Stony Brook University

The Dynamical Basis of the Decanalization of Gene Expression in *Krüppel* Mutant Embryos

A Dissertation Presented

by

Lena Panok

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

August 2013

Stony Brook University

The Graduate School

Lena Panok

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 and Statistics

Professor David Green, Committee Chairperson
Department of Applied Mathematics and Statistics

Professor Yuefan Deng, Committee Member
Department of Applied Mathematics and Statistics

Professor Marco Martens, Outside Committee Member
Department of Mathematics

This dissertation is accepted by the Graduate School.

Charles Taber
Interim Dean of the Graduate School

Abstract of the Dissertation

The Dynamical Basis of the Decanalization of Gene Expression in *Krüppel* Mutant Embryos

by

Lena Panok

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

2013

Canalization is the ability of an organism's phenotype to remain stable under a perturbation in their genotype or the environment. It has been noted that there is no canalization or canalization is to a much smaller degree in mutants than in wild type. Waddington's idea was to understand canalization through studying gene interactions. We investigate the mechanisms of canalization of gap gene patterns in *D.melanogaster* through dynamical analysis of the gene circuit.

To carry out this investigation we construct a dynamical model of the evolution of the concentration of protein products of the gap genes in time. Our model, which is fitted to both wild type and Kr^- data, correctly captures the wild type averaged data and the lowered gap gene expression in Kr^-. Through a dynamical analysis on a simplified version of this model, we investigate pattern formation in both wild type and Kr^-. The analysis is concentrated on the gap genes *hb*, Kr, *gt* and *kni* with outside input from maternal genes *bcd* and Cad. Canalization manifests itself in this model by producing a lower variance, in wild type, of posterior *gt* domain, in comparison to Bcd. Kr^- mutants do not canalize Bcd perturbation. We find the geometric structure that ensures the canalization in wild type. This structure is an unstable manifold that patterns the posterior of the wild type embryos while remaining invariant with respect to changes in Bcd. By continuously changing one parameter in our

model (maximum synthesis rate of Kr), we can smoothly turn down the function of Kr in such a way as to model an allelic series running from homozygous wild type levels to heterozygous, and then to hypomorphic and finally to functional null. We discovered that that there is an abrupt onset of the mutant phenotype and loss of canalization triggered by the loss of the canalizing unstable manifold.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my advisor, Prof. John Reinitz, who let me join his lab even though I had no experience working on code or doing any lab work. The summer I have collected *kni* mutants data set was the most exciting time as a graduate student. Prof. Reinitz provided motivation and suggested different ways to approach the problem whenever I was stuck.

I am very grateful to the members of Reinitz lab, who not only provided great conversations during lunch, but also helpful suggestions and words of encouragement. I am especially indebted to Ah Ram Kim, who turned out not only to be a great friend but also a fantastic teacher. He supervised the first few experiments that I did in the lab, and did not let me blow anything up. Zhihao Lou helped me to understand the fly code and was always happy to help whenever I had problems with simulated annealing. I want to thank Manu for letting me use his algorithm for creating phase space figures, and for patiently answering my many questions in great detail be it over the phone, email or in person.

Our collaborators from St. Petersburg have provided important information that was used in this work. I want to thank Svetlana Surkova for allowing me to use the data she collected on Kr mutants. She also provided an incredibly entertaining company while we were both spending hours at the confocal, when she visited Stony Brook. Vitaly Gursky for his discussions in the course of writing his paper. Konstantin Kozlov for sampling different T^{abc} terms, finding the optimal one and for letting me use the information found.

I would like to thank my family for their support and for not asking me too many times when I will get a real job.

I am deeply grateful to my husband Tom Poole, for his love and support throughout the years. He moved with me to Chicago so that I could continue with my work. I also want to thank him for proofreading this dissertation and for always being there for me in good times and in bad.

Publications

1. S. Surkova, E. Golubkova, Manu, L. Panok, L. Mamon, J.Reinitz and M. Samsonova: **Quantitative Dynamics and Increased Variability of Segmentation Gene Expression in the Drosophila Krüppel and knirps mutants**. Developmental Biololgy, 376:99-112, 2013.

2. V. V. Gursky, L. Panok, E. M. Myasnikova, Manu, M. G. Samsonova, J. Reinitz and A. M. Samsonov: **Mechanisms of gap gene expression canalization in the Drosophila blastoderm**. BMC Systems Biology, 5:118, 2011.

3. E. Myasnikova, S. Surkova, L. Panok, M. Samsonova and J. Reinitz: **Estimation of errors introduced by confocal imaging into the data on segmentation gene expression in Drosophila**. Bioinformatics 25(3): 346-352, 2009.

Chapter 1

Introduction

Determination of patterns is an essential process for the development of an organism. Morphological properties, such as pattern formation, are related to biochemical determinism inside the cell. In the early embryo a cell can become one of several different types of structures, but as development progresses the cell's fate is determined. Driesch separated the two cells from a first cell division in sea urchins and noted that the entire larva was produced from each [9]. Thus each cell at such an early stage had the capabilities to form into a complete organism. On the other hand this is not true in the determined cell. Determination in a cell is manifested by a commitment to a particular fate even if it is experimentally placed in an abnormal environment. In 1918 Spemann demonstrated that if epidermal cells from the late gastrula are transplanted into an area where neural tissues form, those cells still developed into skin, hence their fate was determined and could not be swayed by changes in the surroundings [14].

Robustness of pattern in the face of genotypic and environmental variation is an important property. Organisms remain relatively stable in the face of small environmental and genotypic perturbations. While there are a vast number of different mutations most of them do not produce a mutant phenotype. Furthermore there are countless ways in which the environment changes, yet there is still a limited number of phenotypes. The buffering of

phenotype against small internal and external disturbances is known as canalization.

In order to understand pattern formation and canalization, we can either develop all the physics and chemistry necessary to have a complete theory of the phenomenon on the molecular level, which would be equivalent to pulling the system apart and reconstructing all of its constituents; this would take a very long time, and might not even be possible to achieve. An alternative is to follow in the footsteps of Alan Turing in constructing a mathematical model of pattern formation [74]. Such a model would incorporate key protein interactions and mechanisms, and furthermore locally agree with the properties of the biological system studied. Studying this model *in silico* would allow us to discern the mechanism(s) that ensure the stability of the system with respect to perturbations.

To facilitate understanding of pattern formation as well as the robustness of the pattern we turn to dynamical analysis. Determination in a cell can be related to the mathematical concept of dynamical stability; a cell differentiating into the same type when it is placed in a different environment is said to posses dynamical stability. Furthermore, organisms demonstrate another dynamical systems concept—structural stability. A natural variation in a population results in the same fate map, an assignment of the regions in an embryo to the specific tissues that they will become in an adult organism. Canalization, the ability of the organism's phenotype to remain stable under perturbations in their genotype or the environment, is directly related to developmental trajectories of the system being stable. These parallels between biology and mathematics suggest that we can try to analyze development with the aid of the dynamical systems.

Drosophila melanogaster is an ideal organism to study in order to understand the process of determination of pattern, it has been widely used for genetic analysis over a number of years. Its popularity is partially due to the relatively small size of the organism, numerous visible genetic markers, ability to carry out saturation mutagenesis in order to classify genes as well as availability of a wide range of tools for genetic manipulation. *Drosophila* has a segmented body, where the segments along the anterior-posterior axis are determined by the

segmentation genes—maternal genes, gap genes, pair rule genes and segment-polarity genes. Dynamical and structural stability are present and it is possible to form a dynamical model of the protein concentrations that would mirror the gene expression data using gene circuits [55, 56]. Furthermore it has been illustrated that the wild type *D. melanogaster* canalizes while various mutants do not. Up to now the efforts to find a model that correctly patterns both wild type and mutants as well as showing greater degree of canalization for the wild type have been unsuccessful.

This dissertation will use the ideas of Rene Thom on structurally stable dynamical systems and high temporal resolution data [69, 67] and build upon the dynamical systems models [56, 57, 24, 42] in order to accomplish two goals. The first goal is to find a circuit that correctly patterns both wild type and Kr^- embryos as well as correctly predicts canalization in wild type and decanalization in mutants. The second goal is to use ideas from bifurcation analysis to uncover the mechanism of canalization in wild type and explain why this mechanism breaks when the Kr^- system is considered.

