Canalization of Gap Gene Expression During Early Development in *Drosophila melanogaster*

A Dissertation Presented
by

Manu

to
The Graduate School
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University
December 2007

Stony Brook University
The Graduate School

Manu

We, the dissertation committee for the above candidate for the Doctor of Philosophy degree, hereby recommend acceptance of this dissertation.

Professor John Reinitz, Advisor
Department of Applied Mathematics & Statistics

Professor Alan Tucker, Committee Chairperson
Department of Applied Mathematics & Statistics

Professor Ram P. Srivastav, Committee Member
Department of Applied Mathematics & Statistics

Professor Jin Wang, Committee Member
Department of Applied Mathematics & Statistics

Professor J. Peter Gergen, Outside Committee Member
Department of Biochemistry & Cell Biology

This dissertation is accepted by the Graduate School.

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Canalization of Gap Gene Expression During Early Development in *Drosophila melanogaster*

by

Manu

Doctor of Philosophy
in
Applied Mathematics and Statistics
Stony Brook University
2007

The process of animal development is stable with respect to genotypic and environmental variation. This stability property was first described by C. H. Waddington, who characterized it in terms of a metaphor of canalized flow on the epigenetic landscape of an animal. Recent quantitative studies permit the analysis of canalization at the molecular level in certain systems, where it manifests itself as a reduction in the variation of gene expression over time. In particular, canalization is evident during the segment determination process of *Drosophila melanogaster*, during which gap genes form precisely positioned expression domains controlled by maternal factors and gap gene cross regulation. These studies have shown that besides reduction in variation over time, gap gene expression patterns also have much lower positional variance than Bicoid, a morphogenetic gradient active in the embryo.

This dissertation presents a theoretical and experimental analysis of the origin of canalization in the gap gene system. The theoretical analysis was performed using the method of gene circuits, which permits the representation of gene networks as dynamical systems which reproduce gene expression data with high fidelity. Despite biophysical evidence of the importance of protein synthesis delay in gene expression, I establish that ordinary differential equations are sufficient for implementing a gene circuit that describes the dynamics of the gap gene system.

I further demonstrate that such circuits correctly predict the observed variance of the gap gene borders in the presence of the much larger variance of the Bicoid gradient. Analysis of the regulation of the gap genes in these circuits leads to the prediction that the canalization of Bicoid variation results from gap gene cross-regulation. This prediction was confirmed experimentally.

These circuits also reproduce the reduction in variance of gap gene expression over time. This property of the gene circuit is characterized further by studying the qualitative dynamics of the phase space of the gap gene dynamical system. The dynamical analysis demonstrated that the embryo can be divided into two regions with very different qualitative properties. In the anterior, gap gene expression states are determined and canalized by point attractors. In the posterior, states are determined and canalized by a one-dimensional attracting manifold. Gap gene borders can form by one of the following mechanisms: 1) the movement of an attractor through the phase space; 2) an initial state crossing a boundary between basins of attraction; or 3) reaching different regions of the attracting one-dimensional manifold.

These results imply that stable steady states are not necessary to provide canalization, and furthermore that descriptions which only involve stable steady states are insufficient for capturing the dynamical behavior that occurs in actual developmental systems. These results provide a precise mathematical description of canalization in a specific biological system which is fully supported by quantitative molecular data.

Contents

1 Introduction **1**
- 1.1 Canalization, epigenetic landscape, and structural stability . . 3
  - 1.1.1 Experimental evidence for canalization and genetic assimilation 3
  - 1.1.2 The epigenetic landscape and chreods 4
  - 1.1.3 Structural stability and genotypic canalization 6
  - 1.1.4 Attractors and stability of developmental trajectories . 7
  - 1.1.5 Catastrophes, differentiation and pattern formation . . 8
- 1.2 Segment determination in *Drosophila* 9
- 1.3 Evidence for canalization during segmentation 14
- 1.4 Gene circuits 17

2 Gene circuits and their numerical treatment **20**
- 2.1 Gene Circuits 20
  - 2.1.1 Gap gene circuits 20
  - 2.1.2 Equations 23
  - 2.1.3 Maternal systems in the gene circuit 26
  - 2.1.4 Time classes 26
  - 2.1.5 Numerical implementation of time varying inputs . . . 26
  - 2.1.6 Selection of a representative Bcd profile 27
- 2.2 Numerical treatment of the interphase and mitosis rules 29
  - 2.2.1 Numerical solution of Ordinary Differential Equations . 31
  - 2.2.2 Numerical Stability and Stiffness 38
  - 2.2.3 Stiffness of gene circuits and numerical methods for their solution 41
  - 2.2.4 Solver comparison 42
- 2.3 Gene circuits with production delays 45
  - 2.3.1 Need for a delay model 46
  - 2.3.2 Delay model 48

2.3.3 Numerical Integration of gene circuits with delays . . . 50
2.3.4 Results 56

3 Reduction of Bcd variation by the gap gene system 58
3.1 Optimization and selection of a gap gene circuit 59
3.2 Simulating Bcd variation 62
3.3 Regulatory analysis of gap gene borders 64

4 *Kr/kni* double mutants 77
4.1 Double mutant experiment 78
4.1.1 The cross 78
4.1.2 Staining scheme and confocal imaging 80
4.1.3 Data processing and boundary position measurement . 81
4.2 Results 83

5 Qualitative Dynamics 85
5.1 Diffusionless approximation and positional information 86
5.2 Setting up the analysis 89
5.3 Qualitative dynamics 91
5.3.1 Equilibria and other invariant sets 91
5.3.2 Continuous analysis 95
5.4 Dynamical structure of the phase space 96
5.4.1 Anterior Regime 35% to 53% EL 98
5.4.2 Posterior Region 53% to 72% EL 104
5.5 Stability of trajectories 110
5.5.1 Calculating volumes in time 112
5.5.2 Reduction of initial variation 114
5.5.3 Stability of trajectories in the posterior region 115

6 Conclusions 118
6.1 Reduction of the variance of the *hb* border 119
6.1.1 Antibody data, dosage, and mutants 119
6.1.2 *In vivo* Bcd data 121
6.1.3 Reporter constructs 129
6.1.4 Theories for the accuracy of *hb* 131
6.1.5 Summary 133
6.2 Canalization and dynamical structure of the gap gene system . 135
6.2.1 Genotypic canalization 135
6.2.2 Structural stability 136
6.2.3 Stability of developmental trajectories 137

6.2.4 Boundary formation 141
6.3 Limitations of the gene circuit approach 142
6.3.1 Unmatched Bcd and maternal Hb gradients 142
6.3.2 Simulating mutants 144
6.3.3 Diffusion 145
6.3.4 Comparison of circuits 145
6.4 Future directions 145

A Materials and methods 149
A.1 Integrated data 149
A.1.1 Fixation, staining, and acquisition of images 150
A.1.2 Segmentation 151
A.1.3 Time classification 152
A.1.4 Background removal 153
A.1.5 Registration and averaging 153
A.2 Optimization 154
A.2.1 Cost function and search spaces for parameters 155
A.2.2 Parallel Lam Simulated Annealing 157

List of Figures

1.1 Epigenetic landscape 5
1.2 Canalization of segmentation gene expression patterns 16
1.3 Reduction of Bcd variation by *hb* 17

2.1 Division schedule and time classes 22
2.2 The relative synthesis rate function $g(u)$ 24
2.3 Initial conditions for the gene circuit 25
2.4 Interpolation from data of time-varying external inputs 28
2.5 Selection of a representative Bcd profile 30
2.6 Solver stability regions and stiffness of gene circuits 40
2.7 Instability of the numerical solution in gene circuits 43
2.8 Propagation of discontinuities in delay equations 52
2.9 Gap gene patterns in gene circuits with production delays . . 57

3.1 Gap gene expression patterns in gene circuit `hkgn58c13k1_007` 61
3.2 Simulation of Bcd variation using gene circuits 63
3.3 Regulatory analysis of the posterior border of the anterior *hb* domain 69
3.4 Regulatory analysis of the posterior border of the central *Kr* domain 70
3.5 Regulatory analysis of the anterior border of the abdominal *kni* domain 71
3.6 Regulatory analysis of the posterior border of the abdominal *kni* domain 72
3.7 Regulatory analysis of the anterior border of the posterior *gt* domain 73
3.8 Regulatory analysis of the posterior border of the posterior *gt* domain 74

4.1 Determination of domain boundary position from data 82

4.2 Variation of *hb* and *gt* expression patterns in *Kr*/*kni* double mutants 83

5.1 Anteroposterior region for qualitative dynamics analysis 90
5.2 Interpolation of T6 Cad profile for continuation analysis . . . 96
5.3 Bifurcations in the anterior region 100
5.4 All four three-dimensional projections in an anterior nucleus . 101
5.5 Bifurcations in the posterior region 104
5.6 Maternal Hb is the morphogen in the posterior region 106
5.7 All four three-dimensional projections for a posterior nucleus . 108
5.8 Shifts due to attraction by the manifold U_+^2 111
5.9 Reduction of initial variation in the diffusionless gene circuit . 115
5.10 Tolerance to variation in maternal Hb 117

6.1 Overestimation of positional variance using average concentration profiles 125
6.2 Comparison of *in vivo* Bcd variability with Hb variability . . . 127

List of Tables

2.1 Comparison of solver performance on gene circuits 44
2.2 Estimation of production delays 47
2.3 Parameter values for the BS(2,3) triple 54

3.1 Regulatory parameters of the nonautonomous gap gene circuit `hkgn58c13k1_007` 60
3.2 Kinetic parameters of the nonautonomous gap gene circuit `hkgn58c13k1_007` 61
3.3 Comparison of positional variance of six gap gene borders with positional variance of Bcd 65
3.4 Positional variation of gap gene borders in circuits produced with Bcd gradients other than the median 66

5.1 Position of gap gene boundaries in the circuit with and without diffusion in time class T8 87
5.2 Summary of all equilibria, the A–P region they exist in, and their function 92
5.3 Comparison of bifurcation parameter values determined by Newton-Raphson and continuous analysis 97
5.4 Basins of point attractors in the anterior region, and their selection by maternal Hb 103

A.1 Search Spaces for Gene Circuits 156

Acknowledgments

This dissertation has been made possible with help from a lot of people. First of all, I would like to thank my advisor, John Reinitz. He provided exceptional moral and tangible support at two critical junctures during my dissertation work. The first time was when my whole project was challenged; he pulled out all the stops and encouraged me to stay on in the lab until we met the challenge. The second time, he went to great lengths to try and secure additional funding for me. Also, I really enjoyed our wide-ranging conversations over cigarettes, in which I learnt a lot more than just science. In the future, when we will surely be outcasts as smokers, I will sorely miss doing science this way.

I am especially grateful to Alexander Spirov for letting me flesh out his discovery of variance reduction in gap gene models.

I thank the members of the Reinitz Lab without whom this project would not have been possible. I have benefited enormously from discussions with Johannes Jaeger, who also cleaned up the fly code, making it easier to modify, and bequeathed to us folders of photocopied papers that saved me many trips to the library. Carlos Alonso and Ah Ram Kim trained me at the bench and at the microscope. They, together with Hilde Janssens performed the cross from which I stained embryos later. Vitaly Gursky and Konstantin Kozlov provided the code that was used for the initial characterization of single-nucleus dynamics. Svetlana Surkova kindly let me use her figure of gap gene canalization in this dissertation. Ekaterina Myasnikova helped me eliminate the orientation of embryos as the source of correlation in Bcd parameters.

The larger group of Reinitz Lab collaborators provided invaluable help. I thank Dave Sharp for insightful discussions and starting me off on the dynamical analysis of the gap gene system. Ovidiu Radulescu allowed me to use the results of his bifurcation analysis in this dissertation.

I thank Ram Srivastav for his advice on considering approximations with reduced dimensionality. Peter Gergen helped me plan crosses and gave very detailed and helpful comments on this dissertation.

My experience as an undergrad in Ram Ramaswamy's lab at J.N.U. provided me with the background to carry out this work, and I also thank him for his support when I switched back to research. Special thanks are due to Kapilanjan Krishan, who helped me overcome a mental block, and got me started with comprehending the gap gene phase portraits.

I also thank my smoking buddy, Jon Wood, who was always willing to spare cigarettes, even if I was not in a chatty mood.

I would like to thank my sister, Mansi Srivastava, for forcefully encouraging me to return to academics. I thank my parents, Surendra and Manju Srivastava for their love and support. They encouraged me to do basic science even though their dream was for me to be a mechanical engineer at IIT, and fix their old Fiat.

Finally, I am grateful to my wife, Radha Panini, for her support and patience as I slowly made my way from one crisis to another. She commuted to Manhattan for three years and took care of the little details of life so that I could focus on this work. I also thank her for painstakingly proofreading this dissertation, for being my partner in Halo, and for indulging my technophilia generously.

Chapter 1

Introduction

Animals have the ability, within limits, to be viable even if their genotype or environment deviates from normal. This ability of animals to buffer genotypic and environmental variation is termed canalization. Canalization, so defined, is not only an empirical observation, but is also a requirement for evolution by natural selection. A population must possess a variety of traits for some among them to be selected, and its members must remain viable, despite this variation, in order to reproduce.

The phenotype of an animal is produced during its development by the interaction of its genotype and its environment. Therefore, a necessary condition for canalization is that the course of the development of an animal be stable with respect to environmental perturbations. Further, the canalization of development is a mechanism for the ability of animals to acquire characters that are adapted to environmental conditions. If an environmental stimulus produces a phenotype that has evolutionary advantage, the genotype will adjust in order to produce the new phenotype in a stable manner, and will then continue to produce it even if the environmental stimulus is removed.

Though it is clear that canalization during development is both evident and important, its origins are not so well understood. C.H. Waddington proposed that canalization could be understood in terms of the interactions between the genes of an animal (Waddington, 1942), while René Thom created a mathematical formalism for understanding it (Thom, 1975). Limited progress was made in applying these ideas to an actual biological system since neither the genes, nor their interactions were known in molecular terms.

The advances of the past three decades in molecular biology, genetics, and computing have made it possible to approach the problem of canalization in an actual developmental system. During the early embryogenesis of *Drosophila melanogaster*, the genes that set its body pattern have been very well characterized at the molecular level (see Akam, 1987, for review). The time course of the protein concentrations of these genes has been observed at a high temporal resolution (Surkova et al., 2007; Myasnikova et al., 2002). These data have revealed specific instances of canalization in the segmentation system of *Drosophila*. At the same time, combining such data with dynamical systems models (Reinitz et al., 1995; Reinitz and Sharp, 1995; Jaeger et al., 2004a) has allowed the reconstruction of the developmental trajectories *in silico* with high fidelity. The aim of this dissertation is to use techniques of non-linear dynamical analysis to better understand canalization during the early development of *Drosophila melanogaster*.

1.1 Canalization, epigenetic landscape, and structural stability

1.1.1 Experimental evidence for canalization and genetic assimilation

Waddington coined the term canalization based on the observation that in *Drosophila*, mutants have greater phenotypic variability than the wild type (Waddington, 1942). To give an example relevant to this study, larvae homozygous for an amorphic allele of the segmentation gene *Krüppel*, Kr^1, sometimes show three denticle bands, sometimes four, and one band may or may not have reverse polarity (Wieschaus et al., 1984b). Indeed almost every segmentation gene mutant (see for example, Lehmann, 1988; Schüpbach and Wieschaus, 1986) phenotype is variable.

Recent work (Rutherford and Lindquist, 1998), suggests that the phenotypic variability in mutants mirrors the underlying genotypic variability. Flies heterozygous for mutant alleles of the heat shock gene *Hsp83* were observed to have a large variety of morphological defects. In a detailed set of experiments, Rutherford and Lindquist (1998) showed that the type of defect is dependent on the genotype of the fly, and that the defect is a heritable trait. This established that the phenotypic variation observed in *Hsp83* mutants arises from genotypic variation among individuals. The effects of this genotypic variation are normally buffered in wild type flies by the activity of the Hsp90 protein encoded by *Hsp83*.

The canalization of development is also well supported by experimental evidence. A result of the earliest embryological investigations by Driesch, Spemann and others (Gilbert and Sarkar, 2000) into the nature of development is that under certain circumstances embryos can survive radical surgery. Driesch discovered that if the cells of a 4-cell-stage sea urchin embryo are separated, they will each grow into a complete larva. Thus development is regulated, capable of correcting for perturbations or irregularities.

Further, the genotype of an animal is capable of assimilating phenotypes produced by environmental stimuli (Waddington, 1942, 1959b,a) as shown by the following experiment (Waddington, 1953). *Drosophila melanogaster* flies were selected for a certain phenotype, crossveinless, that appears only when pupae are heat-shocked. After several generations of selection, the phenotype showed even without heat-shock.

1.1.2 The epigenetic landscape and chreods

Waddington invoked a metaphor to illustrate how canalization comes about (Waddington, 1966, 1968). The course of the development of an animal is represented by water flowing in a landscape of hills and valleys (see Figure 1.1). The valleys represent different cell fates. This landscape is called the "epigenetic landscape" because its structure is determined by the interactions of the genes of the animal, and it causes the epigenesis of the animal.

The flow of the water in the valleys is stable with respect to changes in the landscape, that is, with respect to the genotype. The flow is also stable if its path is perturbed by an environmental stimulus, since it would tend to return to the valley. Such a stable developmental trajectory that attracts neighboring perturbed trajectories is called a chreod.

The genetic assimilation experiment described in Section 1.1.1 can now be understood in terms of chreods. The wild type developmental trajectory of

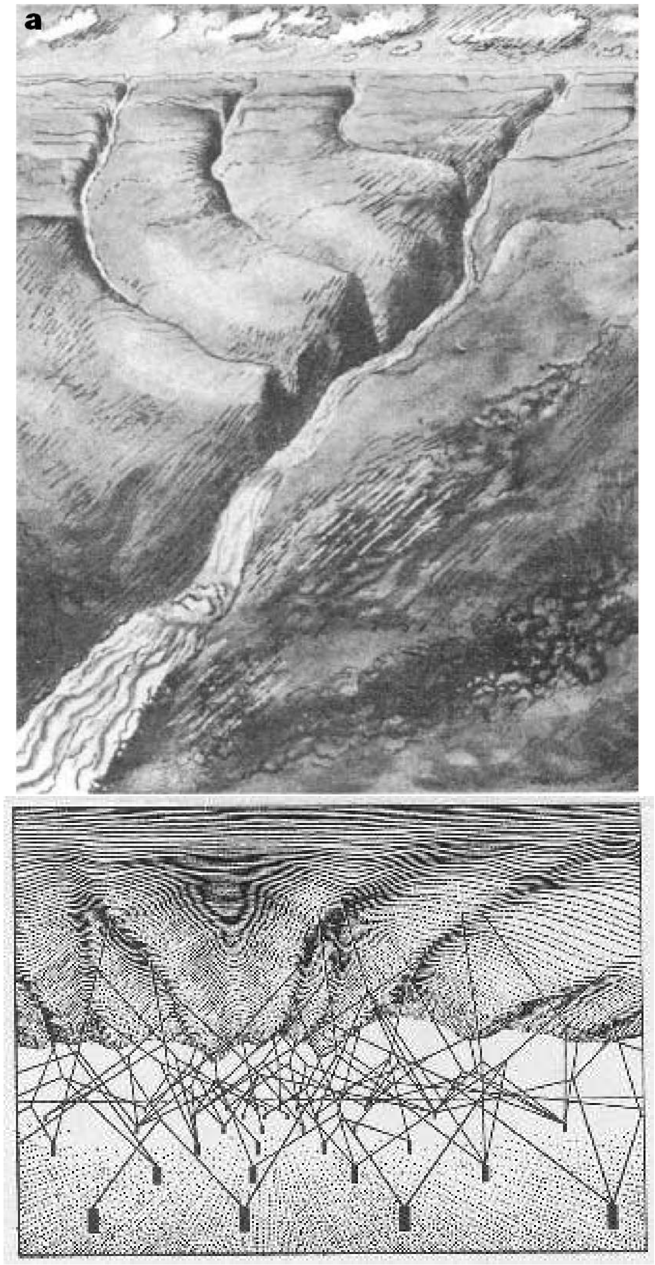

Figure 1.1: Epigenetic landscape. Above, metaphorical representation of the development of an animal by water flowing in a landscape of hills and valleys. Below, the "underbelly" of the landscape reveals that it is determined by the animal's genotype. The ropes represent its genes. Reproduced from `http://www.usc.edu/hsc/dental/odg/jaskoll01.htm`, originally from Waddington (1957).

the flies is a chreod and leads to the development of wings with a posterior crossvein. When the heat-shock is applied, the trajectory shifts and is not a chreod anymore. If the stimulus is maintained across several generations, the requirement of canalization that the developmental trajectory be a chreod leads to a shift in the epigenetic landscape such that the new trajectory is a chreod. In subsequent generations, this new trajectory will then be taken even in the absence of the stimulus.

1.1.3 Structural stability and genotypic canalization

Thom formalized the ideas presented by Waddington (Thom, 1969, 1975, 1983). The state of a cell is described by the concentrations of all its chemical constituents. Let the state variables be $c_1(x,t), c_2(x,t), \ldots, c_k(x,t)$, where there are k chemical species, space is represented by x, and time by t. Also, let the state variables $c_i(x,t)$ be governed by the system of coupled ordinary differential equations

$$\frac{\partial c_i}{\partial t} = X_i(c_1, \ldots, c_k, x, t), \tag{1.1}$$

where $i = 1, 2, \ldots, k$. The space of all possible values of the state variables R^k is called the phase space of the system. The dynamics of the cell are governed by a vector field $X(c_1, \ldots, c_k, x, t) = X_1, X_2, \ldots, X_k$ defined on R^k.

A homeomorphism between two subsets of R^k, A and B is a one-to-one and onto continuous mapping $h : A \rightarrow B$ that has an inverse $h^{-1} : B \rightarrow A$ that is also continuous (Perko, 1996).

X is considered to be structurally stable if, for every vector field X' on R^k sufficiently close to X, there exists a homeomorphism $h_{X'} : R^k \rightarrow R^k$ which transforms every trajectory of X into a trajectory of X' (Thom, 1983). In other words, the topology of the phase space does not qualitatively change in going from X to X'.

The vector field X is determined by the genotype of the animal. Therefore, if the vector field X is structurally stable, the animal is said to have genotypic canalization.

1.1.4 Attractors and stability of developmental trajectories

A necessary condition for buffering of the phenotype against genotypic and environmental variation is that the developmental trajectories are stable, that is, they are chreods (Section 1.1.2). A topological feature of vector fields, attractors, can explain how this stability arises. Attractors are defined below.

For Eq. (1.1), the orbit is defined as the function $\phi(\cdot, x) : R \to R^k$, where $x \in R^k$. In other words, the orbit is the set of points that can be reached from the starting point x.

A subset E of $F \subset R^k$ is called dense in F if every point of F is a limit point of E, or a point of E, or both.

For the vector field X, an attractor A is defined as follows (Thom, 1983).

- A_l maps to itself under X, that is, it is invariant.
- Almost all orbits that start in A_l are dense in A.
- There exists a system of neighborhoods U_i of A, such that all trajectories starting in U_i asymptotically approach A; and if a point $u \in U_i$ meets A as $t \to -\infty$, then u must lie on A.

The union of neighborhoods U_i is called the basin of attraction of A.

Since trajectories in the basin of an attractor of A asymptotically approach A, the final state is stable with respect to small perturbations, and the trajectories are said to canalize. If there is more than one attractor in the phase space, the attractors' basins are separated by surfaces called basin boundaries. Note that attractors are more general than steady states, which are just point attractors. For example, an attractor may be a closed trajectory, also called a limit cycle.

The specification of cell fate by point attractors is called homeostasis (Thom, 1983). Homeostasis as a concept has been rather popular and has been used in many theoretical models of biological systems (Kauffman, 1969; Huang et al., 2007; von Dassow et al., 2000; Umulis et al., 2006; Sánchez and Thieffry, 2001). Although particular components of the state of a cell can reach steady state, the state as a whole will not, since the cell is not immortal.

Therefore, the stability of developmental trajectories is more general than homeostasis. If the developmental trajectory itself is stable, that is, trajectories perturbed by small amount tend to return to it, this developmental trajectory is said to be homeorhetic.

1.1.5 Catastrophes, differentiation and pattern formation

The vector field $X = X(c_1, \ldots, c_k, x, t)$ of Eq. (1.1) is parameterized by space and time. Since X changes with x, t it is possible for X not to be structurally stable for certain values of x and t. These situations are called catastrophes (Thom, 1969, 1983). Catastrophes occur when the Jacobian of X is singular. In other words, at a catastrophe of the dynamical system, the topology of the phase space changes discontinuously. Specifically, the number and type of attractors of the phase space changes.

As the time parameter of X, t is varied, the catastrophes that occur describe the different attractors, or cell types, that arise as the animal develops. In other words, the catastrophes describe differentiation. As the space parameter of X, x is varied, the catastrophes describe the different cell types possible in space. That is, they describe pattern formation.

This mathematical framework captures the key biological ideas of canalization, stability of developmental trajectories, differentiation, and pattern formation. However, it has limited applicability if the state variables to describe the system are not known. I will now exhibit a specific canalizing system with observable state variables, the *Drosophila* blastoderm.

1.2 Segment determination in *Drosophila*

The body plan of the adult fruit fly *Drosophila melanogaster* consists of fourteen repeating units called segments. There are three segments in the head (mandibular, maxillary, and labial), three in the thorax (T1–T3) and eight in the abdomen (A1–A8). Each segment has a different identity which is specified by the Hox genes (Gilbert and Sarkar, 2000). The life cycle of the fly consists of an embryogenesis stage (∼1 day), a larval stage (∼4 days), a pupal stage (∼4 days), and adulthood (∼60 days) (Campos-Ortega and Hartenstein, 1985).

The segments are determined (Simcox and Sang, 1983) in the first three hours of the development of *Drosophila*. We briefly describe the events during this early stage below; for a more comprehensive description of *Drosophila* embryogenesis, see Campos-Ortega and Hartenstein (1985) and Lawrence (1992).

Drosophila melanogaster embryos are approximately ellipsoidal in shape, with an anteroposterior (adult head-to-tail) length of $500\mu m$, and a dorsoventral (adult back-to-abdomen) length of $150\mu m$. During this stage, the embryo is a syncytium, which means that nuclei are not surrounded by cell membranes. Nuclear division (cleavage) begins around 18 minutes after egg laying (Foe and Alberts, 1983), and the syncytium undergoes nine very rapid nuclear divisions. Following the ninth division, a majority of the nuclei migrate to the periphery of the embryo, leaving behind the yolk cells. The embryo is then a blastoderm, and remains in this state for four more cleavage cycles (10–14A). During the middle of cycle 14A, the cell membrane begins to invaginate; this is called cellularization. During late cycle 14A, the embryo undergoes the mid-blastula transition (Foe and Alberts, 1983; Renzis et al., 2007), when maternal mRNA and proteins are degraded, and zygotic transcription increases many fold. At the completion of cellularization (end of cycle 14A), the embryo undergoes a complex set of tissue movements, called gastrulation, that leads to the formation of the germ layers.

During this three-hour period, the segments are determined by a set of genes called the segmentation genes. These genes were discovered in mutagenesis screens for segmentation defects (Nüsslein-Volhard and Wieschaus, 1980; Nüsslein-Volhard et al., 1984; Jürgens et al., 1984; Wieschaus et al., 1984a; Schüpbach and Wieschaus, 1986; Nüsslein-Volhard et al., 1987). The earliest morphological defects are seen only 15 minutes after gastrulation (Nüsslein-Volhard and Wieschaus, 1980), therefore the steps of the segment determination process before gastrulation largely occur at the molecular level. The mRNA transcripts and proteins produced by these genes have been molecularly characterized and visualized. Based on their genetic interactions, these genes can be organized into a causal hierarchy (see Akam, 1987; Ingham, 1988, for review). The genes at lower levels of this hierarchy are expressed at finer spatial resolution along the anteroposterior axis of the embryo. The proteins of the maternal coordinate genes form monotonic gradients from mRNA deposited in the egg by the mother. The gap genes are expressed in broad overlapping domains. The pair-rule genes are expressed in overlapping stripes with double segment periodicity. The segment-polarity genes are expressed in the germ band after gastrulation and form the segment prepattern.

There are three groups of maternal genes (Schüpbach and Wieschaus, 1986). The anterior group specifies the formation of the head and thorax, the posterior group specifies the abdomen, and the terminal group specifies the terminal regions. The anterior system acts primarily through *bicoid* (*bcd*), and is independent of the posterior system and the terminal system in the presumptive germ band of the embryo (Driever and Nüsslein-Volhard, 1988a). *bcd* mRNA is deposited at the anterior tip of the embryo by the mother, and starts to be translated soon after egg laying, forming an exponential profile from anterior to posterior (Driever and Nüsslein-Volhard, 1988b). *bcd* sets up another maternal gradient in the syncytium, *caudal* (*cad*, Mlodzik et al., 1985), by repressing its translation (Rivera-Pomar et al., 1996). *cad* is expressed both maternally and zygotically, and embryos lacking both maternal and zygotic gene function show segment deletions like the gap genes (Macdonald and Struhl, 1986). Another gene with both maternal and zygotic activity is the gap gene *hunchback* (*hb*, Tautz et al., 1987; Tautz, 1988). The posterior maternal system acts through *hb* by repressing the translation of its uniformly distributed maternal mRNA (Irish et al., 1989; Lehmann and Nüsslein-Volhard, 1991). The terminal system acts through two genes with gap gene-like activity, *tailless* (*tll*) and *huckebein* (*hkb*) (Casanova, 1990; Reinitz and Levine, 1990; Duffy and Perrimon, 1994).

The maternal systems establish the expression of the gap genes, which comprise the next level in the segmentation gene hierarchy. This dissertation focusses on the gap genes expressed in the presumptive germ band, *hb*, *Kr*, *gt*, and *kni*. These genes start to be zygotically expressed during the cleavage cycles 10–12 (see Jaeger et al., 2007, and refs. therein), and reach maximum expression during the middle of cycle 14A (Surkova et al., 2007). Their expression patterns are very dynamic, since they are established in 60 mins, and also shift with time (Jaeger et al., 2004b). All of them encode transcription factors (Tautz et al., 1987; Redemann et al., 1988; Nauber et al., 1988; Mohler et al., 1989), that is, their proteins regulate the expression of other genes by binding to DNA. They are transcriptionally regulated by *bcd* (Driever and Nüsslein-Volhard, 1989; Rivera-Pomar et al., 1995; Hoch et al., 1991; Eldon and Pirrotta, 1991), *hb* (Tautz, 1988; Hoch et al., 1991), and *cad* (Rivera-Pomar et al., 1995; Schulz and Tautz, 1995). *bcd* and *cad* act as transcriptional activators, that is, they promote the transcription of a gene when bound to its DNA control region (Driever and Nüsslein-Volhard, 1989; Rivera-Pomar et al., 1995). *hb* can be a repressor (Tautz, 1988), that is, it can inhibit the transcription of a target, or an activator, by acting together with *bcd* (Simpson-Brose et al., 1994). Also, based on their expression patterns in mutants, it was established that the gap genes cross-regulate each other (Gaul et al., 1987; Tautz, 1988; Eldon and Pirrotta, 1991; Kraut and Levine, 1991a; Clyde et al., 2003).

The monotonic maternal gradients of Bcd, Hb, and Cad specify the positions of gap gene domains, and hence, through the segmentation gene hierarchy, ultimately specify the positions of the segment prepattern. Positional information is a hypothesis for the mechanism by which maternal gradients might specify position in the embryo. The hypothesis is that a monotonic gradient, termed a morphogen, which is set up by production at a source, diffusion, and degradation at a sink, instructs patterns to form at particular values of its concentration (Wolpert, 1968). The following experiments seem to confirm that Bcd and maternal Hb provide positional information in the embryo, and are morphogens. First, if the protein concentrations of *bcd* and maternal *hb* are varied by changing the number of their copies (or dosage) in the maternal genome, the downstream gene patterns (gap, pair-rule) shift along the A–P axis (Driever and Nüsslein-Volhard, 1988a; Struhl et al., 1992; Houchmandzadeh et al., 2002). Second, in mutants for all three maternal systems, gap genes are expressed uniformly, that is, no pattern is formed (Struhl et al., 1992).

With the availability of higher-resolution quantitative data (Surkova et al., 2007), and a careful study of Bcd dosage in a large population of embryos (Houchmandzadeh et al., 2002), this statement of positional information had to be revised. First, the positions of gap gene domains and pair-rule stripes in the posterior half of the embryo changes with time, which cannot be explained by threshold-based pattern formation (Jaeger et al., 2004b). Second, though gap genes domains shift with *bcd* dosage, the magnitude of the shifts does not follow a concentration threshold of Bcd (Driever and Nüsslein-Volhard, 1988a; Houchmandzadeh et al., 2002; Bergmann et al., 2007).

The nature of the revision (Jaeger et al., 2004b) was the incorporation of the role of gap gene cross regulation and dynamics in specifying positional information. Using quantitative models that faithfully reproduce the wild type gene expression patterns (Jaeger et al., 2004a; Perkins et al., 2006), it was found that the gaps genes are activated by the maternal gradients of *bcd* and *cad*, with mutually-repressive interactions between them setting their boundaries. With these interactions, the model is able to capture the temporal shifts in the gap gene domains (Jaeger et al., 2004b). It was deduced from the analysis of the model that asymmetric repression between the gap genes is responsible for the anterior shifts of domains. Thus positions of target gene borders are specified not only by the maternal information, but by the internal dynamics of the targets themselves.

1.3 Evidence for canalization during segmentation

This section presents experimental evidence that the segmentation gene system is a model for the study of developmental canalization. Antibodies that allow the visualization of all the major segmentation proteins have been developed (Kosman et al., 1998). Using these antibodies, an atlas of segmentation gene expression patterns with high temporal resolution has been built (Surkova et al., 2007). This atlas of segmentation gene expression provides us with snapshots of the developmental trajectory of a *Drosophila* embryo during the syncytial blastoderm stage. This trajectory eventually leads to the specification of the segment prepattern, and its canalization is therefore biologically significant.

Data from individual embryos show that variation in gap gene expression levels reduces over time, and by gastrulation the patterns are very reproducible from embryo to embryo (see Surkova et al., 2007; Jaeger et al., 2007, and Figure 1.2). Variation in border positions of gap gene domains also show a similar reduction. The effect is even more striking in pair-rule patterns. Unlike gap genes patterns, which vary in expression level but have the same shapes, individual pair-rule patterns have variation in the order in which stripes resolve. By gastrulation, though, pair-rule patterns also have high reproducibility in shape, levels, and border positions from embryo to embryo. These data thus hint at the possibility that the developmental trajectories of gap and pair-rule genes are stable (as defined in Waddington, 1942).

In addition to providing evidence for stability, gene expression data also provides evidence for genotypic canalization. The wild type variation of the *bcd* gradient is much larger than that of the gap genes (see Surkova et al., 2007, and Figure 1.3). This was first demonstrated for *hb* by Houchmandzadeh et al. (2002). The variability of the Bcd gradient is due to intrinsic factors such as the localization of its mRNA (Berleth et al., 1988), diffusion and degradation rates. These intrinsic factors are determined by the genotype of the mother and the zygote. For example the localization of *bcd* mRNA to the anterior pole is controlled by other maternal genes (for example *staufen*, *exuperantia*, *swallow*, see Johnston and Nüsslein-Volhard, 1992; Frohnhöfer and Nüsslein-Volhard, 1987). Therefore, the variability of the Bcd gradient is due to genotypic variability. Houchmandzadeh et al. (2002) also showed that the Bcd gradient changes with temperature, while the border position of Hb does not. This can be regarded as an instance of environmental canalization. However, this case is not considered in this dissertation.

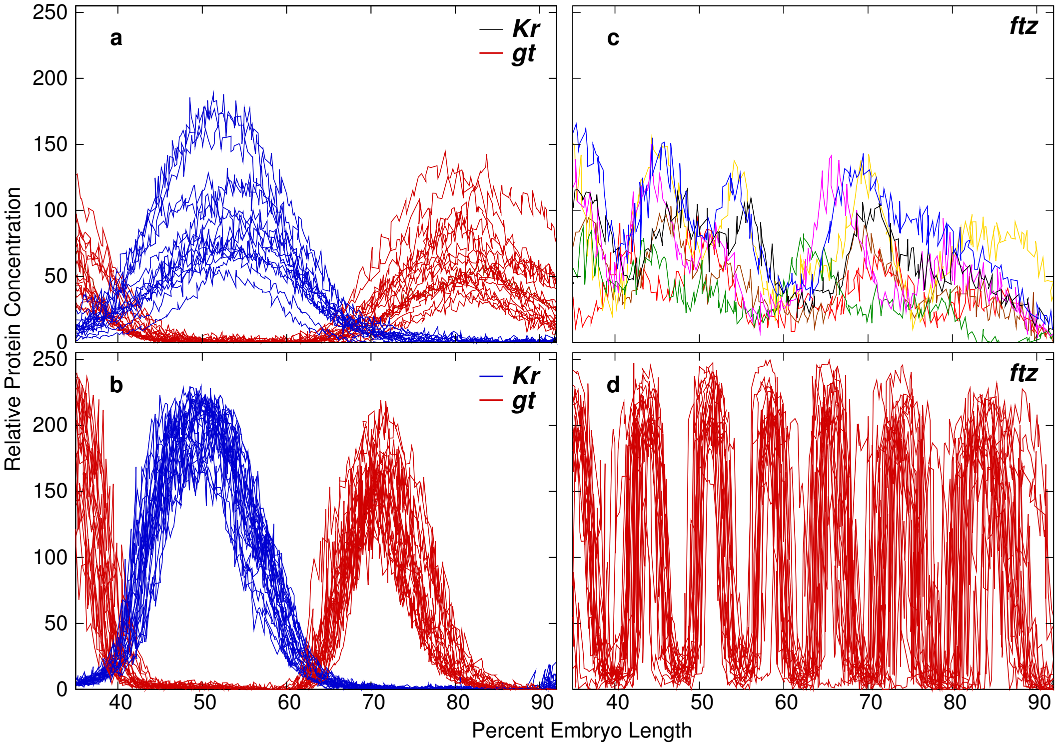

Figure 1.2: Canalization of segmentation gene expression patterns. Early (**a**) and late (**b**) *Kr* and *gt* expression patterns during cycle 14A. The variation in expression levels is much lower in late cycle 14A. Early (**c**) and late (**d**) expression patterns for the pair-rule gene *ftz*. Reproduced here with permission of S. Surkova (Surkova et al., 2007).

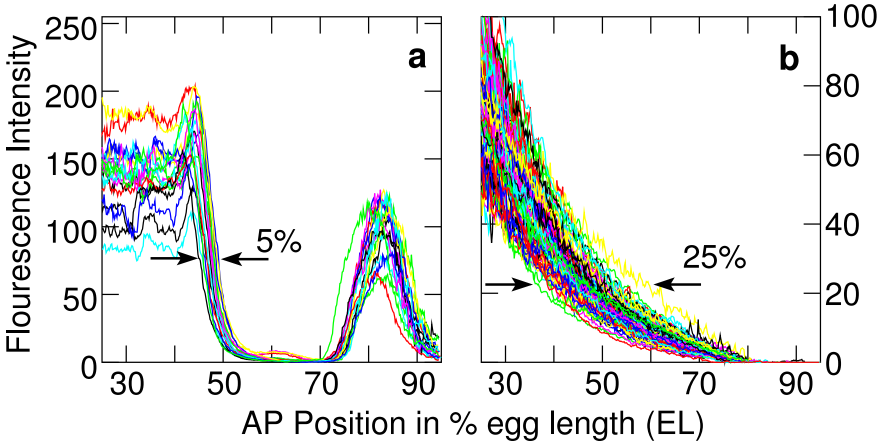

Figure 1.3: Reduction of Bcd variation by *hb*. (**a**) Wild type cleavage cycle 14A Hb expression patterns in 89 embryos. The range of border positions at half-maximum is 5% EL. (**b**) Wild type cleavage cycle 13 Bcd expression patterns in 89 embryos. The range of positions at a fixed threshold is 25% EL. All data are from FlyEx.

1.4 Gene circuits

We can better understand how segmentation genes canalize by analyzing the vector field X (Eq. 1.1), that determines the developmental trajectories. It is possible to construct a dynamical system that reproduces gene expression data (Surkova et al., 2007; Myasnikova et al., 2001) with high accuracy using the gene circuit method (Reinitz and Sharp, 1995; Reinitz et al., 1998; Jaeger et al., 2004a, 2007).

