
- Small, S., Arnosti, D.N. and Levine, M. (1993) *Development* 119, 767–772.
- Small, S., Blair, A. and Levine, M. (1992) *EMBO J.* 11, 4047–4057.
- Small, S., Kraut, R., Hoey, T., Warrior, R. and Levine, M. (1991) *Genes Dev.* 5, 827–839.
- Stanojevic, D., Hoey, T. and Levine, M. (1989) *Nature* 341, 331–335.
- Stanojevic, D., Small, S. and Levine, M. (1991) *Science* 254, 1385–1387.
- Strecker, T.R., Merriam, J.R. and Lengyel, J.A. (1988) *Development* 102, 721–734.
- Tautz, D., Lehmann, R., Schnurch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K. and Jackle, H. (1987) *Nature* 327, 383–389.
- Turing, A.M. (1952) *Transactions R. Soc. Lond. Series B* 237, 37–72.
- Wieschaus, E., Nusslein-Volhard, C. and Jurgens, G. (1984) *Wilhelm Roux's Arch. Dev. Biol.* 193, 296–307.
- Zink, B. and Paro, R. (1989) *Nature* 337, 468–471.