1.1 Canalization

Canalization is the ability of an organism's phenotype to remain stable under a perturbation in their genotype or environment. An organism has an innate ability to remain on a particular developmental pathway even if it is introduced to abnormalities in external factors such as temperature or internal factors such as mutations. Waddington noted that canalization or buffering is especially clear by noticing how much alike individuals from the wild type population are—this led him to proclaim wild type "amazingly constant" [75]. While there are numerous variations in the genetic makeup and the environment that fostered an individual organism, the organismal population produces distinct and well defined body types regardless. This constancy of wild type comes about through buffering against small changes, and it is beneficial for the organisms that are developing in a world where the conditions are

variable.

Waddington's idea was to understand canalization through studying gene interactions [75]. Reasoning that genotype determines the different developmental pathways the cell in an organism can take, Waddington used a metaphor of a landscape to illustrate this point (see figure 1.1.1). The epigenetic landscape's features such as valleys, ridges, etc. are determined by the organism's genotype [76]. The range of all possible phenotypes is illustrated by the full scope of the landscape, with time being represented by a flow through it. Hence the flow through such a landscape represents the trajectory of development of a cell. Epigenetic landscape gives us a way to describe multi-dimensional dynamics of biological system in an intuitive way. It illustrates without the use of complex functions and calculations all the possible states of the system, how they can be achieved as well as how sensitive they are to perturbations. The height of the features in the landscape is directly related to how sensitive the phenotype would be to external and internal variability. Furthermore, the epigenetic landscape illustrates how canalization works. If the cell experiences small genotypic or environmental perturbations the landscape will change a bit, but the ridges and valleys will remain, and the cell, whose position also undergoes a small shift will settle back into the trajectory that leads it through the valley. To come out of the valley would require major environmental and/or genotypic changes. Thus the correct shape of the landscape results in canalization.

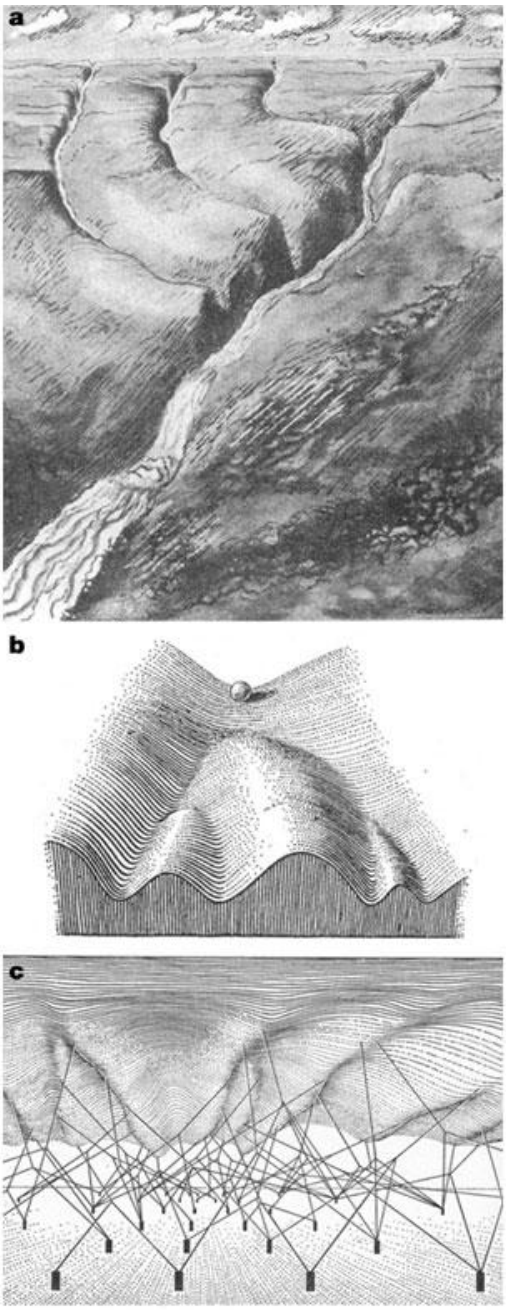
The metaphor bridges a gap between mathematics and biology by making it easier to visualize and describe dynamical concepts, although it does not provide the means to make quantitative statements about our system. Furthermore not all possible biological behaviors are readily translatable into the language of the epigenetic landscape. Oscillatory behavior, such as that of circadian clock do not have an obvious parallel in the landscape unless one visualizes a repetitive landscape that goes on forever. Additionally Waddington's epigenetic landscape has gradient dynamics, where the flow tends to go towards the lowest points in the landscape or the minima of the corresponding gradient function. The problem is that

most biological networks do not have the gradient dynamics.

Waddington observed that the phenomenon of greater canalization of wild type in comparison to mutant was especially evident in *D. melanogaster* [75]. He noted that while there was variability between individual wild type organisms, the mutant population had a much greater variance between its members in the phenotypic expression. For example, the majority of wild type flies have 4 bristles on the scutellum, with rare occurrences of 3 or 5 bristles. On the other hand, Rendel noted that *scute* mutants have a more variable number of bristles. For males there are anywhere between 0 to 3 and 0 to 4 in females [58].

In order to try to explain the above observation invoking the metaphor of a landscape, we can imagine that in going from wild type to mutant organism the landscape changes so significantly that it no longer possess the canalizing properties (i.e. ridge became too small or disappeared altogether, thus allowing the trajectory to flow over the flat surface, whereby perturbations are no longer buffered).

Rene Thom provided a mathematical interpretation of Waddington's ideas [72, 73]. In order to study problems in developmental biology Rene Thom used methods from differential topology and created catastrophe theory, which is part of bifurcation theory. Formalizing the ideas from the epigenetic landscape metaphor, a biological model can be though of as a dynamical system, where the state of the system changes with the passing of time. Parameters of the model define the dynamics of it and hence the shape of the landscape. The full scope of the landscape is the state space and given a set of initial conditions as a starting point, dynamics dictate the trajectory, which is the flow through the landscape. We can describe the dynamics of a system (in our case the state of a cell) as a set of ODEs depending on the concentrations of the system's constituents, external parameters and time. A solution to these ODEs (given the starting condition) will describe the system at any time point. Such a solution corresponds to a flow in the epigenetic landscape. In general, an ODE can not be solved explicitly in a closed form. Using the theory of qualitative dynamics we can find the global geometric behavior of the system, and therefore read off



Reproduced from Jonathan M.W. Slack, Nature Reviews Genetics vol. 3:889-895(2002)

Figure 1.1.1: **(a)** Epigenetic landscape, with water flowing down the valleys. **(b)** Another representation of epigenetic landscape, where instead of flow of water the trajectories are represented by a rolling ball. **(c)** The underside of the landscape. Gene interaction forms all the features of the landscape.

the asymptotic behavior of trajectories without finding an analytic solution. Since we are interested in canalization, we need to consider the case where perturbations are introduced into the parameters of the ODE. In order to possess genotypic canalization the epigenetic landscape of the perturbed system should still have the same shape in terms of ridges and valleys. To minimize variability in the phenotype, perturbed trajectories must converge to a common area. In the language of dynamical systems, we say that perturbed and unperturbed solutions must exhibit the same asymptotic behavior. One way for such behavior to occur is if both solutions belong to the basin of the same attractor. Here we define an attractor to be the set such that if the flow under the dynamics of the system passes in the close vicinity of this set it will tend to go to it. The basin of the attractor is the set such that the trajectories with initial conditions starting in this set approach the attractor asymptotically. Thus if perturbed and unperturbed initial conditions lie in the basin of the same attractor their developmental trajectories will canalize due to attraction to the same set.

On the other hand situations where the system is not structurally stable are also of importance for the development of the cell. For example a fork in the valley of the epigenetic landscape provides a way for the cell to develop into different types. Here a small perturbation would lead to a completely different outcome. Alternatively, the whole topology (location and existence of ridges for example) of the epigenetic landscape can change with small perturbations in the parameters, thus leading the developmental trajectories of the perturbed and unperturbed systems to completely different states. This is important as the distinct outcomes for different cells would result in pattern formation. The study of how perturbations in the parameters cause topological changes such as the type and number of equilibria and attractor sets constitutes bifurcation theory. Bifurcation theory provides a framework for classifying and describing small continuous changes in the parameters that cause significant qualitative changes in behavior of the system. This framework will be used to describe pattern formation and differing degree of canalization for wild type and Kr^{-} *D. melanogaster*.