The gene circuit method consists of three steps. First, a model is formulated. Second, the model's parameters are determined by optimization. Third, the model's behavior is analyzed for biological insight.

The state variables of such a gene circuit are protein concentrations in nuclei that are arranged in a one-dimensional row in the A–P direction along the surface of the syncytium. The circuit operates according to three rules:

interphase, mitosis, and division. The division rule specifies that nuclei divide simultaneously at the end of a mitosis, and daughter nuclei inherit their states from mother nuclei. The interphase and mitosis rules use coarse-grained rate equations to specify the rate of change of the state variables during interphase and mitosis respectively.

The gene circuit's parameters are determined from gene expression data by an optimization method called Parallel Lam Simulated Annealing (PLSA) (see Kirkpatrick et al., 1983; Lam and Delosme, 1988a,b; Chu et al., 1999; Chu, 2001, and Section A.2.2). Once a circuit that reproduces gene expression data has been obtained, its behavior is analyzed to determine the dynamical properties of the gene network being modeled.

This dissertation presents a study of the canalization of four gap genes: *hb*, *Kr*, *gt*, and *kni* using the gene circuit method. In Chapter 2, a gene circuit for these genes is constructed with outside input from *bcd*, *cad*, and *tll*. Several numerical methods for the solution of the gene circuit's equations are evaluated on the criteria of accuracy, stability, and efficiency, and the Bulirsch-Stoer method is chosen from among them. A modified gene circuit that incorporates protein synthesis delays is also constructed using delay differential equations. It is shown that despite the evidence for the importance of such delays (Rothe et al., 1989), gene circuits implemented with ordinary differential equations are sufficient for describing the dynamical behavior of the gap genes.

In Chapter 3, it will be shown that the gene circuit exhibits genotypic canalization. This is done by varying the Bcd profile from data, and monitoring the resulting gap gene patterns. It is established that this property of the gap gene system arises due to epistasis between the gap genes. The experimental verification for this prediction is presented in Chapter 4.

The possibility that gap gene trajectories are stable (Section 1.3) is investigated by qualitative dynamical analysis of each nucleus in Chapter 5. It is shown that the trajectories of the gap genes are stable, and that homeostasis is not sufficient to explain this stability. This analysis also reveals that the embryo can be subdivided into an anterior, and a posterior region that show distinct dynamical behavior. In the posterior, the stability of the trajectories arises from attraction by a one-dimensional manifold. The analysis also suggests different mechanisms for gap gene border formation in dynamical terms.

The results of Chapters 3 and 5 are summarized in Chapter 6. The concepts of genotypic canalization, structural stability, homeostasis, homeorhesis are reevaluated in light of the canalizing gap gene system. The results on canalization with respect to Bcd are compared with earlier experimental and theoretical work. Appendix A presents methods that were not invented in the course of the dissertation, but are used here.

Chapter 2

Gene circuits and their numerical treatment

2.1 Gene Circuits

2.1.1 Gap gene circuits

A gene circuit (Mjolsness et al., 1991; Reinitz et al., 1995; Reinitz and Sharp, 1995; Jaeger et al., 2004b,a) determines the time evolution of protein concentrations in the syncytial blastoderm of *Drosophila melanogaster*. The circuits used in this paper comprise the gap genes *hb*, *Kr*, *gt*, and *kni* of the anteroposterior segmentation system. Their proteins are transcription factors that localize in the nuclei of the blastoderm (see Section 1.2). Therefore, the state variables are the protein concentrations of these genes inside nuclei. Anteroposterior (A–P) and dorsoventral (D–V) patterning systems are largely independent of each other in the presumptive germ band of the blastoderm embryo. This allows us to consider a one-dimensional row of nuclei along the anteroposterior axis of the embryo, from 35% EL to 92% EL. The modeled region extends over 58% EL region of the embryo, from the peak of the third *gt* stripe to the posterior border of the posterior *hb* domain.

The circuit functions according to three rules: interphase, mitosis and division. The first two rules describe continuous dynamics of proteins during interphase and mitosis. During interphase the evolution of protein concentrations is determined by three processes: regulated protein synthesis, protein transport, and protein decay. During mitosis, transcription shuts down and nascent transcripts are destroyed. Therefore, only protein transport and protein decay govern the dynamics in the mitosis rule.

The third rule, division, accounts for the cleavage of the blastoderm. It models mitotic division as a discontinuous change in the state of the system. At the end of a mitosis, each nucleus is replaced with its daughter nuclei. The inter-nuclear distance is halved and the daughter nuclei inherit the state of the mother nucleus. The divisions are carried out according to a division schedule based cleavage cycle data (Figure 2.1).

The gap gene circuits used in this study start at the onset of cleavage cycles 13 and last until gastrulation at the end of cycle 14A (Foe and Alberts, 1983). *Kr*, *gt*, and *kni* are exclusively zygotic, and their proteins first appear in early cycle 13 (Gaul and Jäckle, 1987; Eldon and Pirrotta, 1991; Kraut and Levine, 1991b; Jaeger et al., 2007; Surkova et al., 2007; Myasnikova et al., 2005). *hb*, which is both maternal and zygotic, also shows a manyfold increase in expression in cycle 13 (Houchmandzadeh et al., 2002), indicating commencement of its zygotic expression. Time t is measured in minutes from the start of cleavage cycle 13. The interphase of cycle 13 lasts for 16.0 min, and its mitosis from 16.0 to 21.1 min. At $t = 21.1$ min, the thirteenth division is carried out by

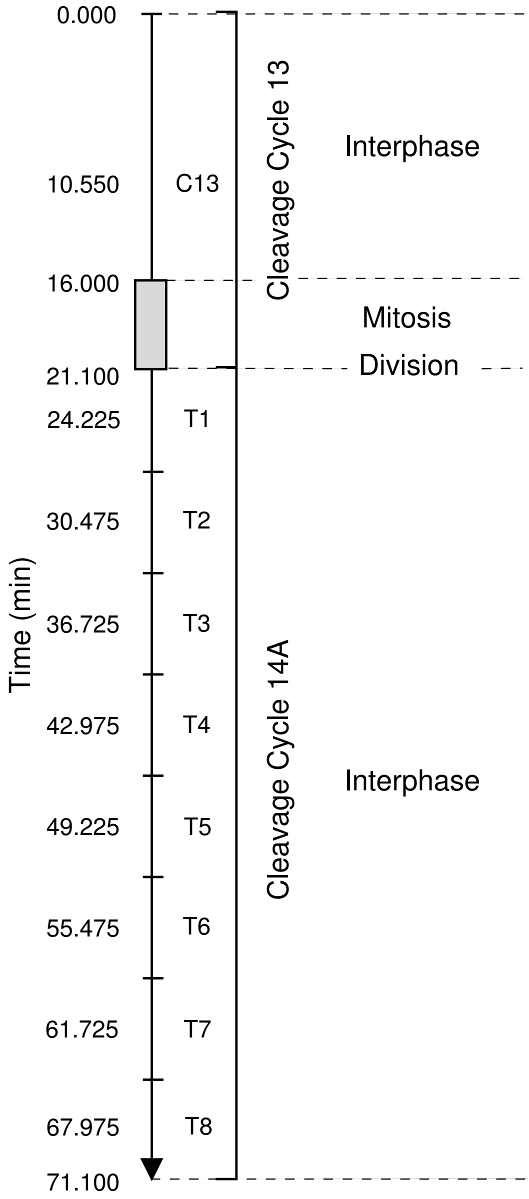

Figure 2.1: Division schedule and time classes. $t = 0$ is at the start of cleavage cycle 13. The model runs until gastrulation $t = 71.1$ min. The dashed lines demarcate time intervals in which different rules of the model apply. The intervals are labeled with rules on the right. The nine time points at which model is compared to data are indicated. There is one time point for cycle 13 (C13), and eight points for cycle 14A (time classes T1–T8).

applying the division rule. Interphase of cycle 14A starts immediately after division, and lasts until gastrulation at $t = 71.1$ min.

2.1.2 Equations

The two continuous rules, interphase and mitosis, use a system of ordinary differential equations (ODEs) to describe the dynamics of protein concentrations. Let there be M nuclei in the modeled region during a particular cleavage cycle. Let i denote a particular nucleus, counting from anterior to posterior. We denote a particular segmentation gene by $a \in 1, \ldots, N$, where N genes are represented in the circuit. $v_i^a(t)$ is the protein concentration of gene a in nucleus i. The time evolution of state variables $v_i^a(t)$ are given by the solution of the system of $M \times N$ ODEs,

$$\begin{aligned} \frac{dv_i^a}{dt} = R^a g\left(\sum_{b=1}^{N} T^{ab} v_i^b + m^a v_i^{\mathrm{Bcd}} + \sum_{\beta=1}^{N_e} E^{a\beta} v_i^{\beta}(t) + h^a\right) \\ + D^a(n)\left[(v_{i-1}^a - v_i^a) + (v_{i+1}^a - v_i^a)\right] - \lambda^a v_i^a. \end{aligned} \tag{2.1}$$

The first term on the right hand side of Eq. (2.1) represents protein synthesis, the second one protein transport through Fickian diffusion and the last term represents first-order protein degradation.

Protein synthesis rate for a gene a is determined by the maximum synthesis rate R^a and the regulatory input to a, u_a. The rate of protein synthesis is the product of the maximum synthesis rate and the regulation-expression function $g(u^a) = \frac{1}{2}\left[\left(u^a/\sqrt{(u^a)^2+1}\right)+1\right]$ (see Figure 2.2).

The regulatory input $u^a = \sum_{b=1}^{N} T^{ab} v_i^b + m^a v_i^{\mathrm{Bcd}} + \sum_{\beta=1}^{N_e} E^{a\beta} v_i^{\beta}(t) + h^a$ in turn accounts for the transcriptional regulation of gene a by transcription factors. Each term corresponds to a distinct type of factor. The first term

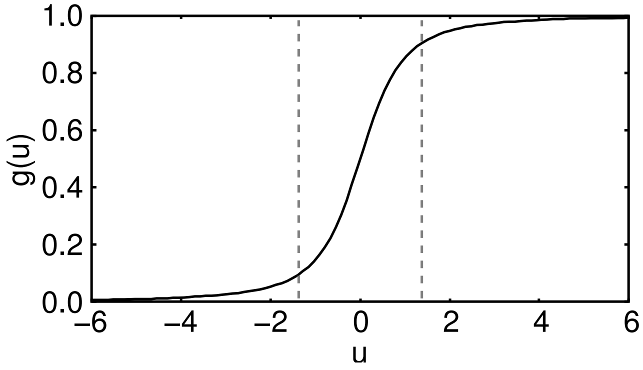

Figure 2.2: The relative synthesis rate function $g(u)$. The dashed vertical lines are at values of total input u at which synthesis rate is 10% or 90% of maximum.

represents the regulation of gene a by the genes $a = 1, \ldots, N$ of the circuit. T^{ab} characterizes the regulatory effect of protein b on the synthesis of gene a. Spatially and temporally homogeneous maternal factors are represented in u^a via its fourth term, h^a. The second term of u^a represents a maternal factor, Bcd, which is spatially inhomogeneous, but constant in time. m^a is the strength of regulation of gene a by Bcd, while v_i^{Bcd} is the concentration of Bcd in nucleus i.

Finally, we incorporate maternal or zygotic factors that are spatially inhomogeneous and change with time through the third term of u^a, $\sum_{\beta=1}^{N_e} E^{a\beta} v_i^{\beta}(t)$. N_e is the number of such factors in the circuit, $E^{a\beta}$ is the regulatory effect of external input β on gene a, and $v_i^{\beta}(t)$ is the concentration of external input β in nucleus i at time t. The gene circuit used in this study has two such factors, *cad* ($v_i^{\mathrm{C}ad}(t)$), and *tll* ($v_i^{\mathrm{T}ll}(t)$). *cad* is both maternal and zygotic. However, its expression pattern is not affected in mutants for *hb*, *Kr*, *gt* and *kni*, except for a slight expansion of the posterior stripe in *hb*$^-$ during late cycle 14 (Mlodzik and Gehring, 1987). Similarly, *tll* expression is not affected in these mutants (Brönner and Jäckle, 1991).

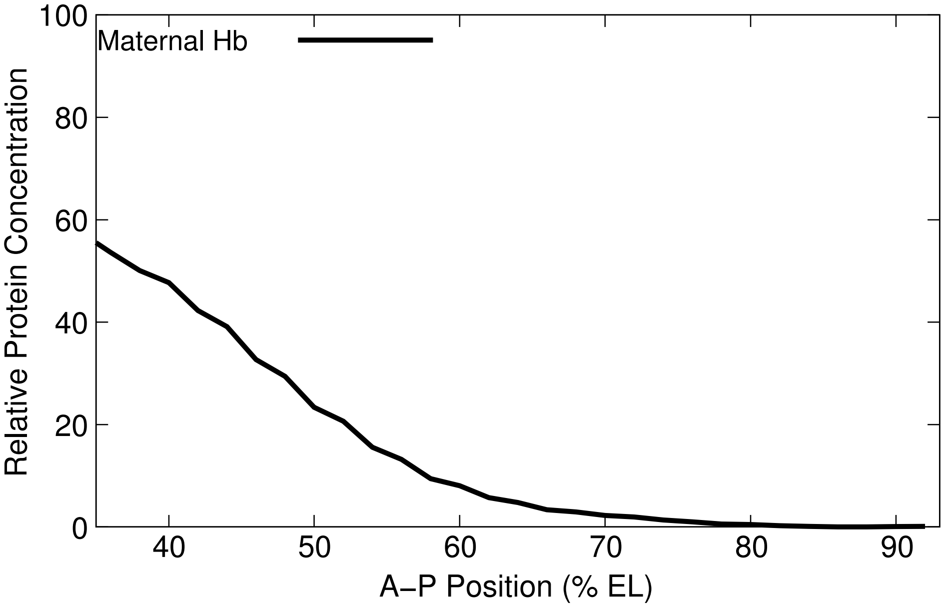

Figure 2.3: Initial conditions for the gene circuit. The only non-zero initial condition (Hb) is shown. The x axis is the modeled anteroposterior region from 35% EL to 92% EL. Kr, Gt and Kni proteins are not detected until cycle 13.

The second term of Eq. (2.1) is protein transport between nuclei represented as spatially-discretized Fickian diffusion. The diffusion parameter D^a is assumed to vary inversely with the square distance between neighboring nuclei. The third term is protein degradation, in which λ^a is the decay rate of the product of gene a. It is related to the protein half life of the product of gene a by $t^a_{1/2} = \ln 2/\lambda^a$.

Since *Kr*, *gt*, and *kni* proteins first appear only in cycle 13, they have initial conditions of zero, that is $v_i^{Kr}(0) = v_i^{gt}(0) = v_i^{kni}(0) = 0$. For *hb*, the expression data from cycle 12 is used as initial condition (Figure 2.3). These data are the maternal component of *hb* since its expression intensifies only in cycle 13.

2.1.3 Maternal systems in the gene circuit

The anterior maternal system is represented explicitly through *bcd* in the regulatory input term u^a of Eq. (2.1). The posterior system is represented through the maternal Hb gradient (Irish et al., 1989; Struhl, 1989) as an initial condition. The terminal system is represented through *tll* (Weigel et al., 1990; Brönner and Jäckle, 1991; Pignoni et al., 1992), the concentration of its protein product is included as time varying input.

2.1.4 Time classes

In order to compare with quantitative gene expression data, we calculate the solution of the model at several time points. The data are classified into one time class during cycle 13 and eight time classes during cycle 14A. Time points at half time through each class are used for the comparison: C13, $t_0 = 10.550$ min; T1, $t_1 = 24.225$ min; T2, $t_2 = 30.475$ min; T3, $t_3 = 36.725$ min; T4, $t_4 = 42.975$ min; T5, $t_5 = 49.225$ min; T6, $t_6 = 55.475$ min; T7, $t_7 = 61.725$ min; T8, $t_8 = 67.975$ min (Figure 2.1).

2.1.5 Numerical implementation of time varying inputs

In order to specify the right hand side of Eq. (2.1) fully, the concentrations of Cad and Tll $v_i^\beta(t)$ must be supplied for any time in the duration of the model. Average concentrations for these genes are known at nine time points $t_k, k = 0, \ldots, 8$. t_0 is the midpoint of cycle 13, and $t_1, \ldots, t_8$ correspond to the eight time classes in cycle 14 (Surkova et al., 2007). Thus for nucleus i and external input β, we know $v_i^\beta(t_k) = v_{i,k}^\beta$, where $k = 0, \ldots, 8$. The concentration of external input β in nucleus i is then determined at arbitrary time t by piecewise linear interpolation,

$$v_i^{\beta}(t) = \frac{(t_{k+1} - t)v_k + (t - t_k)v_{k+1}}{t_{k+1} - t_k}, \quad t_k \leq t \leq t_{k+1}.$$

Fig. 2.4 shows such interpolation at 50% EL for Cad, and at 92% EL for Tll. Higher order methods like cubic splines were not used because they gave rise to artifacts from experimental noise.

2.1.6 Selection of a representative Bcd profile

The Bcd gradient is essentially stationary during cleavage cycles 13 and 14 (Driever and Nüsslein-Volhard, 1988b; Gregor et al., 2007b; Surkova et al., 2007), and hence, v_i^{Bcd} is assumed to be constant in time. It is known through antibody studies (Driever and Nüsslein-Volhard, 1988b; Houchmandzadeh et al., 2002, 2005; Surkova et al., 2007) and a recent GFP-Bcd (Gregor et al., 2007b) study that the Bcd profile is an exponential function of A–P position x, so that

$$v^{\mathrm{B}cd}(x) = A \exp(-\lambda x). \tag{2.2}$$

The arithmetic mean of exponential curves is not exponential. Thus, it is not possible to generate a representative Bcd profile by taking an arithmetic average over embryos. Instead of averaging, a representative individual Bcd profile $v_i^{\mathrm{B}cd}$ was chosen in the following manner. Background signal was removed from Bcd profiles from 88 cycle 13 embryos as described (Myasnikova et al., 2005). Taking the logarithm of Eq. (2.2), we get

$$\log(v^{\mathrm{B}cd}(x)) = \log(A) - \lambda x. \tag{2.3}$$

The background-removed profiles were then fit by linear least-squares to Eq. (2.3).

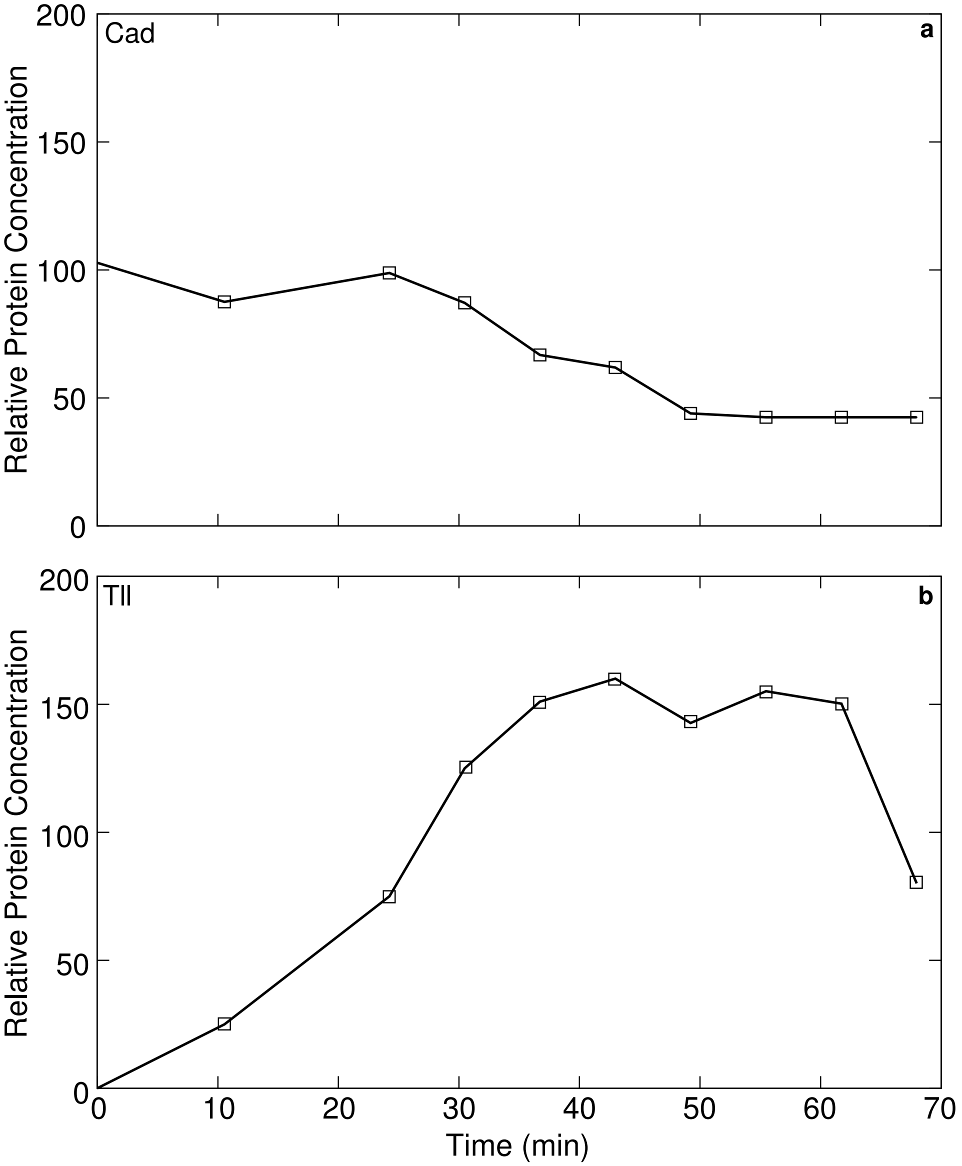

Figure 2.4: Interpolation from data of time-varying external inputs. (**a**) Cad data and linearly interpolated time profile at 50% EL. (**b**) Tll data and linearly interpolated time profile at 92% EL. Square boxes are data at ten time points. Cycle 12 data are used at $t = 0.0$. The other time points are shown in Figure 2.1. Lines are interpolated concentrations.

This procedure yielded two parameters for each profile, λ (length scale), and A (concentration scale). Figure 2.5a is a scatter plot of $\log A$ with λ. A profile was chosen such that its parameters lie in the middle of the scatter plot (Figure 2.5a). Since this study concerns only embryo-to-embryo variation and not nucleus-to-nucleus noise, the exponential fit of the profile was used in the model. Using background-removed profiles directly in the model yields circuits with the same properties as those of circuits with exponential fits.

2.2 Numerical treatment of the interphase and mitosis rules

Gene circuits have two kinds of rules, discrete and continuous (see Section 2.1.1). In our model, division is the only discrete rule. It determines the number of nuclei after a division and the protein concentrations in the nuclei. Interphase and mitosis are continuous rules which determine the time evolution of protein concentrations during interphase and mitosis. They are governed by systems of ordinary differential equations or ODEs (Eq. 2.1).

The numerical treatment of the continuous rules is important since an accurate and efficient solution is required. A numerical solution is always different from the exact solution, and their difference is the error in the numerical estimate. A solution is accurate when the error is small. It is more computationally expensive to calculate solutions with lower error. Since the continuous rules equations are solved tens of millions of times in a single simulated annealing run (see Section A.2), the numerical method has to be computationally efficient. For a method to be useful for gene circuits, it should be capable of satisfying both the accuracy and efficiency requirements simultaneously.

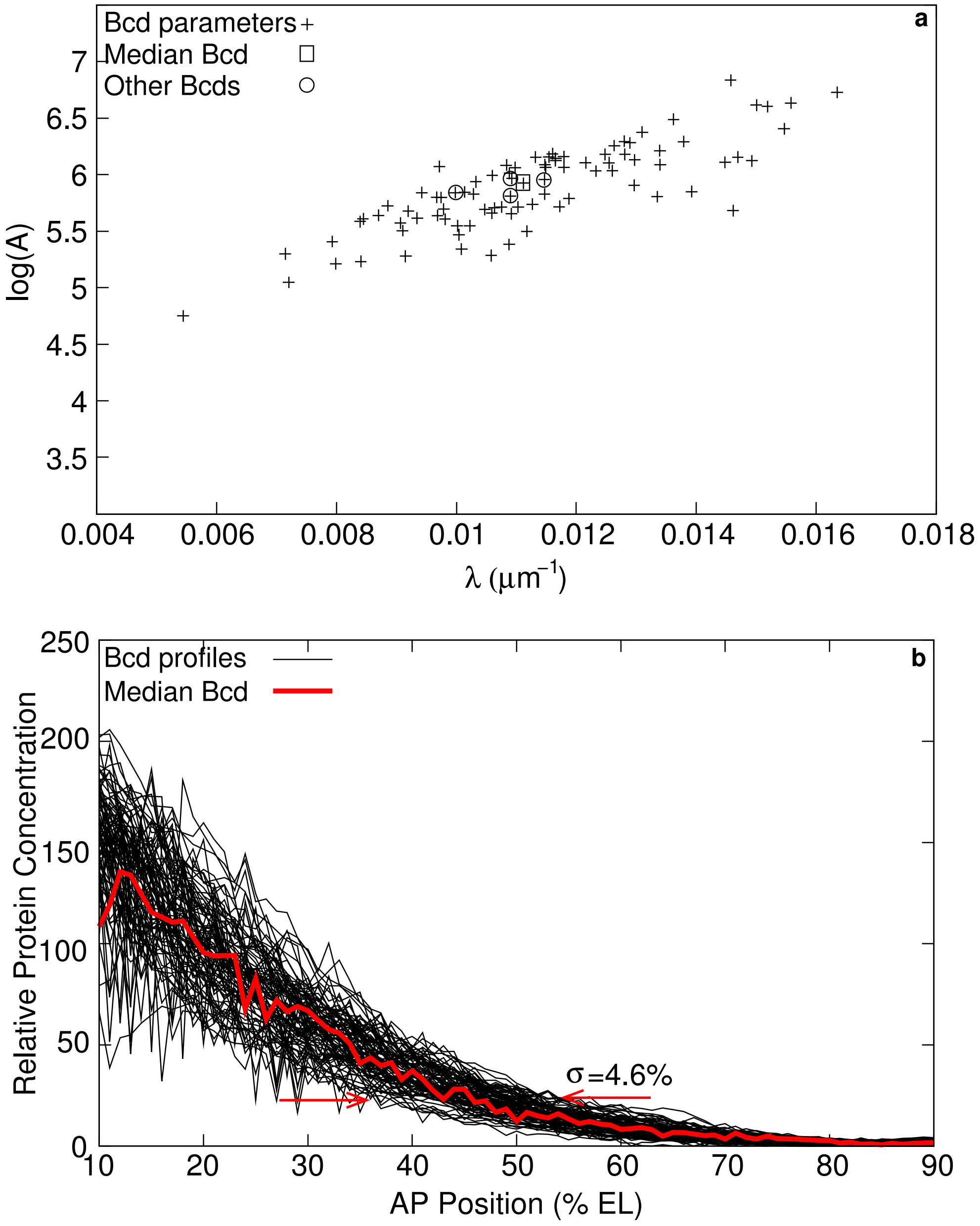

Figure 2.5: Selection of a representative Bcd profile. (**a**) Scatter plot of log amplitude ($\log A$) and slope (λ) of 88 Bcd profiles from cycle 13 embryos. The boxed profile and the circle profiles were investigated further; unless explicitly mentioned analysis in this dissertation use the boxed profile. (**b**) An overlay of all 88 Bcd profiles used in the simulations. The median profile is highlighted in red. The threshold concentration at which Hb border forms in the gene circuit was determined from the median profile. The positions at which these 88 profiles cross that threshold has a range of 20.6% EL, and standard deviation (σ) of 4.6% EL.

Depending on the underlying problem, it is possible for the error in the numerical solutions of systems of ODEs to grow exponentially. A numerical method is said to be unstable in such a situation. Instability is especially likely to occur when solving systems of ODEs whose solutions are comprised of fast- and slow-changing components. Such equations are said to be stiff. Furthermore, the optimization of a gene circuit's parameters by simulated annealing introduces another level of complication. Since the parameters are chosen by random moves, extremely stiff equations are possible.

Therefore, the ideal numerical method has to be accurate, computationally efficient, and stable.

This section begins with a brief review of numerical methods for solving ODEs. I discuss the stability properties of different solvers and study their suitability for a stiff systems of ODEs. I then look at gene circuits and show that this system of equations is stiff. As a result of this stiffness we seek a computationally efficient and stable solver. We compare the performance of four different solvers: fourth order Runge-Kutta, Adams-Bashforth, Bulirsch-Stoer and Bader-Deuflhard. Our results show that the Bulirsch-Stoer solver is the best compromise between stability and computational efficiency.

2.2.1 Numerical solution of Ordinary Differential Equations

Consider the system of first order ordinary differential equations (ODEs)

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}, t), \tag{2.4}$$

where the dependent variable $\mathbf{y}(t)$ is an N-dimensional vector, and the independent variable t is scalar. $\mathbf{y}'$ is the derivative of $\mathbf{y}(t)$ with respect to t. $\mathbf{f}(\mathbf{y}, t)$ is an N-dimensional vector of functions that take $N + 1$ arguments. Given the initial condition $\mathbf{y}(0) = \mathbf{y}_0$, we seek the solution $\mathbf{y}(t)$ on an interval $[0, T]$. Due to the fundamental existence theorem (Hochstadt, 1975), Eq. (2.4) has a unique solution in some domain of $\mathbf{y}$-space provided $\mathbf{f}(\mathbf{y}, t)$ is bounded and Lipschitz in that domain.

Numerical methods for the solution of Eq. (2.4) can be categorized in two ways. The first is according to the number of values of $\mathbf{y}$ used in calculating the solution; one-step methods require only one, while multi-step ones require several. The second categorization is based on whether the equation that determines the approximate solution is explicit (of form $\mathbf{y} = \mathbf{g}(t)$), or implicit (of form $\mathbf{y} = \mathbf{g}(\mathbf{y}, t)$). Implicit methods are in general more stable than explicit methods (Conte and de Boor, 1980; Press et al., 1992; Bulirsch and Stoer, 1992). A brief review of these methods presented below. Without loss of generality, a scalar ODE is considered for all subsequent discussion.

Explicit one-step methods

Given the solution at t, $y(t)$, a one-step numerical method determines an approximation to the solution at $t + h$ through the truncated Taylor series

$$y(t+h) = y(t) + hf(y,t) + \frac{h^2}{2!}f'(y,t) + \ldots + \frac{h^k}{k!}f^{(k)}(y,t). \qquad (2.5)$$

Here, $f^{(k)}(y, t)$ denotes the k^{th} total derivative of $f(y, t)$ with respect to t. The local discretization error is the omitted remainder of the Taylor series expansion. This error is of the order of the first omitted term, or $\mathcal{O}(h^{k+1})$. To keep this error small, h needs to be kept small. The solution is thus calculated in a series of steps $n = 0, 1, \ldots$, such that the n^{th} step advances the solution at

t_n (y_n) to the one at t_{n+1} (y_{n+1}). The simplest such rule is Euler's rule (Conte and de Boor, 1980), which simply truncates the Taylor's series at the second order term. It has local discretization error of $\mathcal{O}(h^2)$.

Any higher-order rule made directly from the Taylor series (Eq. 2.5) is not general. This is because it requires the calculation of higher-order derivatives of f explicitly for each equation. To generalize the rule, multiple function evaluations of $f(y,t)$ are made in $[t_n, t_{n+1}]$, and a formula is constructed such that it mirrors the Taylor expansion until the k^{th} order term. Such a formula gives rise to a $\mathcal{O}(h^k)$ method. For example, a second-order method may be constructed from the general formula that requires two function evaluations,

$$y_{n+1} = y_n + h(a_1 f(y_n, t_n) + a_2 f(y_n + p_1 h f(y_n, t_n), t_n + p_2 h)). \tag{2.6}$$

The special case, $a_1 = 0, a_2 = 1, p_1 = \frac{1}{2}, p_2 = \frac{1}{2}$ gives rise to the modified Euler, or the mid-point method,

$$y_{n+1} = y_n + hf\left(y_n + \frac{h}{2} f(y_n, t_n), t_n + \frac{h}{2}\right). \tag{2.7}$$

The fourth order Runge-Kutta (Bulirsch and Stoer, 1992; Press et al., 1992) method (RK4) requires four function evaluations, and is given by

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4), \tag{2.8}$$

where

$$
\begin{aligned}
k_1 &= hf(y_n, t_n), \\
k_2 &= hf\left(y_n + \frac{k_1}{2}, t_n + \frac{h}{2}\right), \\
k_3 &= hf\left(y_n + \frac{k_2}{2}, t_n + \frac{h}{2}\right), \\
k_4 &= hf(y_n + k_3, t_n + h).
\end{aligned}
\tag{2.9}
$$

Practical implementations of these methods do not use a fixed step-size (h). Rather h is adaptively changed according to estimates of the local discretization error. There are two main advantages of one-step methods. First, they do not make any assumptions about the solution, apart from the existence of the k^{th} derivative for a method of $\mathcal{O}(h^k)$ (Conte and de Boor, 1980). Second, even if there is a discontinuity, Runge-Kutta methods make errors near it, but the solution far away is still accurate (Bulirsch and Stoer, 1992).

Explicit Multi-Step methods

Multi-step methods are derived from the integral representation of the solution of the differential equation

$$\int_{t_n}^{t_{n+1}} y' dt = \int_{t_n}^{t_{n+1}} f(y(t), t) dt,$$

that is,

$$y_{n+1} = y_n + \int_{t_n}^{t_{n+1}} f(y(t), t) dt.$$

In order to evaluate this integral, $f(y, t)$ is approximated by polynomial interpolation. For example, by the Newton backward differences interpolation formula (Conte and de Boor, 1980), we get the fourth order Adams-Bashforth method

$$y_{n+1} = y_n + \frac{h}{24}(55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3}), \tag{2.10}$$

where $f_n = f(y_n, t_n)$. This method requires knowledge of the solution at four points. Therefore, y_1, y_2, and y_3 cannot be calculated by Adams-Bashforth. It needs to be "started" by a one-step method such as Runge-Kutta. The advantage of multi-step methods is that they require fewer function evaluations, since once started, f_{n-1}, f_{n-2} and f_{n-3} are already known. The disadvantage is that these methods only work well for smooth solutions that are well approximated by polynomials.

A particularly simple multi-step method is the modified mid-point method which is based on the mid-point method (Eq. 2.7). It advances $y(t)$ to $y(t+H)$, in a sequence of p substeps of length $h = H/p$. The first substep is made with Euler, and the subsequent steps by midpoint, that is,

$$\begin{aligned} z_0 &\equiv y_n \approx y(t) \\ z_1 &= z_0 + hf(z_0, t) \\ z_{m+1} &= z_{m-1} + 2hf(z_m, t + mh), \quad \text{for} \quad m = 1, 2, \ldots, p-1, \\ y(t+H) \approx y_{n+1}^{(p)} &\equiv \frac{1}{2}(z_p + z_{p-1} + hf(z_p, t+H)). \end{aligned} \tag{2.11}$$

Here z_m is the result of the m^{th} substep, and $y_{n+1}^{(p)}$ is the final approximation to the solution at $t + H$ using p substeps. This is a second order method like the midpoint method (Eq. 2.7), but requires only one function evaluation per substep, since it uses the previous substep's result as the midpoint. Another advantage is that its error series is only in even powers of h,

$$y_{n+1}^{(p)} - y(t+H) = \sum_{i=1}^{\infty} \alpha_i h^{2i}. \tag{2.12}$$

This makes the modified midpoint method particularly suitable for use with Richardson extrapolation to the limit (Conte and de Boor, 1980; Press et al., 1992; Bulirsch and Stoer, 1992). In Richardson extrapolation, a solution is calculated at several step sizes, and then extrapolated to zero step size using a polynomial fit. The Bulirsch-Stoer method (Press et al., 1992) utilizes the modified midpoint method in exactly that way.

The solution is advanced by a "large" step H in p substeps using Eq. (2.11). Successively better estimates for y_{n+1}, $y_{n+1}^{(p)}$ are obtained by increasing p in the sequence

$$p = 2, 4, 6, 8, 10, \ldots \tag{2.13}$$

After each refinement, the extrapolated estimate $y_{n+1}^{\infty} \approx y(t+H)$, and an error estimate ϵ_p are calculated. If even after successive refinements the error estimate does not go below the desired tolerance (say, ϵ), the step-size, H is reduced. Due the even error series of the modified midpoint method (Eq. 2.12), $\epsilon_p \sim H^{2p+1}$, and therefore a new estimate for H is,

$$H_p = H \left(\frac{\epsilon}{\epsilon_p} \right)^{1/(2p+1)}. \tag{2.14}$$

Implicit methods

Implicit methods are those in which the rule for calculating y_{n+1} depends on y_{n+1} itself. For instance, the implicit Euler rule is

$$y_{n+1} = y_n + hf(y_{n+1}, t_{n+1}). \tag{2.15}$$

For non-linear ODEs, an implicit rule such as Eq. (2.15) is solved by an iterative method like Newton-Raphson (Conte and de Boor, 1980), which is computationally expensive and not always guaranteed to converge. A less demanding approach is that of semi-implicit methods (Press et al., 1992), arrived at by linearizing the fully implicit method. For instance, if we linearize Eq. (2.15) about y_n, keeping only the first-order y_n term, we get

$$y_{n+1} = y_n + h\left[f(y_n, t_{n+1}) + J(y_n, t_n)(y_{n+1} - y_n)\right]. \tag{2.16}$$

$J(y_n, t_n)$ is the Jacobian of $f(y_n, t_n)$ with respect to y at y_n. Solving for y_{n+1}, the semi-implicit counterpart of the implicit Euler method is

$$y_{n+1} = y_n + h\left[1 - J^{-1}(y_n, t_n)f(y_n, t_n)\right]. \tag{2.17}$$

The Bader-Deuflhard method (Press et al., 1992) is an example of a semi-implicit method. It is similar to the Bulirsch-Stoer in its use of Richardson extrapolation. However, it uses the semi-implicit modified mid-point rule for the substeps $h = H/p$,

$$\begin{aligned} [1 - hJ(z_m, t_m)]z_{m+1} = {} & [1 + hJ(z_m, t_m)]z_{m-1} \\ & + 2h[f(z_m, t_m) - J(z_m, t_m)z_m], \end{aligned} \tag{2.18}$$

where $m = 2, \ldots, p-1$ and z_m is the intermediate value at the substep m. The first substep is the semi-implicit Euler (Eq. 2.17), while the last step is

$$y_{n+1}^{(p)} = \frac{1}{2}(z_p + z_{p-1}). \tag{2.19}$$

This method, like the modified midpoint method has an even error series (Eq. 2.12), though a different sequence of refinements is used for the Richardson extrapolation and so

$$p = 2, 6, 10, 14, 22, 34, 50, \ldots \tag{2.20}$$

Though not as demanding as implicit methods, semi-implicit methods can be fairly computationally expensive due to the matrix inversion if the number of equations is large.

2.2.2 Numerical Stability and Stiffness

The numerical methods discussed in the previous section approximate the ODE with a difference equation

$$y_{n+1} = g(h; y_{n+1}, y_n, y_{n-1}, \ldots, y_{n-k}, t_n, t_{n-1}, \ldots, t_{n-k}). \tag{2.21}$$

Here, g is some function that depends on the numerical method and $f(y, t)$. As $h \to 0$, the solution of the difference equation, approaches that of the ODE. However, in practice h is finite, and the numerical solution can diverge exponentially from the exact solution. We illustrate this with an example. Consider the solution of

$$y' = \lambda y, y(0) = y_0 \tag{2.22}$$

using the Euler method. With $\lambda = -100$, and $h = 0.03$, the difference equation is

$$\begin{aligned} y_{n+1} &= (1+\lambda h) y_n, \\ \text{or,} \qquad y_{n+1} &= -2 y_n. \end{aligned}$$

Using the initial condition, the solution is $y_n = (-2)^n y_0$, which oscillates between negative and positive values, while the analytical solution is $y_0 e^{-100t}$. Therefore for this example the Euler method is stable in the region $|1+\lambda h| < 1$ (see Figure 2.6a).

In general, the stability region depends on both the ODE being solved and the numerical method. To compare methods on an equal footing however, the test system of Eq. (2.22) is used. For this test system, Eq. (2.21) is a linear homogenous difference equation,

$$a_0 y_{n+1} + a_1 y_n + \ldots + a_{k+1} y_{n-k} = 0, \tag{2.23}$$

with the general solution

$$y_n = c_1 \beta_1(\lambda h)^n + c_2 \beta_2(\lambda h)^n + \ldots + c_{k+1} \beta_{k+1}(\lambda h)^n. \tag{2.24}$$

A method is said to be absolutely stable if $|\beta_i(\lambda h)| < 1, i = 1, \ldots, k+1$.

The difference equation for explicit single-step methods depends only on y_n, that is $k = 0$. As a consequence, their general solution has only one component $c_1 \beta_1(\lambda h)^n$. For the RK4 method, the region of absolute stability (see Figure 2.6**c**) is

$$|\beta_1(\lambda h)| = \left| 1 + \lambda h + \frac{(\lambda h)^2}{2} + \frac{(\lambda h)^3}{6} + \frac{(\lambda h)^4}{24} \right| < 1. \tag{2.25}$$

For methods that are absolutely stable, reducing the step size h sufficiently guarantees stability. The picture is further complicated when solving systems

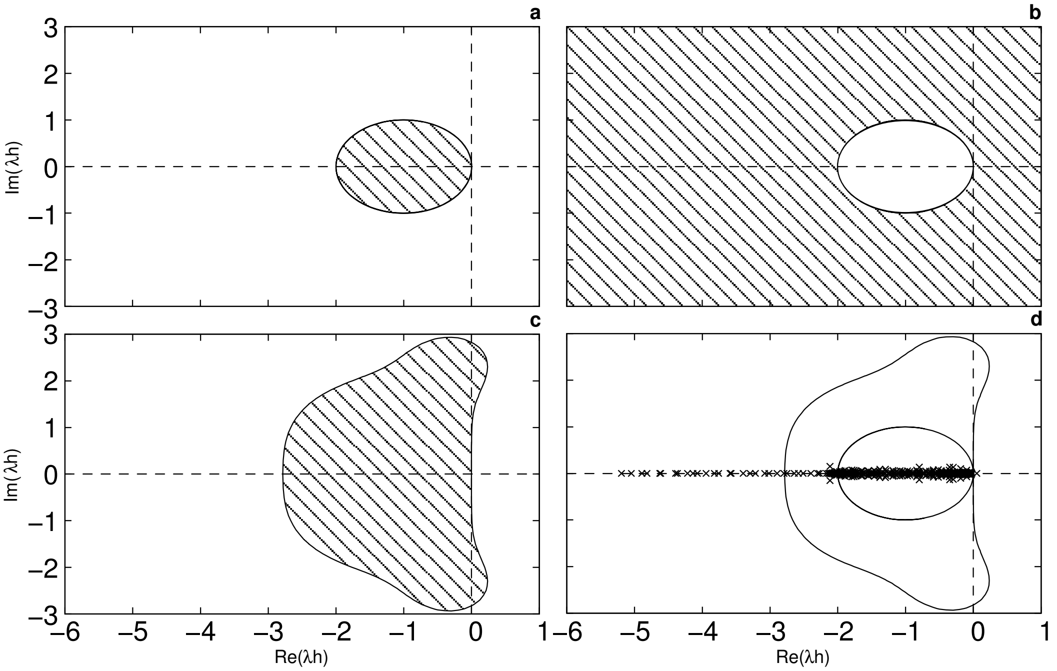

Figure 2.6: Stability regions of (**a**) Euler, (**c**) RK4 for the test system $y' = \lambda y$, and (**b**) Implicit Euler for the test system $y' = -\lambda y$. The graphs show the λh complex plane, where h is the stepsize of the method. Shaded regions are stable. (**d**) The eigenvalues (crosses) of a gene circuit are plotted with the stability regions of Euler and RK4 with stepsize $h = 1$ min. They were calculated for the solution at $t = 71.1$ min.

of ODEs such as the gene circuit equations (Eq. 2.1). If we linearize the solution, about a point y_0, we get the counterpart of the one-dimensional test system (Eq. 2.22)

$$z' = A(y_0)z, \quad z = y - y_0. \tag{2.26}$$

$A(y_0)$ is the Jacobian of the system at y_0. Now the stability condition needs to be satisfied for all eigenvalues $\lambda_j, j = 1, \ldots, N$, that is $|\beta_i(\lambda_j h)| < 1, \forall i, j$. If the largest and smallest eigenvalues differ by a large factor, the step size h needs to be very small for absolute stability. Usually one looks at the ratio $\frac{|\lambda_{max}|}{|\lambda_{min}|}$, λ_{max} is the largest eigenvalue and λ_{min} is the smallest. If $\frac{|\lambda_{max}|}{|\lambda_{min}|} \gg 1$, the system of ODEs is said to be stiff.

Implicit methods are especially suited for solving stiff problems. They have much larger stability regions than those of explicit methods. As an example, consider the implicit Euler method (Eq. 2.15). For the test system $y' = -\lambda y$, $y_{n+1} = \frac{1}{1+\lambda h} y_n$, the stability region is $|1 + \lambda h| > 1$ (Figure 2.6b), which covers most of the complex plane. As mentioned before, this stability comes at the cost of increased computational demands for solving a non-linear implicit equation at each step. Semi-implicit methods are not guaranteed to be stable, but usually are (Press et al., 1992).

2.2.3 Stiffness of gene circuits and numerical methods for their solution

Figure 2.6d shows the spread of the eigenvalues of a gene circuit juxtaposed over the stability regions of Euler and RK4. The eigenvalues are divided into two groups that are spread far apart. One group falls into the stability region of RK4 and the other group is outside. In fact, $\frac{|\lambda_{max}|}{|\lambda_{min}|} \approx 10^3$ and hence the gene circuit is stiff.

Therefore, RK4 at step size 1.0 min is expected to be unstable for this problem. A gene circuit was produced by simulated annealing using RK4 with step size 1.0 min. Figure 2.7a shows the output of this gene circuit. On lowering the stepsize to 0.1, the output changes qualitatively, resulting in biologically incorrect patterns (see Figure 2.7b). This shows that RK4 is not a reliable numerical method for gene circuits.

In order to determine the numerical method best suited to the gene circuit (Eq. 2.1), several methods were tested. The following methods were tried: the explicit single-step RK4 method (Eq. 2.8), Adams-Bashforth, an explicit multi-step method (Eq. 2.10), Bulirsch-Stoer, an explicit adaptive stepsize method (see Section 2.2.1), and a semi-implicit Bader-Deuflhard method (see Section 2.2.1). The fixed stepsize solvers were included merely for comparison, as it is risky to use them for stiff systems of ODEs (Press et al., 1992). An explicit solver with stepsize control, however, is capable of handling stiff systems of ODEs by reducing stepsize to remain in the stable region of the solver. This comes at the cost of increased computational effort. Similarly, implicit and semi-implicit solvers have the additional computational burden of an iterative method or matrix inversion respectively. Since the optimal solver has to be both efficient and stable, we did not assume *a priori* which is better.

2.2.4 Solver comparison

The comparison of the four methods is shown in Table 2.1. The root mean squared difference between data and gene circuit (rms score, Eq. A.1) was used as a measure of accuracy. The best possible solution was calculated by

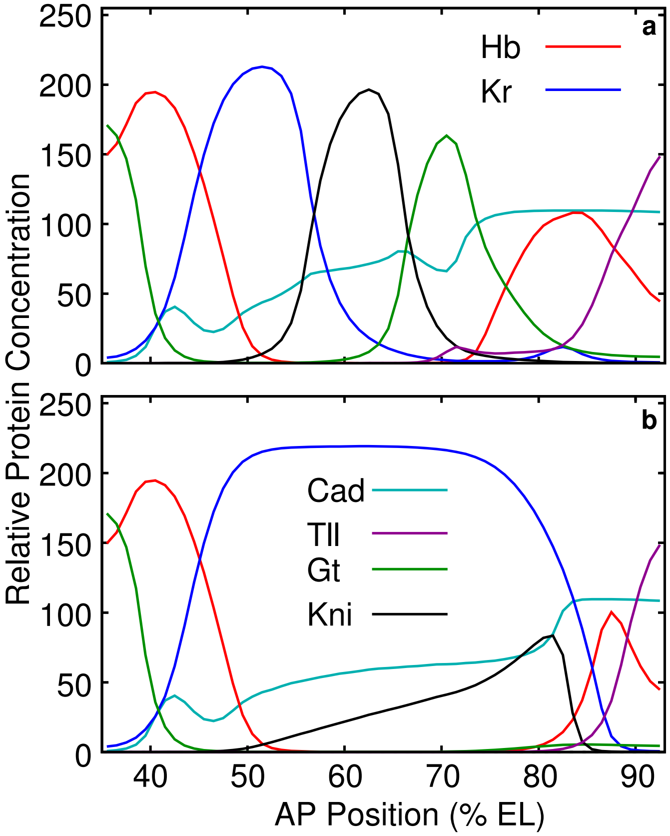

Figure 2.7: Instability of the numerical solution in gene circuits. A gene circuit was fit to data using RK4 with stepsize $h = 1$ min. (**a**) Gap gene patterns produced by the circuit at time 71.1 mins. (**b**) Gap gene patterns produced by the circuit if stepsize is reduced to $h = 0.1$ min.

Solver	**Step/Acc**	**RMS score**	**Time (ms)**	**Function evaluations**
RK4	1.0	22.2825	7.93	284
RK4	0.1	10.1712	68.48	2844
Adams-Bashforth	1.0	25.2379	3.97	160
Adams-Bashforth	0.1	10.1713	43.16	1440
Bulirsch-Stoer	10^{-3}	10.1699	7.87	337
Bulirsch-Stoer	10^{-4}	10.1714	16.58	538
Bader-Deuflhard	10^{-3}	10.1712	3809.79	476
Bader-Deuflhard	10^{-4}	10.1713	4677.13	682

Table 2.1: Comparison of solver performance on gene circuits. The first column lists the solvers that were tested. The second column is the stepsize for fixed-stepsize methods, or accuracy requirement (at each step) for adaptive-stepsize methods. The rms score is listed in the third column. The rms score with Bulirsch-Stoer with accuracy requirement 10^{-12} is 10.171220555381. The fourth column gives the CPU time to solve a gene circuit (average over 1000 repetitions) on a 2.6 GHz Pentium 4 Xeon processor. CPU time was calculated with the `getrusage()` function call of the `libc6` Linux library. The last column gives the number of times the RHS of Eq. (2.1) was evaluated to solve the gene circuit.

lowering the stepsize/accuracy requirement until the methods converged. Both Bulirsch-Stoer and Bader-Deuflhard are within 0.1% of the best solution at an accuracy requirement 10^{-3}. Both fixed stepsize solvers require a stepsize of 0.1 to be accurate, and take many more function evaluations and much more time than Bulirsch-Stoer. Although Bader-Deuflhard makes a comparable number of function evaluations, each function evaluation is much more expensive, making it the least-efficient method. Bulirsch-Stoer is the most efficient of all methods tested by a wide margin, and it is used in all gene circuits in this thesis (except circuits with production delays, Section 2.3).

2.3 Gene circuits with production delays

One key approximation of the interphase rule of the gene circuit 2.1 is that synthesis is instantaneous. Synthesis of a protein in eukaryotes, however, involves several steps. First the RNA polymerase II transcribes the DNA base by base, to produce mRNA. Introns are spliced out of the mRNA to produce mature mRNA, which is then transported out of the nucleus for translation. The mature mRNA is translated by the ribosome into protein codon by codon. Thus, we can expect that there will be some delay from the start of transcription to the time when the finished protein becomes available as a transcription factor.

There is evidence that such delays have a biological function. The gene *knirps-like* (*knrl*), (Oro et al., 1988), encodes a protein with the same DNA-binding motif as *kni*, but lacks segmentation function (Rothe et al., 1989). This is due to the 23kb transcriptional unit of *knrl* as compared to the 3kb unit of *kni* (Rothe et al., 1989). A *knrl* complementary DNA rescues *kni* phenotype (Rothe et al., 1992), indicating that transcription and splicing delays are responsible for the different function of the two genes. In mammalian cell lines, synthesis delays have been proposed as the mechanism for oscillations in the expression of Hes1, p53, and NF-κB (Monk, 2003).

In this section we first consider the case for relaxing the instantaneous-synthesis approximation in gene circuits. Next we propose a model based on delay differential equations, and discuss the details of its numerical implementation. Finally we present delay-equation-based gene circuits and discuss their suitability for describing the gap gene system.

2.3.1 Need for a delay model

Although the mRNA of gap genes is detected as early as cycles 11 and 12 (Knipple et al., 1985; Tautz et al., 1987; Rothe et al., 1989; Jaeger et al., 2007), proteins are not detected until early cycle 13 (Gaul et al., 1987; Eldon and Pirrotta, 1991; Kraut and Levine, 1991a; Surkova et al., 2007). Therefore, even though transcription of gap genes is taking place, significant amounts of gap proteins are not being translated.

We made very rough estimates of the transcriptional and translational delays, based on transcript and protein lengths from FlyBase (Crosby et al., 2007) and GenBank (Benson et al., 2007). There are two reports of the translocation rate of RNA Polymerase II, one in *Drosophila*, and the other in sea urchin cells. In *Drosophila*, the *Ubx* gene took 55 minutes to transcribe (Shermoen and O'Farrell, 1991). The transcription unit of *Ubx* is 77kb long, giving an estimate of 23 base pairs per second. In sea urchin cells, the transcription rate was measured to be around eighteen base pairs per second at 25 °C (Davidson, 1986) . For bacterial polymerase, the estimate is 40 base pairs per second at 37 °C (Lewin, 2000). For simplicity, a value of twenty base pairs per second

Gene	**Transcript length (bp)**	**Txn time τ (min)**	**Protein length (aa)**	**Translation time τ (min)**	**Total τ (min)**
hb P1	6502	5.4	758	1.1	9.5–12.5
hb P2	3284	2.8	758	1.1	6.9–9.9
Kr	2423	2.0	466	0.8	5.8–8.8
gt	1963	1.6	448	0.8	5.4–8.4
kni	3012	2.5	429	0.7	6.2–9.2

Table 2.2: Estimation of transcription (Txn) and translation times for gap genes based on unprocessed transcript length and protein length respectively.

was used in our estimates.

The translation rate in bacteria is estimated to be around 15 amino acids per second at 37 °C (Lewin, 2000). A value of 10 amino acids per second at 25 °C was used calculate translation time. Splicing delays were estimated to be around three minutes in the early *Drosophila* embryo based on electron microscopy observations (Beyer and Osheim, 1988). The nuclear latency of β-globin in mammalian cells was measured to be between 2.5 and 4.4 minutes (Audibert et al., 2002). Based on these estimates we used three minutes as a lower bound and six minutes as the upper bound for splicing and nuclear export delays. The full set of estimates for all the gap genes are shown in Table 2.2.

Delays of 6–10 minutes are significant compared to the time scale of the model, which encompasses about 100 minutes from the end of nuclear division 12 to the onset of gastrulation. Furthermore, the effects of these delays are amplified by the short duration of cleavage cycles 10 and 11 (Foe and Alberts, 1983). Shermoen and O'Farrell (1991) visualized *Ubx* transcripts using *in situ* hybridization and report the disappearance of unfinished *Ubx* transcripts during mitosis. This prevents the synthesis of *Ubx* protein until much later. In fact there is global shutdown of transcription during mitosis (Gottesfeld and Forbes, 1997).

Finally, gene circuits do not produce the early patterns of gap gene in cycle 13 accurately (Jaeger et al., 2004a; Jaeger, 2005). Though gap gene domains have well-defined boundaries when they first appear in cycle 13 (Jaeger et al., 2007; Surkova et al., 2007), gene circuits produce patterns that are derepressed initially, but then get refined by gap-gap cross-regulation (Jaeger et al., 2004a). As gap transcripts are also produced in a localized manner (Jaeger et al., 2007) in cycles 10-12, this suggests that production delays might be required in the model in order to obtain correct behavior of early patterns. In the following section, we propose a modification of the gene circuit model that incorporates such delays.

2.3.2 Delay model

In this section, a gene circuit with production delays will be constructed by modifying its continuous rules (see Section 2.1.2). The segmentation genes do not regulate the cell cycle, hence the division rule remains the same as in Section 2.1.1. In order to incorporate delays, it is convenient to rewrite the interphase rule as

$$\begin{aligned}\frac{dv_i^a}{dt} = R^a\theta(t)g\left(\sum_{b=1}^{N} T^{ab}v_i^b + m^a v_i^{\mathrm{Bcd}} + \sum_{\beta=1}^{N_e} E^{a\beta}v_i^\beta + h^a\right)& \\ + D^a(n)(v_{i+1}^a - 2v_i^a + v_{i-1}^a) - \lambda^a v_i^a,&\end{aligned} \tag{2.27}$$

where, except for $\theta(t)$, all the terms are the same as in Eq. (2.1).

$\theta(t)$ is a square-wave of unit magnitude in time that represents the shutdown of transcription during mitoses because

$$\theta(t) = \begin{cases} 0 & \text{before ninth mitotic division,} \\ 0 & \text{if } M_{\text{start}} \leq t \leq M_{\text{stop}}, \\ 1 & \text{otherwise.} \end{cases} \tag{2.28}$$

Here M_{start} and M_{stop} are times at which mitoses start and end respectively. Note that since no zygotic transcription is observed for gap genes before cycle 10 (Jaeger et al., 2007), $\theta(t)$ is set to 0 before the ninth division. Hence the interphase and mitosis rules have been combined into a single equation. This is advantageous because with delays the synthesis rate at the beginning of an interphase will depend on the state of the system during the preceding mitosis.

It is assumed that the time taken for synthesis (transcription, splicing, export, and translation) does not vary in time or depend on the protein concentrations of the gap genes. Also, diffusion and decay are assumed to be much faster than synthesis. They will be represented as instantaneous steps (as in Eq. 2.1). Letting the production delay for protein a be τ_a, the rate of change of its concentration in nucleus i, v_i^a is represented by the delay differential equation (DDE, Bellman and Cooke, 1993)

$$\begin{aligned} \frac{dv_i^a}{dt} = R^a \theta(t - \tau^a) \left(\sum_{b=1}^{N} T^{ab} v_i^b(t - \tau^a) + m^a v_i^{\text{Bcd}} + \sum_{\beta=1}^{N_e} E^{a\beta} v_i^{\beta}(t - \tau^a) + h^a \right) \\ + D^a(n)(v_{i+1}^a - 2v_i^a + v_{i-1}^a) - \lambda^a v_i^a. \end{aligned} \tag{2.29}$$

To completely specify the delay differential equation, the history S must be provided, which is a generalization of the initial conditions. It is specified as

$$v_i^a(t) = S_i^a(t), \quad \text{for} \quad -\tau \leq t \leq 0, \quad \tau = \max_a\{\tau^a\}.$$

In Eq. (2.29), the rate of protein synthesis depends on the concentrations of the products of genes b at a point τ^a minutes in the past. The numerical treatment of such equations is considered next.

2.3.3 Numerical Integration of gene circuits with delays

The numerical integration of the delay model (Eq. 2.29) is relatively simple compared to the general case, since the time delays are constant. The method of steps (Bellman and Cooke, 1993) was employed to solve the gene circuit with production delays.

The method of steps

Consider the scalar DDE

$$y' = f(t, y(t), y(t-\tau)), \qquad t \geq t_0 \tag{2.30}$$

with the history specified by

$$y(t) = S(t), \qquad t < t_0. \tag{2.31}$$

The method of steps solves Eq. (2.30) over successive intervals $[t_0 + (l-1)\tau, t_0 + l\tau], \quad l = 1, 2, 3, \ldots$ by setting

$$
\begin{aligned}
y_1'(t) &= f(t, y_1(t), S(t-\tau)), \quad t \in [t_0, t_0+\tau], \\
y_2'(t) &= f(t, y_2(t), y_1(t-\tau)), \quad t \in [t_0+\tau, t_0+2\tau], \\
&\vdots \\
y_m'(t) &= f(t, y_m(t), y_{m-1}(t-\tau)), \quad t \in [t_0+(m-1)\tau, t_0+m\tau].
\end{aligned} \tag{2.32}
$$

This yields the solution $y(t) = y_l(t)$ for $t \in [t_0 + (l-1)\tau, t_0 + l\tau]$.

A general property of DDEs is that the discontinuities in the history $S(t)$ propagate to the points $t_0 + l\tau$. For instance, it is not unusual for f and the left derivative of $S(t)$ to be different at t_0, that is, $f(t_0, y(t_0), S(t_0 - \tau)) \neq S'(t_0^-)$, introducing a first-order discontinuity that propagates. For example, Figure 2.8 shows the solution of Eq. (2.30) with $\tau = 1$, and $S(t) = 1, t < 0$. The first order discontinuity propagates as higher order discontinuities to points $1, 2, 3, \ldots$, making the solution smoother.

Next, a numerical solver is presented that uses the method of steps to solve DDEs.

A numerical method for DDEs

The specific implementation of the solver is based on the `dde23` solver in MATLAB (Shampine and Thompson, 2001). Since Eq. (2.29) has multiple delays, the equation is of the type

$$y'(x) = f(x, y(x), y(x-\tau_1), y(x-\tau_2), \ldots, y(x-\tau_k)), \tag{2.33}$$

on $[a, b]$, for constant delays τ_j such that $\tau = \min(\tau_1, \ldots, \tau_k) > 0$. The history needs to be specified as $y(x) = S(x)$ for $x \leq a$. No generality is lost in considering a scalar equation, as the formulas are valid for systems of DDEs

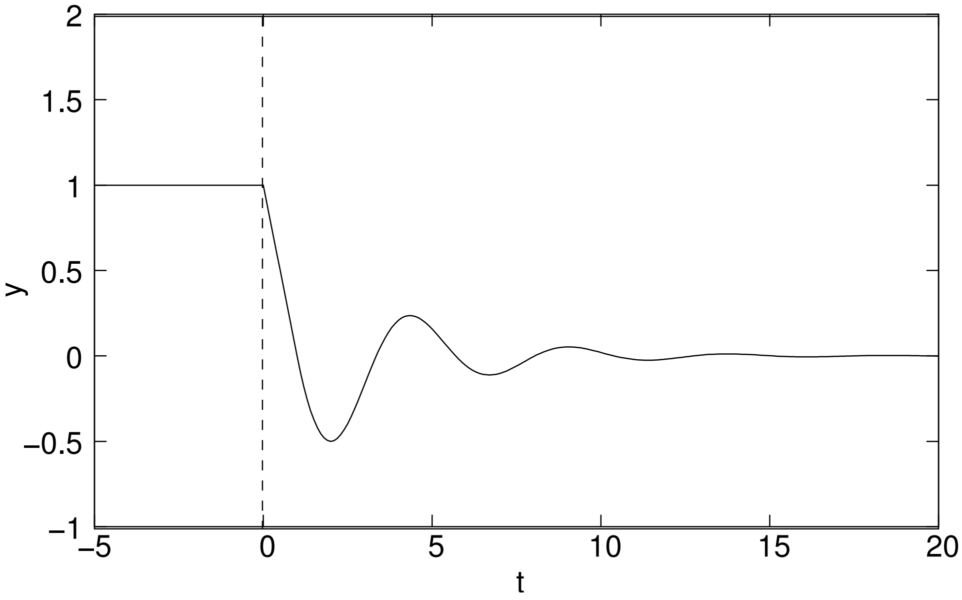

Figure 2.8: Propagation of discontinuities in delay equations. The solution to the delay equation $y' = -y(t-1)$, with history $S(t) = 1, \quad t < 0$.

also. The key features of this solver are determined by the following general properties of delay equations.

1. Need for dense output. Suppose f is being evaluated at some grid point x_n. It is likely that its arguments, $x_n - \tau_1, x_n - \tau_2, \ldots, x_n - \tau_k$, are not grid points. Thus, an approximation to the solution at non-grid-points is needed.

2. Ability to handle discontinuities. As discussed above, the discontinuity in y' at $x = a$ will propagate as discontinuities in higher derivatives at $(a+\tau_1, a+2\tau_1, \ldots), (a+\tau_2, a+2\tau_2, \ldots)$ and so on. The order of the solver formulas depends on the smoothness of the solution, and therefore the solver should either be able to side step these discontinuities, or ignore them if they are in a high enough derivative.

3. Implicit formulae. If at some point, the step-size $h_n > \tau$, the formulae become implicit and the solver must be capable of handling them.

In view of these required features, `dde23` is based on what is called the BS(2,3) Runge-Kutta triple (Bogacki and Shampine, 1989), which is modified to handle discontinuities and implicit formulae. The BS(2,3) triple uses continuous extensions (Dormand, 1996) to provide dense output.

The triple consists of three explicit formulas. The first formula advances the solution, while the second one provides a lower-order estimate of the solution, so that an error estimate can be made. The third formula provides a polynomial interpolation so that the solution can be approximated at non-grid-points.

Suppose an approximation $\hat{y}_n$ to $y(x)$ at x_n is known, and the solution needs to be advanced to $\hat{y}_{n+1}$ at $x_{n+1} = x_n + h_n$. The formula to advance the solution is a three stage Runge-Kutta method. The intermediate stages y_{ni} are evaluated at $x_{ni} = x_n + c_i h_n$, where, $i = 1, 2, 3, 4$. The parameters for the BS(2,3) triple are listed in Table 2.3. The intermediate stages for the method are given by

$$y_{ni} = \hat{y}_n + h_n \sum_{j=1}^{i-1} a_{ij} f_{nj},$$

where $f_{ni} = f(x_{ni}, y_{ni})$. Then the solution at $x_n + h_n$ is given by

$$\hat{y}_{n+1} = \hat{y}_n + h_n \sum_{i=1}^{3} \hat{b}_i f_{ni},$$

which provides a third order estimate to $y(x_n)$. The second formula is a four-stage formula that provides a second-order estimate. It is given by

$\mathbf{c_i}$	$\mathbf{a_{ij}}$			$\hat{\mathbf{b}}_\mathbf{i}$	$\mathbf{b_i}$	$\mathbf{b_i^*}(\sigma)$
0				$\frac{2}{9}$	$\frac{7}{24}$	$1 - \frac{4}{3}\sigma + \frac{5}{9}\sigma^2$
$\frac{1}{2}$	$\frac{1}{2}$			$\frac{1}{3}$	$\frac{1}{4}$	$\sigma(1 - \frac{2}{3}\sigma)$
$\frac{3}{4}$	0	$\frac{3}{4}$		$\frac{4}{9}$	$\frac{1}{3}$	$\frac{4}{3}\sigma(1 - \frac{2}{3}\sigma)$
1	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{4}{9}$	0	$\frac{1}{8}$	$\sigma(\sigma - 1)$

Table 2.3: Parameter values for the BS(2,3) triple.

$$y_{n+1} = \hat{y}_n + h_n \sum_{i=1}^{4} b_i f_{ni}.$$

The parameters (Table 2.3) are chosen such the fourth stage in the above formula is given by the third-order estimate itself, that is $y_{n4} = \hat{y}_{n+1}$. This trick saves three function evaluations, and is known as the "First Same As Last" device.

The third formula provides the continuous extension of the solution from $\hat{y}_n$ to $\hat{y}_{n+1}$. It is given by

$$y^*_{n+\sigma} = y_n + h_n \sum_{i=1}^{4} b^*_i(\sigma) f_{ni}.$$

σ assumes values between 0 and 1. This last formula provides the dense output, or intermediate output at non-grid-points required for a delay solver.

The step-size is controlled by

$$h_{n+1} = h_n \left(\frac{\epsilon}{\delta_{n+1}} \right)^{1/3}, \tag{2.34}$$

where δ_{n+1} is the error estimated from the second-order formula, and ϵ is the desired accuracy.

In `dde23`, the intermediate stages of the BS(2,3) triple f_{ni} are retained even when the step from x_n to $x_n + h_n$ is complete. This is because later steps of the solver might require the solution in this interval due to delays.

The other two features required of a delay solver, that is, provision for discontinuities, and for implicit equations are handled as follows. Since the delay gene circuit has constant delays, a list of time points is precalculated where discontinuities are expected to arise. Then, using the method of steps, the equation is only solved on intervals between these points.

Secondly the formulas for the BS(2,3) triple can become implicit if the step size is bigger than the smallest delay. Function iteration (Press et al., 1992) is used to solve the implicit equations. If the iteration does not converge, the step size is reduced, making the formulas explicit.

Customization for gene circuits

Due to the combination of discrete and continuous rules of gene circuits (Section 2.1.1), further customization of the `dde23` solver described above was made in two steps. In the first step, the precalculation of discontinuities included the first-order discontinuities at the beginning of mitoses and interphases when the right hand side of Eq. (2.29) changes discontinuously.

The second complication involved the discrete rule of division. As mentioned at the end of the previous section, the intermediate calculations of the BS(2,3) triple are retained in case they are required in the calculation at a later point. In gene circuits, however, the number of equations is not fixed, and changes at each division. Hence, situations are possible, in which, to calculate a daughter nucleus's concentration, the mother nucleus's concentration is required. This was handled by applying the division rule to the BS(2,3) intermediates. Note that for a general gene circuit, it is not possible to assume a maximum delay, and it is quite possible to require the intermediates from cleavage cycle 11 for the solution in cleavage cycle 14A.

2.3.4 Results

A delay gene circuit was constructed with six gap genes, *cad*, *hb*, *Kr*, *gt*, *kni*, and *tll*. No external inputs were included in this circuit. For *Kr*, *gt*, *kni*, and *tll*, the history was zero, as they are only detected first in cleavage cycle 13 (Section 2.1.2). For *hb* and *cad*, data from cleavage cycles 11 and 12 were provided as history.

This circuit was fit to integrated gap gene data (Surkova et al., 2007) using two different approaches. In the first, the delay values were fixed to empirical values calculated in Table 2.2. In the second approach, delays τ^a were determined by the optimization. Figure 2.9 shows the early and late patterns for the two circuits.

The empirical-delays gene circuit has an rms score of 13.95, and accordingly has several patterning defects (see Figure 2.9). The optimized-delays circuit has score 10.97, but still has minor patterning defects. The early patterns in both circuits do not form clearly-defined *Kr* or *gt* domains. This suggests delay gene circuits with simple production delays (Eq. 2.29) are not sufficient to accurately reproduce early gap gene patterns. Circuits without delays were used for all other computation in this dissertation.

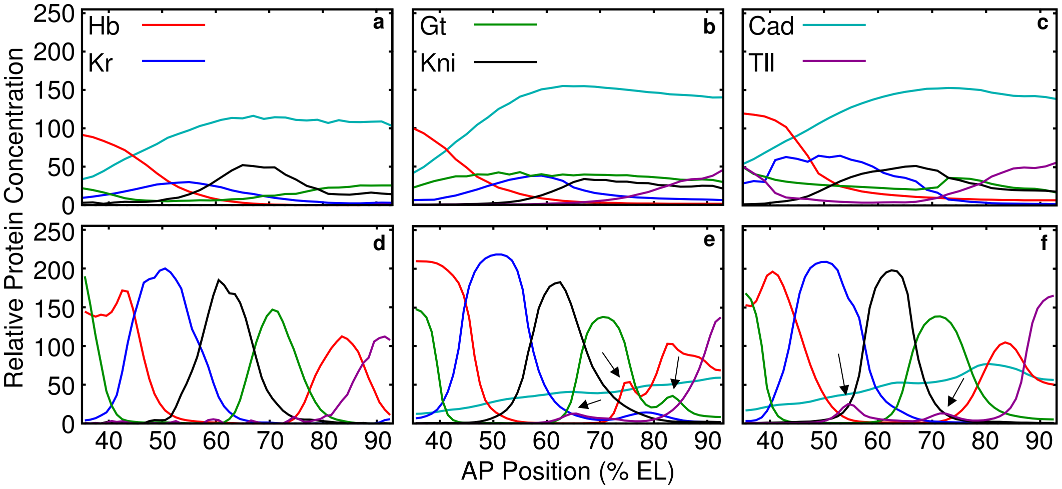

Figure 2.9: Integrated gap gene data (**a**) early (middle of cycle 13), and (**d**) late (time class T8 of cycle 14A). (**b**, **e**) Early and late gap gene patterns produced by the production-delay gene circuit with empirically estimated delays (Table 2.2). This circuit does not produce correct early patterns; *gt* is derepressed in the middle (compare with panel **a**), and *kni* posterior boundary does not form. Also, there are major patterning defects (arrows) in the late patterns. (**c**, **f**) Early and late patterns produced by the gene circuit in which delay values were determined by optimization. This circuit does not produce correct early patterns also; *Kr* anterior boundary does not form, and *gt* is derepressed in the middle. The late patterns have minor defects (arrows) in *tll*, which is not represented correctly in this circuit (see Section 2.1.1).

Chapter 3

Reduction of Bcd variation by the gap gene system

In Section 1.3 (Figure 1.3), recent experimental results (Surkova et al., 2007; Houchmandzadeh et al., 2002) revealed by quantitative gene expression data were reviewed. These data establish that gap gene borders have reduced variance compared to the Bcd gradient. Since the Bcd gradient is established by intrinsic factors determined by the genotype of the fly, these results present an instance of genotypic canalization.

In this chapter, the effects of Bcd variation on the gap genes will be simulated using gene circuits. In the framework of the model, all interactions are subsumed in the structure of the regulatory input u^a (Section 2.1.2). Hence, Bcd variation is simulated by using Bcd data from individual embryos, while keeping the regulatory parameters of the gene circuits fixed.

The first section describes a particular circuit that was obtained by optimization. In Section 3.2, we simulate Bcd variation using Bcd data from FlyEx and show that in this circuit, gap gene borders have reduced variation as compared to Bcd. Finally, this property is investigated by analyzing the regulation of gap gene borders in this circuit.

3.1 Optimization and selection of a gap gene circuit

The gap gene circuit formulated in Section 2.1.1 was fit to integrated gap gene data (Surkova et al., 2007) using Parallel Lam Simulated Annealing (PLSA) (Chu et al., 1999; Chu, 2001). PLSA minimizes the root mean squared (rms) difference (see Section A.2.1) between model output and data. The minimized rms difference for a circuit is a measure of the quality of fit, and is called the rms score. This procedure produces many candidate circuits, and we selected a circuit using a three-step method (Jaeger et al., 2004a). First, only circuits with an rms score less than 12.0 were considered. These circuits were screened further for patterning defects, and any circuit with major defects was discarded. Finally, experimental (Clyde et al., 2003) and theoretical (Perkins et al., 2006) investigations have shown that *Kr* represses *hb*, therefore, we only select circuits for further consideration if they show this property.

With this screening process, we obtained 23 circuits out of a total of 65 optimizations. This set of circuits have the same network topology as the circuits studied in earlier work (Jaeger et al., 2004a; Perkins et al., 2006). Bcd and Cad are activators of *hb*, *Kr*, *gt*, and *kni*. *tll* is an activator of *hb*, and a repressor of the other gap genes. The interaction between the four gap genes is one of mutual repression, with two exceptions: (1) *gt* is an activator of *hb* in all the circuits obtained. (2) *kni* is an activator of *gt* in about half the circuits, and a repressor in the other half. This network topology is discussed in depth

	Regulator gene b						
Target gene a	*bcd*	*cad*	*tll*	*hb*	*Kr*	*gt*	*kni*
hb	0.025	0.004	0.003	0.021	–0.001	0.022	–0.112
Kr	0.118	0.021	–0.203	–0.026	0.035	–0.042	–0.062
gt	0.256	0.023	–0.011	–0.028	–0.202	0.007	0.003
kni	0.012	0.020	–0.187	–0.082	0.000	–0.017	0.013

Table 3.1: Regulatory parameters of the nonautonomous gap gene circuit `hkgn58c13k1_007`. The first, second, and third columns show m^a, $E^{a \leftarrow cad}$, and $E^{a \leftarrow tll}$ respectively. The columns of the 4×4 matrix T^{ab} are shown in the last four columns of the table. The parameters are explained in Eq. (2.1).

elsewhere (Jaeger et al., 2004a; Perkins et al., 2006). These results also hold true for circuits obtained using other Bicoid profiles from the middle of the parameter scatter in Figure 2.5a (black circles).

We chose one circuit (`hkgn58c13k1_007`) which has rms score 10.76 for further analysis. Its parameters are given in Tables 3.1 and 3.2. The chosen circuit's gap gene patterns (Figure 3.1) are consistent with data except for two minor defects. The first one is a bulge on the anterior border of the posterior *hb* domain. The second is that the posterior border of posterior *hb* domain does not form fully. Some circuits reported in previous work (Jaeger et al., 2004a) also suffer from these defects. The first defect is due to very low levels of spurious *tll* expression in the middle part of the embryo stemming from imperfect background removal. The second is due to the omission of the gene *hkb* (Weigel et al., 1990; Brönner and Jäckle, 1991; Pignoni et al., 1992) from the model. *hkb* is a terminal gap gene expressed near the posterior pole.

Parameter	Gene a hb	 Kr	 gt	 kni
R^a	15.000	10.354	15.000	15.000
D^a	0.166	0.200	0.103	0.200
$t^a_{1/2}$	9.529	15.908	9.438	13.062

Table 3.2: Kinetic parameters of the nonautonomous gap gene circuit `hkgn58c13k1_007`. h^a was kept fixed at -2.50 during optimization.

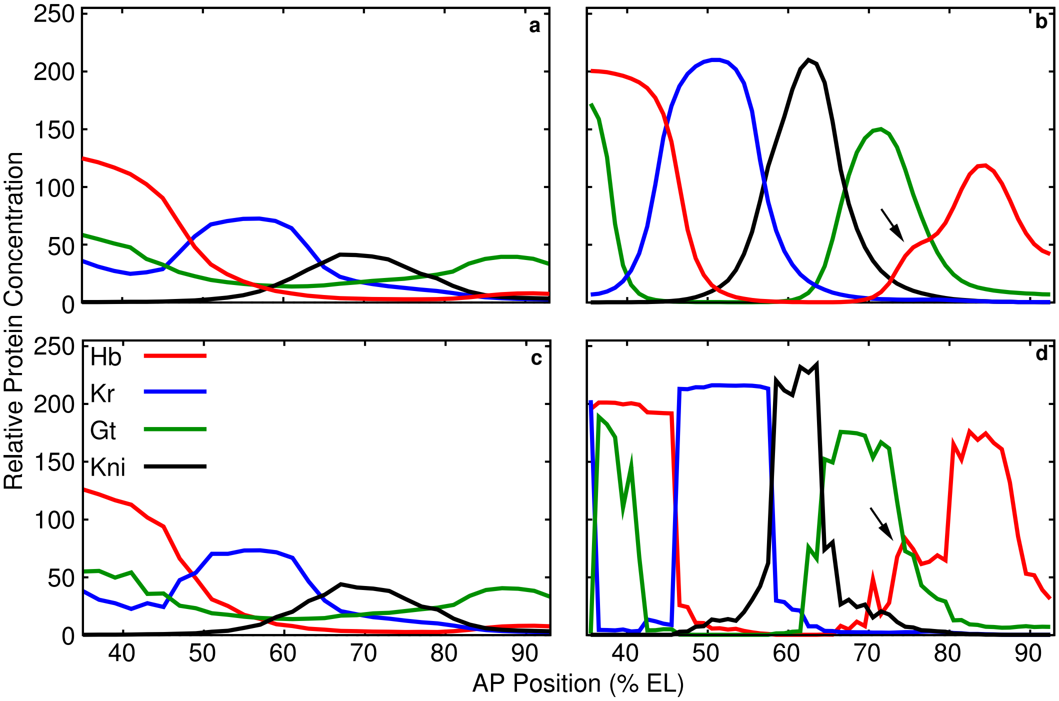

Figure 3.1: Gap gene expression patterns in gene circuit `hkgn58c13k1_007`. Early (**a**) and late (**b**) gap gene expression patterns produced by the gene circuit. The arrow shows the main patterning defect, which is related to experimental noise in Tll data. Early (**c**) and late (**d**) gap gene expression patterns in the same circuit if diffusion is turned off ($D^a = 0$ for all proteins).

3.2 Simulating Bcd variation

We simulated the effects of embryo-to-embryo variation by using a family of 88 bicoid profiles (shown in Figure 2.5) in the circuit derived from the median Bcd (highlighted in Figure 2.5a by a red circle). All parameters of the circuit were kept fixed, and only the Bcd input to the circuit was changed. The positional variance of the gap gene borders was compared with that of Bcd profiles. The position of a gap gene domain's border was calculated as the point where the gene's concentration was half its maximum level in the domain. The positional variance of a border is the standard deviation of its position in the 88 simulations. The positional variance of the family of Bcd profiles is the standard deviation of the position at which each profile crosses a threshold concentration. The threshold is the concentration at which a particular gap gene border forms in the circuit with the median Bcd.