1.2 Molecular Patterning of *Drosophila Melanogaster*

This section will serve as a brief introduction to embryogenesis and segmentation of *Drosophila melanogaster* (for a more in depth discussion of *D.melanogaster* development see [35, 3]). Particular attention will be drawn to the first three hours after egg deposition as segment determination occurs in this time interval.

Drosophila melanogaster, like all insects, is segmented. There are a total of 14 segments. Depending on its location each segment has its own distinct identity. Three segments make up the head. Three segments make up the thorax. The rest of the eight segments make up the abdomen.

The eggs are 500 μm long and 150 μm wide, which is fairly large compared to the length of an adult female, which is 2500 μm (males are slightly smaller). After the egg is fertilized the newly formed zygotic nucleus undergoes a total of 13 divisions. The time between two consecutive divisions is known as a cleavage cycle. Cleavage cycle i starts after division $i-1$ and ends before ith division. Initially, the nucleus in the egg undergoes 9 rapid mitotic divisions, each lasting about 10 minutes, within the central yolky regions of the embryo. During the tenth nuclear division most of the nuclei migrate to the periphery of the embryo. At this stage the embryo is said to be a syncytial blastoderm because the nuclei are not separated by cell membranes and in fact the only cell membrane is the one surrounding the egg itself. As the nuclei are contained within a common cytoplasm proteins and other substances can readily diffuse between the nuclei. After the 13th division, in the middle of cycle 14 (C14A), cell membranes form around the nuclei. This is called the cellular blastoderm, and at the end we have a single layer of somatic nuclei, that are separated from each other and are located at the periphery of the embryo. After this stage is completed, the embryo undergoes gastrulation, to form the germ layers ectoderm, endoderm and mesoderm. These will eventually form the tissues and organs of *Drosophila melanogaster*. Cycle 14A is the longest one, lasting about 50 minutes. As this cycle is extensive and is also the focus of our study, it can be further subdivided into eight time classes (T1-T8), each lasting 6.5

minutes [68].

The segments that constitute the body plan of an adult fruit are determined by the segmentation genes [50, 51, 49, 78, 62, 26]. These segmentation genes can be classified into maternal factor genes and three types of zygotic segmentation genes, depending on the mutant phenotype produced by saturation mutagenesis. Maternal factor genes come from the mother and establish the anterior-posterior axis in the embryo. The three types of zygotic segmentation genes are: gap, pair-rule and segment polarity genes. Mutation in the gap genes leads to gaps of several segments in an embryo, pair-rule genes affect every other segment whereas the segment polarity genes affect each segment.

The segmentation genes can be separated into a hierarchical structure (see figure 1.2.1), where the upstream genes regulate the downstream genes [1, 19]. The domain of gene expression becomes more refined as we go from maternal to gap to pair-rule and finally to segment polarity genes. Maternal factor genes *bicoid (bcd), caudal (cad) and nanos (nos)* form concentration gradients across the embryo and regulate the next ones downstream, which are the gap genes. Gap genes such as *hunchback (hb), Kruppel (Kr), giant (gt)* and *knirps (kni)* have broad expression domains. The gap genes along with maternal factor genes regulate pair-rule genes, expression of which forms the first repetitive pattern. Pair-rule genes such as *paired (prd), fushi tarazu (ftz), even-skipped (eve), odd-skipped (odd)*, which are expressed in patterns of 7 stripes. The next downstream sets of genes are the segment polarity genes such as *wingless (wg)* and *engrailed* (*en*) are expressed in 14 domains; these eventually lead to the formation of the segments in the embryo.

Prior to fertilization maternal factors are deposited in the egg, and as the nuclei are not separated by membranes, these freely diffuse throughout the embryo. The anterior-posterior body plan of the embryo is set up by three groups of maternal genes, that form spatial protein gradients across the embryo. These groups effect the formation of anterior (head and thorax), posterior (abdomen) and terminal regions.

The formation of the anterior structures in *D. melanogaster* is primarily influenced by

bicoid (*bcd*) [10]. Progeny of Bcd mutant mothers lack the head and thorax [13]. Maternal transcription factor Bcd forms a protein gradient that is a negative exponential, starting high in the anterior pole, where its mRNA was originally deposited by the mother, and dissipating in the posterior [11]. Besides forming its own gradient Bcd forces the formation of Cad gradient. Ubiquitous *cad* mRNA translation is blocked by Bcd. Since Bcd is high in the anterior *cad* is repressed in the anterior and as the *bcd* expression decreases exponentially towards the posterior *cad*'s expression becomes stronger. Maternal *hunchback* (*hb*) mRNA is distributed throughout the embryo [71]. Nanos (Nos), which is located in the posterior inhibits the translation of *hb* whereas *bcd* activates it [70, 61, 12, 66]. As Nos forms a gradient with high levels in the posterior, Hb forms a complimentary gradient with low level in the posterior and high levels in anterior of the embryo [37]. Finally, in the terminal regions of the embryo maternal input acts through regulation for the gap genes *tailless* (*tll*) and *huckebein* (*hkb*) [77].

Thus the early patterns are formed by the morphogen gradient of the maternal factors, where each part of the embryo gets a different concentration of the protein, and in turn responds accordingly to form initial broad gap gene expression. The location where gap genes are expressed becomes more precise once all of the gap gene form their own short-range concentration gradients. The way zygotic genes interact is by making transcription factors that either activate or repress the other genes (themselves included) present in the embryo.

Zygotic expression of the above mentioned genes starts in cleavage cycle 10-12 [23]. Their gene expression starts out low, increases with time and finally peaks in the middle of cycle 14A [68]. The maternal protein gradients of Bcd, Hb and Cad provide the regulatory inputs for gap and pair-ruled genes. While the maternal factors provide the initial input to forming gap gene domains, gap-gap cross regulation is needed to create the sharp borders [23].

This dissertation focuses on regulation of four gap genes: *hb*, Kr, *gt* and *kni*, and more specifically on understanding what happens to this system when Kr is not present.

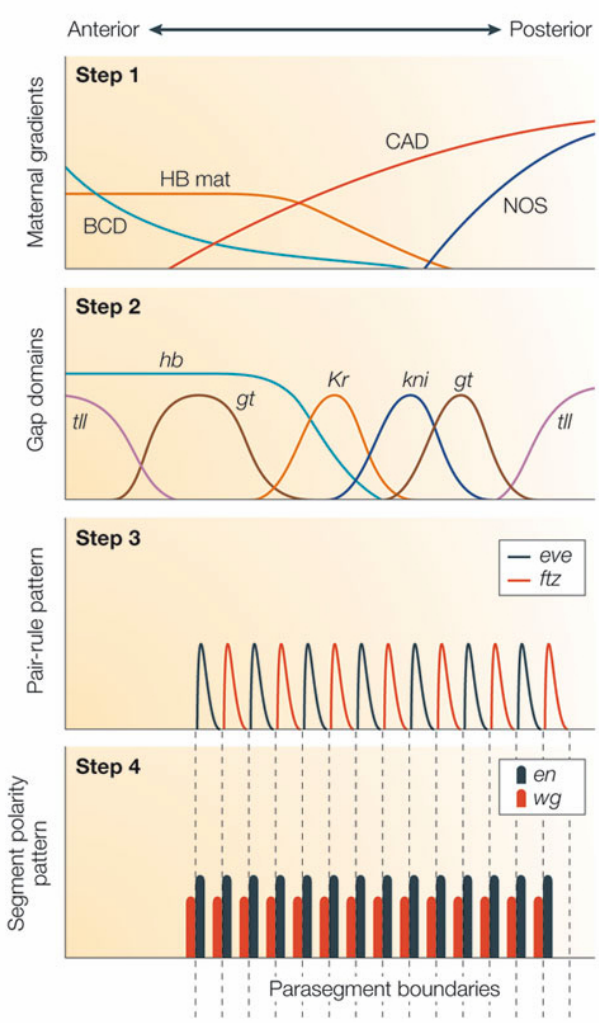


Figure 1.2.1: Reproduced from Andrew D. Peel et. al., Nature Reviews Genetics vol 6: 905-916 (2005). Hierarchy of segmentation genes in *D. melanogaster*.