The model correctly predicts the positional variance of six gap gene borders: the posterior border of the anterior *hb* domain, the posterior border of the central *Kr* domain, both borders of the abdominal *kni* domain and both borders of the posterior *gt* domain (see Table 3.3 and Figure 3.2). Furthermore, all six borders show variation comparable to data. For instance, the posterior border of the anterior *hb* domain has a standard deviation of 1.3% EL in the simulations and 1.1% EL in data, both of which are much less than the standard deviation of the Bcd threshold position, 4.6% EL. We performed the *in silico* control experiment of putting *hb* under exclusive control of *bcd* by turning off gap gene cross regulation of *hb* in the model. In this simulation, the posterior border of the anterior *hb* domain had the same variation as the Bcd threshold position.

From the 23 circuits that had consistent network topology, 15 showed

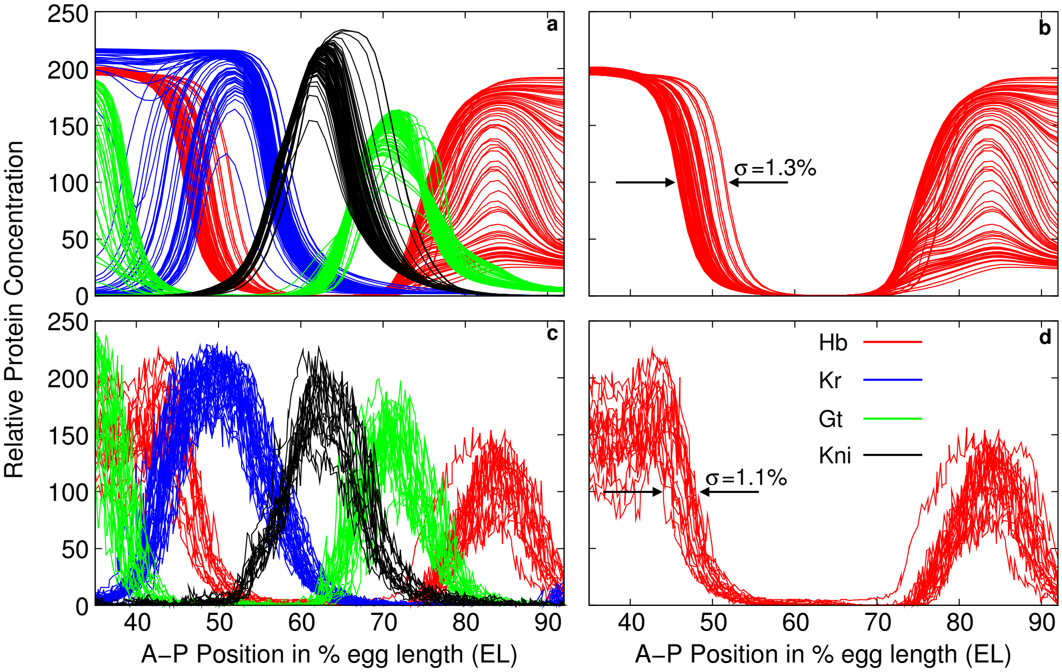

Figure 3.2: Simulation of Bcd variation using gene circuits. (**a, b**) Model, and (**c, d**) Data. (**a**) *hb*, *Kr*, *gt*, *kni* patterns three minutes before gastrulation (time class T8) in 88 simulations using a single gene circuit. A unique Bcd profile from a set of 88 cycle 13 embryos was used in each simulation. The *hb* threshold has a standard deviation (σ) of 4.6% EL (range of 21% EL) in this family of Bcd profiles (see Figure 2.5). With the exception of four borders, variation in gap gene border positions is comparable to data, and much less than Bcd variation. The four exceptions are posterior border of the third *gt* domain, anterior border of the central *Kr* domain, and the borders of the posterior *hb* domain. (**b**) Model output for the posterior border of the anterior *hb* domain. This border's position in model has $\sigma = 1.3\%$ EL (range 5.6% EL). (**c**) Gap gene expression data from a 10% D–V strip in 92 late cleavage cycle 14 (time class T8) wild type embryos. Embryos were stained for *eve*, and two gap genes (*hb*, *Kr*, *gt*, *kni*). There are 18 profiles for *hb*, 33 for *Kr*, 20 for *gt*, and 17 for *kni*. (**d**) *hb* border position data has $\sigma = 1.1\%$ EL, and range 5.9% EL.

positional accuracy for these six borders. Also, the ability of the gene circuit to filter out Bcd's positional variation does not depend on the particular Bcd profile used to fit the circuit. Circuits produced with other profiles from the middle of the parameter scatter (black circles in Figure 2.5b) also showed accuracy when subjected to Bcd variation (see Table 3.4). The gene circuit has this filtration property irrespective of whether smoothed or unsmoothed profiles are used to simulate Bcd variation (Table 3.4).

Four borders show large variation: The posterior border of the third *gt* domain, the anterior border of the central *Kr* domain, and both the borders of the posterior *hb* domain (see Figure 3.2). All these borders are at the edges of the modeled region, where the model is missing key regulators. In the anterior, the head gap genes, *orthodenticle*, *empty spiracles*, and *buttonhead* (Dalton et al., 1989; Finkelstein and Perrimon, 1990; Wimmer et al., 1995) are excluded. They are known to affect the anterior pattern of trunk gap genes (Eldon and Pirrotta, 1991), and respond to Bcd in a concentration-dependent manner (Gao and Finkelstein, 1998). As previously mentioned, in the posterior terminal region, *hkb* has been left out.

3.3 Regulatory analysis of gap gene borders

We performed regulatory analysis of all the gap gene borders that show low positional variation with respect to *bcd*. This analysis is done by determining which regulators of a gene a are driving the synthesis term $R^a g(u_i^a)$ to form a border. Let i_B be the indices of the nuclei over which the border of gene a forms. Since the sigmoid $g(u^a)$ is approximately linear between 10% and 90% expression levels (Figure 2.2), we can write $g(u_{i_B}^a) \approx u_{i_B}^a = \sum_{b=1}^{N} T^{ab} v_{i_B}^b + m^a v_{i_B}^{\mathrm{Bcd}} + \sum_{\beta=1}^{N_e} E^{a\beta} v_{i_B}^{\beta} + h^a$ for the purposes of this analysis.

Boundary	**Bcd threshold position (% EL)**		**Gap gene boundary position (% EL)**			**N/TC (data)**
	range	std dev	range (model)	std dev (model)	std dev (data)	
hb posterior	20.6	4.6	5.9	1.3	1.10	71/T4–T7
Kr posterior	30.7	6.0	5.1	0.7	1.04	46/T5
kni anterior	30.7	6.0	4.4	1.0	1.00	22/T5
kni posterior	44.4	7.7	7.1	1.3	1.18	22/T5
gt anterior	43.1	7.5	7.3	1.2	0.97	23/T5
gt posterior	56.8	9.3	6.2	1.1	1.48	23/T5

Table 3.3: Comparison of positional variance of six gap gene borders with positional variance of Bcd. The first column lists the borders. The range and standard deviation of the Bcd threshold position in the family of 88 Bcds are shown in the second and third columns, respectively. The range and standard deviation of the positions of the borders in the corresponding simulations are shown in the fourth and fifth columns, respectively. The sixth column lists the standard deviation of border positions in our dataset (Surkova et al., 2007). The last column shows the number of embryos and their time class (TC) used to calculate the positional variation of the borders in data. Time class T4-T7 data were used for *hb*. Since the other borders shift in cycle 14A (Jaeger et al., 2004b), only time class T5 data were used for them.

Boundary	`k1_007`	`k12_004`		`k13_015`		`k14_015`		`k15_004`	
	raw	fit	raw	fit	raw	fit	raw	fit	raw
hb posterior	1.45	1.27	1.42	0.75	0.91	0.94	1.07	0.98	1.16
Kr posterior	0.78	0.58	0.74	0.51	0.58	0.41	0.48	0.59	0.61
kni anterior	0.96	0.62	0.89	1.00	0.99	0.38	0.68	1.13	0.99
kni posterior	1.05	1.17	0.86	0.89	0.86	0.90	1.03	1.28	1.15
gt anterior	1.11	0.95	0.99	1.04	1.39	1.30	1.49	1.37	1.44
gt posterior	1.56	1.30	1.51	0.57	1.59	1.44	1.93	1.08	1.83

Table 3.4: Positional variation of gap gene borders in circuits produced with Bcd gradients other than the median. Numbers are standard deviations in % EL. Top row shows the circuits. `k1_007` is the circuit `hkgn58c13k1_007` analyzed in this study. The others are circuits fit using Bcd profiles highlighted in Figure 2.5b with black circles. First column shows the six borders that show low positional variation in gene circuit simulations. Bcd variation was simulated in each circuit with two families of Bcd profiles. The first family consisted of raw, background-removed Bcd profiles (even-numbered columns). The second family consisted of the exponential fits of the profiles of the first family made using Eq. (2.3) (odd-numbered columns). The positional variation of the borders in circuit `hkgn58c13k1_007` using exponential fits is given in Table 3.3.

Consider the posterior border of a gap gene a. Let a's protein be at 90% expression level at $i = i_1$, and at 10% level at $i = i_M$. The analysis is similar for anterior borders, but with $i = i_1$ at the 10% level, and $i = i_M$ at the 90% level. The total change in u^a is $u^a_{i_1} - u^a_{i_M}$. Since this change is just a sum of regulatory contributions, we can divide it into its constituent parts,

$$u^a_{i_1} - u^a_{i_M} = \sum_{b=1}^{N} T^{ab}(v^b_{i_1} - v^b_{i_M}) + m^a(v^{\mathrm{Bcd}}_{i_1} - v^{\mathrm{Bcd}}_{i_M}) + \sum_{\beta=1}^{N_e} E^{a\beta}(v^\beta_{i_1} - v^\beta_{i_M}). \quad (3.1)$$

By comparing the magnitudes of the different regulatory terms, we can determine which regulators are driving the formation of the boundary.

It is possible to simplify this analysis by eliminating many regulators that cannot set the border. Autoregulation, for instance, cannot set the boundary position. It can only make a boundary sharper. Since this is a posterior border, the concentration of a reduces as one goes from anterior (i_1) to posterior (i_M). An activator whose concentration is increasing from i_1 to i_M will tend to counteract the reduction in a's concentration. Thus an activator whose gradient is in the opposite direction to the border cannot aid its formation. Similarly a repressor whose concentration is decreasing from i_1 to i_M, that is, a repressor gradient in the same direction as the border, will not set its position.

Such simplification is also possible for anterior borders (increasing from i_1 to i_M). Activator gradients in the opposite direction to the border, and repressor gradients in the same direction as the anterior border cannot set its position.

Following this method we analyzed the six borders that show accuracy in the nonautonomous gap gene model. For graphical representation, we plot the spatial derivative of u^a and the regulatory contributions to a gene, such that the area between curves is proportional to the regulatory contribution of an input. The analysis is shown in Figure 3.3a for the posterior border of the *hb* anterior domain, Figure 3.4a for the posterior border of the central *Kr* domain, Figures 3.5a and 3.6a for the anterior and posterior *kni* borders respectively, and Figures 3.7a and 3.8a for the *gt* anterior and posterior borders respectively. The results are consistent with previous work (Jaeger et al., 2004a; Perkins et al., 2006). We briefly summarize them below; see captions of these figures and (Jaeger et al., 2004a; Perkins et al., 2006) for details.

In general, these borders are set up by an activator (either *bcd* or *cad*) and two repressors. There are two tiers of repression. Genes with complementary domains in the modeled region, *Kr* and *gt* for example, have strong mutual repression (Figure 3.1b). Genes with overlapping domains in the modeled region, *gt* and *kni* for example, have weak mutual repression. The only exception to this rule is that the anterior border of the posterior *gt* domain has only one repressor, *Kr*. This is well supported by the very large anterior shift of this border in Kr^- embryos (Kraut and Levine, 1991a).

We studied the effects of varying the Bcd profile in the gap gene circuit as follows. For all borders except *gt*'s, the positions at which the circuits formed a particular boundary were divided into 1% EL bins. Circuits with boundary positions falling in the same bin were grouped together. We averaged the repression levels at the border for each group. Similarly, we averaged the *bcd* activation at the border for each group.

The pooled Bcd activation and gap repression levels at gap gene borders are plotted for the posterior border of anterior *hb* domain (Figure 3.3b,c), the posterior border of the central *Kr* domain (Figure 3.4b,c), the anterior border of the abdominal *kni* domain (Figure 3.5b,c), and the posterior border of the

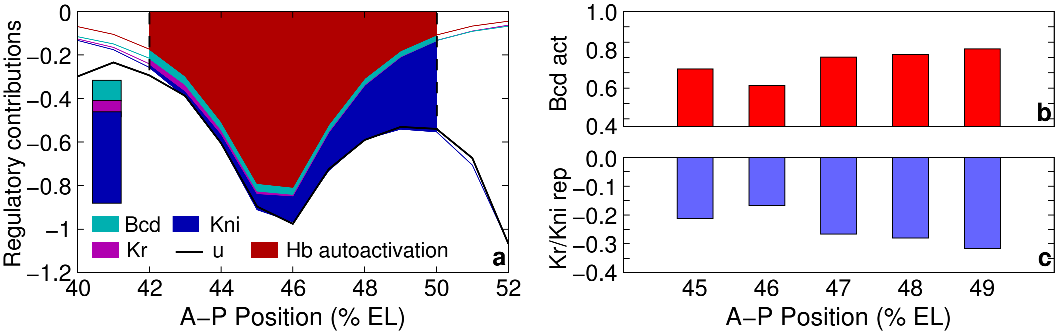

Figure 3.3: Regulatory analysis of the posterior border of the anterior *hb* domain. (**a**) Spatial derivatives of regulatory contributions to the synthesis term of *hb* on its border. Dashed vertical lines are positions where Hb is at 90% and 10% of peak level. The solid black line is the discretized spatial derivative of the total regulatory input u^{hb} to *hb*. The area above u^{hb} is the total change in u^{hb} that drives the formation of the border. The colored lines are spatial derivatives of components of u^{hb} such that the area between curves corresponds to the regulatory contribution of a particular input. *bcd* activation is cyan, *Kr* repression is magenta, *kni* repression is blue, and *Hb* autoactivation is red. Together these four account for all of the change in u^{hb} since they cover the area above u^{hb}. Of these inputs, *hb* autoactivation does not set the border, but merely sharpens it. The colored bar shows the relative contribution of the remaining three external inputs to *hb*. *Kr* and *kni* provide a relatively large contribution in setting the posterior border of the anterior *hb* domain. (**b, c**) Balance of activation and repression at the posterior border of the anterior *hb* domain in 88 simulations. Circuits that form the border within a 1% EL bin were grouped together. Each bar shows the average Bcd activation (red) and combines Kr/Kni repression (blue) for each group. x-axis is bin position.

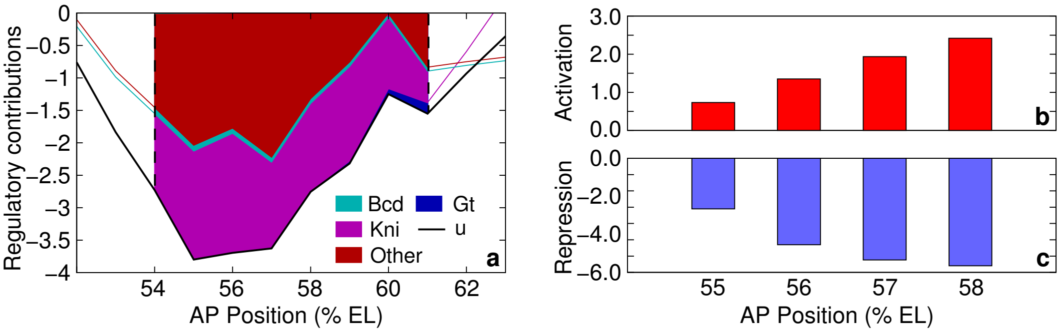

Figure 3.4: Regulatory analysis of the posterior border of the central *Kr* domain. (**a**) Spatial derivatives of regulatory contributions to the synthesis term of *Kr*. Dashed vertical lines are positions where Kr is at 90% and 10% of peak level. The black curve is the discretized spatial derivative of u^{Kr}, and the colored area above it is total change in u^{Kr} that drives Kr from a 90% expression level to a 10% level. The colored lines are spatial derivatives of individual regulators. The colored area between the lines is the total change in an individual regulatory input. Regulators that cannot set the *Kr* boundary are lumped together (red area). The *Kr* posterior border is set by Bcd activation (cyan), and repression from Kni (magenta) and Gt (blue). (**b, c**) Balance of activation (**b**) and repression (**c**) at the border in 88 simulations. Circuits that form the border within a 1% EL bin were grouped together. Each bar shows the average Bcd activation (red) and Kni/Gt repression (blue) for each group. x-axis is the bin position.

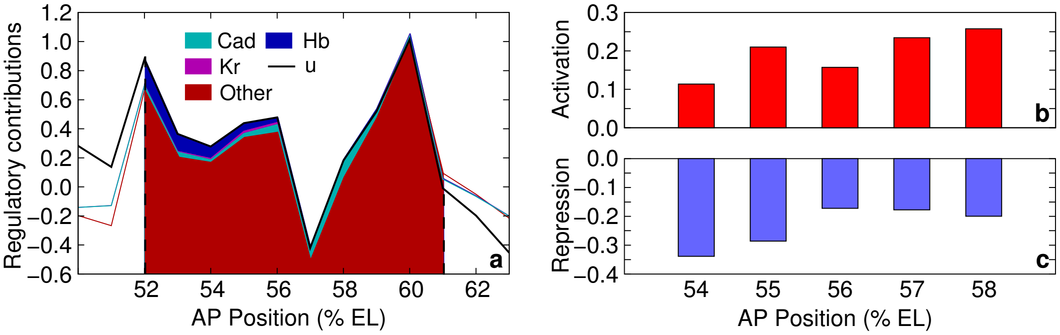

Figure 3.5: Regulatory analysis of the anterior border of the abdominal *kni* domain. (**a**) Spatial derivatives of regulatory contributions to the synthesis term of *kni*. Dashed vertical lines are positions where Kni is at 90% and 10% of peak level. The black curve is the discretized spatial derivative of u^{kni}, and the colored area below it is total change in u^{kni} that drives Kni from a 10% expression level to a 90% level. The colored lines are spatial derivatives of individual regulators. The colored area between the lines is the total change in an individual regulatory input. Regulators that cannot set the *kni* boundary are lumped together (red area). *kni* anterior border is set by Cad activation (cyan), and repression from Kr (magenta) and Hb (blue). (**b,c**) Balance of activation (**b**) and repression (**c**) at the *kni* anterior border in 88 simulations. Circuits that form the *kni* border within a 1% EL bin were grouped together. Each bar shows the average Bcd activation (red) and Kr/Hb repression (blue) for each group. x-axis is the bin position.

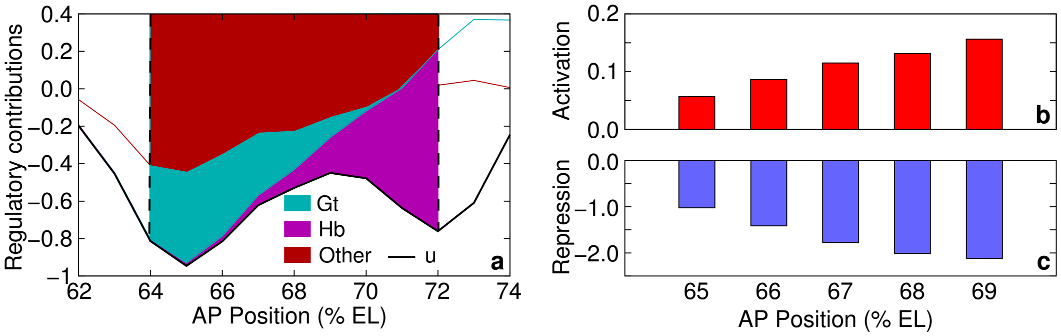

Figure 3.6: Regulatory analysis of the posterior border of the abdominal *kni* domain. (**a**) Spatial derivatives of regulatory contributions to the synthesis term of *kni*. Dashed vertical lines are positions where Kni is at 90% and 10% of peak level. The black curve is the discretized spatial derivative of u^{kni}, and the colored area above it is total change in u^{kni} that drives Kni from a 90% expression level to a 10% level. The colored lines are spatial derivatives of individual regulators. The colored area between the lines is the total change in an individual regulatory input. Regulators that cannot set the *kni* boundary are lumped together (red area). The border is set by repression from Gt (cyan) and Hb (magenta). (**b, c**) Balance of activation (**b**) and repression (**c**) at the border in 88 simulations. Circuits that form the *kni* border within a 1% EL bin were grouped together. Each bar shows the average Bcd activation (red) and Gt/Hb repression (blue) for each group. x-axis is the bin position.

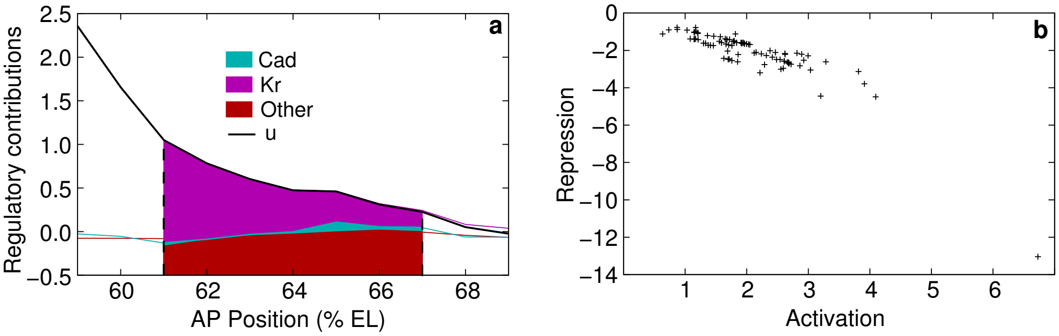

Figure 3.7: Regulatory analysis of the anterior border of the posterior *gt* domain. (**a**) Analysis of the circuit with median Bcd. Spatial derivatives of regulatory contributions to the synthesis term of *gt*. Dashed vertical lines are positions where Gt is at 10% and 90% of peak level. The black curve is the discretized spatial derivative of u^{gt}, and the colored area below it is total change in u^{gt} that drives Gt from a 10% expression level to a 90% level. The colored lines are spatial derivatives of individual regulators. The colored area between the lines is the total change in an individual regulatory input. Regulators that cannot set the *gt* boundary are lumped together (red area). *gt* anterior border is set by Cad activation (cyan) and Kr repression (magenta). (**b**) Balance of activation (Bcd) and repression (Kr) at the *gt* anterior border in 88 simulations.

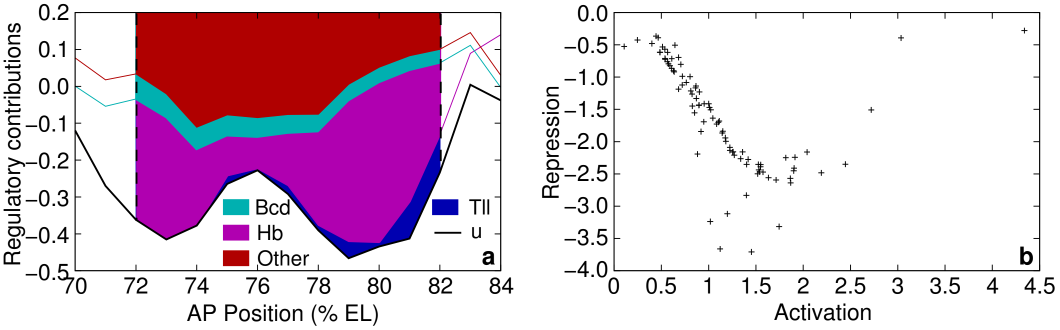

Figure 3.8: Regulatory analysis of the posterior border of the posterior *gt* domain. (**a**) Spatial derivatives of regulatory contributions to the synthesis term of *gt*. Dashed vertical lines are positions where Gt is at 90% and 10% of peak level. The black curve is the discretized spatial derivative of u^{gt}, and the colored area above it is total change in u^{gt} that drives Gt from a 90% expression level to a 10% level. The colored lines are spatial derivatives of individual regulators. The colored area between the lines is the total change in an individual regulatory input. Regulators that cannot set the *gt* boundary are lumped together (red area). *gt* posterior border is set by Bcd activation (cyan), and repression from Hb (magenta) and Tll (blue). (**b**) Balance of activation (Bcd) and repression (Hb/Tll) at the *gt* posterior border in 88 simulations.

abdominal *kni* domain (Figure 3.6b,c).

We assayed the average Bcd activation levels in each 1% EL pool of circuits. The Bcd activation levels increase with A–P position for three borders, showing that greater Bcd levels cause borders to form at more posterior positions. These borders are the posterior border of the anterior *hb* domain (Figure 3.3b), the posterior border of the central *Kr* domain (Figure 3.4b), and the posterior border of the abdominal *kni* domain (Figure 3.6b. The average repression levels at these borders show a concomitant increase with position in each pool as well (panels c of Figures 3.3, 3.4, and 3.6), and hence modulate the effect of Bcd.

At the *kni* anterior border, while pools of circuits with higher Bcd levels cause the border to form further posterior, the effect is modulated by repression levels decreasing with A–P position. This is suggestive of the position being set repressively by the domains immediately anterior of the abdominal *kni* domain, namely, *Kr* and *hb*.

At the two borders of *gt*, Bcd activation levels are no longer related to position. In general the relationship between Bcd levels and position becomes weaker at more posterior positions. The borders are set primarily by repression by other gap genes. However, the borders still form at a balance of activation and repression levels (panel b of Figures 3.7 and 3.8).

The correlation of Bcd level with position at more anterior borders suggests that Bcd provides positional information at these positions in the embryo. However, the shift induced by Bcd is limited by a simultaneous change in repression by other gap genes. With more posterior positions, the relationship between Bcd level and position becomes weaker, until at the borders of the posterior *gt* domain, there is none at all. Hence, Bcd does not provide positional information to the posterior. The borders in this region are set primarily by repression by other gap genes.

The epistasis of the gap genes provides genotypic canalization. This is an instance of the epigenetic landscape being structured by the interactions of the gap genes such that variations in the genotype, that is, Bcd, do not affect the developmental trajectory.

The canalization of the gap gene system also implies that gap gene mutants will exhibit greater variability (see Section 1.1). Houchmandzadeh et al. (2002) performed just such an experiment. They measured the variability of the posterior border of the anterior *hb* domain in Kr^-, kni^-, and gt^-. The border was found to have exactly the same variance as in wild type. This experiment directly contradicts the results of this chapter. If the border is set by *Kr* and *kni*, then removing either of them should result in increased variability. In the next chapter, an experiment is presented that attempts to resolve this dilemma.

Chapter 4

Kr/kni double mutants

In the previous chapter we showed that the gap gene circuit (Section 2.1) is able to reproduce the canalization of gap gene boundary positions. We also found that this property of the circuit arises due to the epigenetic balance between activating and repressing factors at gap gene boundaries (Section 3.3). In particular, it was shown that the posterior border of the anterior *hb* domain is buffered from *bcd* variation due to the repressive influence of *Kr* and *kni*. This contradicts the result of Houchmandzadeh et al. (2002) that the posterior border of the anterior *hb* domain has the same variation as wild type in embryos mutant for *Kr* or *kni*.

It is possible that both *Kr* and *kni* are required to balance the variation induced by *bcd*. We checked this possibility by measuring the variation of the posterior border of the anterior *hb* domain in embryos mutant for both *Kr* and *kni*. In this chapter, we describe the cross carried out to make such mutant embryos, the acquisition of expression data from these embryos, and the methods for calculating positional variation.

4.1 Double mutant experiment

4.1.1 The cross

We made embryos null for both Kr and Kni proteins by the cross described below. For *Kr*, we used flies carrying an amorphic allele of *Kr*, Kr^1, that does not have any Kr protein. For *kni*, the flies had a deficiency (deletion), *Df(3L)ri–79c*, in the *kni* locus, from chromosome segment 77B7–77F5 (FlyBase, Crosby et al., 2007).

The *Kr* allele and the *kni* deletion are recessive lethal. Therefore, flies carrying them are kept in what are known as balanced stocks. In such flies, the chromosome homologous to the one carrying the mutation is a balancer chromosome. Balancer chromosomes have multiple inversions that prevent recombination with their homolog. This way the mutation cannot be lost through recombination. The balancer also has dominant markers, that allow the unambiguous identification of flies that receive the recessive mutation in a cross. In the balancers used here, the dominant marker is also recessive lethal, which means that the flies can be maintained as a heterozygous stock.

Because the Kr^1, and *Df(3l)ri–79c* flies are kept in a balanced stock, $\frac{1}{4}$ of their embryos are homozygous. In a double mutant therefore, $\frac{1}{16}$ of the embryos are homozygous for both Kr^1 and *Df(3l)ri–79c.* Also due to the recessive lethality of these mutations, a method is required to unambiguously differentiate the genotypes of the embryos. This was accomplished by staining for both Kr and Kni proteins. Double mutants will not show any signal.

The cross was carried out as follows:

$$\begin{array}{ccc} \dfrac{Kr^1}{SM6a} & \times & \dfrac{Df(3L)ri\text{–}79c}{TM3,\,Sb^1} \\ \text{☿} & & \text{♂} \\ & \Big\downarrow & \end{array} \tag{Cross 1}$$

$$\frac{Kr^1}{+};\frac{Df(3L)ri\text{–}79c}{+} \quad \frac{Kr^1}{+};\frac{TM3,Sb^1}{+} \quad \frac{SM6a}{+};\frac{Df(3L)ri\text{–}79c}{+} \quad \frac{SM6a}{+};\frac{TM3,Sb^1}{+}$$

$$\begin{array}{ccc} \dfrac{Kr^1}{+};\dfrac{Df(3L)ri\text{–}79c}{+} & \times & \dfrac{Kr^1}{+};\dfrac{Df(3L)ri\text{–}79c}{+} \\ \text{♀} & & \text{♂} \\ & \Big\downarrow & \end{array} \tag{Cross 2}$$

$$\frac{Kr^1}{Kr^1};\frac{Df(3L)ri\text{–}79c}{Df(3L)ri\text{–}79c} \quad \text{embryos}$$

Here, *SM6a* is the balancer for the second chromosome and carries the dominant visible marker Cy^1. Flies that have this balancer have curled-up wings and can be easily identified. *TM3* is the balancer for the third chromosome, and carries the marker Sb^1, that is easily identifiable as short, thick bristles on the back of the fly. The symbol + signifies a wild type chromosome.

In the first cross (Cross 1), the visible markers allowed us to pick out flies that were heterozygous for both Kr^1 and *Df(3L)ri–79c*. The second cross (Cross 2), was carried out in a fly cage. The embryos were collected and fixed according to standard protocols (Kosman et al., 1998). These embryos were expected to have the genotype $\frac{Kr^1}{Kr^1}; \frac{Df(3L)ri\text{–}79c}{Df(3L)ri\text{–}79c}$ with a frequency 1/16.

4.1.2 Staining scheme and confocal imaging

The embryos resulting from the cross described above were immunofluorescently stained for the proteins Kr, Kni, Gt, and Hb using the standard protocol (Kosman et al., 1998). The primary antibodies used were guinea pig anti-Kr, guinea pig anti-Kni, rabbit anti-Gt, and rat anti-Hb. The secondary antibodies, all from goat, were anti-guinea pig, anti rabbit, and anti-rat conjugated to Alexa Fluor 488, Alexa Fluor 555, and Alexa Fluor 647 respectively. Histone was also stained for as described in Section A.1.1. Oregon R. embryos were also stained in an identical manner, and in parallel with the mutants. It was ensured that all sera and reagents were added to mutants and wild type in the same dilution.

The confocal imaging of these embryos was done as described in Section A.1.1 with one important difference. The microscope's gain and offset were standardized on wild type embryos. This allows us to compare the intensity of mutant expression patterns with our wild type dataset (see Surkova et al., 2007, and Section A.1). The standard was remade for each staining, and only one standard was used to record data from all embryos in a staining. To set the gain for a given protein, we identified four to five wild type embryos in the time class that the protein has maximal expression in (T4 for Hb, T7 for Gt). For each embryo we set the gain such that a few pixels saturated, that is, had value 255. Next we compared the gain setting for the wild type embryos identified for standardization. The brightest embryo requires the least gain to saturate the pixels, and hence was chosen as the standard embryo.

The first channel contained both Kr and Kni signal. Absence of signal in that channel allowed us to identify embryos homozygous for both Kr^1 and *Df(3L)ri–79c* unambiguously. Such embryos in lateral orientation were imaged for Gt, Hb, and histone staining in a transverse section. The membrane was also imaged in the sagittal plane on the dorsal and ventral sides for time classification.

4.1.3 Data processing and boundary position measurement

The segmentation and background removal for these embryos was done in exactly the same manner as was done for the segmentation genes' integrated dataset (see Sections A.1.2 and A.1.4). For time classification, we had to rely purely on morphological cues such as size and shape of nuclei, invagination of membrane, (Section A.1.3), as the embryos were not stained for *eve.* No registration or averaging was done (Section A.1.5) since we are interested in the variation of boundary positions in these data.

Boundary position measurement was done in two steps. First, we fit a quadratic spline to the one-dimensional profile derived from a 10% D–V strip of the embryo (Myasnikova et al., 2001). Using the spline, we located the peaks of the *gt* domain, or the parasegment four stripe in late *hb* pattern (see Figure 4.1). Then the domain border, or the posterior border of the anterior *hb* domain was calculated as the point on the border where expression was half the peak level.

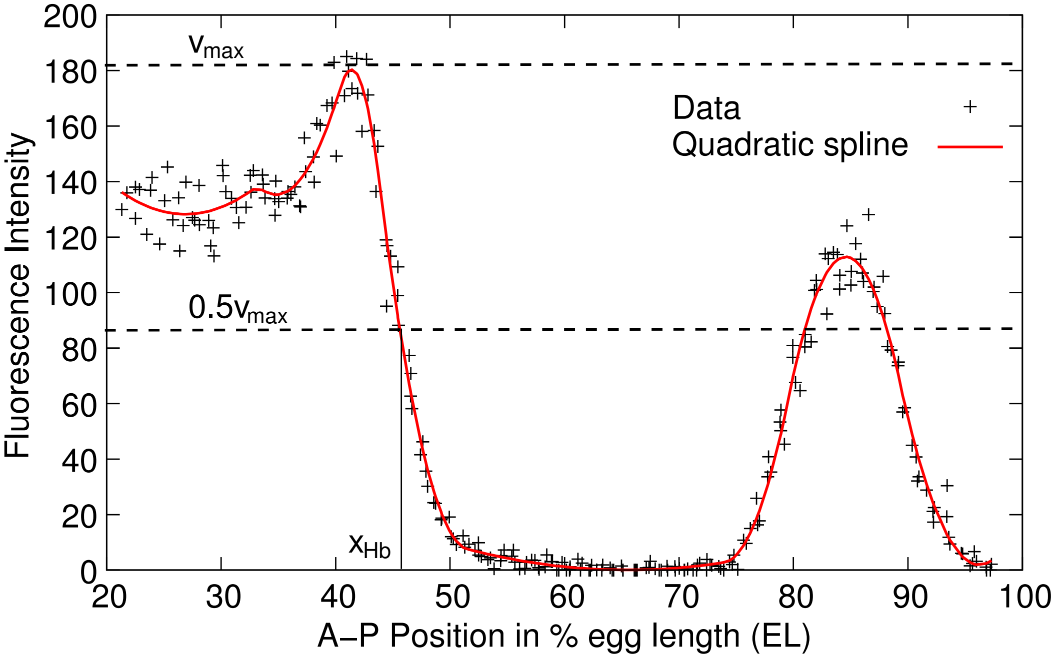

Figure 4.1: Determination of domain boundary position from data. Hb expression data from a 10% D–V strip are shown as points. A quadratic spline fit to these data is shown in red. v_{max} is the local maximum immediately anterior to the posterior border of the anterior *hb* domain. The border position, x_{Hb}, is calculated as the point where the spline crosses $v_{max}/2$.

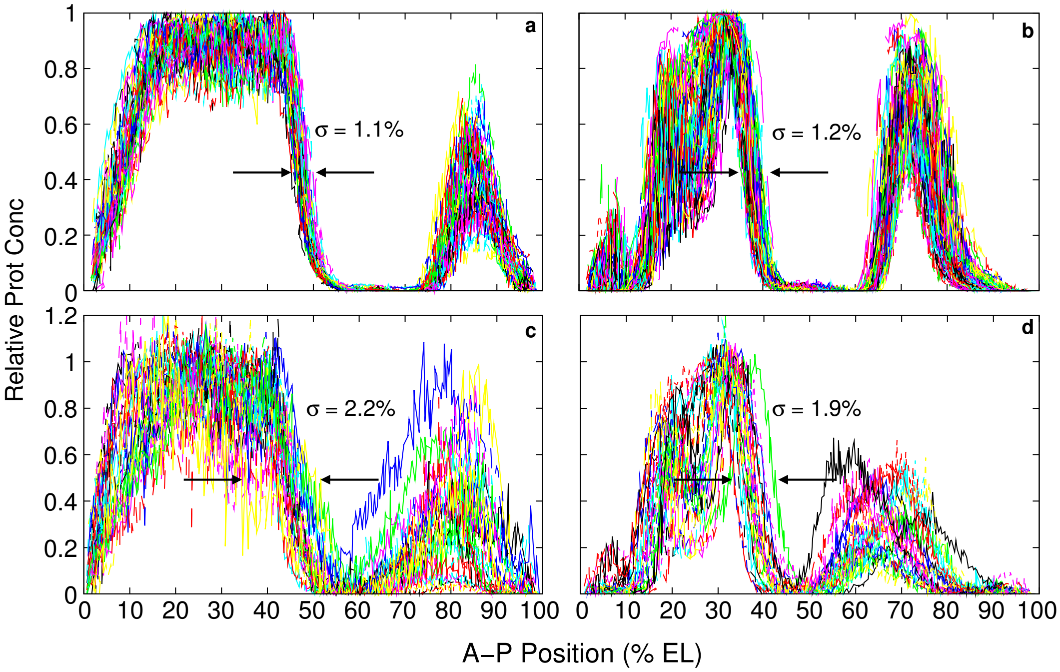

Figure 4.2: Variation of *hb* and *gt* expression patterns in *Kr*/*kni* double mutants. (**a**) Wild type Hb expression in 68 embryos. (**b**) Wild type Gt in 92 embryos. (**c**) Hb in 28 *Kr*/*kni* double mutant embryos. (**d**) Gt in 28 *Kr*/*kni* double mutant embryos. All data are from a 10% D–V strip along the A–P axis. All embryos are from time classes T4–T7. Wild type data are from FlyEx.

4.2 Results

The expression patterns of *hb* and *gt* in Kr^-/kni^- embryos in time classes T4–T7 is shown in Figure 4.2. We found that the standard deviation of the position of the posterior border of the anterior *hb* domain was 2.2% EL compared to a wild type value of 1.1% EL. The range increased to 9.3% EL from 4.9% EL in wild type. The standard deviation of the *gt* third stripe's position also increases from 1.2% EL to 1.9% EL.

In *Kr*/*kni* double mutants, *hb* mRNA data show derepression of *hb* expression between the PS4 stripe and the posterior domain during late cycle 14A

(Clyde et al., 2003). PS4 stripe and posterior domain expression arises from the promoter P1 of *hb*, and is not thought to be under the control of Bcd.

This leads to the question of whether the increased variability in *hb* presented here reflects Bcd variability or not. These protein data are from time classes T4–T7, which reflect the transcription of ∼12 minutes earlier (Section 2.3.1 and Table 2.2) due to transcription delays. Therefore, these protein expression patterns are from the transcripts produced from T2–T5. The transcripts that form the PS4 stripe start to appear only during mid cycle 14A (Schröder et al., 1988), and it is visible in integrated protein data only in T7 (Surkova et al., 2007). Furthermore, while the region between the PS4 stripe and the posterior Hb domain is completely derepressed in the data presented in Clyde et al. (2003), the anterior and posterior domains are separate in the data shown in Figure 4.2. Therefore, it may be concluded that the variation of Hb seen in the double mutants reflects the variability of the Bcd gradient, rather than the derepression of *hb* between the PS4 stripe and the posterior domain.

The results of this chapter show that the variability of the posterior border of the anterior *hb* domain is under combined maternal and zygotic control, and not under the exclusive control of Bcd.

Chapter 5

Qualitative Dynamics

In Chapters 3 and 4, it was shown that gene circuits show genotypic canalization with respect to Bcd. The stability of developmental trajectories is a necessary condition for canalization (Section 1.1). Quantitative gene expression data show that the variation of gap gene expression patterns reduces over time in cleavage cycle 14A (Section 1.3). In this chapter, the reduction of initial variation is studied using the techniques of qualitative dynamics. It is shown that the reduction in initial variation is due to the presence of stable developmental trajectories, satisfying a necessary condition for canalization.

Qualitative dynamics is a technique for studying nonlinear dynamical systems. It is usually not possible to express solutions of systems of nonlinear first-order ODEs in terms of exact functions. However insight can be gained into the behavior of solutions by studying the structure of the ODE's phase space (Hirsch et al., 2004; Perko, 1996). The phase space contains subsets that are mapped to themselves, and hence as a whole are invariant under the flow defined by the ODE. The simplest such set is a single point, an equilibrium solution of the ODE. More complicated invariants that have geometrical structure are also possible. The geometry of these invariant sets reflects the underlying structure of the phase space that guides the solutions and constrains their behavior. Thus, calculating such invariant sets, starting with equilibria, allows one to understand the solutions' behavior in the absence of analytical solutions. Such an analysis is termed qualitative dynamics (Thom, 1983).

Eq. (2.1) is a set of $M \times N$ coupled ODEs. In the gap gene circuit used in this study (Section 2.1.1), there are 58 nuclei and 4 gap genes. Such a large number of dimensions pose a significant visualization challenge for the analysis of the phase space. In order to make this analysis tractable, we considered the approximation that there is no diffusion, that is, $D^a = 0$ in Eq. (2.1).

5.1 Diffusionless approximation and positional information

It has been noted before that in autonomous gap gene circuits, diffusion does not play a role in patterning except for smoothing gap domains (Jaeger et al., 2004b). The same holds true for nonautonomous circuits. Diffusion was turned off in the circuit by setting the diffusion coefficients D^a to zero, while keeping all other parameters constant. Figure 3.1 shows the comparison of the gap gene patterns produced by the circuits and their diffusionless approximation.

Setting $D^a = 0$ in Eq. (2.1) uncouples the nuclei. The system of $M \times N$ coupled ODEs reduces to a set of M independent systems of N ODEs

$$\frac{dv_i^a}{dt} = R^a g\left(\sum_{b=1}^{N} T^{ab} v_i^b + m^a v_i^{\mathrm{Bcd}} + \sum_{\beta=1}^{N_e} E^{a\beta} v_i^{\beta}(t) + h^a\right) - \lambda^a v_i^a. \tag{5.1}$$

The diffusionless circuit (Eq. 5.1) has the correct progression of gap-gene

Boundary	**Position with diffusion (% EL)**	**Position without diffusion (% EL)**
hb posterior	46	45–46
Kr anterior	43	45–46
Kr posterior	56	57–58
kni anterior	56	57–58
kni posterior	66	63–64
gt anterior	65	63–64

Table 5.1: Position of gap gene boundaries in the circuit with (first column) and without (second column) diffusion in time class T8. The position of a domain boundary is the nucleus in which the concentration is half its maximum value in the domain. For the diffusionless case, the domain boundaries are steep (see Figure 3.1). Hence, the pair of nuclei between which the concentration changes the most are shown in the second column.

domains from anterior to posterior. The domain border positions are within 3% EL (Table 5.1) of those in the circuit with diffusion. It also captures the key dynamical feature of domain shifts (Jaeger et al., 2004b). There are three main differences in gap gene patterns between circuits having, or not having diffusion. First, whereas in circuits with diffusion, adjacent gap gene domains overlap by about 5% EL during late cycle 14A, they are almost mutually exclusive in the diffusionless circuit. Second, patterns are spatially smoother in the presence of diffusion. Third, in the diffusionless circuit the most anterior nucleus (35% EL) has an incorrect state (see Section 5.4.1 for discussion) that is smoothed out in the presence of diffusion.

In Eq. (5.1), each nucleus i has a system of N ODEs describing its state. Since i depends on A–P position x, we can write its state as $v_i^a = v^a(x,t)$, where $a = 1, \ldots, N$. We can also rewrite $v_i^{\mathrm{B}cd} = v^{\mathrm{B}cd}(x)$, and $v_i^{\beta}(t) = v^{\beta}(x,t)$. Eq. (5.1) then becomes

$$\frac{dv^a}{dt} = R^a g\left(\sum_{b=1}^{N} T^{ab}v^b + m^a v^{\mathrm{Bcd}}(x) + \sum_{\beta=1}^{N_e} E^{a\beta}v^{\beta}(x,t) + h^a\right) - \lambda^a v^a. \quad (5.2)$$

In Eq. (5.2), A–P position only appears through the terms $v^{\mathrm{B}cd}(x)$ and $v^{\beta}(x,t)$. In other words, in the absence of diffusion position is directly parameterized by the concentrations of Bcd, Cad, and Tll, $v^{\mathrm{B}cd}(x)$, $v^{\mathrm{C}ad}(x,t)$, and $v^{\mathrm{T}ll}(x,t)$. There is one more source of spatial asymmetry that provides positional information in the gene circuit: the initial conditions specified by the maternal Hb gradient (see Section 2.1.2 and Figure 5.10b). As mentioned before (Section 2.1.3), these factors represent the three maternal systems in the circuit and specify the anteroposterior position of a nucleus.

The inputs to the synthesis term and the initial conditions have distinct roles in positional specification. Bcd, Cad, and Tll concentrations act as parameters of the dynamical system defined by Eq. (5.2). The dynamical structure of the phase space is determined by these inputs. In contrast, the initial conditions function by selecting a particular trajectory from the many that are possible in the phase space.

The dynamical analysis of gap gene circuits proceeds from anterior to posterior by varying three parameters—Bcd, Cad and Tll—according to nuclear position. To further simplify the analysis, a circuit without *tll* was considered. The concentration of Tll was set to zero in the diffusionless gap gene circuit (Eq. 5.2), giving the equation,

$$\frac{dv^a}{dt} = R^a g\left(\sum_{b=1}^{N} T^{ab}v^b + m^a v^{\mathrm{Bcd}}(x) + E^{a\leftarrow\mathrm{Cad}}v^{\mathrm{Cad}}(x,t) + h^a\right) - \lambda^a v^a. \quad (5.3)$$

The patterns (Figure 5.1) of such a circuit (Eq. 5.3) are consistent with a *tll*/*hkb* double mutant (Eldon and Pirrotta, 1991, *hkb* is not represented in the model). This analysis in only valid in the region from 35% EL to 72% EL, since *tll* is required for correct pattern formation in the region posterior to 72% EL. Thus, the three-parameter analysis is simplified to a two-parameter one.

5.2 Setting up the analysis

In the previous section the full 232-dimensional gap gene circuit equations was reduced to 58 independent 4-dimensional systems of ODEs (Eq. 5.3) parameterized by Bcd and Cad concentrations. These equations are nonautonomous, hence the dynamical structure of the gene circuit changes with time. We are interested in studying the dynamical structure during late cycle 14A, since that is when the variation in gap gene patterns is least (Figure 1.2). However, there is a two-fold decrease in Bcd concentration in time classes T7 and T8 (Surkova et al., 2007), and a stationary approximation is not valid after time class T6. For this reason, the dynamical structure of the system in time class T6 was studied. This is accomplished by rewriting Eq. (5.3) such that the system is nonautonomous until time class T6, and autonomous thereafter. Eq. (5.3) is rewritten as

$$\frac{dv^a}{dt} = \begin{cases} R^a g\left(\sum_{b=1}^{N} T^{ab}v^b + m^a v^{\mathrm{Bcd}}(x) + E^{a\leftarrow\mathrm{Cad}}v^{\mathrm{Cad}}(x,t) + h^a\right) \\ \qquad -\lambda^a v^a, \quad \text{if} \quad t < t_6, \\ R^a g\left(\sum_{b=1}^{N} T^{ab}v^b + m^a v^{\mathrm{Bcd}}(x) + E^{a\leftarrow\mathrm{Cad}}v^{\mathrm{Cad}}(x,t_6) + h^a\right) \\ \qquad -\lambda^a v^a, \quad \text{if} \quad t \geq t_6. \end{cases} \tag{5.4}$$

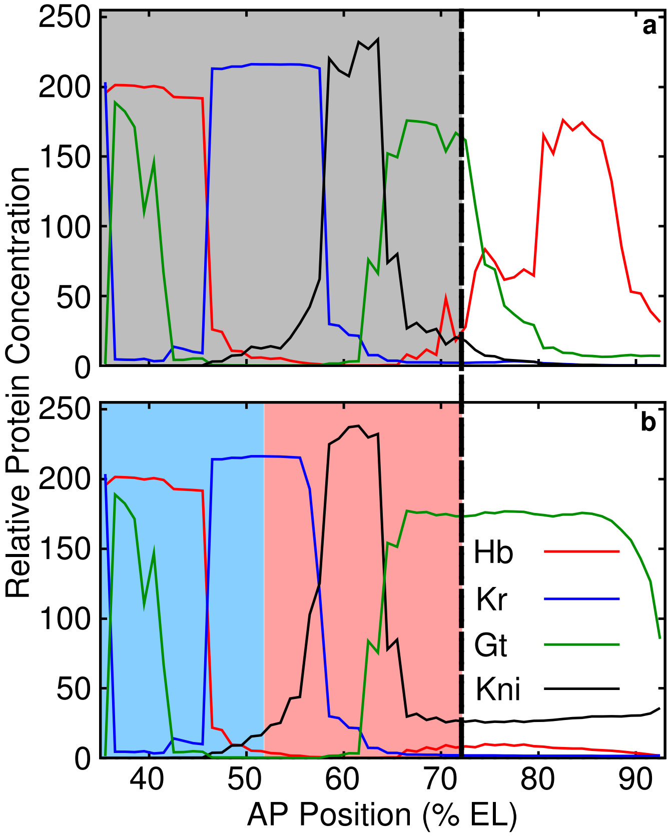

Figure 5.1: Anteroposterior region for qualitative dynamics analysis. (**a**) Patterns in the diffusionless circuit (`hkgn58c13k1_007`) that includes Tll as a nonautonomous input. (**b**) Patterns in the circuit if Tll is removed from it. The dynamics divide into two distinct regimes, anterior (highlighted in blue), and posterior (highlighted in red). Grayed area (panel **a**) anterior to dashed line shows the region (35%–72% EL) in which gap gene patterns are not affected by Tll.

Here, t_6 is the midpoint of time class T6 (see Section 2.1.4). Note that for $t \geq t_6$, $v^{\mathrm{Cad}}(x,t) \equiv v^{\mathrm{Cad}}(x,t_6)$, that is, the concentration of Cad doesn't change after T6.

The trajectories of the dynamical system were calculated with the nonautonomous set of equations (Eq. 5.4). To determine the dynamical structure at $t = t_6$, the first step is to calculate the equilibria, that is, points $\hat{v}^a$ that satisfy the condition

$$\frac{d\hat{v}^a}{dt} = 0, \quad \text{for} \quad t \geq t_6. \tag{5.5}$$

This is equivalent to setting the autonomous part of Eq. (5.4) to zero. Let $\hat{v}^{\mathrm{Cad}} = v^{\mathrm{Cad}}(x,t_6)$. Then the equilibria $\hat{v}^a$ are the solutions of a system of nonlinear coupled equations, given by

$$R^a g\left(\sum_{b=1}^{N} T^{ab}\hat{v}^b + m^a v^{\mathrm{Bcd}}(x) + E^{a\leftarrow \mathrm{Cad}}\hat{v}^{\mathrm{Cad}}(x) + h^a\right) - \lambda^a \hat{v}^a = 0. \tag{5.6}$$

5.3 Qualitative dynamics

5.3.1 Equilibria and other invariant sets

Qualitative dynamics was used to analyze the phase space for each nucleus in the region from 35% to 72% EL. For the pair of Bcd and Cad concentrations ($v^{Bcd}(x)$ and $\hat{v}^{Cad}(x)$) that position the nucleus at A–P position x, the equilibria $\hat{v} = (\hat{v}^{\mathrm{Hb}}, \hat{v}^{\mathrm{K}r}, \hat{v}^{\mathrm{Gt}}, \hat{v}^{\mathrm{K}ni})$ were calculated by solving Eq. (5.6), using the Newton-Raphson method (Conte and de Boor, 1980; Press et al., 1992). The ability of the Newton-Raphson algorithm to find zeroes depends on the starting point. To ensure that all equilibria were obtained, the Newton-Raphson iteration was started from n^4 points on a uniform grid in the four-dimensional

Equilibrium	**Existence (% EL)**	**Comments**
$A^1_{0,4}$	35–72	*hb,gt*-on attractor until 41%, *hb*-on after 41%
$A^2_{0,4}$	35–72	*hb,Kr*-on attractor
$A^3_{0,4}$	35–53	*Kr*-on attractor
$S^1_{1,3}$	35–72	
$S^2_{1,3}$	35–72	
$S^3_{1,3}$	35–37	Annihilated at 36.96% with $S^4_{2,2}$
$S^4_{2,2}$	35–37	Annihilated at 36.96% with $S^3_{1,3}$
$S^5_{1,3}$	46–54	Created at 46.14% with $S^6_{2,2}$ and annihilated at 53.64% with $A^3_{0,4}$
$S^6_{2,2}$	46–72	Created at 46.14% with $S^5_{1,3}$
$S^7_{1,3}$	53–72	Created at 53.32% with $S^8_{2,2}$
$S^8_{2,2}$	53–72	Created at 53.32% with $S^7_{1,3}$

Table 5.2: Summary of all equilibria, the A–P region they exist in, and their function. Creation and annihilation are with respect to motion from anterior to posterior. Bifurcation values were determined by the method described in Section 5.3.2.

box $(0, 250) \times (0, 250) \times (0, 250) \times (0, 250)$. This box contains the upper bounds R^a/λ^a of the solutions of Eq. (5.3). Newton-Raphson was run with a tolerance of 10^{-6}, and only a negligible number of starting points failed to converge. n was increased from 3 to 51. For all nuclei, no new equilibria were found after $n = 10$.

The equilibria were classified according to the eigenvalues of the linearized system of equations $\frac{dy}{dt} = J(\hat{v}^a)y$, where $y = v^a - \hat{v}^a$ is a 4-dimensional vector, and $J(\hat{v}^a)$ is the Jacobian of the autonomous part of Eq. (5.4) at the equilibrium point $\hat{v}^a$. The eigenvalues are complex numbers in general, and the real part determines the stability of the equilibrium. An equilibrium which has all eigenvalues with negative real parts is called an attractor. An equilibrium which has at least one eigenvalue with a positive real part and none that are zero is called a non-degenerate hyperbolic equilibrium. Degenerate equilibria, which have at least one zero eigenvalue, occur when J is singular. This situation is called a bifurcation, and number and stability of equilibria of the dynamical system change.

A non-degenerate hyperbolic equilibrium has $p < 4$ eigenvalues with positive real parts and $q = 4 - p$ eigenvalues with negative real parts. The j^{th} such point is denoted by $S^j_{p,q}$. p is called the *index* of the hyperbolic equilibrium. Associated with such a point are two invariant sets, called the *global stable manifold* and the *global unstable manifold* (Perko, 1996). The global stable manifold is the set of all points such that trajectories starting from them have the hyperbolic equilibrium as their limit as $t \to \infty$, and is of dimension $4 - p$. The global unstable manifold is the set of all starting points whose trajectories have the hyperbolic equilibrium as their limit as $t \to -\infty$, and is of dimension p. The stable and unstable manifolds of saddles of index 1 ($S^j_{1,3}$) are of particular interest. The three-dimensional stable manifold of such points forms the boundaries for basins of attraction of point attractors (Guckenheimer and Vladimirsky, 2004). A stable manifold of dimension 3 is very computationally expensive to calculate (Krauskopf and Osinga, 2005; Dellnitz et al., 2001). By comparison, calculating the one-dimensional unstable manifold of $S^j_{1,3}$ is straightforward (Guckenheimer and Vladimirsky, 2004). Two starting points were chosen, one displaced by 10^{-1} from $S^j_{1,3}$ in the direction of the eigenvector of the positive eigenvalue; the other displaced by the same amount in the opposite direction. The union of trajectories resulting from these starting points, U^j_+, and U^j_- is an approximation to the one-dimensional unstable manifold of

$S^j_{1,3}$.

Note that even though the unstable manifolds of $S^j_{1,3}$ were calculated for the autonomous part of Eq. (5.4), their definition is more general, and holds for nonautonomous systems of equations too. The same holds true for stable manifolds.

An equilibrium point that has all eigenvalues with negative real parts is called a node or a point attractor (Hirsch et al., 2004; Perko, 1996). The jth such point is denoted by $A^j_{0,4}$. There exists a set of points such that any trajectory starting within it approaches the node as $t \to \infty$ (Perko, 1996). This set is called the *basin of attraction* of the node, and the node is an *attractor* of this set.

Since the Hb axis forms the biological set of initial conditions (no Kr, Gt, and Kni protein is detected before cycle 13, see Section 2.1.2), we only characterized the intersection of the Hb axis with the basin of attraction of the nodes. In all subsequent discussion, the term "basin of attraction" is used with the special meaning that it is the intersection of the four-dimensional basin of attraction with the Hb axis. The trajectories were calculated using the hybrid nonautonomous-autonomous system (Eq. 5.4). They were started from uniformly spaced starting points on the Hb axis and integrated until $|v^a_i(t) - A^j_{0,4}| < 10^{-6}$. The set of starting points that came close to $A^j_{0,4}$ formed the preliminary characterization of its basin.

Next the open interval (B^j_1, B^j_2) on the Hb axis was explicitly calculated, within which all starting points reached the point attractor $A^j_{0,4}$. B^j_1 and B^j_2 are starting points (on the Hb axis) of singular trajectories that reach a hyperbolic equilibrium point of index 1. The set of all singular trajectories that reach a hyperbolic equilibrium of index 1 defines the boundary between basins of attraction (Guckenheimer and Vladimirsky, 2004) of the attractors of the system. Thus the B_1^j and B_2^j are the intersection of the basin boundaries with the Hb axis.

5.3.2 Continuous analysis

The calculations described in the previous section involve varying the A–P position x in Eq. (5.6) at integer values that correspond to nuclear position. This is because the concentration of Cad ($\hat{v}^{\mathrm{Cad}}(x)$) is known only at those positions in data. This leads to the question of whether appearances or disappearances of equilibria as x is varied are important for the dynamics, or artifacts of small fluctuations in the Cad data.

To answer this question, a continuation analysis was performed (O. Radulescu, unpublished data). The T6 Cad data were interpolated with a cubic polynomial (Figure 5.2). This allows the continuous variation of x in Eq. (5.6). Note that the Bcd profile has already been parameterized (Eq. 2.2) with position and can be varied continuously.

In this analysis, x was varied continuously, and bifurcations are detected when the determinant of the Jacobian became zero. Then, the bifurcation's type was determined. All the bifurcations were of saddle-node type. The bifurcations determined in this manner cross-validated the discrete analysis of the previous section—the two analyses found the same equilibria and bifurcations. Since the continuation analysis and the discrete analysis used different Cad profiles (Figure 5.2), the A–P positions at which these bifurcations occur are slightly different. The comparison is shown in Table 5.3. For simplicity, the bifurcation positions determined by continuation analysis are used in the discussion of the dynamical structure.

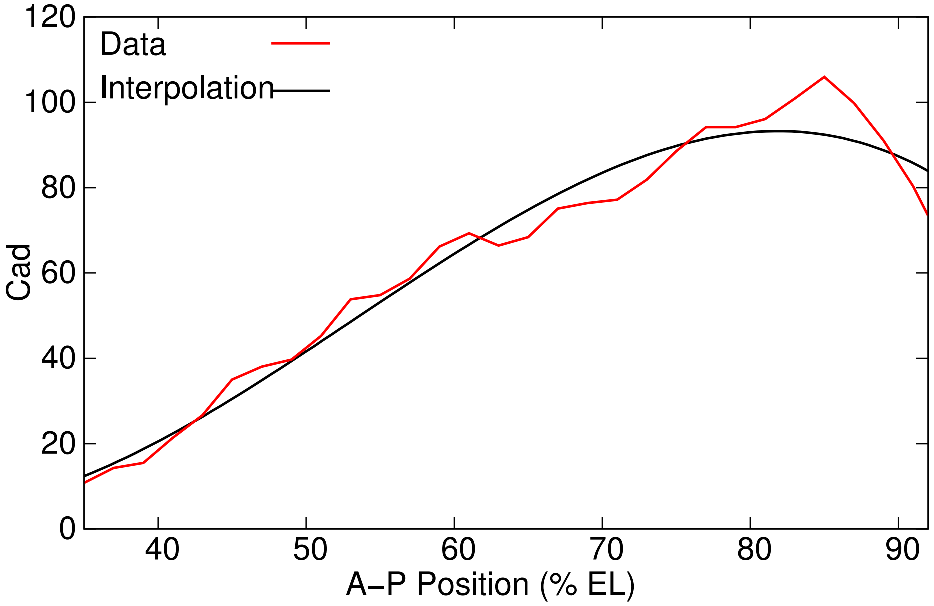

Figure 5.2: Interpolation of T6 Cad profile for continuation analysis. The interpolant (black curve) is the cubic polynomial $-0.0075x^3 + 0.2264x^2 + 2.3611x + 9.8004$.

5.4 Dynamical structure of the phase space

The region of interest from 35% EL to 72% EL can be divided into two regions based on the biological properties of the qualitative dynamics. In the anterior region, pattern formation is largely controlled by Bcd, while in the posterior it is controlled by maternal Hb. In previous work (Jaeger et al., 2004b), it was shown that dynamical shifts occur in the posterior of the embryo, but not in the anterior. The regions identified here correspond exactly to the regions of that study.

These regions are separated by a saddle-node bifurcation at 53.64% EL. Although there are other bifurcations (Table 5.3) within these regions, the bifurcation at 53.64% EL is the only one that involves an attractor, and changes the overall dynamical picture.

Equilibria involved	**Type**	**Newton-Raphson value (between %EL)**	**Continuous analysis value (at %EL)**
$S^3_{1,3}$–$S^4_{2,2}$	Annihilation	35–37	36.96
$S^5_{1,3}$–$S^6_{2,2}$	Creation	43–45	46.14
$A^3_{0,4}$–$S^5_{1,3}$	Annihilation	51–53	53.64
$S^7_{1,3}$–$S^8_{2,2}$	Creation	53–55	53.32

Table 5.3: Comparison of bifurcation parameter values determined by Newton-Raphson and continuous analysis. In the continuous analysis, A–P position x was varied from anterior to posterior. All bifurcations are saddle-node bifurcations. The type of saddle-node bifurcation is shown in the second column. The third column shows the nuclei positions between which the equilibria either appear or disappear. The last column shows the A–P positions at which these bifurcations occur for continuous approximation to the *cad* profile.

The anterior region extends from 35% EL to 53% EL, which corresponds roughly to the region between the peak of the third anterior Gt domain and the peak of the Kr abdominal domain (Figure 5.1b). The posterior region stretches from 53% EL to 72% EL, that is, from the peak of the Kr domain to the peak of the Gt posterior domain.

The dynamical regimes in the two regions are distinguished by three features. The first is the number of point attractors. The anterior region has three point attractors ($A^1_{0,4}$, $A^2_{0,4}$, $A^3_{0,4}$) and the posterior region has only two ($A^1_{0,4}$, $A^2_{0,4}$). The second distinguishing feature is the degree of sensitivity of the dynamical structure with respect to anteroposterior position. In the anterior region, as nuclear position changes, the positions of the equilibria change significantly and the sizes of basins of attraction change. The qualitative posterior dynamical structure is invariant with nuclear position, save for a single bifurcation and movements of the hyperbolic equilibria. The final difference between the two regions is in the time taken by trajectories to relax to point attractors. Whereas in the anterior region all trajectories are close to their attractors by gastrulation, the posterior region has trajectories that reach their point attractors after the biologically observable period has ended.

In this section it is shown through dynamical analysis that while the trajectories of anterior nuclei are stable due to attraction by point attractors, nuclear trajectories in posterior nuclei are stable due to attraction by a one-dimensional manifold of a hyperbolic equilibrium. It is also shown that boundaries form by three different mechanisms. (1) By the continuous dependence of an attractor on a morphogen (the posterior border of the third *gt* stripe), (2) by the crossing of a basin boundary by the initial condition (the posterior border of the anterior *hb* domain), and (3) by reaching an attracting manifold at different points (all boundaries in the posterior region). Finally, it was found that the dynamical shifts occur in the posterior region due to the attracting manifold.

Two conventions are adopted for all subsequent discussion. In the analysis of the phase space it was found that attractors were either at high concentrations or low, but not in intermediate values. Therefore, they were given descriptive names based on which proteins were at high levels (*on*) and which proteins were at low levels (*off*). Additionally, genes that are *off* in the qualitative description of equilibria are not mentioned explicitly. For example, if a point attractor is at *hb*-on, *Kr*-off, *gt*-on, and *kni*-off, it is referred to as the "*hb,gt*-on" attractor.

5.4.1 Anterior Regime 35% to 53% EL

The phase space of nuclei in the anterior region has three point attractors: $A^1_{0,4}$, $A^2_{0,4}$, and $A^3_{0,4}$. $A^1_{0,4}$ is in *hb,gt*-on state at 35% EL. Gt expression at this attractor reduces smoothly with AP position (compare panels b and c of Figure 5.3), until it becomes *hb*-on at 43% EL, and remains in that state at further posterior positions. By contrast, $A^2_{0,4}$ and $A^3_{0,4}$ are in the same state in all the nuclei of the anterior region. $A^2_{0,4}$ is in the *hb,Kr*-on state and $A^3_{0,4}$ is in the *Kr*-on state.

At 35% EL, there are four hyperbolic equilibria: $S^1_{1,3}$, $S^2_{1,3}$, $S^3_{1,3}$, and $S^4_{2,2}$ (see Figure 5.3a). At 36.96% EL, $S^3_{1,3}$ and $S^4_{2,2}$ annihilate each other through a saddle-node bifurcation (compare panels a and b of Figure 5.3), leaving only $S^1_{1,3}$ and $S^2_{1,3}$. Two new saddles, $S^5_{1,3}$ and $S^6_{2,2}$, are created by another saddle-node bifurcation (compare panels c and d of Figure 5.3) at AP position 46.14%, and remain for the rest of the anterior region. $S^1_{1,3}$ is contained in the Hb-Kr plane, at the *hb*-on edge (Figure 5.3). *Kr* expression at this equilibrium point increases with AP position, that is, the saddle moves towards the *hb,Kr*-on attractor $A^2_{0,4}$. The $S^2_{1,3}$ hyperbolic equilibrium is also contained in the Hb-Kr plane, at the *Kr*-on edge. With increasingly posterior AP position, *hb* expression increases; $S^2_{1,3}$ also moves towards $A^2_{0,4}$.

The stable manifolds of the hyperbolic equilibria of index 1 ($S^1_{1,3}$, $S^2_{1,3}$) form the boundaries of the basins of attraction of the point attractors ($A^1_{0,4}$, $A^2_{0,4}$, $A^3_{0,4}$). By calculating singular trajectories from the Hb axis that reach such hyperbolic equilibria, it was found that the basins of $A^1_{0,4}$ and $A^2_{0,4}$ are separated by the stable manifold of $S^1_{1,3}$. Similarly, the basins of $A^2_{0,4}$ and $A^3_{0,4}$ are separated by the stable manifold of $S^2_{1,3}$. The basins of the three attractors in the anterior region were determined (see Table 5.4). Trajectories within a basin reach close to their attractors by gastrulation (see Figure 5.4), and hence show stability.

The anteroposterior progression of states in the diffusionless model is then

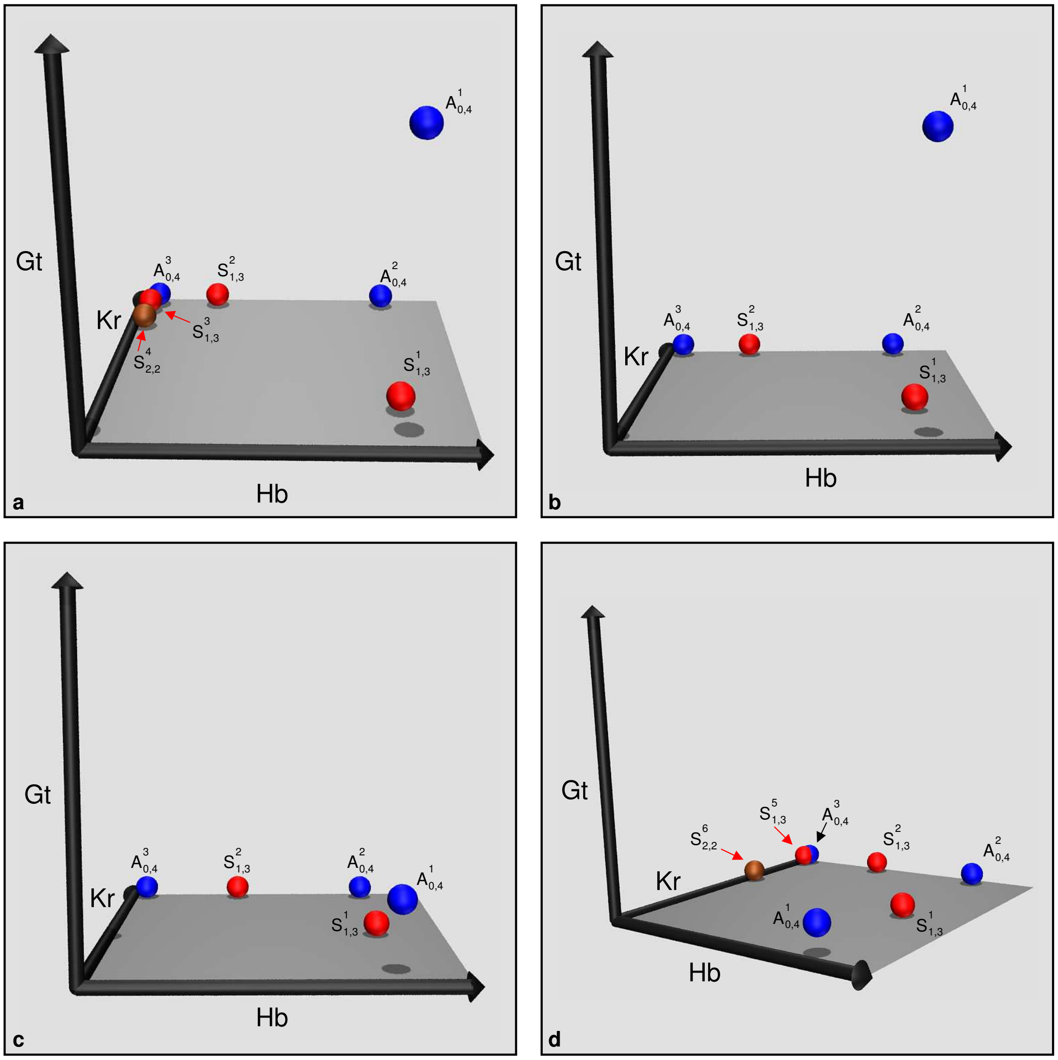

Figure 5.3: Bifurcations in the anterior region. Hb-Kr-Gt projection of equilibria diagrams at (**a**) 35% EL, (**b**) 37% EL, (**c**) 43% EL, and (**d**) 45% EL.The axes originate from $(-10, -10, -10)$, and have length 250 in relative concentration units. The xy-plane is shown in gray. To aid perception of depth, shadows from a light source directly above the xy-plane are rendered as dark gray traces on the xy-plane. Equilibria are represented by spheres of radius 10. Point attractors are blue, hyperbolic equilibria of index 1 are red, and hyperbolic equilibria of index 2 are brown. Red arrows in panel (**a**) point to saddles, $S^3_{1,3}$ and $S^4_{2,2}$, that disappear through a saddle-node bifurcation between 35% EL and 37% EL. In panels (**b**) and (c), the $A^1_{0,4}$ attractor goes from *hb,gt*-on state to *hb*-on state. Red arrows in panel (**d**) point to two saddles, $S^5_{1,3}$ and $S^6_{2,2}$ created by a saddle node bifurcation between 43% EL and 45% EL. $S^5_{1,3}$ and $A^3_{0,4}$ disappear through a saddle node bifurcation at 53% EL that separates the anterior and posterior regimes.

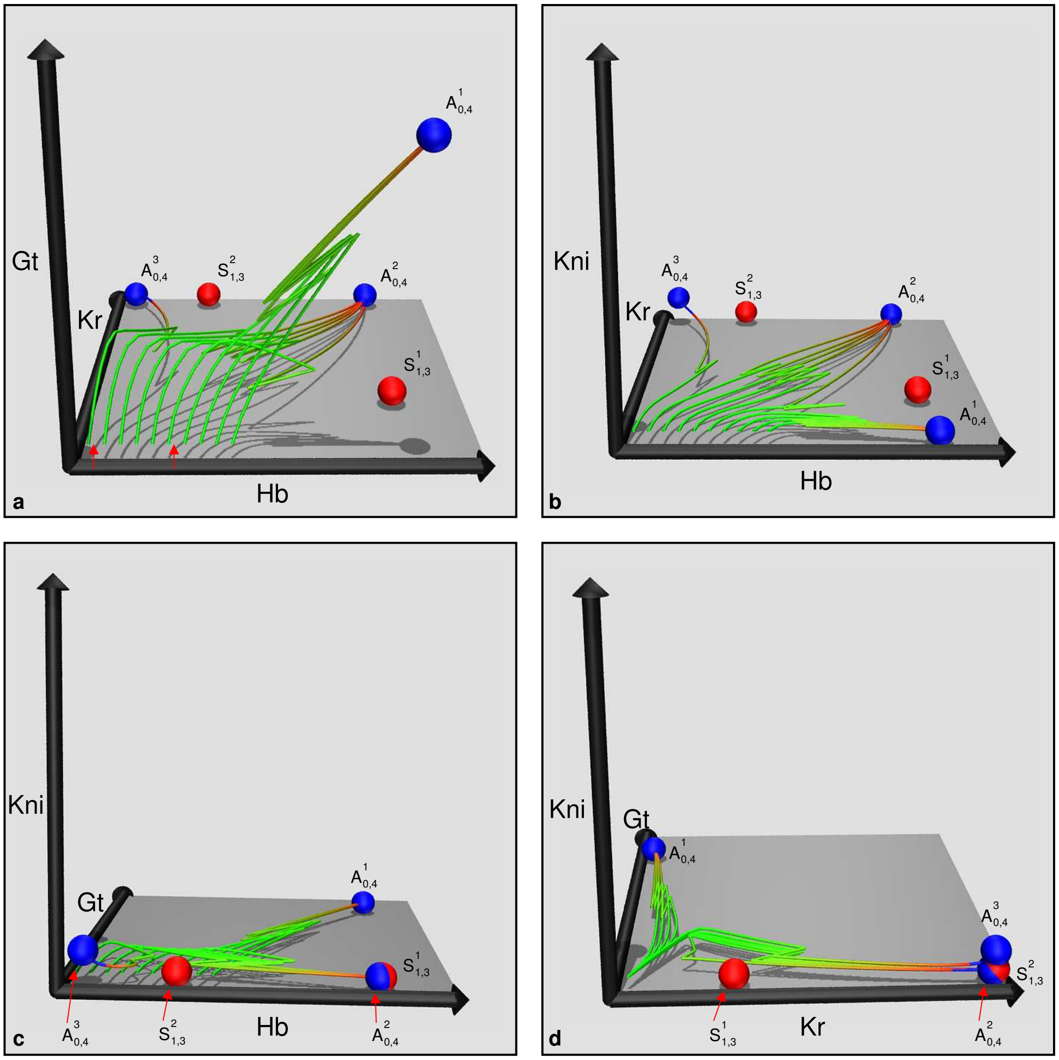

Figure 5.4: All four three-dimensional projections of the phase portrait at 37% EL. (**a**) Hb-Kr-Gt projection; red arrows are basin boundaries. (**b**) Hb-Kr-Kni projection. (**c**) Hb-Gt-Kni projection. (**d**) Kr-Gt-Kni projection. The axes, xy-plane, and equilibria are as in Figure 5.3. 10 trajectories are shown with starting points equally distributed on the Hb axis between 0–100. Time is represented as a color gradient along the trajectories, with start of cycle 13 as green, and gastrulation as red; trajectories are blue after gastrulation. The temporary reversals in trajectories are mitoses, during which trajectories move towards the origin.

determined by which basin the maternal *hb* gradient passes through. As an example, we illustrate the case of the nucleus at AP position 37%. The basin of attraction (range of $v^{\mathrm{H}b}(t = 0)$) of $A^3_{0,4}$ is $(0.00, 2.05)$. $A^2_{0,4}$ has a basin of $(2.05, 52.12)$, and $A^1_{0,4}$'s basin is $(52.12, 100.00)$. In our circuit `hkgn58c13k1_007`, the value of maternal *hb* in this nucleus is 53.63. This places the nucleus in the basin of attraction of $A^1_{0,4}$, which corresponds to *hb,gt*-on. The last column of Table 5.4 shows the selection of states for all nuclei in the anterior region of circuit `hkgn58c13k1_007`.

The nuclei remain in the basin of $A^1_{0,4}$ until 43% EL. As A–P position is varied from anterior to posterior, $A^1_{0,4}$ moves from a *hb,gt*-on state to a *hb*-on state. Therefore the posterior border of the third anterior *gt* domain forms by the movement of this attractor. This is the first mechanism of border formation revealed by qualitative dynamics.

At 45% EL the nucleus passes from the basin of $A^1_{0,4}$ into the basin of $A^3_{0,4}$ (*Kr*-on) at 45% EL. This exemplifies the second mechanism of border formation in the circuit. The posterior border of the anterior *hb* domain and the *Kr* anterior border form by the nucleus crossing a basin boundary in a multistable dynamical system.

Except for the anteriormost nucleus (35% EL), the selection of nuclear states by the combination of the dynamical structure and initial conditions (Table 5.4) correctly accounts for all the nuclei in the anterior region. For the anomalous nucleus, the value of maternal Hb puts it in the $A^2_{0,4}$ (*hb,Kr*-on) basin instead of the $A^1_{0,4}$ basin, suggesting that the dynamical structure is not correct in this nucleus. This is a minor problem, since the nucleus behaves correctly in the circuit with diffusion (see Section 5.1 and Figure 3.1).

Nuc. (%EL)	**Basin** $A^3_{0,4}$	**Basin** $A^2_{0,4}$	**Basin** $A^1_{0,4}$	**Mat. Hb conc.**	**Basin selected**
35	NA	$(0, 58.20)$	$(58.20, 100)$	57.43	$A^2_{0,4}$ (*hb,Kr*-on)
37	$(0, 2.05)$	$(2.05, 52.12)$	$(52.12, 100)$	53.63	$A^1_{0,4}$ (*hb,gt*-on)
39	$(0, 10.61)$	$(10.61, 47.87)$	$(47.87, 100)$	50.05	$A^1_{0,4}$ (*hb,gt*-on)
41	$(0, 19.01)$	$(19.01, 42.98)$	$(42.98, 100)$	47.73	$A^1_{0,4}$ (*hb,gt*-on)
43	$(0, 26.29)$	$(26.29, 41.04)$	$(41.04, 100)$	42.27	$A^1_{0,4}$ (*hb*-on)
45	$(0, 30.79)$	$(30.79, 37.46)$	$(37.46, 100)$	39.12	$A^1_{0,4}$ (*hb*-on)
47	$0, 34.18)$	$(34.18, 37.76)$	$(37.76, 100)$	32.69	$A^3_{0,4}$ (*Kr*-on)
49	$(0, 35.24)$	$(35.24, 36.91)$	$(36.91, 100)$	29.40	$A^3_{0,4}$ (*Kr*-on)
51	$(0.94, 38.31)$	$(38.31, 39.26)$	$(39.26, 100)$	23.39	$A^3_{0,4}$ (*Kr*-on)

Table 5.4: Basins of point attractors in the anterior region, and their selection by maternal Hb. The last column shows the attractor in whose basin the initial condition lies. The diffusionless circuit remains in the basin of the $A^1_{0,4}$ attractor in the region where the *gt* border forms. The border forms by the movement of $A^1_{0,4}$ from *hb,gt*-on state to *hb*-on state. The *hb* and *Kr* borders form by switching basins instead (from $A^1_{0,4}$ to $A^3_{0,4}$).