1.3 Understanding Pattern Formation and Canalization of Gap Gene Expression Pattern Through Dynamical Analysis

Using high resolution data [69] along with a quantitative model that correctly represents the patterns of wild type gap gene expression [24] and with the aid of techniques from dynamical analysis it is possible to study the structure of the phase space of our system. Phase space is a collection of all the possible states that the system can take. The model determines the state that the set is in at any time by solving a system of equations representing the dynamics inside each nucleus—protein synthesis, degradation and diffusion. Furthermore, this approach allows us to analyze phenomenons such as patterning, precise border formation and canalization [42, 44, 43, 16].

Previous work on the model considered a simplified system, where diffusion and *tll* were not present. Furthermore, the analysis focused on the middle part of the embryo, excluding the terminal regions. It has been shown that the formation of the gap gene boundaries comes about as a result of both maternal regulation as well as gap gene cross regulation [21, 44]. The dynamical analysis approach allows us to precisely identify the different mechanisms responsible for formation of boundaries [42, 16]. These mechanisms further illustrate the roles of maternal factors in boundary formation. The following two mechanisms are important in the anterior of the embryo. The first one is the point attractor's dependence on Bcd concentration. The second one is the selection of a particular point attractor based on the levels of maternal Hb (which serves as the initial condition). The last mechanism is important in the posterior of the embryo—it is the dependence of the state of the system by gastrulation on the 1-dimensional attracting set on the maternal Hb.

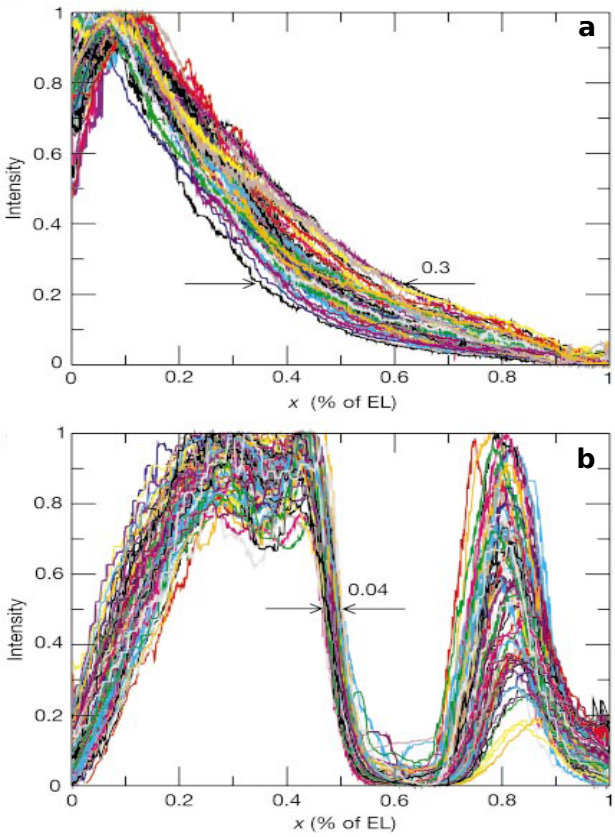
Pattern formation is controlled by different mechanisms in the anterior and posterior of the embryo [42]. In the anterior pattern formation is mostly influenced by Bcd, whereas in the posterior maternal Hb controls it. Furthermore, Manu has shown that the two regions

are separated by a bifurcation. This bifurcation has biological significance. As mentioned before *bcd* mutants are missing the segments in the anterior, right up to the location of this bifurcation (see figure 4.2.1). Hence the dynamics of the model captures the fundamentally important switch between two different modes of patterning. Manu further showed that the anterior and posterior parts of the embryo exhibited different mechanisms for pattern formation. In the anterior attraction to point attractors created the final patterns, whereas in the posterior it was the one-dimensional manifold that attracted the trajectories.

This analysis allows us to explore canalization of gap gene patterns under Bcd perturbation. Canalization at the gap gene level has been experimentally demonstrated [68, 18]. Variability of the maternal input—Bcd was shown to be much greater than the variability in the borders of the gap gene domains [68]. An earlier experiment by Houchmandzadeh [18] showed that embryo-to-embryo variability of Bcd gradient was high while the corresponding posterior boundary of anterior *hb* displayed a reproducible profile between the 100 embryos sampled (see figure 1.3.1). Furthermore Hb levels showed canalization when environmental perturbations in terms of temperature were introduced. The temperature was varied between 9-29° C. The Bcd profile had high dependence on temperature while Hb border displayed the same variability as before (see figure 1.3.2).

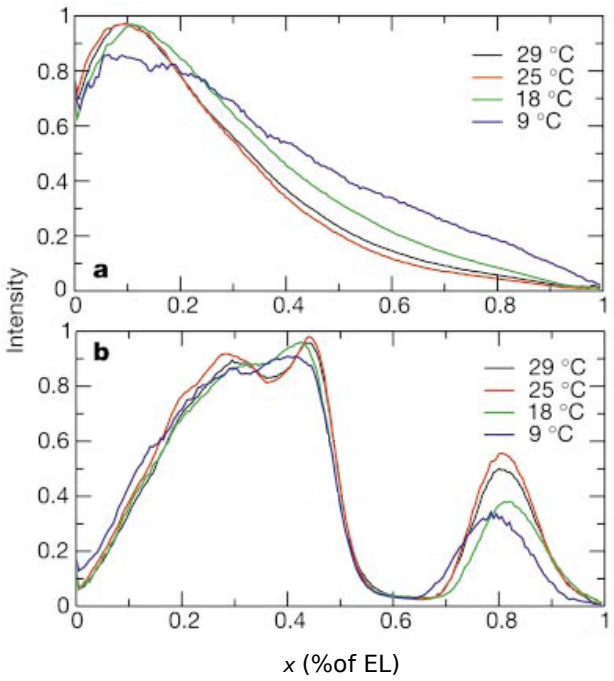
The circuits in [43, 44] correctly predict the reduction in variability from maternal Bcd to gap gene expression. This reduction was shown to be dependent on gap gene cross regulation. It was found that each of the two regions possesses structural stability, but even more importantly there is a stability in trajectories as well; the latter allows for canalization to be quantitative.

While these gene circuits correctly predict the canalization of wild type expression, the efforts to model null mutants (Kr^- mutants in particular) have not been successful. Previous to the work described in this dissertation it has been possible to have two separate models—one for wild type and another one for Kr^-. Those two models would correctly predict the patterns of their corresponding genotype, but they would have different parameters. What



Reproduced from Bahram Houchmandzadeh, Nature vol 415:798-802 (2002)

Figure 1.3.1: Variability over 100 embryos in **(a)** Bcd and **(b)** Hb. The arrows demonstrate the spread in position where the protein concentration crosses a particular threshold. For **(a)** Bcd crosses 0.23 of maximal intensity and correspondingly **(b)** Hb crosses 0.5 of maximal intensity.



Reproduced from Bahram Houchmandzadeh, Nature vol 415:798-802(2002)

Figure 1.3.2: Variability in **(a)** Bcd and **(b)** Hb gradient due to temperature. **(a)** Average profile of Bcd gradient for different temperatures. **(b)** Average profile of Hb gradient for different temperatures.

is needed is one model with unique parameters such that it fits both wild type and Kr^- data. Thus the model should correctly represent wild type and when the parameter that describes the maximum synthesis rate for Kr is set to 0 it should correctly represent Kr^- patterns.

This dissertation will focus on understanding the decanalization of gap gene expression in Kr^- embryos through bifurcation analysis. A review of the protocol to acquire experimental data is presented in Chapter 2. Furthermore chapter 2 contains the segmentation gene expression in Kr^- embryos, documented in [67], along with the kni^- experimental results undertaken in the course of this dissertation, which are also part of [67]. Chapter 3 provides an overview of the theoretical methods from the dynamical analysis that were used as well as the summary of gap gene circuits and optimization methods for finding the parameters for these circuits. Chapter 4 presents the bifurcation analysis of model from [42] with respect to maternal Bcd and Cad inputs. Chapter 5 illustrates the pattern formation for both wild type and Kr^-. Chapter 6 describes the mechanism of canalization in wild type as well as the cause of failure to canalize in Kr^-. Chapter 7 summarizes the results of chapter 5 and 6 as well as discusses other modeling approaches to pattern formation in *D. melanogaster*.

Chapter 2

Experimental Data Acquisition

The aim of this dissertation is to explain the mechanism of canalization in wild type as well as to understand why the mutant gap gene expression becomes decanalized. For this reason data from mutant embryos is needed. Such data was obtained from Kr^- and kni^- as well as $Kr;kni$ double mutant by Surkova [67], the author [67], and Manu [42] respectively. While data from kni^- illustrated important effects the mutation had on the gap gene pattern there were not enough embryos to launch an in depth analysis into the phenomenon of canalization. On the other hand Kr^- data was much denser thus this dissertation will focus on Kr mutants. A summary of data from both Kr^- and kni^- will be presented at the end of this chapter.

In this section the acquisition of data for the gene expression of kni^- will be described, as this genotype was experimentally assayed by the author. In order to accurately compare kni^- to wild type Oregon R., the procedure for collection, fixation and staining of embryos was done in parallel for both to ensure that the same experimental conditions held and the same dilutions were used. The protocols for collection, fixation, staining and image processing were developed in the lab by others and used in this work.

2.1 Collection, fixation and staining of embryos

kni^- embryos came from a balanced stock, Df(3L)ri-XT1, ru[1]st[1]e[1]ca[1]/TM3, Ser[1], in which one chromosome has a deficiency for *kni*. The embryos were collected every 1-4 hours on apple juice agar plates. Both wild type and kni^- embryos were fixed and immunofluorescently stained according to the protocol described in a previous publication [28]. Primary antibodies used were against Kni, Hb and Eve. These were followed by fluorescently labeled secondary antibodies, Alexa Fluor 488 anti-guinea pig, Alexa Fluor 555 anti-rabbit and Alexa Fluor 647 anti-rat. For nuclear staining, mouse anti-histone H3 (primary), biotin-conjugated anti-mouse (secondary) antibodies were used. Streptavidin Alexa Flour 700 conjugate was then used as tertiary.

2.2 Confocal microscopy

The fluorescently stained embryos in lateral orientation were imaged on a Leica TCS SP2 confocal microscope (see figure 2.5.1 (A)-(F)). Embryos from cycle 14A were used. kni^- embryos were identified by having no expression in the Kni channel. In order to obtain comparable confocal images for wild type and kni^-, intensity was standardized on the wild type embryos (as those showed higher levels of expression). The gain in the microscope was set such that the brightest expression had several pixels saturated. The offset was set such that the background that was not part of the embryo was removed. The gain and offset were fixed and both wild type and kni^- were imaged with the same settings. For each staining the gain and offset were recalculated. A Differential Interference Contrast (DIC) image was taken to aid in the temporal classification of the embryo.

2.3 Image segmentation

Confocal images were segmented to obtain tabulated concentration data for each nucleus [25] (see figure 2.5.1 for a confocal image of a wild type embryo (A), (C), (E) and its segmented data of intensities of Eve, Hb and Kni along the A-P axis). First an embryo mask was generated. An embryo mask is a binary representation of the scanned image, where pixels have a value of 255 if they belong inside the embryo and 0 otherwise. The mask was generated using nuclear staining or the maximum intensities for each nucleus from all channels and thresholding this maximal image. The embryo mask was applied to the image and the region outside of it was not considered.

A nuclear mask was then constructed. This mask sets the location of the nuclei by allocating nonzero pixels exclusively to the nuclei. To construct the nuclear mask, the watershed method [15] was used on the maximal image—the pixels with lower intensities then those occupied by the nuclei were denoted as watershed areas, thus forming a boundary around the nuclei. To detect edges the Shen-Castan edge detection method was implemented [64].

The nuclear mask provides us with a way of locating the positions of nuclei along A-P and D-V axes. Thus it allows us to obtain tabulated positions of nuclei as well as average intensities in each channel for each nucleus.

2.4 Background removal

Non-specific binding of the antibodies produced background. For each individual embryo, background needed to be removed. In order to achieve zero background and not to lose any information about the intensities, the two dimensional paraboloid, which was found by Myasnikova [48] to best approximate the background noise was fitted to data and subsequently subtracted. Then the data was normalized such that the intensities that were at background level or below were set to 0, and the maximum intensity, 255 stayed the same.

2.5 Time classification and registration

The next step was to classify the embryos into different time classes. Since the flies were laying eggs for 3 hours, the acquired embryos were of different ages. As the patterns changed with time, we want to average embryos from the same time class. The goal was to find averaged gap gene expression at different times, providing us with the dynamics of gene expression. In order to classify the embryos in cycle 14A DIC images of the membrane as well as nuclear morphology and the patterns of the *eve* gene were used.

The shape of nuclei provided the visual clues to the age of the embryos. For early time classes the nuclei were round and the membrane was not observable. By T3 the nuclei became elongated. Starting at T4 membrane became visible. From T5 to T8 the membrane moved from the apical to the basal side of the nuclei.

The above visual clues were combined with the expression of *eve* patterns. The early time classes were classified according to the number of *eve* stripes that were present and the later time classes, T5-T8, were classified according to the features of the 7 striped eve pattern. At T1 *eve* had a broad domain. T2-T3 the stripes started coming up with more anterior ones first, by T4 all of the stripes were present but their height was low. From T5 until T8 *eve* stripes grew in intensity and became similar in height. At T7 the height of all the stripes was similar. T8 was differentiated from T7 by the saw tooth pattern of the stripes.

The next step was to do image registration [47]. The patterns from all the embryos needed to be aligned prior to the averaging. Stripe peaks and interstripe valleys were used here as the characteristic features of *eve* expression pattern that were aligned. Those features were extracted by fast dyadic wavelet transform. The latter allowed us to find the locations of extrema in the *eve* expression by finding where the first derivative of the data was zero. To align the extrema of *eve* expression an affine coordinate transformation was applied such that the distance from each extrema in an embryo to the average position of that extrema was minimized. The other channels were aligned based on the Eve channel.

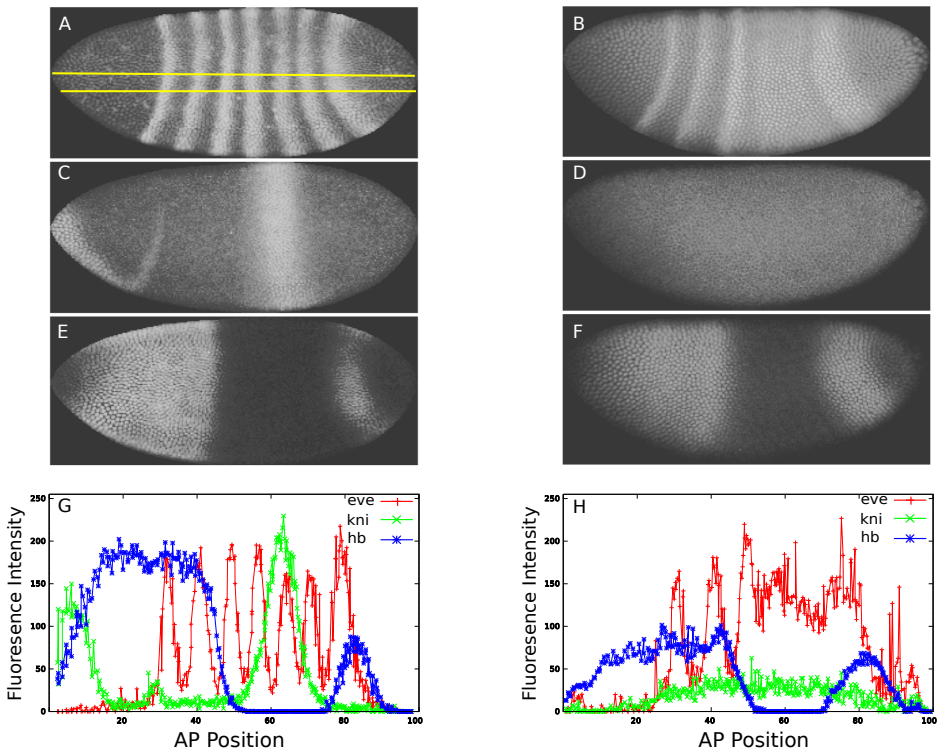


Figure 2.5.1: Confocal image of *eve* expression in **(A)** wild type and **(B)** kni^-. Yellow lines show the middle 10% strip used. Confocal image of *kni* expression in **(C)** wild type and **(D)** kni^-. Confocal image of *hb* in **(E)** wild type and **(F)** kni^-. **(G)** Segmented data from individual wild type embryo. **(H)** Segmented data from individual kni^- embryo.

Before averaging the embryos were subdivided into 100 bins along the anterior-posterior axis, each nucleus being almost exactly 1% EL. The averaging occurred at the bin level for each of the segmentation proteins. Middle 10% of dorso-ventral values were used.