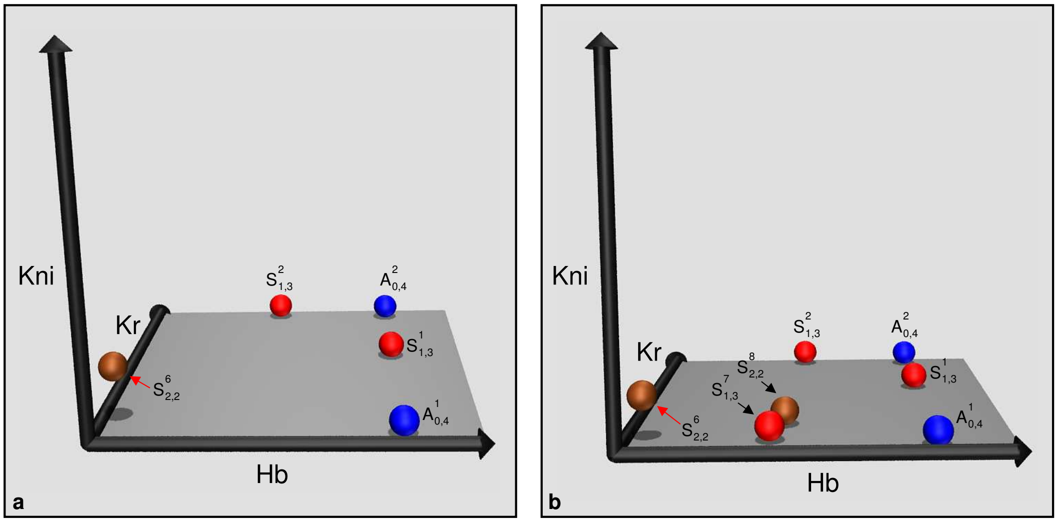

Figure 5.5: Bifurcations in the posterior region. Hb-Kr-Kni projection of phase portraits at (**a**) 53% EL and (**b**) 55% EL. The axes, xy-plane, and equilibria are as in Figure 5.3. See Table 5.3 for bifurcation parameter values determined by continuation. Black arrows point to saddles, $S^7_{1,3}$ and $S^8_{2,2}$, that are created via a saddle-node bifurcation between 53% EL and 55% EL.

5.4.2 Posterior Region 53% to 72% EL

A saddle-node bifurcation at 53.64% EL sets up the system for posterior regime dynamics. $A^3_{0,4}$ and $S^5_{1,3}$ annihilate each other (compare Figures 5.3d and 5.5a), leaving only two attractors, $A^1_{0,4}$ (*hb*-on) and $A^2_{0,4}$ (*hb,Kr*-on). Also, saddles $S^1_{1,3}$, $S^2_{1,3}$, and $S^6_{2,2}$ carry over from the anterior region. Two new saddles $S^7_{1,3}$ and $S^8_{2,2}$ are created through a saddle-node bifurcation at 53.32% EL (see Figure 5.5). Following this bifurcation, the qualitative dynamical structure remains invariant for the rest of the posterior region.

Since the dynamical structure is parameterized by Bcd and Cad concentration, its invariance merely reflects the relative constancy of Bcd and Cad in the region. The Bcd profile becomes shallow at these positions due to its exponential decay with AP position. Unlike the anterior region, where *bcd* drives boundary formation by changing the dynamical structure, the posterior region's pattern formation process is independent of *bcd*. Therefore, in the posterior region only one maternal gradient, maternal *hb*, drives pattern formation. In fact, changing only the maternal Hb concentration in a posterior region nucleus (Bcd concentration is fixed), produces correct gap gene patterns in the posterior region (see Figure 5.6).

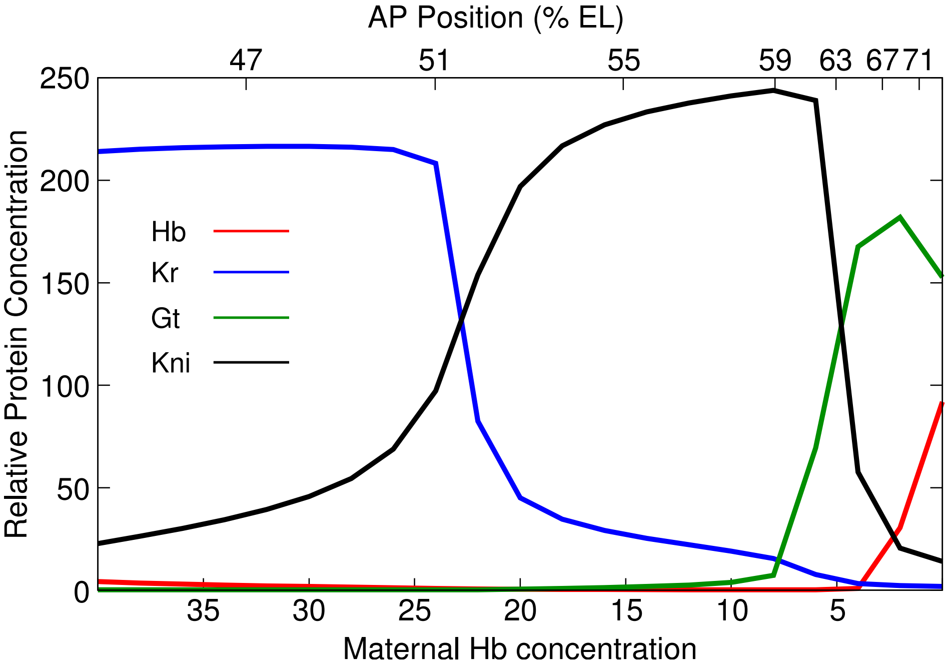

Figure 5.6: Maternal Hb is the morphogen in the posterior region. Gap gene concentrations in a single nucleus as a function of initial Hb concentration. The initial Hb concentration was varied uniformly in the nucleus at 63% EL (Kni peak at gastrulation), keeping the Bcd and Cad inputs constant. The nucleus produces all gap gene states in the posterior region from *Kr* peak (53% EL) to *gt* peak (72% EL) as initial Hb concentration is decreased from 40 to 0. The shapes of the "domains" are distorted since maternal Hb has faster than linear decay with position; as a consequence anterior "domains" are exaggerated. The x-axis on the top shows AP positions determined from the values of maternal Hb, showing that domains are in corrected proportions spatially. Posterior region nuclei form domains by responding to maternal Hb without any instruction for Bcd.

This prediction finds confirmation in an experiment which demonstrated that maternal *hb* is a morphogen (Struhl et al., 1992). This was accomplished by showing that gap gene patterns in embryos from $bcd^- tsl^-$ females respond to maternal *hb* in a dosage-dependent manner. These embryos simulate the posterior dynamical region since, (1) Bcd protein is absent and hence constant (simulating low, slowly changing Bcd levels in the posterior region). (2) The terminal system is absent (*tll* and *hkb* are not expressed in the posterior region). With wild type maternal *hb*, these embryos produce the posterior region's sequence of gap gene patterns, from the Kr peak in the anterior to the Gt peak in the posterior. Since there is no other maternal gradient present, maternal *hb* is driving pattern formation in these embryos (and the posterior dynamical region).

In order to characterize the mechanism by which maternal Hb patterns the posterior, we consider the dynamical structure of these nuclei in greater detail. Except for a very small interval (basin of $A^2_{0,4}$), all initial conditions form the basin of $A^1_{0,4}$. $A^2_{0,4}$'s basin divides $A^1_{0,4}$'s basin into two intervals. The first interval is between zero and the stable manifold of $S^2_{1,3}$, and the second interval is from the stable manifold of $S^1_{1,3}$ to 100.0. The behavior of trajectories originating in the two intervals is very different. While trajectories originating in the second interval (*direct route*) reach close to $A^1_{0,4}$ by gastrulation (see Figure 5.7), the trajectories in the first interval (*indirect route*) reach $A^1_{0,4}$ only asymptotically. In fact, maternal Hb concentrations in the circuit place all posterior region nuclei in the first interval of $A^1_{0,4}$'s basin. For example, at 53% EL, $A^2_{0,4}$'s basin is $(39.09, 39.62)$, and $A^1_{0,4}$'s basin is divided into the intervals $(0.0, 39.09)$ and $(39.62, 100.0)$. The first interval gives rise to indirect trajectories, while the second one gives rise to direct ones.

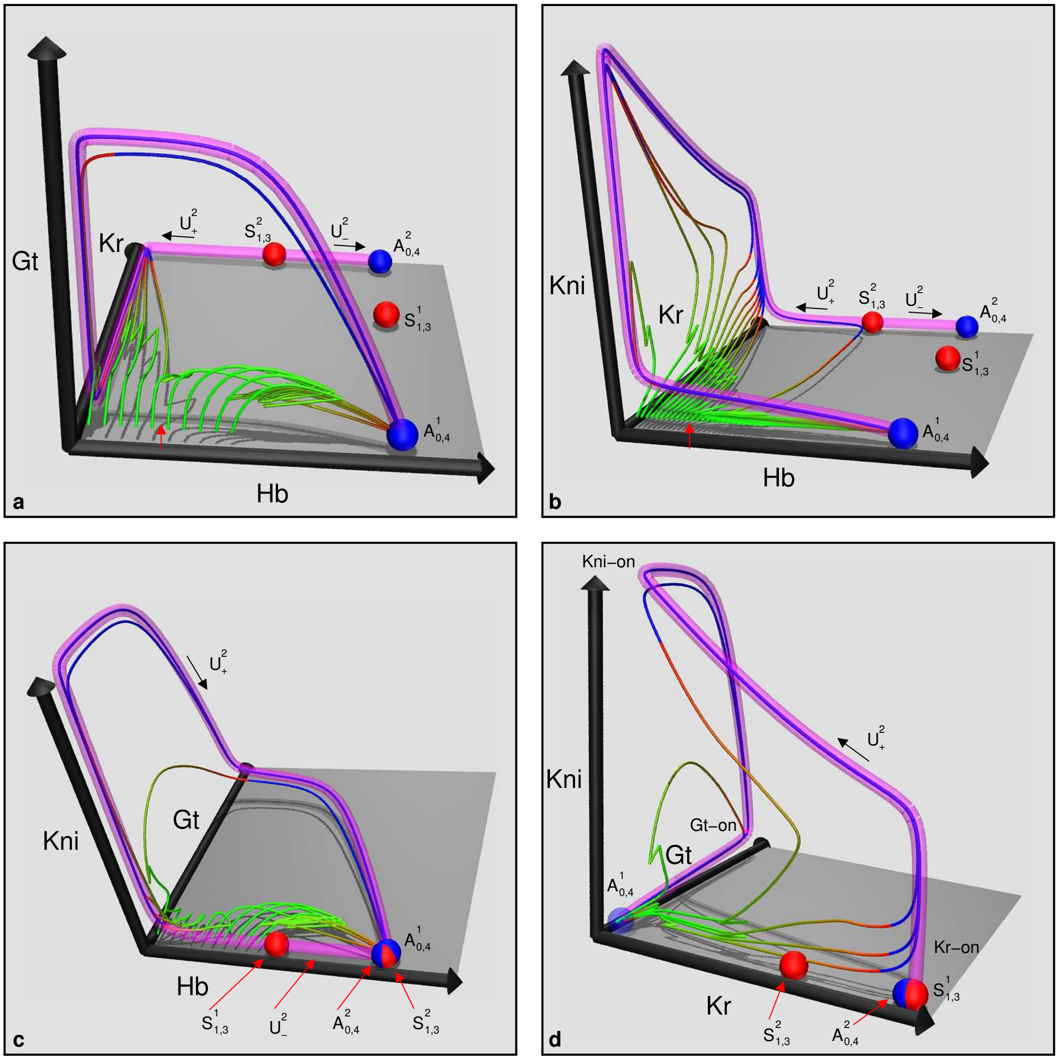

Figure 5.7: All four three-dimensional projections of the phase portrait at 57% EL. Axes, xy-plane, and equilibria are as in Figure 5.3. All hyperbolic equilibria are not shown (see Figure 5.5). The unstable manifold of saddle $S^2_{1,3}$, U^2 is shown as a translucent magenta tube of radius 5. (**a**) Hb-Kr-Gt projection. Red arrow shows the separation of the indirect route trajectories from direct route ones. (**b**) Hb-Kr-Kni projection. Red arrow shows the separation of the indirect route trajectories from direct route ones. (**c**) Hb-Gt-Kni projection. (**d**) Kr-Gt-Kni projection. U^2_+ traverses the anteroposterior progression of gap gene states in the posterior region—*Kr*-on to *kni*-on to *gt*-on.

Since nuclear states in the posterior region are determined by the indirect route trajectories, all of which lie in the same basin (of $A^1_{0,4}$), patterning by the crossing of basin boundaries is ruled out. Since $A^1_{0,4}$ does not move with AP position, the boundary doesn't form by movement of an attractor either. Instead, the indirect route trajectories converge to an invariant trajectory (Figure 5.7), U^2_+, that traverses all the posterior region states (*Kr*-on, *kni*-on, *gt*-on, and intermediate values), before reaching $A^1_{0,4}$. U^2_+ is one of the two trajectories that constitute the one-dimensional unstable manifold, U^2, of the hyperbolic equilibrium $S^2_{1,3}$. The other trajectory, U^2_-, goes from $S^2_{1,3}$ to $A^2_{0,4}$ (Figure 5.7).

U^2_+ is an attracting manifold, since all the indirect route trajectories converge to it. Moreover, the trajectories reach close to U^2_+ by gastrulation (see Figure 5.7). In other words, in the posterior region, this attracting invariant manifold plays the same role as the point attractors do in the anterior region, and is the reason for the stability of indirect route trajectories. The state achieved by such a trajectory has a continuous dependance on initial Hb value. High values of maternal Hb (20–40) in a nucleus lead to a *Kr*-on state, intermediate levels (12–20) lead to a *kni*-on state and lower values (4–8) lead to a *gt*-on state. Since maternal Hb monotonically reduces with AP position, this leads to the gap gene expression patterns in the posterior region (see Figure 5.6). This is the third mechanism of pattern formation in the model: attraction by an invariant manifold with dependence on initial conditions.

Shifts in the Posterior Region

The `hkgn58c13k1_007` circuit and its diffusionless counterpart show the dynamic shifts of posterior gap gene domains reported in Jaeger et al. (2004b).

The shifts only happen in the posterior region (53% EL–72% EL). The dynamical analysis reveals that the shifts are a consequence of two features of the posterior region's dynamical structure. First, the U_+^2 trajectory goes through the anteroposterior sequence of gap domains: Kr-on, Kni-on, and Gt-on (see Figure 5.5d). Second, U_+^2 is an attracting manifold. Taken together, these points imply that as a nucleus is pulled into U_+^2, it will go through anterior states before reaching its gastrulation state. We illustrate this mechanism with the nucleus through which the Kr posterior border and the Kni anterior border pass as they shift anteriorly (Figure 5.8(b,d)). Figure 5.8a shows the trajectory of the nucleus at 59% EL. The trajectory starts on the Hb axis, and is attracted into U_+^2, reaching close to the *kni*-on state at gastrulation. However in approaching U_+^2, it first goes through intermediate states. First Kr increases, and then Kni increases with a concomitant reduction in Kr (see Figure 5.8c). A similar mechanism applies to nuclei at the Kni posterior and Gt anterior boundaries.

Finally, we observe that the U_+^2 trajectory is a result of asymmetric repression of anterior domain genes by posterior domain proteins. A nucleus in Kr-on state goes to Kni-on state because Kni's repression of *Kr* is stronger than that of *kni* by Kr. Similarly Kni-on state goes to Gt-on due to asymmetric repression of *kni* by Gt, and Gt-on state leads to Hb-on because of asymmetric repression of *gt* by Hb. This analysis is consistent with the results in Jaeger et al. (2004b).

5.5 Stability of trajectories

In Section 5.4 it was shown that the trajectories are stable due to attraction by either point attractors or a one-dimensional manifold. In other words, they

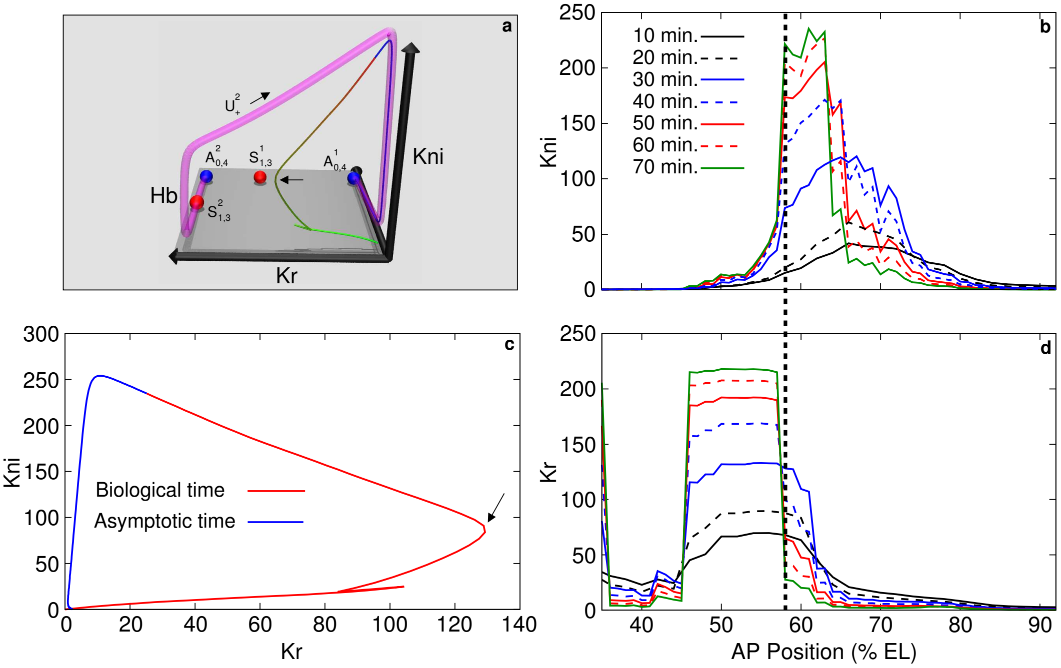

Figure 5.8: Shifts due to attraction by the manifold U^2_+. Dynamics of the nucleus at 59% EL, through which the Kr posterior and Kni anterior boundary pass. (**b,d**) The patterns of Kr, Kni in the diffusionless model show anterior shifts. The nucleus at 59% EL is indicated with a dashed vertical line. (**a**) A three dimensional projection of the phase portrait of the nucleus in the Hb, Kr, Kni hyperplane. The invariant manifold U^2 is shown as a magenta tube. The trajectory in the nucleus is plotted in a continuous color gradient from green ($t = 0$ min) to red ($t = 71$ min, gastrulation). Times after gastrulation are depicted as blue. The nucleus passes through intermediate states (indicated with arrow) with high Kr concentrations before reaching the peak of *kni* domain. This registers as an anterior shift in the posterior *Kr* and anterior *kni* borders. (**c**) A two dimensional projection of the trajectory in the Kr, Kni plane. The trajectory (red) starts at the origin. It attains a high Kr value at $t = 45$ mins (arrow) before approaching the Kni peak. The temporary reversal in the trajectory is a mitosis, during which the trajectory moves towards the origin. Time after gastrulation is shown in blue.

reduce variation in initial conditions. It is however important to establish the extent to which this variation is reduced, and whether it is biologically significant.

The biological initial conditions for a nucleus are $v^{Kr}(0) = 0$, $v^{gt}(0) = 0$, $v^{kni}(0) = 0$, and $v^{hb}(0)$ is specified by the maternal Hb concentration in the nucleus (Section 2.1.2, Figure 2.3). Early gap gene mRNA data (Jaeger et al., 2007) and protein data (Surkova et al., 2007) show large variation in expression levels. This early variability can be modeled through the initial conditions of the model.

From gap gene protein expression data during cleavage cycle 13 an upper bound for expression levels of a protein was determined. Based on these data a "box" of initial conditions was constructed in the four-dimensional phase space. The volume of the box forms an upper bound for initial variation observed in gap gene expression levels. Then, this box was evolved forward in time using the dynamical system, and its volume was tracked in time. The amount of volume reduction by gastrulation is a measure of how well this system reduces initial variation.

5.5.1 Calculating volumes in time

Given the diffusionless gap gene dynamical system in nucleus i,

$$\frac{dv^a}{dt} = R^a g\left(\sum_{b=1}^{N} T^{ab}v^b + m^a v_i^{\mathrm{Bcd}} + \sum_{\beta=1}^{N_e} E^{a\beta}v_i^{\beta}(t) + h^a\right) - \lambda^a v^a,$$

we want to calculate the volume at time, V_t, of a box of initial conditions $\mathcal{B}_0$. The variables have the same meaning as in Eq. (5.1), except the index i has been dropped from v_i^a. Let us represent the right hand side as $f(v_t, t)$, where

$v_t = (v_t^1, v_t^2, v_t^3, v_t^4)$ is the four-dimensional state vector at time t.

The ODE is discretized into a map,

$$v_{t+dt} = \gamma(v_t, t), \tag{5.7}$$

where $\gamma(v_t, t) = v_t + f(v_t, t)dt$. Regard the map at time $t + (n-1)dt$,

$$v_{t+ndt} = \gamma(v_{t+(n-1)dt}, t + (n-1)dt),$$

as a curvilinear coordinate transformation of the phase space at $t + (n-1)dt$ to $t + ndt$, then the infinitesimal volume at time $t + ndt$,

$$dV_{t+ndt} = dv^1_{t+ndt} dv^2_{t+ndt} dv^3_{t+ndt} dv^4_{t+ndt}$$

can be written as,

$$dV_{t+ndt} = J(v_{t+(n-1)dt}, t + (n-1)dt) dV_{t+(n-1)dt}. \tag{5.8}$$

Here, $dV_{t+(n-1)dt}$ is the infinitesimal volume at $t + (n-1)dt$, and $J(v_t, t)$ is the Jacobian of the map at time t. Applying Eq. (5.8) repeatedly, the infinitesimal volume at t, dV_t, can be written in terms of the initial infinitesimal volume, dV_0 as

$$dV_t = J(v_{t-(n-1)dt}, t - (n-1)dt) J(v_{t-(n-2)dt}, t - (n-2)dt) \ldots J(v_0, 0) dV_0.$$

If the initial box $\mathcal{B}_0$ evolves to $\mathcal{B}_t$ at time t, $\mathcal{B}_t$'s volume is

$$\begin{aligned} V_t &= \iiiint\limits_{\mathcal{B}_t} dV_t \\ &= \iiiint\limits_{\mathcal{B}_0} [J(v_{t-(n-1)dt}, t - (n-1)dt) \ldots J(v_0, 0)] dv_0^1 dv_0^2 dv_0^3 dv_0^4. \end{aligned}$$

The integral on the right hand side was evaluated using the multidimensional trapezoidal rule (Press et al., 1992), successively refining the grid on the initial box $\mathcal{B}_0$ until the integral converged. The time-step for the discretized map $\gamma(v_t, t)$ was chosen small enough such that the Euler-method solution of Eq. (5.1) converged.

5.5.2 Reduction of initial variation

First, the upper bound for the variation in initial conditions needs to be determined. For anterior nuclei the basin boundary for the attractor gives the bounds for Hb variation. For instance at 37% EL, the trajectories are attracted by $A^1_{0,4}$ (Table 5.4), whose basin is $(52.12, 100.0)$. For the other gap genes whose proteins are only expressed in cycle 13, the upper bound is determined from FlyEx data. At 37% EL, Kni and Kr expression was zero in all embryos, while the maximum Gt expression observed was 60 (embryo as14 from FlyEx). Hence, for the nucleus at 37% EL the initial box $(52, 100) \times (0, 1) \times (0, 60) \times (0, 1)$ was evolved forward in time. Its volume reduces by a factor of $\sim 10^8$ by gastrulation (Figure 5.9). The average shrinkage in a dimension is given by $V^{1/4}$ which reduces by a factor of ~ 0.05. For the posterior region nucleus at 57% EL, the initial box is $(0, 20) \times (0, 80) \times (0, 80) \times (0, 80)$. At this A–P position maximum expression for Kr is 80 (embryo iz4), Kni is 75 (embryo dm16), Gt is 50 (embryo ra10). Maternal Hb, that is Hb data from cleavage cycle 12, has a range 0–20 in the posterior region in all embryos. Volume reduces by a factor of $\sim 10^{-6}$ by gastrulation, and each dimension shrinks by an average factor of ~ 0.1.

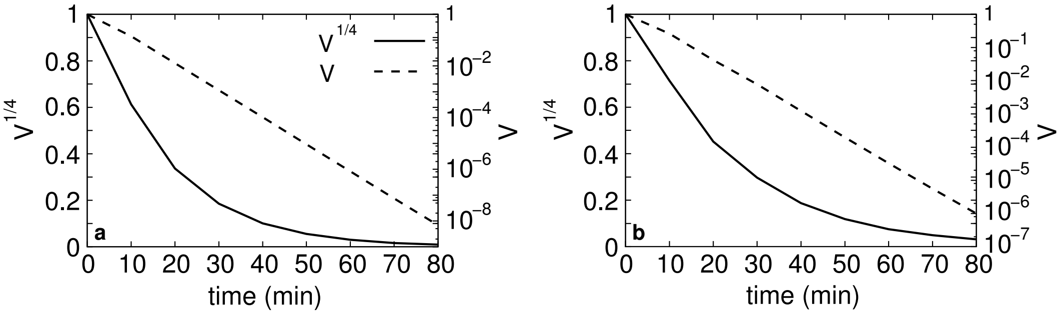

Figure 5.9: Reduction of initial variation. (**a**) Reduction in the initial volume of a box $(52,100)\times(0,1)\times(0,60)\times(0,1)$ in the anterior region nucleus at 37% EL. (**b**) Reduction in the initial volume of a box $(0,20)\times(0,80)\times(0,80)\times(0,80)$ in posterior region nucleus at 57% EL. Dashed line is volume in log-scale, solid line is $V^{1/4}$.

5.5.3 Stability of trajectories in the posterior region

In the anterior region, reduction in initial variation happens due to the convergence of trajectories to a point attractor. The size of an attractor's basin is a natural tolerance range for variation in initial conditions. In the posterior region, since the trajectories are attracted by the one-dimensional manifold U_+^2, there is a residual dependence on initial Hb concentration. This sensitivity to maternal Hb concentration is required in order to pattern the posterior (Figure 5.6). Despite this sensitivity, there is a further reduction of variation in initial conditions. The indirect trajectory interval of $A_{0,4}^1$'s basin can be divided into subintervals, such that trajectories from within a subinterval reach the same state at gastrulation.

Unlike the anterior region, where basin boundaries provided a natural tolerance range, the posterior phase portraits don't have any feature that can be exploited to calculate these subintervals or tolerances. They were calculated empirically from gap gene variation data in FlyEx. The variation in expression levels of Hb, Kr, Gt, and Kni in time class T8 was determined at three positions on the AP axis (corresponding to the peaks of Kr, Kni, and Gt). Next, the model was solved with different values of initial Hb concentration as starting points. The tolerance was calculated as the range of Hb initial concentrations that produced model output with the same variation in expression levels as data. The results are shown in Figure 5.10. It was found that there is significant reduction of maternal Hb variation in posterior nuclei. For instance, at Kr peak, where *Kr* expression level varies by 30% in time class T8, the tolerance for initial variation is 150%.

In this section it has been established that the gap gene system reduces initial variation. Therefore its developmental trajectories are stable. In the anterior region, homeostasis (Section 1.1.4), that is, attraction by point attractors is responsible for this stability. However, in the posterior region, a one-dimensional manifold attracts the trajectories. This implies that something more general than homeostasis is operative in the posterior region of the embryo. Furthermore, this section established that the indirect route trajectories produce a given gap gene state on U_+^2 within a tolerance range of maternal Hb concentration. This suggests that these trajectories are showing homeorhesis, as small perturbations within this tolerance are removed as they approach U_+^2. In other words, the trajectories with starting points in the tolerance range are chreods.

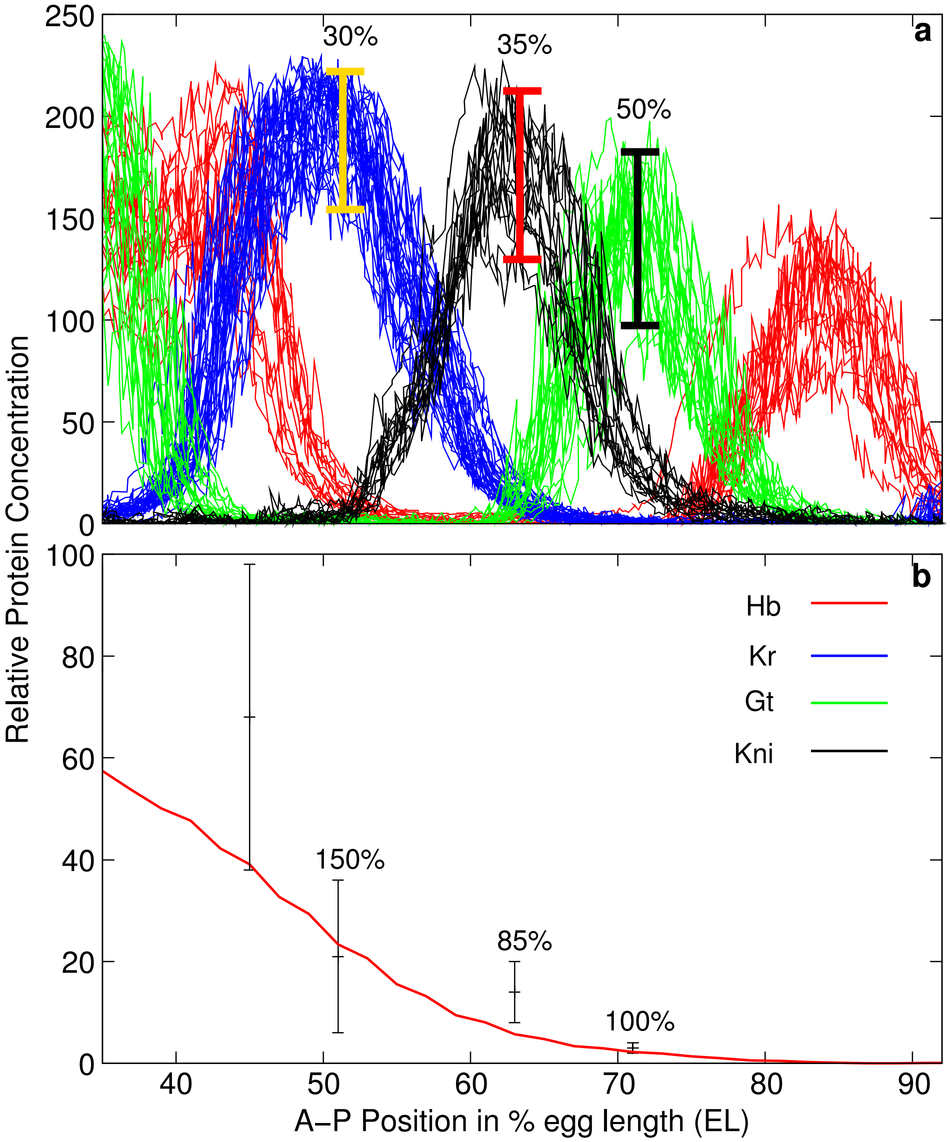

Figure 5.10: Tolerance to variation in initial data. Tolerance to initial Hb variation (**b**) was determined from variation in time class T8 data (**a**). (**a**) The variation in expression levels at the Kr peak is 30% (yellow bar), at the Kni peak is 35% (red bar), and at the Gt peak is 50% (black bar). (**b**) The tolerance range for maternal hb is shown as error bars for one anterior region position (45% EL) and three posterior region positions (*Kr*, *kni*, and *gt* peaks). Maternal Hb is shown in red. The tolerance range at 45% EL is the extent of the attractor's ($A^1_{0,4}$) basin. For the posterior region positions, the range was determined as the set of initial Hb starting points in the gene circuit that produce the variation levels determined in panel (**a**). The tolerance to initial variation is 150% at *Kr* peak, 85% at *kni* peak and 100% at *gt* peak.

Chapter 6

Conclusions

There are two major results from the work reported in this dissertation. First, it has been established that the reduction in gap gene borders' variation with respect to Bcd is due to gap gene cross regulation, which has resolved a controversy over the variance of the posterior border of the anterior *hb* domain. Second, it has been shown that gap gene expression patterns in the posterior are stable with respect to initial variation due to the existence of a chreod in the dynamical system.

The first section of this chapter discusses the reduction of the variance of the posterior border of the anterior *hb* domain with respect to Bcd. Recent experimental and theoretical work is reviewed and evaluated in light of the results presented in Chapters 3 and 4. In the second section, the results from the qualitative dynamics analysis (Chapter 5) are elaborated upon. The ideas of canalization, structural stability, and homeorhesis are evaluated in the context of the dynamical structure of the gap gene system. The third section contains a discussion of the limitations of this study, while the last section proposes future experimental and theoretical studies suggested by this work.

For the rest of this chapter, we refer to the posterior border of the anterior *hb* domain as the *hb* border.

6.1 Reduction of the variance of the *hb* border

The reduction of the variance of the *hb* border with respect to Bcd, first discovered by Houchmandzadeh et al. (2002), and confirmed by Surkova et al. (2007), has received considerable attention. On the one hand, a great number of theoretical models have been built in order to explain the effect, and on the other the result's validity has been challenged.

This section begins with a review of some pertinent points from Houchmandzadeh et al. (2002). Then, recent work (Gregor et al., 2007a; Crauk and Dostatni, 2005) suggesting that the variance of the *hb* border is accounted for by the variance of the Bcd gradient is critically examined. I argue that though the variance of Bcd gradient is lower than previously thought, it is still twice that of the *hb* border. Finally, other theoretical models to explain this effect are discussed.

6.1.1 Antibody data, dosage, and mutants

By immunofluorescently staining fixed embryos for Bcd and Hb proteins simultaneously, Houchmandzadeh et al. (2002) showed that the range of positions where the Bcd gradient crossed a particular threshold is 30% EL, equivalent to $\sigma_{\text{Bcd}} = 7\%$ EL (cf. Figure 2.5, Surkova et al., 2007). The position of the *hb* border had a range of 4% EL ($\sigma_{\text{Hb}} = 1\%$ EL, cf. Table 3.3).

Additional data are presented in Houchmandzadeh et al. (2002) which supports the conclusion that the *hb* border position is not determined exclusively by Bcd concentration. In embryos from mothers with one dose, or more than two doses of *bcd*, the *hb* border position is expected to shift anteriorly or posteriorly respectively. This is because the threshold occurs at a different position in these embryos. The *hb* border does shift in the predicted direction, but by an amount smaller than the spatial shift of a fixed Bcd threshold. For example, in embryos derived from mothers with one copy of the *bcd* gene, the *hb* border is expected to shift by 19% EL to the anterior, however the average shift observed is about half that, 8% EL to the anterior. σ_{Hb} is still 1% EL (Houchmandzadeh et al., 2002, 2005).

Further, the authors showed that while the position of *hb* border scales with egg size, the position of the Bcd threshold does not. Since Bcd provides a measure of distance from the anterior pole, it was postulated that a similar gradient from the posterior that regulates *hb* might provide a measure of distance from the posterior pole, and hence explain the scaling property.

Houchmandzadeh et al. (2002) performed a systematic search for genes that affect the variability of the *hb* border. Several maternal and zygotic mutants were tested. As mentioned at the end of Chapter 3, the variability of the *hb* border in gap gene mutants is $\sigma_{\mathrm{Hb}} = 1.0$–$1.1\%$ EL, essentially the same as wild type. The largest zygotic effect was observed in embryos missing the chromosome arm 3L, in which $\sigma_{\mathrm{Hb}} = 2\%$ EL.

Among maternal mutants, obvious candidates for posterior gradients like *oskar* and *nanos* were tested. It was found that the posterior group of maternal genes (see Section 1.2) increased the variability of the *hb* border by a factor of 1.6.

The biggest effect was seen in two maternal genes, *swallow* (Frohnhöfer and Nüsslein-Volhard, 1987), and *staufen* (Driever and Nüsslein-Volhard, 1988a;

Johnston et al., 1991), that are involved in the localization of *bcd* mRNA to the anterior pole and affect the Bcd gradient (Driever and Nüsslein-Volhard, 1988a).

There was one weakness in the data analysis of Houchmandzadeh et al. (2002). In calculating the variance of the Bcd threshold, the Bcd profiles were normalized so that their maximum fluorescence level was set to one, and the minimum level was set to zero. In other words, in a profile $A\exp(-\lambda x), \quad A \equiv 1$ (Eq. 2.2), leaving λ as the only source of variance in the profile. As shown in Figure 2.5a, A and λ are correlated, which means that such a normalization would exaggerate the variance of the Bcd threshold position. In the unnormalized Bcd data (Surkova et al., 2007) used in this dissertation $\sigma_{\mathrm{Bcd}} = 4.6\%$ EL, lower than the value of 7% EL obtained by Houchmandzadeh et al. (2002), but still much higher than σ_{Hb}.

6.1.2 *In vivo* Bcd data

Reevaluating Bcd and Hb variability

The reason for normalization of the Bcd data in Houchmandzadeh et al. (2002) is that such data are collected over the course of several experiments. The microscope gain is standardized separately for each experiment. This implies that the fluorescence levels in such data have higher variability than the intrinsic variability of the protein concentration in the embryo. Note that the measurement of the Hb protein is much less affected by this experimental error, since its expression domains have a characteristic temporal profile (Surkova et al., 2007) that allows reliable standardization across experiments. The position of the *hb* border is also much less affected by the concentration scale, since the border is steep (see Figure 3.2c, and Surkova et al., 2007).

Gregor et al. (2007a) addressed the lack of an absolute concentration scale for Bcd by measuring the Bcd gradient *in vivo*. This was done by constructing a fully functional fusion of the Bcd protein with enhanced green fluorescent protein (eGFP) (Gregor et al., 2007b). The Bcd-GFP construct rescues the phenotype of embryos that lack endogenous Bcd protein. The progeny develop to adulthood, and the position of the cephalic furrow is unchanged from wild type. The Bcd-GFP protein therefore has the same function as the endogenous Bcd.

Gregor et al. (2007a) calculated the variability of the Bcd gradient in three steps. In the first step, *in vivo* Bcd profiles were measured in several embryos placed on the same slide in the course of a single experiment. This solved the problem of reliable microscope standardization across experiments. In the second step, the A–P axis was divided into 2% EL bins. The relative error in the Bcd concentration, that is, the ratio of the standard deviation to mean was calculated for each bin. In the third step, this relative error in concentration was translated into the relative error in position by multiplying it with the inverse of the slope of the average of all the Bcd profiles, that is

$$\sigma_x = \sigma_c \left| \frac{d\bar{c}}{dx} \right|^{-1}. \tag{6.1}$$

Here, σ_c is the standard deviation of the concentration in a bin, $\bar{c}$ is the average concentration profile, and σ_x is the positional error. With these three steps $\sigma_{\mathrm{Bcd}} \sim 2\%$ EL for the *in vivo* profiles.

An additional step of subtracting estimated experimental errors was applied that further reduced the estimate of σ_{Bcd} to 1–2% EL. The experimental errors were estimated from live embryos. For example, the imaging noise was determined by repeatedly imaging Bcd-GFP in live embryos and calculating the variance of the observed fluorescence levels. This step of subtracting errors is problematic, since the error has not been determined independently of the observed variable.

A further problem with interpreting the σ_{Bcd} determined by these Bcd-GFP measurements is that the *hb* border accuracy was not estimated in these embryos. The variance of Bcd profiles does not convey any meaning by itself, but can only be looked at in comparison to the variance of downstream gene borders.

Since a Hb-GFP fusion is not available, one has to resort to fixed tissue data to make such a comparison. Gregor et al. (2007a) proposed a new normalization method to resolve the problems associated with the one employed in Houchmandzadeh et al. (2002). This method does not try to minimize the difference between profiles at any particular A–P position. Instead, it scales individual profiles such that the sum of squared difference between normalized profiles and the mean profile over all A–P positions is minimized. Then the relative error in concentration is translated into relative error in position by Eq. (6.1). The authors reanalyzed the data of Houchmandzadeh et al. (2002) using this new normalization method, and found that $\sigma_{\mathrm{Bcd}} \sim 3\%$ EL. This is still three times the $\sigma_{\mathrm{Hb}} = 1\%$ EL observed by Houchmandzadeh et al. (2002).