2.6 Segmentation gene expression in Kr^- and kni^-

2.6.1 Kr^-

In this section some relevant results of Surkova et. al. [67] regarding gap gene expression in the posterior part of Kr^- embryos will be briefly discussed. Surkova acquired a quantitative data set for Kr^- through the methods described above. This allowed her to discern the average gene expression in mutants, as well as to compare it to that of the wild type.

Surkova found that the posterior *gt* expression was lower than in wild type, particularly the difference is most striking in early and late cycle 14A. The Kr^- posterior *gt* domain is broader than its wild type counterpart. Even though the domain exhibits the same tendency as wild type to contract with time, by the time of the gastrulation it is still about 5% EL broader than the wild type. There is also a major difference in mutants in terms of the degree of the movement of *gt* posterior domain. Kr^- embryos exhibit more than 13% EL anterior shift in posterior *gt*, even moving into the locus of *kni* domain (see second column of figure 2.6.1).

kni expression in Kr^- has an even more dramatic reduction that than of the *gt* domain. Surkova noted that the maximum amplitude of *kni* domain in Kr^- is about a tenth of its wild type counterpart, which makes *kni* expression be dwarfed even in comparison with the reduced expression of *gt*. In addition to this reduction, the expression starts to decline after time class 5, and by the time of gastrulation it almost reaches the background level. The anterior shift in *kni* is smaller than that of the *gt* domain. When compared to the wild type *kni* domain, the location of the *kni* domain in Kr^- is 3% EL more anterior (see third column of figure 2.6.1). Surkova also noted that in wild type the gap gene expression levels decrease or stop increasing late in cycle 14A, in contrast to Kr^- where these levels reach their maximum earlier and go down to very low levels by late cycle 14A.

When analyzing each individual embryos separately, Surkova was able to find insight into the variability of the patterns of gap genes. She found that the posterior *gt* domain is very variable early in cycle 14A when compared to wild type. Just like in wild type the variability does decrease with time. On the other hand, the variability of *kni* domain in Kr^- was as low as in wild type. According to Surkova the posterior part of Drosophila embryo seems to be less robust to mutations in trunk gap genes, such as Kr. This is especially evident from the high variability of the posterior *gt* domain early on.

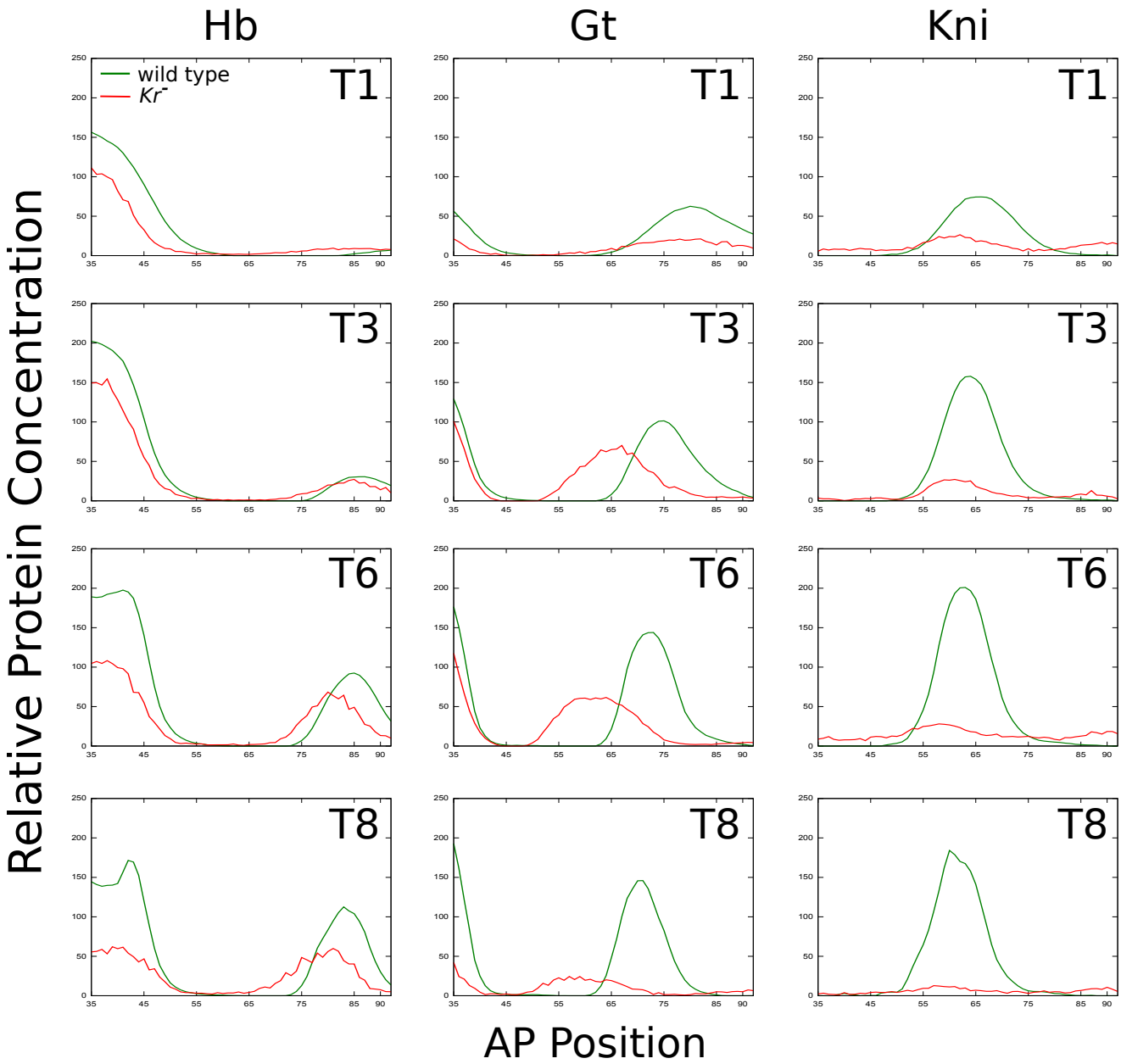


Figure 2.6.1: Comparison of the wild type and Kr^- averaged gap gene expression in the region 35-92% EL at selected time classes. Data is taken from Surkova et. al. [67]

2.6.2 *kni*$^{-}$

Results from *kni*$^{-}$ obtained by the author in the course of this dissertation work are described below. A total of 64 *kni*$^{-}$ embryos were acquired. As these were stained only for gap gene *hb* and pair-rule *eve* (staining for *kni* was only done to determine *kni* mutants), the results will be based on *hb* and *eve* expression. For a more detailed discussion of *kni*$^{-}$ refer to [67].

Overall level of *hb* were lower in *kni*$^{-}$ than in wild type embryos. The level of anterior *hb* region peaked around T5 and then was reduced quickly to the level of the early times. For all the time classes this domain had lower intensity than its wild type counterpart. The posterior *hb* domain formed later than in wild type [68], it became visible only around T3 (see column two of figure 2.6.2).

There were even more drastic differences between *kni*$^{-}$ and wild type *eve* expression. First of all there was a delay in the formation of *eve* stripes in *kni*$^{-}$. Stripes started to come up only around T3, prior to that time *eve* formed a broad domain. Since the *kni* domain was located over stripes 4-6, this is where the most significant difference were noted. Stripes 4, 5 and 6 were not differentiated, they were merged together in a domain of high expression. *kni*$^{-}$ embryos possessed only stripes 1, 2, 3 and 7, whereas wild type had all 7 stripes (see column one of figure 2.6.2).

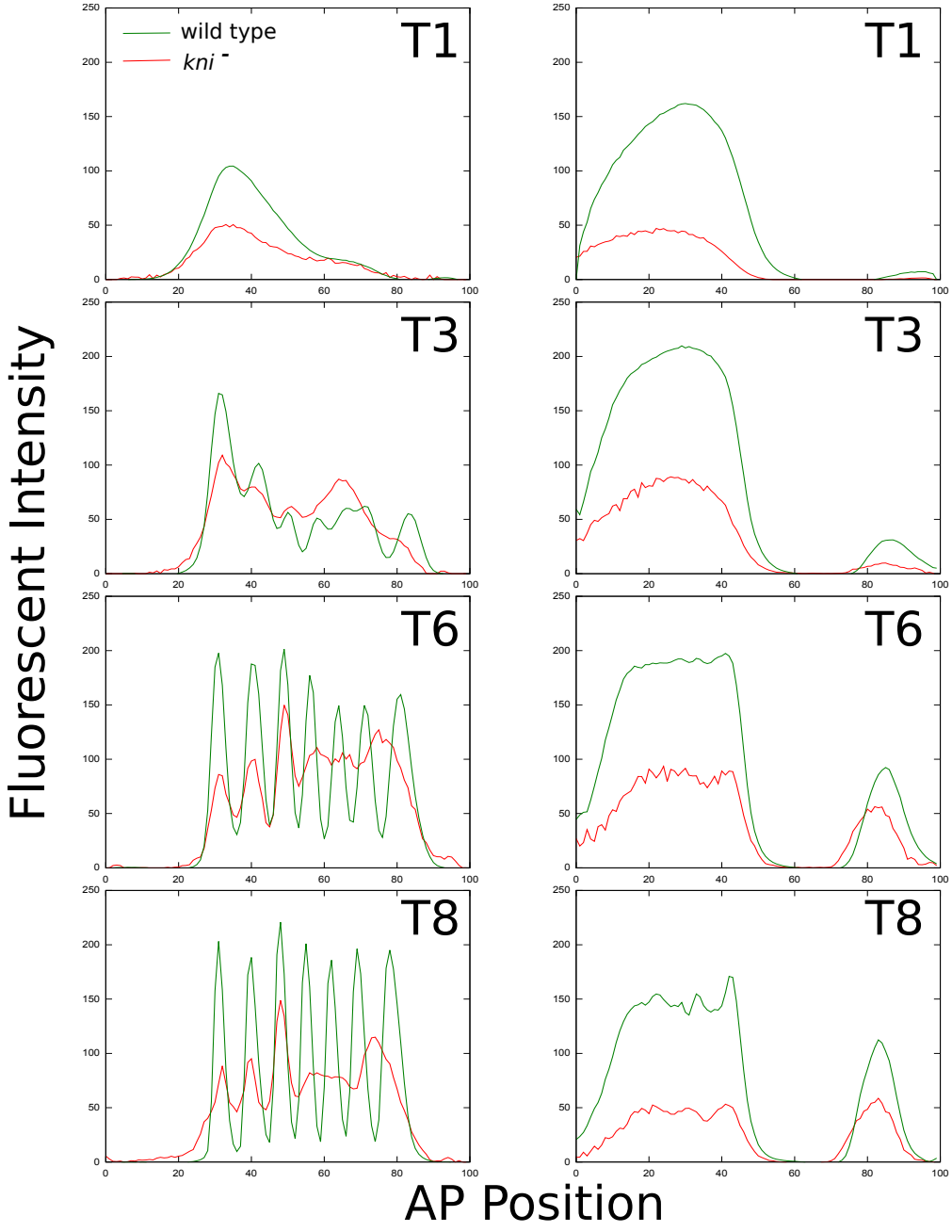


Figure 2.6.2: Comparison of averaged gene expression between wild type and kni^- at indicated time classes. Left column is the dynamics of *eve* expression for wild type (green) and kni^- (red). Right column is the dynamics of *hb* expression.

Chapter 3

The Mathematical Representation of the Gap Gene System

This chapter will review the methods used in the analysis of canalization and pattern formation. In particular gap gene circuits and finding optimal parameters for the model will be discussed. This will be followed by simplifications to the above model to make analysis accessible. Finally relevant material from dynamical analysis that will be used in subsequent chapters is introduced.

3.1 Gap Gene Circuits

We want to model and understand the time evolution of the gap gene protein concentrations in the embryo of *D. melanogaster*. During the blastoderm stage the embryo has a defined anterior-posterior and dorsal-ventral axes, and the gene expression along the two axes are fairly independent of each other in the trunk area of the embryo. We consider the anterior-posterior pattern development and treat the embryo as a 1-dimensional row of nuclei from 35% EL to 92% EL. We will model the evolution from cycle 13 to cycle 14A, consisting of interphase, followed by mitosis, followed by the only discontinuous process—division, and then back to interphase. Interphase is modeled by synthesis, diffusion and decay terms.

Mitosis prepares the cells for division by organizing the genetic information for the progeny (two daughter nuclei for each nucleus), which will be identical to the predecessor nucleus. Mitosis is modeled by turning off protein synthesis and utilizing only the diffusion and decay of proteins. Division follows mitosis, and at the end of division there are twice as many nuclei and the distance between neighboring nuclei is halved.

We will limit the time scope in this analysis to a total of 71.1 minutes, from the beginning of cycle 13 to the end of cycle 14. The time is measured in minutes after the completion of the 12th nuclear division. Cycle 13 ends and C14A begins at 21.1 minutes. C14A has a finer subdivision into time classes. The time classification is based on the membrane invagination and Eve protein patterns as was discussed in the previous chapter. The model [22, 24, 56, 57] for the expression and regulation of the gap genes of *Drosophila melanogaster* comprises of a system of ODEs describing the evolution in time of the concentration of proteins corresponding to the gap genes *hb, Kr, gt* and *kni*, given by

$$\frac{dv_i^a}{dt} = R^a g(u^a) + D^a(n)\Big[(v_{i-1}^a - v_i^a) + (v_{i+1}^a - v_i^a)\Big] - \lambda^a v^a. \tag{3.1.1}$$

In 3.1.1 the first term is the protein synthesis. The second term is the diffusion, and the third terms is protein degradation. Here i stands for the number of the nucleus starting from anterior to posterior. Note that i is also the equivalent of % EL(egg length) in cycle 14. The gap genes are denoted by the indices a and b. v_i^a stands for the concentration of protein a in the i^{th} nucleus, where $a \in \{hb, Kr, gt, kni\}$. The time evolution of v_i^a is the solution of (3.1.1) and is our main object of interest.

R^a is the parameter for the maximum synthesis rate that can be attained. This rate is mediated by a regulation-expression function $g(u^a)$ which essentially determines what percent of R^a is used. $g(u^a)$ is a sigmoidal function that takes values in the range from 0 to 1. Explicitly $g(u^a) = \frac{1}{2}(\frac{u^a}{\sqrt{(u^a)^2+1}} + 1)$. Note that if we are considering a mutant in gene a,

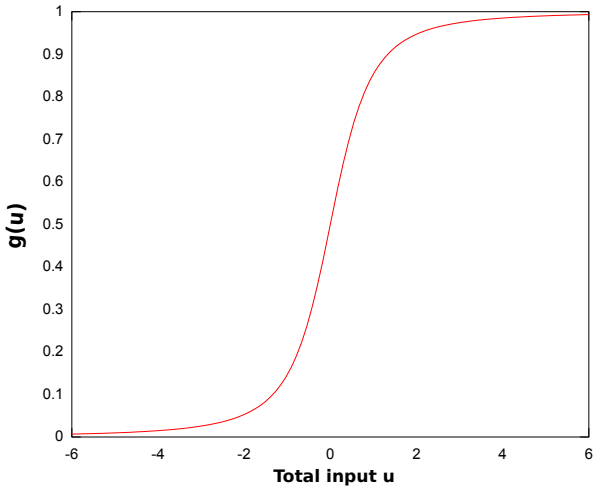


Figure 3.1.1: Sigmoidal regulation expression function. X-axis represents the combined input from all terms. Y-axis shows what fraction of R^a is used.

then the corresponding R^a would be 0.

We have two options for u^a. The first option is

$$u^a = \sum_{b=1}^{N} T^{ab} v_i^b + m^a v_i^{\mathrm{Bcd}}(x) + E^{a\beta} v_i^{\beta}(t) + h^a, \tag{3.1.2}$$

and we denote the corresponding equation (3.1.1) as model 1. The second option is

$$u^a = \sum_{b=1}^{N} T^{ab} v_i^b + \sum_{b,c}^{N} T^{abc} v_i^b v_i^c + m^a v_i^{\mathrm{Bcd}}(x) + E^{a\beta} v_i^{\beta}(t) + h^a, \tag{3.1.3}$$

and the corresponding equation (3.1.1) is model 2.