Next, Gregor et al. (2007a) acquired a new fixed-tissue dataset for Bcd and Hb in which a large number of embryos were fixed, stained and scanned together to minimize experimental error. By applying the data analysis methodology described above, it was found that the positional error estimate for Bcd reduced from $\sigma_{\mathrm{Bcd}} \sim 3\%$ EL to $\sigma_{\mathrm{Bcd}} \sim 2\%$ EL. However, in the new data, σ_{Hb} increased to $\sim$2% EL from the 1% EL estimated by Houchmandzadeh et al.

(2002). This led the authors to claim that Bcd variability accounted for Hb variability within the limits of experimental error.

This is a rather contradictory result, as an experiment with greater control over experimental error has yielded a larger estimate for σ_{Hb}, rather than a smaller one. Careful consideration of Eq. (6.1) explains this puzzling result. Eq. (6.1) relies on the average concentration profile $\bar{c}(x)$ to convert concentration error into positional error. The slope of the average profile $\bar{c}(x)$ is always less than the slopes of individual profiles. This means that its inverse will be larger, leading to a larger estimate of the positional error σ_x. This is illustrated by a simple example in Figure 6.1. Two linear profiles $c_1(x)$ and $c_2(x)$, each with slope $1/S_l$, and maximum 1, are separated by a distance S_p on the x-axis. Therefore, the positional error between these two profiles is their spread, S_p. Their average profile $\bar{c}(x) = \frac{c_1(x)+c_2(x)}{2}$ has slope $\frac{1}{S_l+S_p}$. At the point marked X, the concentration error is 1, therefore, using Eq. (6.1), one obtains a spread of $S_l + S_p > S_p$. The overestimation of σ_x will be greater if the profiles are sharper, which explains why the estimate for a shallow profile like Bcd reduces, while it increases for a sharp one like Hb. Evidently, Figure 6.1 is too simple an example, and the magnitude of the increase depends on the details of the profiles and their spread. By applying Eq. (6.1) to the Hb profiles in Figure 3.2a, we obtained $\sigma_x = 1.72\%$ EL at 46% EL, and $\sigma_x = 2.37\%$ EL at 47% EL. The positional variance for the *hb* border using the half-maximum method (Section 3.2) is 1.3% EL.

Since the gap genes in general, and *hb* specifically, are recessive lethal, the variability of their concentrations is less important than the variability of their boundary positions. Therefore, it is better to compare the positional variance of the Bcd-GFP data directly with *hb* border position variance in

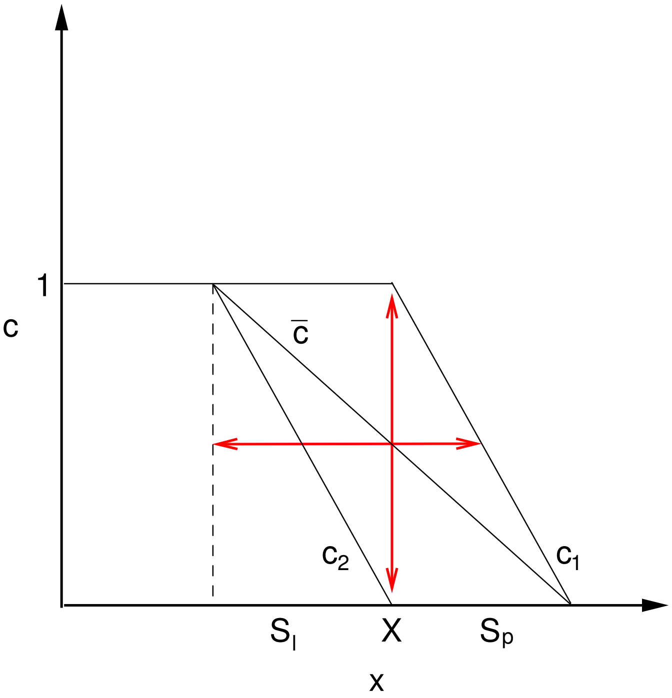

Figure 6.1: Overestimation of positional variance using average concentration profiles. Two concentration profiles c_1, c_2 are parallel to each other and have slope $1/S_l$. They are separated by S_p. The average profile $\bar{c}$ has slope $\frac{1}{S_l+S_p}$. At point X, the concentration error is 1, giving an exaggerated estimate of positional error $S_l + S_p > S_p$.

fixed-tissue Hb data. Since the numerical Bcd-GFP data are not available, a rough comparison is shown between Figure 5A in Gregor et al. (2007a) and Hb data from FlyEx in Figure 6.2. The spread of Bcd concentrations is ~10% EL, while that of the *hb* border is ~5% EL, giving a discrepancy factor of two. Therefore we find that Bcd variability does not fully account for Hb variability. A direct comparison between Bcd-GFP data and Hb data (such as Figure 6.2) is not made in Gregor et al. (2007a).

Bcd variability due to low molecular numbers

Gregor et al. (2007a) were also able to calculate the absolute protein concentration in a nucleus by calibrating their measurements against a known eGFP concentration. They estimated that in the middle of the embryo, where the Hb border forms, the number of Bcd molecules is $\sim 4.8/\mu\mathrm{m}^3$. The effects of molecular noise become evident at such low concentrations.

A formula first used in the theory of bacterial chemotaxis (Berg and Purcell, 1977) allows one to estimate the extent of the variability due to low molecular numbers. The calculation used to derive the formula takes into account the effect of the random arrival of the molecules at the receptors of a cell immersed in a solution of a chemical species. The cell can sense the local concentration c by the capture and release of the molecules of the chemical by its receptors. Assuming that the cell is a perfect counter of molecules, the error it makes in sensing c is given by

$$\frac{\sigma_c}{c} \sim \frac{1}{(TcaD)^{1/2}}. \tag{6.2}$$

Here, T is the amount of time spent counting, a is the receptor size, and D is the diffusion constant of the chemical species. In other words Eq. (6.2) is

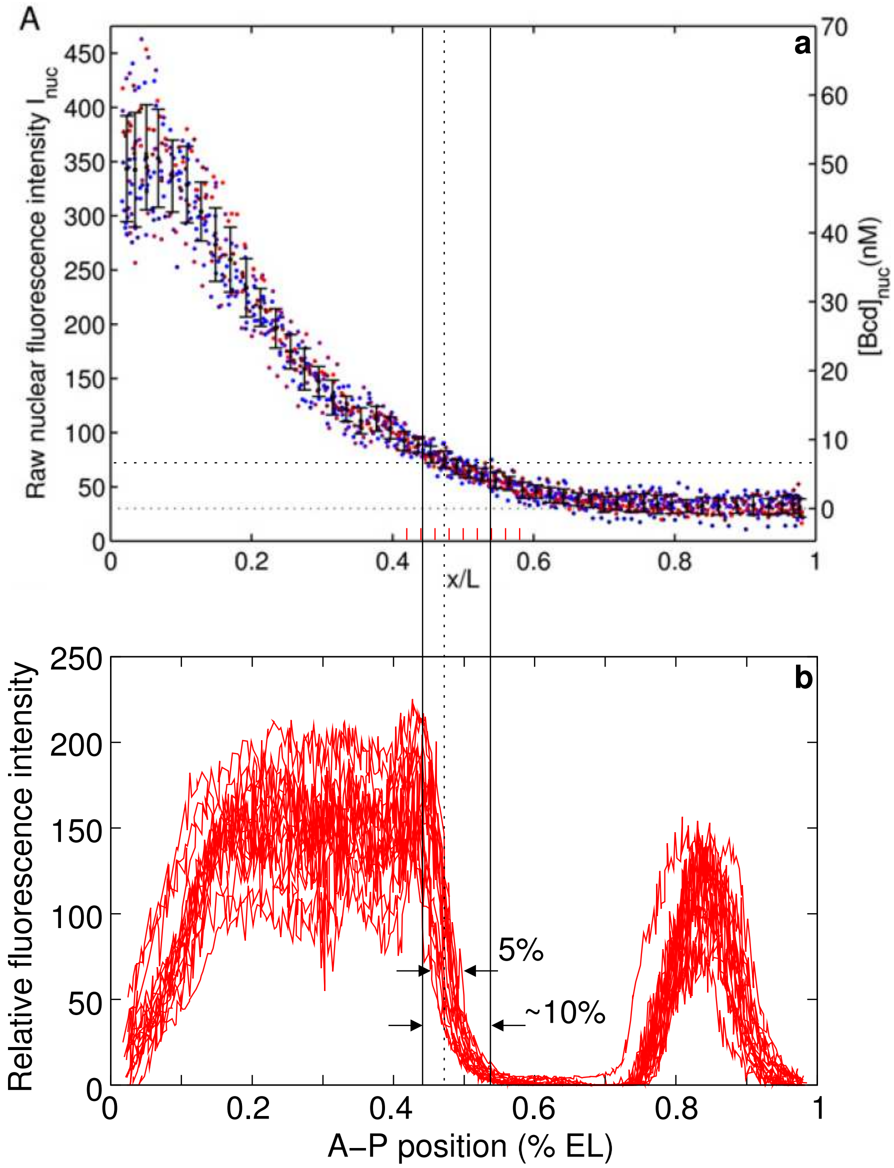

Figure 6.2: Comparison of *in vivo* Bcd variability with Hb variability. (**a**) Panel A from Figure 5 of Gregor et al. (2007a) showing 15 Bcd-GFP concentration profiles. (**b**) 18 Hb profiles in time class T8 from FlyEx (Surkova et al., 2007). The dotted vertical line is at the average position of *hb* border. The dotted horizontal line passes through the point where the dotted vertical line crosses the middle of the Bcd scatter. The solid black lines delineate the horizontal spread of Bcd. The red tick marks have been placed uniformly using the Linux graphics program Xfig.

a lower bound for the molecular noise at the receptor due to random arrival of molecules, without making any assumptions about the internal mechanism with which the counting is done.

Gregor et al. (2007a) applied this calculation to estimate the effects of low molecular numbers at the *hb* promoter. Using the value of the diffusion constant for Bcd ($\sim 1\mu\mathrm{m}^2/\mathrm{s}$) determined in Gregor et al. (2007b), the size of a 10bp long binding site ($\sim$3nm), and the absolute concentration of Bcd in the nuclei at *hb* border, the authors estimated that the uncertainty in Bcd concentration was too high for the accurate specification of the *hb* border. In order to reconcile the large uncertainty in Bcd concentration with their claim that Bcd variability accounts for Hb variability, it was hypothesized that spatial averaging by some other chemical species could provide additional noise reduction. This averaging would work in a manner such that the Hb level in a nucleus reflects an average over the Bcd concentration in many nuclei. This hypothesis is inconsistent with the Berg-Purcell framework (Berg and Purcell, 1977), since it assumes that receptors of different chemical species are independent, and that there is no communication between them.

In summary, the work of Gregor et al. (2007a) has shown that the variability of Bcd is much lower than 7% EL, in fact it is closer to 2% EL. Also, the Berg-Purcell calculation (Eq. 6.2) shows that the relative error in Bcd due to molecular noise depends on its concentration as $\sim\frac{1}{\sqrt{c}}$. As a consequence, the relative error is lesser in the anterior, where the concentration of Bcd is high, and is greater in the posterior, where the concentration of Bcd is low. Finally, Gregor et al. (2007a) have interpreted their data to mean that Hb variance is accounted for by Bcd variance.

However, the raw data (Figure 6.2) suggest that there is still a discrepancy factor of two between Bcd and Hb variability. This discrepancy shows a remarkable consistency with the results of Chapter 4, since the removal of zygotic input to *hb* doubles the variance of its border from 1.1% EL to 2.2% EL. This suggests that in Kr^-/kni^- embryos, Hb is responding only to Bcd and shows the same variability as the underlying Bcd gradient.

6.1.3 Reporter constructs

The Bcd gradient has been assayed by means of reporter constructs in many experiments (Driever et al., 1989; Struhl et al., 1989; Simpson-Brose et al., 1994). In two recent experiments such data were acquired in large numbers of embryos, allowing the estimation of Bcd variability (Bergmann et al., 2007; Crauk and Dostatni, 2005). In the first experiment (Bergmann et al., 2007), a reporter construct containing three copies of a 123bp fragment from the P2 Bcd-dependent *hb* promoter (hb123x3-LacZ, Struhl et al., 1989) was used to assay the Bcd gradient. LacZ from this reporter is expressed in an anterior cap similar to the 2.9kb transcript from the P2 promoter (Schröder et al., 1988), with a boundary at ~40% EL. It was found (Bergmann et al., 2007, Figure 4) that its variability is ~10% EL (range of positions).

The second reporter construct (Bcd3-LacZ) contains just three Bcd sites, and is expressed at $\sim 30\% EL$ (Crauk and Dostatni, 2005). However, its boundary positions have a range of only ~5% EL. This led Crauk and Dostatni (2005) to claim that the Bcd gradient has the same variability as Hb.

However, once the error in Bcd is taken into account (Eq. 6.2), these results are fully consistent with the rest of the experimental evidence. The variability of Bcd at the *hb* border position, 46% EL is $\sim 10\% EL$. The hb123x3-LacZ construct's boundary is formed only slightly anterior to this position, and therefore has a similar variability. The Bcd3-LacZ construct forms its boundary at ~15% EL to the anterior of the *hb* border. From the GFP measurements of the Bcd gradient (Gregor et al., 2007a, , Figure 2), the concentration of Bcd protein at 30% EL is roughly five times the concentration at the *hb* border forms. Since the error in Bcd goes as $\frac{1}{\sqrt{c}}$, a factor of five increase in concentration would roughly halve the error, which is precisely what is observed for the Bcd3-LacZ construct.

These measurements from reporter constructs add to the evidence that the variability of Bcd is twice that of Hb. Recall from Section 6.1.1 that the largest increase of a factor of two in Hb variability was observed in embryos deficient for the chromosome arm 3L, suggesting that it is the upper bound for Hb variability. This upper bound is reached in *Kr*/*kni* double mutant embryos (Chapter 4), in which Hb responds only to the Bcd gradient. Finally, the hb123x3-LacZ construct that was shown to respond only to Bcd (Struhl et al., 1989), also has twice the variability of the *hb* border in wild type. The border of the Bcd3-LacZ construct has the same variability as the *hb* border, but the resemblance is coincidental. This construct's border forms at ~30% EL, and its lower variability arises from the lower relative error in Bcd concentration at that position. To summarize, the results of three independent and different experiments all lead to the same conclusions: (1) Bcd cannot account for the variability of Hb, and (2) the variability of Bcd is twice that of Hb.

Since the discovery of the discrepancy between Bcd and Hb variability (Houchmandzadeh et al., 2002), a large amount of theoretical work has been done to explain this phenomenon. In the following section, these theoretical models are reviewed.

6.1.4 Theories for the accuracy of *hb*

Posterior gradient models

Following the hypothesis of a posterior gradient discussed in Section 6.1.1, two models were proposed that incorporated such a gradient (Houchmandzadeh et al., 2005; Howard and ten Wolde, 2005). The first model proposes an posterior gradient that is a transcriptional regulator of *hb* and inhibits the activity of Bcd at the *hb* border. This gradient is assumed to be correlated with the Bcd gradient, and provides posterior information at the *hb* border. The second model proposes a gradient of corepressor molecules that bind to Bcd protein molecules and deactivate them. Another work (McHale et al., 2006) independently proposed both of these mechanisms.

Houchmandzadeh et al. (2005) assume that a maternal posterior gradient inhibits the activation of *hb* synthesis transcriptionally. Furthermore, it is required that the *hb* border is established at a fixed ratio of the concentrations of Bcd and the hypothetical gradient. It was found that if this ratio is 1, then the border forms at 50% EL and does not depend on the length scales of the gradients. Hence, such a model shows scaling with egg length.

If the ratio of Bcd concentration and the posterior gradient concentration is not 1 where the target gene's border forms, the model does not exhibit scaling. Since the ratio of Bcd concentration to repressor concentration changes with A–P position, this model cannot explain the accuracy of the more than one gap gene border (Figure 3.2c) at a time.

In the model of Howard and ten Wolde (2005), a corepressor gradient from the posterior reduces the activity of Bcd in the posterior by binding and thus deactivating Bcd molecules. The Bcd and the corepressor gradients are modeled explicitly with synthesis at point sources, Fickian diffusion and first-order degradation. If one assumes that the degradation rates of Bcd and the corepressor are correlated, then the model shows scaling with embryo length. However, there are no unbound Bcd molecules available posterior to 50% EL for activating downstream genes in that region. This is incompatible with experimental evidence that Bcd activates *kni* in the posterior (Rivera-Pomar et al., 1995).

A set of experiments (Lucchetta et al., 2005) challenge the idea of a simple posterior gradient mechanism for the accuracy of the *hb* border. Lucchetta et al. (2005) maintained the anterior and posterior halves of a *Drosophila* embryo at 20 °C and 27 °C, respectively. If the specification of the *hb* border relies on the precise balance between Bcd and a posterior gradient, this would disrupt the accuracy of the *hb* border. It was found, however, that the accuracy of the *hb* border was not disrupted in such experiments.

The main problem these models face is that such a gradient has not been found. The gradient of Nos does not qualify as it regulates Hb translationally (Lehmann and Nüsslein-Volhard, 1991), but not transcriptionally. Furthermore, in contrast to the results presented in Chapter 3, these models cannot explain the simultaneous low variance of multiple gap gene borders across a large A–P region of the embryo. These models can only account for the low positional variance of one border (Houchmandzadeh et al., 2005), or of the borders in the anterior region of the embryo (Howard and ten Wolde, 2005).

Active transport

Following the result of Houchmandzadeh et al. (2002) that *stau* affects the variability of the *hb* border, Aegerter-Wilmsen et al. (2005) proposed that an active transport mechanism might explain its accuracy. This is based on the role of *stau* in the localization of *bcd* and other maternal mRNAs through active transport in the oocyte (Johnston et al., 1991). *stau* binds an RNA-binding domain in the 3' UTR of *bcd* RNA (Johnston and Nüsslein-Volhard, 1992). It is unlikely that the *hb* mRNA has this domain since maternal *hb* mRNA is distributed uniformly in the embryo (Tautz and Pfeifle, 1989).

Pre-steady state Bcd gradient

Bergmann et al. (2007) propose that reduced shifts of gap and pair rule patterns upon varying *bcd* dosage (Section 6.1.1, Houchmandzadeh et al., 2002) can be explained by the determination and fixing of gap gene borders during the time when Bcd is not stationary. They calculated the transient solution for Bcd before steady state based on localized production, Fickian diffusion, and protein degradation. It was shown that on changing the synthesis rate, the expected shift in the Bcd profile would increase in time, until it is maximum at steady state. The *in vivo* measurements (Gregor et al., 2007a) of the Bcd gradient show that it reaches steady state before cleavage cycle 10, before zygotic transcription of *hb* and other gap genes begins (see Section 2.1.1). Therefore, the interpretation of the transient Bcd gradient by the gap genes cannot account for the reduced shift in gap gene expression patterns.

6.1.5 Summary

The *in vivo* results of Gregor et al. (2007a) provide the most accurate estimates of Bcd variability so far. The measurement of absolute Bcd concentration in the embryo leads to the important conclusion that molecular noise is an important source of error in the middle and posterior of the embryo. The authors have attempted to show through data analysis that Hb variability is greater than what has been observed in similar experiments (Houchmandzadeh et al., 2002; Surkova et al., 2007). At the same time it has been argued that Bcd variability is low despite the limitations of molecular noise. As discussed before, the theoretical argument of spatial averaging used to support this conclusion is invalid in the framework of the Berg-Purcell theory (Berg and Purcell, 1977).

If all the experimental results that have been produced so far are considered together, a consistent interpretation of the data emerges. The Bcd-GFP data provide the best estimate so far of Bcd variability at the *hb* border (∼2% EL). The Hb data (Houchmandzadeh et al., 2002; Surkova et al., 2007) show a variability of ∼1% EL, half that of Bcd. In zygotic mutants, the upper bound for *hb* border variability is at the level of Bcd variability (Section 6.1.1). Reporter constructs that are under the exclusive control of Bcd have the same variability as Bcd (Section 6.1.3). If there are any lingering doubts about experimental error, they are removed by *bcd* dosage experiments (Houchmandzadeh et al., 2002, see Section 6.1.1) in which average shifts are calculated. The shift of the *hb* border in embryos from mothers with a single dose of *bcd* is about half of what is expected from a fixed Bcd threshold, confirming the discrepancy factor of two between Bcd and Hb variabilities.

The work presented in this dissertation shows that gene circuits correctly predict the reduction of variance of six gap gene borders (Section 3.2). This property of gene circuits does not depend on the use of a particular Bcd profile, or the use of an exponential fit of a Bcd profile. The circuits show reduction of variance if they are produced using other Bcd profiles, or if raw background-removed Bcd data are used to simulate Bcd variation. Other theoretical models predict the accuracy of a single border only, and rely on a posterior gradient that has not been found yet. The variance reduction in gene circuits happens due to dynamic gap gene cross regulation in response to Bcd variation (Section 3.3). These interactions are known and have been characterized very well in genetic (Driever et al., 1989; Clyde et al., 2003; Hoch et al., 1991; Eldon and Pirrotta, 1991; Rivera-Pomar et al., 1995; Kraut and Levine, 1991a; Gaul et al., 1987) and theoretical studies (Reinitz et al., 1998; Jaeger et al., 2004a; Perkins et al., 2006).

The *Kr*/*kni* double mutant experiment described in Chapter 4 has shown conclusively that the *hb* border position is not just under maternal control, but is under zygotic control as well. The positional variance of the *hb* border doubles when the zygotic input is removed in these mutant embryos (Section 4.2), confirming that Bcd variance is twice that of the *hb* border.

6.2 Canalization and dynamical structure of the gap gene system

This dissertation presents a dynamical analysis of an actual developmental subsystem with observable state variables. With the results of Chapters 3, 4, and 5, it is now possible to evaluate the ideas of structural stability, genotypic canalization, homeostasis, and homeorhesis in the context of this developmental system.

6.2.1 Genotypic canalization

In Section 1.3 it was proposed that the reduction of variance of gap gene border positions with respect to Bcd is an instance of genotypic canalization. Using this criterion, gap gene circuits show canalization (Section 3.2).

Having obtained a gene circuit that canalizes, it is possible to analyze its properties that permit this behavior. Regulatory analysis (Section 3.3) of the gene circuit shows that the gap gene borders have lower variability than Bcd due to epistasis among the gap genes. Activation and repression levels are in balance at the boundaries of their expression domains. Moreover, it is shown that the nature of the epistasis is such that regulators of gene respond to changing levels of Bcd in a manner that damps the perturbations caused by Bcd at the border of the gene's expression domain.

The correspondence between Bcd levels and gap gene boundary positions reduces with increasing A–P position, though the epistatic balance between activation and repression is still maintained at the borders in the posterior (Figures 3.7 and 3.8). This supports earlier results (Jaeger et al., 2004b) that Bcd does not provide positional information in the posterior. Consideration of the dynamical structure of the phase space of posterior nuclei also leads to the same conclusion, as it remains qualitatively the same in the posterior region (see Section 5.4.2).

6.2.2 Structural stability

Does this property of genotypic canalization in the gene circuit arise from structural stability of the gap gene dynamics? Based on the analysis in Chapter 5, it is possible to say that the gap gene dynamical system is at least locally structurally stable. Although there are several bifurcations that occur with the A–P position of a nucleus (Section 5.4, Table 5.3), there is only one that involves an attractor. At the level of the present analysis (Section 5.4), the other bifurcations between hyperbolic equilibria do not seem to affect the overall dynamics, though they may have some subtle, as yet unknown effects.

The bifurcation that annihilates the attractor $A^3_{0,4}$ changes the dynamics drastically, setting up the attracting manifold U^2_+ which controls the trajectories in posterior region nuclei. This bifurcation also divides the embryo into two regions with very different biological properties (Jaeger et al., 2004b). This bifurcation can be classified as a fold catastrophe (Thom, 1969, 1983), since an attractor, $A^3_{0,4}$, disappears and the trajectories that were previously attracted by it go to a different attractor, $A^1_{0,4}$, instead.

Within in the A–P regions separated by the catastrophe, the phase space of nuclei does not change qualitatively. Hence the phase space of the gap gene dynamical system can be regarded as having local structural stability. Note that the structural stability of the phase space implies that patterns will be qualitatively preserved under variation of the Bcd concentration.

The genotypic canalization studied in Chapter 3 is not only qualitative, but quantitative as well. Therefore, qualitative structural stability, though demonstrated in gene expression data and the dynamical system, does not fully capture the kind of genotypic canalization exhibited by the gap gene system.

6.2.3 Stability of developmental trajectories

In Section 1.1 it was argued that a necessary condition for canalization is that the phenotype of an animal be stable against variable conditions during development. Homeostasis, that is, the specification of the cell fates by steady states offers a mechanism for such stability.

Many models describe different biological states in terms of point attractors. The following are a few examples. Reinitz and Vaisnys (1990) describe the lytic and lysogenic states of phage lambda as point attractors. von Dassow et al. (2000) model the establishment and maintenance of *Drosophila* segment polarity expression patterns in steady state. Umulis et al. (2006) propose a model with positive feedback in which initial dorsoventral patterning in *Drosophila* occurs due to a choice between two steady states of the Decapentaplegic protein. Huang et al. (2007) describe the differentiation of bone marrow-derived progenitor into a red blood cell fate or a leukocyte fate by means of a pitch-fork bifurcation.

There have been three attempts to specifically model gap gene expression patterns as steady states. The first one (Sánchez and Thieffry, 2001) uses discrete logical variables, while the other two use continuous variables (Alves and Dilão, 2006; Bergmann et al., 2007).

Gap gene models with homeostasis

Sánchez and Thieffry (2001) used multi-level logical variables to model the gap genes. A state transition table defines the dynamics of the system, and logical states that are stable under the action of the transition table represent gap gene patterns in the blastoderm. Boundaries of gap gene expression patterns are formed by the establishment of the stable states in a different logical state due to maternal input. The patterns produced by this model have many defects, the most notable being that the posterior borders of the anterior *hb* domain and the central *Kr* domain coincide.

The key difference between this approach and the gene circuit approach is that the latter does not assume the existence of stable steady states, consequently, it is more general and captures dynamical features that the logical analysis cannot. For instance, in the gene circuit phase space, though $A^1_{0,4}$ is a point attractor, it moves continuously as Bcd is varied (Chapter 5). Finally, one idea from the logical analysis finds firmer footing in qualitative dynamics. In their analysis Sánchez and Thieffry (2001) found feedback loops that dominate the dynamics of the system. One such loop operative in the posterior region of the embryo is the negative feedback loop *gt-kni-Kr*. These are precisely the dynamics that occur on the attracting manifold U_{+}^{2} (Section 5.4.2).

In the work of Alves and Dilão (2006), gap gene patterns are described by equilibrium binding of each transcription factor to a single site in a gene's promoter. The transient patterns are determined by first order ODEs based on the law of mass action. The gap gene patterns are the steady state solutions of the ODEs, and are reached in biological time. It was found, however, that the correspondence between such solutions and gap gene expression data from FlyEx was weak. Apart from the patterns being qualitatively different, the anterior *hb* domain, the anterior *gt* domain, and the anterior border of the *Kr* domain all fall outside the error range of the data.

Bergmann et al. (2007) present a model for gap genes that relies on individual Hill functions for each regulator, which is equivalent to having individual thresholds for each regulator. This model is designed such that gap gene patterns once established, do not depend on the morphogen. This is achieved by reaching steady state around cycle 13. The posterior boundaries of the third anterior *gt* domain and the anterior *hb* coincide in this model (cf. Figure 3.1b), which is incorrect. Also the gap gene domains are mutually exclusive throughout the duration of the model. Note that even though in the diffusionless case of gene circuits have almost mutually exclusive domains in late cycle 14 (Figure 3.1d), they are not so earlier (Figure 3.1c). Also, the gene circuit has overlapping domains in the presence of diffusion.

Homeorhesis and chreods

As shown above, descriptions of gap genes that have only involved homeostasis do not faithfully capture the dynamics of the gap gene system, or their patterns.

So far, gene circuits have been the most successful at capturing the dynamics of the gap gene system (Perkins et al., 2006). The qualitative dynamical analysis (Section 5.4) shows that gene circuits do not rely exclusively on steady states to make their patterns. At the same time, their trajectories are canalized by the attracting manifold U_+^2. This leads to the question of whether U_+^2 is a chreod.

The attracting manifold U_+^2 (Figure 5.7) is a invariant trajectory that reduces variation in maternal Hb by attracting all indirect route trajectories from the basin of attraction of $A_{0,4}^1$. However, U_+^2 itself is not a biological trajectory since its starting point is in the unstable eigenspace of $S_{1,3}^2$ and not in the biological set of initial conditions that lie on the Hb axis (Section 2.1).

It is also incorrect to regard any one indirect route trajectory as the stable developmental trajectory, since different maternal Hb initial conditions specify different gap gene states (Figure 5.6). Figure 5.10 shows that within a tolerance range that was determined empirically, trajectories canalize to a given state. These trajectories start from biological starting points, and show stability by attraction to U_+^2. Therefore, the set of trajectories within a tolerance range, taken together, may be regarded as a chreod.

6.2.4 Boundary formation

Gap gene boundary formation takes place due to maternal regulation and gap-gap cross-regulation (Jaeger et al., 2004a, 2007; Perkins et al., 2006). The factors that control a boundary are determined either by genetic studies or by regulatory analysis of gene circuits. Regulatory analysis (see Reinitz and Sharp, 1995; Jaeger et al., 2004a, and Section 3.3) is a local analysis in the sense that regulatory interactions are studied on a single trajectory in the phase space, one time point at a time. The qualitative dynamical analysis of Chapter 5 offers a way to understand boundary formation in terms of the global dynamical picture.

This analysis clarifies the roles of maternal factors in setting up different gap gene boundaries. Boundaries in the gap gene system form by one of three methods. The first is by the dependence of the state of a point attractor on Bcd (posterior border of third anterior *gt* domain). The second is by selection between different point attractors by the initial condition (posterior border of anterior *hb* domain and anterior border of the central *Kr* domain), and the third is by selection of a gastrulation state on the attracting manifold by the initial conditions (all borders in the posterior region).

Although the dependence of anterior segmentation genes on Bcd and maternal Hb, and of posterior ones on maternal Hb is already known (Driever and Nüsslein-Volhard, 1989; Simpson-Brose et al., 1994; Hülskamp et al., 1990; Struhl et al., 1992), and has been captured by dynamical models (Reinitz et al., 1995; Jaeger et al., 2004a, 2007), the global analysis reveals the dynamical mechanism by which these interactions happen.

The global dynamical analysis has some advantages over the local analysis. First, the whole trajectory is visualized at once, providing information that can only be obtained in the regulatory analysis by tedious repetition at several time points. Second, as was seen in Chapter 5, although the phase space is high-dimensional, the dynamics are confined to lower-dimensional manifolds. Such manifolds can only be discerned in a global analysis. Related to this, the local analysis only observes the dynamics on such restricted manifolds, while the global analysis reveals the whole phase space. This could be useful for understanding mutants, in which the trajectories are in other regions of the phase space not observed in wild type.

6.3 Limitations of the gene circuit approach

6.3.1 Unmatched Bcd and maternal Hb gradients

A general problem with gap gene circuits studied here (Sections 2.1.1 and 2.3.4), and earlier studies (Jaeger et al., 2004b,a) is that early gap gene domains are derepressed initially but refine later due to gap-gap cross-repression. The motivation for implementing a delay model was to address this problem (Section 2.3.1). However, despite evidence for the functional importance of protein synthesis delays (see Section 2.3.1), gene circuits with such delays failed to produce accurate early gap gene expression patterns. This leads to the question of whether the specific representation of the synthesis term used in this dissertation and earlier work (see Section 2.1.1 and Reinitz and Sharp, 1995; Jaeger et al., 2004b) is capable of giving correct early patterns given the maternal gradients Bcd, Hb, and Cad.

This question was considered by Jaeger et al. (2007) by constructing gene circuits with gap gene mRNAs as state variables. These started in cleavage cycle 10 and lasted until early cleavage cycle 13. In these models, there are no gap-gap interactions, and pattern formation is driven solely by maternal Bcd, Hb, and Cad. In the early part of cleavage cycle 13, these circuits are essentially equivalent to gene circuits with production delays (Section 2.3.2), in which the effects of gap-gap cross regulation don't appear until the first ten minutes of cycle 13. In addition, different synthesis terms were tried to see if they work better than the one in Eq. (2.1).

The gene circuits obtained by Jaeger et al. (2007) using the same synthesis term and maternal gradients as were used in this dissertation (Eq. 2.1) are only able to form the borders of the abdominal *kni* domain correctly. *gt* shows derepression between the third anterior domain and the abdominal domain, while *Kr* central domain forms either one of the anterior or posterior borders, but not both. The gene circuits that have production delays have the same patterning defects (Figure 2.9c). Using other synthesis terms Jaeger et al. (2007) were able to obtain correct patterns for *gt*, but not for *Kr*. These results led to the conclusion that Bcd activation and maternal Hb repression are not correctly balanced in the early gap gene circuits.

A similar imbalance between Bcd and Hb was observed in the dynamical analysis of gene circuits (Chapter 5). The most anterior nucleus has an incorrect state (Section 5.4.1 and Table 5.4) since the maternal Hb concentration falls outside the basin of the biologically correct state, *hb,gt*-on. To require that the Bcd activation and Hb repression be carefully balanced seems biologically incorrect since the two gradients are established by two independent maternal systems (Driever and Nüsslein-Volhard, 1988a). However, it was shown in Section 5.4.2 that a given Bcd gradient imposes tolerance ranges within which the maternal Hb concentrations must occur (Figure 5.10) for correct pattern formation. Hence the Bcd and Hb gradients do not have to be strictly correlated, but be within a certain tolerance of each other. There is certainly such a limit in the posterior, where ubiquitous maternal Hb in embryos from *nos*$^-$ mothers causes segmentation defects. Note that the proposed occurrence of Bcd and Hb gradients within a certain tolerance is different from posterior gradient models (Section 6.1.4). The two gradients don't have to be strictly correlated in their geometrical properties or in the mechanisms that establish them.

6.3.2 Simulating mutants

In Chapter 3 it was shown that gene circuits show reduction of the variance of gap gene borders with respect to Bcd due to gap-gap cross regulation (see Figure 3.3 for example). This prediction was verified in *Kr*/*kni* double mutants (Chapter 4), in which the posterior border of the anterior *hb* domain has increased variation with respect to Bcd. However, the variance of this border is unchanged in *Kr*$^-$ or *kni*$^-$ (Houchmandzadeh et al., 2002, and unpublished data of S. Surkova).

In *Kr*$^-$, the anterior border of the posterior *gt* domain shifts anteriorly (Kraut and Levine, 1991a), by about 15% A–P (unpublished data of Svetlana Surkova). Hence it is possible that *gt* and *kni* might together regulate the posterior border of the anterior *hb* domain. Another hypothesis is that a single regulator might be sufficient to regulate *hb*.

So far, the only zygotic mutant that has been simulated correctly by gene circuits is *tll*$^-$/*hkb*$^-$ (see Section 5.1 and Figure 5.1). Therefore it is currently not possible to test the above hypotheses by simulating *Kr*$^-$ or *kni*$^-$ embryos. This might be because the wild type patterns impose a lesser constraint on the parameters of the gene circuit compared to mutant patterns. However, efforts to constrain the parameters more by optimizing on the wild type and Kr^- datasets together have not succeeded either.

6.3.3 Diffusion

The analysis of Chapter 5 relies upon the correctness of the diffusionless approximation (Section 5.1). Although the progression of the patterns and boundary placement largely agree (Figure 3.1), the question of whether the dynamical structure of the nuclei, and the results it implies (Section 5.4) are preserved in the presence of diffusion. As suggested by Thom (1983), it likely that dynamical structure in the two cases will be topologically equivalent as the diffusion term is a linear perturbation, and hence, structurally stable. This still remains to be explicitly verified, since a dynamical analysis of the gene circuit in the presence of diffusion has not been attempted here.

6.3.4 Comparison of circuits

It is important to point out that the dynamical analysis of Chapter 5 has been performed on only one set of parameters, that of gene circuit `hkgn58c13k1_007`. In Section 3.2, it was verified that the property of the circuit to reduce positional variance does not depend on a particular circuit, or a particular Bcd profile. A similar comparison of these circuits needs to be done with respect to dynamical structure.

6.4 Future directions

As pointed out in Section 6.3.1, one possible reason that the gene circuits presented here (Section 2.1.1 and Section 2.3.2) and early gap circuits (Jaeger et al., 2007) do not have correct early patterns is that the Bcd and maternal Hb gradients are mismatched. Since maternal Bcd, Hb and Cad used in these models were patterns averaged over several embryos (Surkova et al., 2007), they might not be representative of gradients in individual embryos (Section 2.1.6), nor would they have the same spatial relationship as in an individual embryo. This situation can be remedied by staining for Bcd, Hb and Cad simultaneously and using all three gradients from a single embryo in the model. The embryo picked would have to be representative. This might be accomplished by employing the methods used to pick a median Bcd profile (Section 2.1.6). Only the Hb gradient would need to be parameterized since the maternal Cad gradient is established by Bcd (Rivera-Pomar et al., 1996).

As mentioned in Section 6.3.2, the gene circuit can simulate only one mutant genotype successfully so far, and efforts to fit a Kr^1 dataset have not succeeded. These data show counterintuitive effects, where the removal of a repressor, *Kr*, leads to the reduction in levels of other genes. It is possible that the current representation of interactions that only occur between a pair of regulators is incorrect, and the synthesis term needs to be revised along the lines of Jaeger et al. (2007). Also, for correct prediction of maternal mutants, regulation between maternal gradients (Irish et al., 1989; Rivera-Pomar et al., 1996) needs to be explicitly represented. For instance in nos^-, the abdominal domain of Cad is absent due to ubiquitous maternal Hb.

Simulation of genotypic canalization in this dissertation has involved only the two parameters that specify a Bcd profile. For a complete simulation, all the regulatory parameters might be varied. V. Gursky (personal communication) has been performing such simulations that suggest that gene circuits show canalization, within limits, with respect to perturbations of the regulatory parameters.

For a complete dynamical characterization of the gap gene system, the three-dimensional stable manifolds of the hyperbolic equilibria $S^1_{1,3}$ and $S^2_{1,3}$ need to be calculated. Currently, only one method, using computational set theoretic techniques (Dellnitz et al., 2001), has been implemented that can do such a calculation. However, this method is impractical for gene circuit problem since simultaneous representation of both very large and small scales is required. In principle, other methods (Henderson, 2005; Guckenheimer and Vladimirsky, 2004; Krauskopf and Osinga, 2003) can be used for manifold of dimension greater than two, but implementations only exist for two dimensions. Therefore, new algorithms need to be developed for calculating the stable manifolds for the gene circuit.

In the dynamical analysis of Section 5.4 three mechanisms of boundary formation were identified. Since the first mechanism involves the movement of an attractor with changing Bcd levels, this leads to a prediction. If one constructed an embryo with uniform Bcd at high levels, the model predicts that the posterior boundary of the third anterior *gt* domain will coincide with the posterior boundary of the anterior *hb* domain. With uniform Bcd, $A^1_{0,4}$ will remain in the *hb,gt*-on state at all A–P positions, hence the posterior boundary of the third *gt* domain will not form by its usual mechanism. Instead, it will form by the second mechanism, namely the switching of a nucleus from the basin of $A^1_{0,4}$ (*hb*,*gt*-on) to that of $A^3_{0,4}$ (*Kr*-on). Hence the *hb* border and the posterior border of the third anterior *gt* domain will coincide.

Perhaps the biggest advance in understanding canalization and the stability of developmental trajectories will come with the construction of fusions of all segmentation gene proteins with GFP. It will become possible to directly observe the trajectories with higher temporal resolution than is possible at present. Knowledge of absolute concentrations will allow us to construct more accurate dynamical models that can correctly predict mutants. With these tools it will be possible to understand the organization of the epigenetic landscape that confers such stability to the phenotype of animals.

Appendix A

Materials and methods

This appendix describes, for completeness, the lab methodology used, but not developed in the course of the dissertation work.

The first section describes the methodology with which data on segmentation genes is acquired and processed. These techniques were used to produce the integrated dataset for optimizing gene circuits, and also the data on wild type gap gene variability (Surkova et al., 2007). The same techniques were also used for the acquisition and processing of the *Kr*/*kni* double mutant data that are described in Chapter 4.

The second section describes the Parallel Lam simulated Annealing algorithm used to optimize gene circuits (Chapter 3).

A.1 Integrated data

Different steps of the data acquisition and processing pipeline are applied depending on whether the data are used for building a representative data set of average segmentation gene expression patterns (Surkova et al., 2007), or for studying variability.

A.1.1 Fixation, staining, and acquisition of images

Wild type Oregon R. embryos, 1–4 hours old, are collected on fructose-sucrose agar plates. They are fixed and fluorescently stained for Eve protein and two other segmentation gene proteins according to standard methods and antibodies as described in Kosman et al. (1998). The secondary antibodies used are Alexa Fluor 488 anti-guinea pig, Alexa Fluor 555 anti-rabbit, and Alexa Fluor 647 anti-rat. A monoclonal antibody raised in mice against the histones H1, H2a, H2b, H3, and H4 is used for nuclear staining. In this case, a biotin-conjugated anti-mouse antibody is used as secondary, with a Streptavidin Alexa Fluor 700 conjugate used as tertiary for visualization.

Laterally oriented embryos are scanned in a confocal microscope. The fluorophores are excited using a single laser at a time to prevent leakage between channels. The microscope used is a Leica TCS SP2 confocal with HC PL APO 20x objective. This microscope has tunable wavelength detection windows of 500–545, 560–645, and 650–715 nm. The laser wavelengths are 488 nm (Ar), 543 nm (HeNe), 633 nm (HeNe).

A single embryo is scanned multiple times frame-by-frame and the frames are averaged to suppress optical noise in the photo multiplier tube. Two z-sections are scanned separately to account for non-uniform localization of protein along the basal-apical axis. For a given staining, the microscope is standardized on the embryos showing the greatest signal for a particular gene. Gain is set such that a few maximally-bright pixels are saturated (had 8-bit intensity 255). The offset, which sets the zero, is used to remove optical background in the optical field around the embryo. For time classification (see

Section A.1.3), a differential interference contrast (DIC) image of the dorsal and lateral edges of the embryo is also taken at 2.5x magnification in the sagittal plane.

A.1.2 Segmentation

Segmentation is the process by which average nuclear intensities are extracted from a bitmap image obtained from the microscope (Janssens et al., 2005). Embryo images are aligned in an anterior-left, dorsal-up orientation such that the major axis of the embryo is horizontal. A whole-embryo mask, in which pixels are 255 inside the embryo, and 0 outside it, is constructed using a threshold-based method, and the image is cropped according to the mask.

The images from all channels scanned (including the histone channel) are combined to construct an image in which each nucleus's intensity is the maximum intensity observed in any channel. This maximal image is then used to construct nuclear masks. In a nuclear mask, pixels of nuclei have intensity 255, while pixels in the cytoplasm have intensity 0. To achieve this, first, the watershed method is applied to the maximal image. This method marks regions with low intensity compared to nuclei as watershed regions, defining the limits of nuclei. Then the Shen-Castan edge-detector is applied to detect the edges of nuclei (see Janssens et al., 2005, and references therein for details).

The nuclear mask allows for the calculation of the positions of nuclei relative to the A–P and D–V axes (x and y, due to the alignment of the embryo). The pixel intensities in individual channels are averaged according to the nuclear mask, giving a table of positions of nuclei, and the average intensities in them.

A.1.3 Time classification

The embryos are anywhere between one hour to four hours old at the time of fixation. Therefore in a given experiment, a distribution of embryos of varying ages is obtained. However, for the purposes of averaging expression patterns to build a dataset, or for studying variability, only similarly aged embryos should be compared. For this reason embryos are staged according to age (Myasnikova et al., 2002). First, embryos are classified into cleavage cycles. Data is only acquired for the syncytial blastoderm stage from cleavage cycle 10 to 14A. Transverse sections of cycle 10 embryos have $\sim$130 nuclei, cycle 11 embryos have $\sim$260, cycle 12 embryos have $\sim$450, cycle 13 embryos have $\sim$1000, and cycle 14A embryos have $\sim$2000. Cycles 10 to 13 are short (8–12 mins) (Foe and Alberts, 1983), and embryos in these stages are not further classified. Cycle 14A lasts for about 50 mins, and expression patterns change rapidly. Hence, it is subdivided further into eight time classes T1–T8. Two criteria are used: (1) patterns of the pair-rule gene *eve*, (2) morphology of nuclei and invagination of the cell membrane.

Time classification using *eve* expression patterns is based on careful visual inspection. T1–T4 can be distinguished by the number of *eve* stripes present in the pattern. T5–T8 are distinguished based on features of the overall *eve* pattern. This method has been cross-validated by an automated classification method that uses complex-valued neural networks (Aizenberg et al., 2002), and support-vector regression (Myasnikova et al., 2002).

Time classification using membrane morphology relies on 2.5x DIC images of the dorsal and lateral edges of embryos in a sagittal section. During T1–T2, nuclei are round and the membrane is not visible. In T3, the nuclei elongate slightly. In T4, the elongation continues, and the membrane becomes clearly visible at the apical end. In T5, the membrane is in the middle of nuclei, while in T6 it is at their basal end. In T7, it goes beyond the nuclei. T8 is distinguishable by irregular positioning of the nuclei since the embryo is starting to gastrulate.

A.1.4 Background removal

Immunofluorescently-stained images always have some background signal due to non-specific binding of the antibodies. By staining for a particular protein in a mutant background, it was determined that the background is well-approximated by a quadratic paraboloid (Myasnikova et al., 2005). For an individual wild type embryo, the background signal is determined by fitting a quadratic paraboloid to the segmented data. This signal is then subtracted and the data renormalized by a linear transformation such that non-expressing nuclei have zero intensity, and maximally-expressing nuclei have intensity 255.

At this stage of data processing, one of two things can be done depending on the final application of the data. For studying variation, no further processing is needed. For building an integrated dataset of representative wild type gene expression patterns, the final step of registration and averaging has to be applied.

A.1.5 Registration and averaging

Segmentation gene expression patterns vary from individual to individual. Though this variation is low compared to the maternal environment (see Section 1.3), it is enough to render a naive averaging of patterns useless. Registration is the process by which patterns from individuals are aligned so that they can be averaged meaningfully to build an integrated dataset. This is done by determining features of the *eve* expression pattern and aligning patterns of different embryos such that the features line up. The features are extracted by a fast dyadic wavelet transform, that yields points where the first derivative of the *eve* expression pattern is zero. Once the features are extracted, an affine transformation is applied to the expression patterns such that the distance between a feature's position and the average feature position is minimized for each feature. The same affine transformation is applied to all the channels recorded from the embryo, so that the relative placement of different genes' expression patterns is maintained. For details of the registration method described above, see Myasnikova et al. (2001).

The final step is to construct the average dataset by averaging the registered patterns of all embryos in a given time class. Since the number of nuclei and their positions vary from embryo to embryo, the embryo is divided up into equal-sized bins along the A–P axis. Then for a given segmentation protein, the intensity is averaged over all the nuclei in a bin and all embryos are stained for that protein.

A.2 Optimization

The parameters of the gene circuit were determined (Eq. 2.1) by Parallel Lam Simulated Annealing (Chu et al., 1999; Chu, 2001, PLSA). The squared difference between model output and integrated data (Section A.1) was minimized by changing the model parameters according to PLSA. In this section, the cost function and the annealing search spaces are defined first. Then the PLSA algorithm is described briefly.

A.2.1 Cost function and search spaces for parameters

The cost function is given by

$$E = \sum \left(v_i^a(t)_{\text{model}} - v_i^a(t)_{\text{data}}\right)^2 + E_{\text{penalty}}, \tag{A.1}$$

where, $v_i^a(t)_{\text{model}}$ is the protein concentration of gene a in nucleus i at time t. It is the solution of Eq. (2.1). t is the midpoint of a time class (cycle 13, or one of the eight time classes in cycle 14, T1–T8). $v_i^a(t)_{\text{data}}$, is the integrated data for protein a in nucleus i in the corresponding time class. The summation is over all time classes t, nuclei i, and genes a represented in the model, and for which we have data. Let the total number of such points be N_d. Then, the root mean square score is given by

$$\sqrt{\frac{E}{N_d}}. \tag{A.2}$$

The rms score is a measure of the quality of fit. Although, in a given optimization, minimizing E also leads to the minimization of the rms score, the optimization algorithm is specifically implemented for minimizing E, and not the rms score.

We can restrict the solution of the optimization problem (the parameters of the gene circuit) by specifying constraints on their values. These constraints are called search spaces. Table A.1 lists the search spaces for the kinetic parameters R^a, D^a, $t_{1/2}^a$, τ^a used in fitting the gene circuits used in this dissertation.

For the regulatory parameters T^{ab}, m^a, $E^{a\beta}$, and h^a, a penalty function E_{penalty} was specified instead. The penalty function provides a flexible search space of these parameters so that the saturated part of the threshold function $g(u^a)$ (Figure 2.2) is not sampled endlessly. The penalty function is given by

Parameter	**Lower bound**	**Upper bound**
R^a (min^{-1})	10.0	15.0
D^a (min^{-1})	0.0	0.2
$t^a_{1/2}$ (min)	5.0	18.0
τ^a (min)	2.0	12.0

Table A.1: Search Spaces for Gene Circuits. These parameters are defined in Section 2.1.2 and Section 2.3.2. Note that τ^a was only used in delay models.

$$E_{\text{penalty}} = \begin{cases} \exp\left(\Lambda \Pi_{u^a}\right) - \exp(1) & \text{iff} \quad \Lambda \Pi_{u^a} > 1, \\ 0 & \text{otherwise,} \end{cases} \tag{A.3}$$

where

$$\Pi_{u^a} = \left(\sum_{ab} (T^{ab} v^b_{\text{max}})^2 + (E^{a\beta} v^\beta_{\text{max}})^2 + \left(m^a v^{\text{Bcd}}_{\text{max}}\right)^2 + (h^a)^2 \right).$$

Here, v^b_{max} is the maximum concentration of protein b in any nucleus, v^β_{max} is the maximum concentration of nonautonomous input β, and $v^{\text{Bcd}}_{\text{max}}$ is the maximum concentration of *bcd*. Λ is a free parameter that controls how far from the regulated region the parameters are allowed to go. From Eq. (A.3), it can be seen that the penalty will be active when

$$\sum_{ab} (T^{ab} v^b_{\text{max}})^2 + (E^{a\beta} v^\beta_{\text{max}})^2 + \left(m^a v^{\text{Bcd}}_{\text{max}}\right)^2 + (h^a)^2 > \frac{1}{\Lambda}. \tag{A.4}$$

We chose $\Lambda = 10^{-3}$ for all optimizations in this thesis.

The further the dynamics are from the regulated region of $g(u^a)$, the larger the left hand side of Eq. (A.4), hence greater the penalty. Since we are minimizing the cost function (Eq. A.1), the penalty term puts pressure on the parameters to remain in the regulated region of $g(u^a)$.

A.2.2 Parallel Lam Simulated Annealing

Simulated Annealing

Simulated annealing (Kirkpatrick et al., 1983) is an optimization method based on the Metropolis algorithm (Metropolis et al., 1953). The Metropolis algorithm is a method to calculate an ensemble average of a system at thermodynamic equilibrium. Let the state variables be specified by a vector x, and E be the energy of a state. Given a temperature T, the Metropolis algorithm samples the Boltzmann probability distribution by the following sequence of steps:

1. Compute $E = E_{\text{old}}$ at state x_{old}.
2. Propose a new state x_{new} ("make a move").
3. Compute $E = E_{\text{new}}$ at x_{new}.
4. If $\Delta E = E_{\text{new}} - E_{\text{old}} < 0$, $x_{old} = x_{new}$ ("accept the move").
5. If $\Delta E = E_{\text{new}} - E_{\text{old}} > 0$, $x_{old} = x_{new}$ with Boltzmann probability $\exp(-\Delta E/T)$, otherwise discard x_{new} ("reject the move")

Simulated annealing (SA) adapts the Metropolis algorithm to simulate slow cooling (annealing) by slowly lowering the temperature at each step. The rule that determines how much the temperature is lowered at each step is called the cooling schedule. It is used for optimization by letting the parameters of the system be the state variables x_i of the simulated thermodynamic system.

Lam simulated annealing

Lam simulated annealing (LSA Lam and Delosme, 1988a,b) provides an adaptive cooling schedule, so that it can lower temperature faster in regions of the state space where the cost function has simpler structure. LSA also provides a scheme to control move generation so that the state space is sampled most effectively.

Let $s_k = 1/T_k$ be the inverse temperature at the k^{th} iteration of Lam SA. The inverse temperature in the next step is determined so that the system remains in quasi-equilibrium. It is determined by the equation

$$s_{k+1} = s_k + \lambda \left(\frac{1}{\sigma(s_k)} \right) \left(\frac{1}{s_k^2 \sigma^2(s_k)} \right) \left(\frac{4\rho_0(s_k)(1-\rho_0(s_k))^2}{(2-\rho_0(s_k))^2} \right), \tag{A.5}$$

where $\sigma(s_k)$ is the standard deviation of the energy E at step k, and $\rho_0(s_k)$ is the acceptance ratio, that is, the ratio of accepted moves to the total number of moves proposed.

λ is a quality parameter, such that low values of it lead to better fits with greater computational effort. $\sigma(s_k)$ is the standard deviation of E at temperature s_k, and its inverse measures the distance of the system from equilibrium. $\left(\frac{1}{s_k^2\sigma^2(s_k)}\right)$ is the specific heat. The last term, $\left(\frac{4\rho_0(s_k)(1-\rho_0(s_k))^2}{(2-\rho_0(s_k))^2}\right)$, is proportional to the variance of the distribution of proposed moves, and measures how well the state space is being sampled. It is at a maximum when the acceptance ratio $\rho_0(s_k) \sim 0.44$.

In order to sample the state space most effectively, the acceptance ratio must be maintained near 0.44. This is done by controlling the size of the moves. For parameter x_i, the moves are drawn from an exponential distribution with mean $\overline{\theta}^i$ by the equation

$$x_i^{new} = x_i \pm \overline{\theta}^{\,i} \ln \xi, \tag{A.6}$$

where ξ is drawn uniformly from $[0, 1]$. $\overline{\theta}^{\,i}$ thus controls the size of the move for a parameter x_i and is adjusted by the feedback equation

$$\ln \overline{\theta}^{\,i}_{new} = \ln \overline{\theta}^{\,i}_{old} + 3.0(\rho_0 - 0.44). \tag{A.7}$$

Thus if the current acceptance ratio is above 0.44, the average move size is increased, which will cause more rejected moves, and vice versa.

Finally, in order to evaluate Eq. (A.5), the algorithm requires the acceptance ratio, $\rho(s)$, and the standard deviation of the energy, $\sigma(s)$. The acceptance ratio changes slowly with temperature, and can be estimated from previous moves as the ratio of accepted to proposed moves. Lam and Delosme (1988b) provided estimators for the mean and standard deviation of energy that are determined from the energies of past moves by least squares fitting every 100 moves.

Parallel Lam Simulated Annealing

PLSA exploits the inherent ergodicity of the Metropolis for parallelization (Chu et al., 1999; Chu, 2001). The state of the system is independently sampled by P processors using the Lam schedule for cooling. The increased speed comes about by increasing the rate at which the temperature is lowered (P times faster). This is made possible by the increased rate at which statistics are collected by these P processors.

Due to the faster lowering of temperature, the processors will eventually go out of quasi-equilibrium. To rectify this problem, the states of different processors are mixed. This is done after every $100m/P$ Metropolis iteration on a single processor. Each processor polls every other processor and retrieves their states' energy E. A processor adopts the state of another processor p with probability P_p given by

$$P_p = \frac{e^{-E_p/T}}{\sum_i e^{-E_i/T}}. \tag{A.8}$$

This ensures that quasi-equilibrium is maintained.

The frequency of mixing, m, has to be chosen carefully. Immediately after mixing of states, the processors' states are correlated. If another mixing is done too soon, they will remain correlated throughout the annealing run, and the state space will not be sampled effectively. Thus the lower bound for m is the number of steps it takes for the cross-correlation function between processors to vanish.

If there is too long a gap between mixing, the processors' states will not be in quasi-equilibrium. Therefore, the upper bound for m is the number of evaluations of E taken for the variance of mean energy between processors to exceed an empirically-determined threshold of 7%. The lower and upper bounds for m are determined in a tuning run Chu (2001), and m is set to a value between them. For optimizations carried out in this thesis, m was chosen to be 13.

Starting and stopping an optimization

A random state x_0 is chosen as a starting point. Then the Metropolis algorithm is run for 10^5 E evaluations at a temperature of 10^6. This is done so that there is no memory of the initial state. Initial Lam statistics (Chu, 2001) are collected over an additional $10^5/P$ evaluations, after which a normal PLSA or

LSA optimization starts.

An optimization is stopped when E changes less than 10^{-3} over 500 evaluations.

Bibliography

Aegerter-Wilmsen, T., Aegerter, C. M., and Bisseling, T. (2005). Model for the robust establishment of precise proportions in the early *drosophila* embryo. *Journal of Theoretical Biology*, 234:13–19.

Aizenberg, I., Myasnikova, E., Samsonova, M., and Reinitz, J. (2002). Temporal classification of *Drosopila* segmentation gene expression patterns by the multi-valued neural recognition method. *Mathematical Biosciences*, 176:145–159.

Akam, M. (1987). The molecular basis for metameric pattern in the *Drosophila* embryo. *Development*, 101:1–22.

Alves, F. and Dilão, R. (2006). Modeling segmental patterning in *Drosophila*: Maternal and gap genes. *Journal of Theoretical Biology*, 241:342–359.

Audibert, A., Weil, D., and Dautry, F. (2002). In vivo kinetics of mrna splicing and transport in mammalian cells. *Molecular Cell Biol.*, 22:6706–6718.

Bellman, R. E. and Cooke, K. L. (1993). *Differential-Difference Equations*, volume 6. Mathematics in Science and Engineering, Academic Press.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2007). Genbank. *Nucleic Acids Research*, 35:D21–D25. doi:10.1093/nar/gkl986.

Berg, H. C. and Purcell, E. (1977). Physics of chemoreception. *Biophysical Journal*, 20:193–219.

Bergmann, S., Sandler, O., Sberro, H., Shnider, S., Schejter, E., Shilo, B., and Barkai, N. (2007). Pre-steady-state decoding of the Bicoid morphogen gradient. *PLoS Biology*, 5(2):232–242.

Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M., and Nüsslein-Volhard, C. (1988). The role of localization of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo. *The EMBO Journal*, 7:1749–1756.

Beyer, A. L. and Osheim, Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes and Development*, 2:754–765.

Bogacki, P. and Shampine, L. F. (1989). A 3(2) pair of Runge-Kutta formulas. *Applied Mathematics Letters*, 2(4):321–325.

Brönner, G. and Jäckle, H. (1991). Control and function of terminal gap gene activity in the posterior pole region of the *Drosophila* embryo. *Mechanisms of Development*, 35:205–211.

Bulirsch, R. and Stoer, J. (1992). *Introduction to Numerical Analysis.* Springer-Verlag, New York, second edition.

Campos-Ortega, J. A. and Hartenstein, V. (1985). *The Embryonic Development of* Drosophila melanogaster. Springer, Heidelberg, Germany.

Casanova, J. (1990). Pattern formation under the control of the terminal system in the *Drosophila* embryo. *Development*, 110:621–628.

Chu, K.-W. (2001). *Optimal Parallelization of Simulated Annealing by State Mixing.* PhD Thesis, Department of Applied Mathematics and Statistics, Stony Brook University.

Chu, K. W., Deng, Y., and Reinitz, J. (1999). Parallel simulated annealing by mixing of states. *The Journal of Computational Physics*, 148:646–662.

Clyde, D. E., Corado, M. S., Wu, X., Pare, A., Papatsenko, D., and Small, S. (2003). A self-organizing system of repressor gradients establishes segmental complexity in *Drosophila*. *Nature*, 426:849–853.

Conte, S. D. and de Boor, C. (1980). *Elementary Numerical Analysis.* McGraw-Hill Book Company, New York, New York.

Crauk, O. and Dostatni, N. (2005). Bicoid determines sharp and precise target gene expression in the *drosophila* embryo. *Current Biology*, 15:1888–1898.

Crosby, M., Goodman, J., Strelets, V., Zhang, P., Gelbartand, W., and the FlyBase Consortium (2007). Flybase: genomes by the dozen. *Nucleic Acids Research*, 35:D486–D491. doi:10.1093/nar/gkl827.

Dalton, D., Chadwick, R., and McGinnis, W. (1989). Expression and embryonic function of *empty spiracles*: A *Drosophila* homeo box gene with two patterning functions on the anterior-posterior axis of the embryo. *Genes and Development*, 3:1940–1956.

Davidson, E. H. (1986). *Gene Activity in Early Development*. Academic Press, third edition.

Dellnitz, M., Froyland, G., and Junge, O. (2001). The algorithms behind gaio - set oriented numerical methods for dynamical systems. In Fiedler, B., editor, *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*, pages 145–174.

Dormand, J. R. (1996). *Numerical Methods for Differential Equations*. CRC Press, Boca Raton, FL.

Driever, W. and Nüsslein-Volhard, C. (1988a). The Bicoid protein determines position in the *drosophila* embryo in a concentration-dependent manner. *Cell*, 54:95–104.

Driever, W. and Nüsslein-Volhard, C. (1988b). A gradient of Bicoid protein in *Drosophila* embryos. *Cell*, 54:83–93.

Driever, W. and Nüsslein-Volhard, C. (1989). The Bicoid protein is a positive regulator of *hunchback* transcription in the early *drosophila* embryo. *Nature*, 337:138–143.

Driever, W., Thoma, G., and Nüsslein-Volhard, C. (1989). Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the Bicoid morphogen. *Nature*, 340:363–367.

Duffy, J. B. and Perrimon, N. (1994). The torso pathway in *Drosophila*: Lessons on receptor Tyrosine Kinase signaling and pattern formation. *Developmental Biology*, 166:380–395.

Eldon, E. D. and Pirrotta, V. (1991). Interactions of the *Drosophila* gap gene *giant* with maternal and zygotic pattern-forming genes. *Development*, 111:367–378.

Finkelstein, R. and Perrimon, N. (1990). The *orthodenticle* gene is regulated by *bicoid* and *torso* and specifies *drosophila* head development. *Nature*, 346:485–488.

Foe, V. E. and Alberts, B. M. (1983). Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis. *The Journal of Cell Science*, 61:31–70.

Frohnhöfer, H. G. and Nüsslein-Volhard, C. (1987). Maternal genes required for the anterior localization of *bicoid* activity in the embryo of *Drosophila*. *Genes and Development*, 1:880–890.

Gao, Q. and Finkelstein, R. (1998). Targeting gene expression to the head: the *Drosophila orthodenticle* gene is a direct target of the bicoid morphogen. *Development*, 125:4185–4193.

Gaul, U. and Jäckle, H. (1987). Pole region-dependent repression of the *drosophila* gap gene *krüppel* by maternal gene products. *Cell*, 51:549–555.

Gaul, U., Seifert, E., Schuh, R., and Jäckle, H. (1987). Analysis of *krüppel* protein distribution during early *Drosophila* development reveals posttranscriptional regulation. *Cell*, 50:639–647.

Gilbert, S. F. and Sarkar, S. (2000). Embracing complexity: Organicism for the 21st century. *Developmental Dynamics*, 219:1–9.

Gottesfeld, J. M. and Forbes, D. J. (1997). Mitotic repression of the transcriptional machinery. *Trends in Biochemical Sciences*, 22:197–202.

Gregor, T., Tank, D. W., Wieschaus, E. F., and Bialek, W. (2007a). Probing the limits to positional information. *Cell*, 130:153–164.

Gregor, T., Wieschaus, E. F., McGregor, A. P., Bialek, W., and Tank, D. W. (2007b). Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell*, 130:141–152.

Guckenheimer, J. and Vladimirsky, A. (2004). A fast method for approximating invariant manifolds. *SIAM Journal of Applied Dynamical Systems*, 3:232–260.

Henderson, M. E. (2005). Computing invariant manifolds by integrating fat trajectories. *SIAM Journal of Applied Dynamical Systems*, 4(4):832–882.

Hirsch, M., Smale, S., and Devaney, R. (2004). *Differential Equations, Dynamical Systems, and an Introduction to Chaos.* Academic Press, Boston.

Hoch, M., Seifert, E., and Jäckle, H. (1991). Gene expression mediated by cis-acting sequences of the *Krüppel* gene in response to the *drosophila* morphogens Bicoid and Hunchback. *The EMBO Journal*, 10:2267–2278.

Hochstadt, H. (1975). *Differential Equations: A Modern Approach.* Dover Publications, New York.

Houchmandzadeh, B., Wieschaus, E., and Leibler, S. (2002). Establishment of developmental precision and proportions in the early *Drosophila* embryo. *Nature*, 415:798–802.

Houchmandzadeh, B., Wieschaus, E., and Leibler, S. (2005). Precise domain specification in the developing *Drosophila* embryo. *Physical Review E*, 72. Art. No. 061920 Part 1.

Howard, M. and ten Wolde, P. R. (2005). Finding the center reliably: Robust patterns of developmental gene expression. *Physical Review Letters*, 95:208103.

Huang, S., Guo, Y., May, G., and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental Biology*, 305:695–713.

Hülskamp, M., Pfeifle, C., and Tautz, D. (1990). A morphogenetic gradient of Hunchback protein organizes the expression of the gap genes *krüppel* and *knirps* in the early *drosophila* embryo. *Nature*, 346:577–580.

Ingham, P. W. (1988). The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature*, 335:25–34.

Irish, V., Lehmann, R., and Akam, M. (1989). The *Drosophila* posterior-group gene *nanos* functions by repressing *hunchback* activity. *Nature*, 338:646–648.

Jaeger, J. (2005). *Dynamic Regulatory Analysis of the Gap Gene Network in* Drosophila melanogaster. PhD Thesis, Department of Genetics, Stony Brook University.

Jaeger, J., Blagov, M., Kosman, D., Kozlov, K. N., Manu, Myasnikova, E., Surkova, S., Vanario-Alonso, C. E., Samsonova, M., Sharp, D. H., and Reinitz, J. (2004a). Dynamical analysis of regulatory interactions in the gap gene system of *Drosophila melanogaster*. *Genetics*, 167:1721–1737.

Jaeger, J., Sharp, D. H., and Reinitz, J. (2007). Known maternal gradients are not sufficient for the establishment of gap domains in *Drosophila melanogaster*. *Mechanisms of Development*, 124:108–128.

Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K. N., Manu, Myasnikova, E., Vanario-Alonso, C. E., Samsonova, M., Sharp, D. H., and Reinitz, J. (2004b). Dynamic control of positional information in the early *Drosophila* embryo. *Nature*, 430:368–371.

Janssens, H., Kosman, D., Vanario-Alonso, C. E., Jaeger, J., Samsonova, M., and Reinitz, J. (2005). A high-throughput method for quantifying gene expression data from early *Drosophila* embryos. *Development, Genes and Evolution*, 215:374–381.

Johnston, D. S., Beuchle, D., and Nüsslein-Volhard, C. (1991). *staufen*, a gene required to localize maternal RNAs in the *Drosophila* egg. *Cell*, 66:51–63.

Johnston, D. S. and Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the *drosophila* embryo. *Cell*, 68:201–219.

Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C., and Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. II. Zygotic loci on the third chromosome. *Roux's Archives of Developmental Biology*, 193:283–295.

Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *The Journal of Theoretical Biology*, 22:437–467.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220:671–680.

Knipple, D. C., Seifert, E., Rosenberg, U. B., Preiss, A., and Jäckle, H. (1985). Spatial and temporal patterns of *krüppel* gene expression in early *Drosophila* embryos. *Nature*, 317:40–44.

Kosman, D., Small, S., and Reinitz, J. (1998). Rapid preparation of a panel of polyclonal antibodies to *Drosophila* segmentation proteins. *Development, Genes and Evolution*, 208:290–294.

Krauskopf, B. and Osinga, H. (2003). Computing geodesic level sets on global (un)stable manifolds of vector fields. *SIAM Journal of Applied Dynamical Systems*, 4(2):546–569.

Krauskopf, B. and Osinga, H. (2005). A survey of methods for computing (un)stable manifolds of vector fields. *International Journal of Bifurcation and Chaos*, 15:763–791.

Kraut, R. and Levine, M. (1991a). Mutually repressive interactions between the gap genes *giant* and *krüppel* define middle body regions of the *Drosophila* embryo. *Development*, 111:611–621.

Kraut, R. and Levine, M. (1991b). Spatial regulation of the gap gene *giant* during *Drosophila* development. *Development*, 111:601–609.

Lam, J. and Delosme, J.-M. (1988a). An efficient simulated annealing schedule: Derivation. Technical Report 8816, Yale Electrical Engineering Department, New Haven, CT.

Lam, J. and Delosme, J.-M. (1988b). An efficient simulated annealing schedule: Implementation and evaluation. Technical Report 8817, Yale Electrical Engineering Department, New Haven, CT.

Lawrence, P. A. (1992). *The Making of a Fly.* Blackwell Scientific Publications, Oxford, UK.

Lehmann, R. (1988). Phenotypic comparison between maternal and zygotic genes controlling the segmental pattern of the *drosophila* embryo. *Development (Supplement)*, 104:17–27.

Lehmann, R. and Nüsslein-Volhard, C. (1991). The maternal gene *nanos* has a central role in posterior pattern formation of the *Drosophila* embryo. *Development*, 112:679–691.

Lewin, B. (2000). *Genes VII.* Oxford University Press, Oxford.

Lucchetta, E. M., Lee, J., Fu, L., Patel, N., and Ismagilov, R. (2005). Dynamics of *Drosophila* embryonic patterning network perturbed in space and time using microfluidics. *Nature*, 434:1134–1138.

Macdonald, P. M. and Struhl, G. (1986). A molecular gradient in early *Drosophila* embryos and its role in specifying the body pattern. *Nature*, 324:537–545.

McHale, P., Rappel, W.-J., and Levine, H. (2006). Embryonic pattern scaling achieved by oppositely directed morphogen gradients. *Physical Biology*, 3:107–120.

Metropolis, N., Rosenbluth, A., Rosenbluth, M. N., Teller, A., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21:1087–1092.

Mjolsness, E., Sharp, D. H., and Reinitz, J. (1991). A connectionist model of development. *The Journal of Theoretical Biology*, 152:429–453.

Mlodzik, M., Fjose, A., and Gehring, W. J. (1985). Isolation of *caudal*, a *Drosophila* homeo box- containing gene with maternal expression, whose transcripts form a concentration gradient at pre-blastoderm stage. *The EMBO Journal*, 4:2961–2969.

Mlodzik, M. and Gehring, W. J. (1987). Hierarchy of the genetic interactions that specify the anteroposterior segmentation pattern of the *drosophila* embryo as monitored by *caudal* protein expression. *Development*, 101:421–435.

Mohler, J., Eldon, E. D., and Pirrotta, V. (1989). A novel spatial transcription pattern associated with the segmentation gene, *giant*, of *drosophila. The EMBO Journal*, 8:1539–1548.

Monk, N. A. M. (2003). Oscillatory expression of hes1, p53, and nf-κb driven by transcriptional time delays. *Current Biology*, 13:1–20.

Myasnikova, E., Samsonova, A., Kozlov, K., Samsonova, M., and Reinitz, J. (2001). Registration of the expression patterns of *Drosophila* segmentation genes by two independent methods. *Bioinformatics*, 17:3–12.

Myasnikova, E., Samsonova, A., Samsonova, M., and Reinitz, J. (2002). Support vector regression applied to the determination of the developmental age of a *Drosophila* embryo from its segmentation gene expression patterns. *Bioinformatics*, 18 (Supplement):S87–S95.

Myasnikova, E., Samsonova, M., Kosman, D., and Reinitz, J. (2005). Removal of background signal from *in situ* data on the expression of segmentation genes in *Drosophila. Development, Genes and Evolution*, 215:320–326.

Nauber, U., Pankratz, M. J., Kienlin, A., Seifert, E., Klemm, U., and Jäckle, H. (1988). Abdominal segmentation of the *Drosophila* embryo requires a hormone receptor-like protein encoded by the gap gene *knirps. Nature*, 336:489–492.

Nüsslein-Volhard, C., Frohnhofer, H. G., and Lehmann, R. (1987). Determination of anteroposterior polarity in *drosophila. Science*, 238:1675–1687.

Nüsslein-Volhard, C. and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature*, 287:795–801.

Nüsslein-Volhard, C., Wieschaus, E., and Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. I. Zygotic loci on the second chromosome. *Roux's Archives of Developmental Biology*, 193:267–282.

Oro, A. E., Ong, E. S., Margolis, J. S., Posakony, J. W., McKeown, M., and Evans, R. M. (1988). The *Drosophila* gene *knirps-related* is a member of the steroid-receptor gene superfamily. *Nature*, 336:493–496.

Perkins, T. J., Jaeger, J., Reinitz, J., and Glass, L. (2006). Reverse engineering the gap gene network of *drosophila melanogaster*. *PLoS Computational Biology*, 2:e51. Epub 2006 May 19. http://compbiol.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pcbi.0020051.

Perko, L. (1996). *Differential Equations and Dynamical Systems*. Springer-Verlag, New York.

Pignoni, F., Steingrimsson, E., and Lengyel, J. A. (1992). *bicoid* and the terminal system activate *tailless* expression in the early *Drosophila* embyro. *Development*, 115:239–251.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C*. Cambridge University Press, Cambridge, U.K., second edition.

Redemann, N., Gaul, U., and Jäckle, H. (1988). Disruption of a putative Cys-zinc interaction eliminates the biological activity of the *krüppel* finger protein. *Nature*, 332:90–92.

Reinitz, J., Kosman, D., Vanario-Alonso, C. E., and Sharp, D. H. (1998). Stripe forming architecture of the gap gene system. *Developmental Genetics*, 23:11–27.

Reinitz, J. and Levine, M. (1990). Control of the initiation of homeotic gene expression by the gap genes *giant* and *tailless* in *drosophila*. *Developmental Biology*, 140:57–72.

Reinitz, J., Mjolsness, E., and Sharp, D. H. (1995). Cooperative control of positional information in *drosophila* by *bicoid* and maternal *hunchback*. *The Journal of Experimental Zoology*, 271:47–56.

Reinitz, J. and Sharp, D. H. (1995). Mechanism of *eve* stripe formation. *Mechanisms of Development*, 49:133–158.

Reinitz, J. and Vaisnys, J. R. (1990). Theoretical and experimental analysis of the phage lambda genetic switch implies missing levels of cooperativity. *The Journal of Theoretical Biology*, 145:295–318.

Renzis, S., Elemento, O., and Wieschaus, S. T. E. (2007). Unmasking activation of the zygotic genome using chromosomal deletions in the *Drosophila* embryo. *PLoS Biology*, 5:e117.

Rivera-Pomar, R., Lu, X., Perrimon, N., Taubert, H., and Jäckle, H. (1995). Activation of posterior gap gene expression in the *Drosophila* blastoderm. *Nature*, 376:253–256.

Rivera-Pomar, R., Niessing, D., Schmidt-Ott, U., Gehring, W. J., and Jäckle, H. (1996). RNA binding and translational suppression by Bicoid. *Nature*, 379:746–749.

Rothe, M., Nauber, U., and Jäckle, H. (1989). Three hormone receptor-like *Drosophila* genes encode an identical DNA-binding finger. *The EMBO Journal*, 8:3087–3094.

Rothe, M., Pehl, M., Taubert, H., and Jäckle, H. (1992). Loss of gene function through rapid mitotic cycles in the *Drosophila* embryo. *Nature*, 359:156–159.

Rutherford, S. L. and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature*, 396:336–342.

Sánchez, L. and Thieffry, D. (2001). A logical analysis of the *Drosophila* gap-gene system. *The Journal of Theoretical Biology*, 211:115–141.

Schröder, C., Tautz, D., Seifert, E., and Jäckle, H. (1988). Differential regulation of the two transcripts from the *Drosophila* gap segmentation gene *hunchback*. *The EMBO Journal*, 7:2881–2887.

Schulz, C. and Tautz, D. (1995). Zygotic *caudal* regulation by *hunchback* and its role in abdominal segment formation of the *drosophila* embryo. *Development*, 121:1023–1028.

Schüpbach, T. and Wieschaus, E. (1986). Maternal effect mutations altering the anterior-posterior pattern of the *Drosophila* embryo. *Roux Archives of Developmental Biology*, 195:302–317.

Shampine, L. F. and Thompson, S. (2001). Solving DDEs in MATLAB. *Applied Numerical Mathematics*, 37:441–458.

Shermoen, A. W. and O'Farrell, P. H. (1991). Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. *Cell*, 97:303–310.

Simcox, A. A. and Sang, J. H. (1983). When does determination occur in *Drosophila* embryos? *Developmental Biology*, 97:212–221.

Simpson-Brose, M., Treisman, J., and Desplan, C. (1994). Synergy between the Hunchback and Bicoid morphogens is required for anterior patterning in *drosophila*. *Cell*, 78:855–865.

Struhl, G. (1989). Differing strategies for organizing anterior and posterior body pattern in *Drosophila* embryos. *Nature*, 338:741–744.

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

Struhl, G., Struhl, K., and Macdonald, P. M. (1989). The gradient morphogen Bicoid is a concentration-dependent transcriptional activator. *Cell*, 57:1259–1273.

Surkova, S., Kosman, D., Kozlov, K., Manu, Myasnikova, E., Samsonova, A., Spirov, A., Vanario-Alonso, C. E., Samsonova, M., and Reinitz, J. (2007). Characterization of the *drosophila* segment determination morphome. *Developmental Biology*. Submitted.

Tautz, D. (1988). Regulation of the *Drosophila* segmentation gene *hunchback* by two maternal morphogenetic centres. *Nature*, 332:281–284.

Tautz, D., Lehmann, R., Schnürch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K., and Jäckle, H. (1987). Finger protein of novel structure encoded by *hunchback*, a second member of the gap class of *drosophila* segmentation genes. *Nature*, 327:383–389.

Tautz, D. and Pfeifle, C. (1989). A non-radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. *Chromosoma*, 98:81–85.

Thom, R. (1969). Topological model in biology. *Topology*, 8:313–335.

Thom, R. (1975). *Structural stability and morphogenesis.* W. A. Benjamin, Reading, Massachusetts, USA.

Thom, R. (1983). *Mathematical Models of Morphogenesis.* Ellis Horwood Limited, West Sussex, England.

Umulis, D. M., Serpe, M., O'Connor, M. B., and Othmer, H. G. (2006). Robust, bistable patterning of the dorsal surface of the *Drosophila* embryo. *Proceedings of the National Academy of Sciences USA*, 103(31):11613–11618.

von Dassow, G., Meir, E., Munro, E. M., and Odell, G. M. (2000). The segment polarity network is a robust development module. *Nature*, 406:188–192.

Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature*, 150:563–565.

Waddington, C. H. (1953). Genetic assimilation of an acquired character. *Evolution*, 7(2):118–126.

Waddington, C. H. (1957). *The Strategy of Genes.* George Allen & Unwin, London.

Waddington, C. H. (1959a). Canalization of development and genetic assimilation of acquired characters. *Nature*, 183:1654–1655.

Waddington, C. H. (1959b). Evolutionary systems - animal and human. *Nature*, 183:1634–1638.

Waddington, C. H. (1966). *Principles of Development and Differentiation.* Macmillan, New York.

Waddington, C. H. (1968). The basic ideas of biology. In *Towards a Theoretical Biology*, volume 1, pages 1–32. Aldine Publishing Company, Chicago.

Weigel, D., Jürgens, G., Klingler, M., and Jäckle, H. (1990). Two gap genes mediate maternal terminal pattern information in *Drosophila.* *Science*, 248:495–498.

Wieschaus, E., Nüsslein-Volhard, C., and Jürgens, G. (1984a). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster.* III. Zygotic loci on the X-chromosome and fourth chromosome. *Roux's Archives of Developmental Biology*, 1983:296–307.

Wieschaus, E., Nüsslein-Volhard, C., and Kluding, H. (1984b). *Krüppel*, a gene whose activity is required early in the zygotic genome for normal embryonic segmentation. *Developmental Biology*, 104:172–186.

Wimmer, E. A., Simpson-Brose, M., Cohen, S. M., Desplan, C., and Jäckle, H. (1995). Trans- and cis-acting requirements for blastodermal expression of the head gap gene *buttonhead. Mechanisms of Development*, 53:235–245.

Wolpert, L. (1968). The french flag problem: A contribution to the discussion on pattern development and regulation. In Waddington, C. H., editor, *Towards a Theoretical Biology*, volume 1, pages 125–133.