N is the number of zygotic genes to be modeled, in our case 4. Both models contain the parameters T^{ab}. This is the genetic interconnect matrix, the entries of which are indicative of the interaction of gap gene a and gap gene b, for example, whether it is of an activation or a repression. The second model has additional terms, which are elements of T^{abc}—the second order genetic interconnect matrix. This matrix represents how the combination of maternal factors or gap gene protein products denoted by b and c affect gap gene protein product a.

m^a is the maternal connection strength, which represents the regulatory effect of Bcd on gap gene a. v_i^{Bcd} is the concentration of Bcd in nucleus i, this concentration forms a gradient across the embryo, having a high concentration in the anterior region and tapering off towards the posterior. There is no dependence on time for this maternal input. On the other hand, we have external input that does vary in time as well as in space—the concentration of such input in nucleus i at time t is $v_i^\beta(t)$ and $E^{a\beta}$ is the external input strength of β on gene a. In our study we include two such external inputs - *cad* and *tll*. These factors affect gap gene expression but are themselves not affected by gap genes products [46, 2], hence we do not need extra equations to account for the evolution of their protein concentrations. h^a represents contribution of maternal factors that are homogeneous with respect to both time and space.

Diffusion between nuclei is represented by the D^a parameter. The rate of diffusion is assumed to be proportional to the inverse square distance between the neighboring nuclei. The last term represents protein decay. The decay rate for the protein product of gap gene a is inversely related to its half life $t^a_{1/2}$, $\lambda^a = \frac{\ln 2}{t^a_{1/2}}$.

For each nucleus we need to specify several inputs to the model, namely v^{Bcd}, v^{Cad} and v^{Tll}. The Bcd gradient has the property that it is fairly invariant in time during cycle 13 and most of C14A. Thus we need to specify only one Bcd input for each nucleus for all time. Individual embryos have an exponential Bcd profile and so the Bcd data can be closely approximated by a function of the form Ce^{-ax}. Simple averaging of exponential functions produces a function that is no longer exponential. This problem was overcome in the following way [42]: as two parameters (C, a) define an exponential function Ce^{-ax}, we use the median values of (C, a) from an ensemble of Bcd profiles.

As concentrations of Cad and Tll vary in time and space for each nucleus, we have to provide the values of v_i^β for all possible t. The average data has been collected by Surkova et. al. [68] for each of the eight time classes in cycle 14 as well as for the midpoint in cycle 13. Following [42] we will use linear interpolation between two time classes t_i and t_{i+1} (where

$t_i < t < t_{i+1}$) to get the concentration of external input v_i^β at time t.

Equations 3.1.1 are nonlinear due to the form that $g(u^a)$ takes. A consequence of this is that analytic solutions are not possible (other than in special cases). We can however apply numerical methods to solve our system of ODEs. It was found [42, 22] that the most efficient algorithm for our case would be the Bulirsch-Stoer [54]. This algorithm is used throughout this work to calculate the solutions to system of ODEs.

3.1.1 Optimization of parameters of the model

The parameters of 3.1.1 are optimized such that the model is as close to the data as possible. The cost function that is minimized is

$$E = \sum (v_i^a(t) - v_i^a(t)_{\text{data}})^2 + \text{Penalty} \tag{3.1.4}$$

where $v_i^a(t)$ is the solution of the ODE corresponding to the concentration of the protein product of gene a at time class t and $v_i^a(t)_{\text{data}}$ is the actual data for protein a from the same time class. RMS is defined to be

$$\sqrt{\frac{E}{N_{\text{data}}}} \tag{3.1.5}$$

where N_{data} is the total number of data points.

We utilize two ways of constraining the search space for parameters. Parameters R^a, D^a, and $t_{1/2}^a$ are constrained in an interval that is the same for all four gap genes. Specifically, only the region (0, 15) is sampled for R^a, (0, 0.2) for D^a and (5, 18) for $t_{1/2}^a$. For all the u^a terms—input to $g(u^a)$, there is a collective penalty function that ensures that we do not end up spending all our search time in the saturated part of the thresholding function,

$$\text{Penalty} = \begin{cases} e^{\Lambda U_a} - e^1 & \text{if } \Lambda U_a > 1 \\ 0 & \text{otherwise} \end{cases}$$

where $U_a = \sum_{(a,b)} \left(T^{ab} v_{max}^b)^2 + (E^{a\beta} v_{max}^\beta)^2 + (m^a v_{max}^{\text{Bcd}})^2 + (h^a)^2 \right)$. Here v_{max}^X is the maximum

concentration for protein product of the individual gene. Λ controls how far into the saturated part of $g(u^a)$ our search ventures. We set Λ to 0.001.

To find the parameters that would minimize cost function 3.1.4, we need to sample the large space of parameters of our system. We use PLSA, which is a parallelization scheme of simulated annealing based on Lam's Schedule [33, 34]. Simulated annealing is an optimization method introduced by Kirkpatrick [27] that employs the Metropolis algorithm [45]. The cost function 3.1.4 is used as the energy function for Metropolis algorithm and the state variables are the parameters of our system. The Metropolis algorithm is used to sample a Boltzmann distribution in the following way. We start from any configuration of the state variables X_0 which has energy E_0. A move is proposed in our state space and our new state is X, which has energy E, now we have two choices: either to accept this move or to reject it and go back to the original state. The probability that this move will be accepted is $\min(e^{-(E-E_0)/T}, 1)$, for a given temperature T. The idea behind this algorithm is to start with the system at very high temperature and then to slowly cool it until the lowest energy state is reached, which corresponds to the minimum of the cost function. This algorithm has an advantage of being able to "climb out" from the local minima so that by the time T is so low that the system is "frozen" we are close to global minimum which would be the optimal solution and provide us with best fitting parameters for our model. Lam [33, 34] perfected this algorithm further by creating an efficient cooling schedule for T. The optimal cooling schedule that Lam proposed is

$$s_{n+1} = s_n + \left(\frac{\lambda}{\sigma(s_n)}\right)\left(\frac{1}{s_n^2\sigma^2(s_n)}\right)\left(\frac{4\rho_0(s_n)(1-\rho_0(s_n))^2}{(2-\rho_0(s_n))^2}\right) \tag{3.1.6}$$

where $s_n = \frac{1}{T_n}$, T_n is the temperature after n iterations, $\sigma(s_n)$ is the variance of s_n, and ρ_0 is the acceptance ratio which is the proportion of the moves accepted. λ is a user specified parameter, it has the property that when it is low the quality of the answer produced is better at the price of higher computational time. $s_n^2\sigma^2(s_n)$ is the specific heat, which is a rate of change of $\mu(s_n)$ with respect to the temperature, where $\mu(s_n)$ is the mean energy at

inverse temperature s_n. The last term in the schedule is $\frac{4\rho_0(s_n)(1-\rho_0(s_n))^2}{(2-\rho_0(s_n))^2}$, it is a measure of how effective the move generation is in sampling the state space. It is a maximum at $\rho_0 \approx 0.44$. To keep ρ_0 near its optimum value the size of moves must be managed. Larger sizes of moves lower the acceptance rate and conversely for smaller move sizes. As temperature is lowered move size becomes smaller. μ and σ are estimated adaptively [33, 34, 57].

Chu [5, 4] designed a parallelization scheme for sampling the state by P processors using the Lam cooling schedule. This was done to increase the speed of cooling (which in turn would make calculations faster), as we have synchronized communication between P processors, each can cool P times faster.

For our problem the time for each run was anywhere from 2-10 days. Initial temperature was taken to be either 10^6 or 10^7, and 10 processors were used.

3.1.2 Simplification of the model

In this subsection the model is simplified enough to be used for dynamical analysis and yet still be capable of reproducing the correct pattern. These approximations will be employed for the analysis presented in Chapters 5 and 6

Our model considers 58 nuclei from 35% EL to 92% EL with each nucleus comprising of four ODEs, one for each of the gap genes. The nuclei are coupled through the diffusion term and gap genes are coupled through the T^{ab} and T^{abc} terms. Hence we have 232 coupled equations. This is not an easy computational problem. Fortunately, we can make several approximations that will simplify our task without causing major defects to the model.

Firstly, we turn off diffusion by setting D to 0. It has been shown [21, 44, 43] that diffusion smoothes gap domains but is not responsible for setting the borders (see figure 5.1.2 panels (A)-(D)). Turning off diffusion uncouples the equations, hence we are left with 58 sets of four coupled ODEs, a much more feasible problem.

The nuclear position is defined by the concentration of Bcd, Cad and Tll. To simplify the analysis Tailless (Tll) is removed from it. Tll input is nonautonomous, making calculations

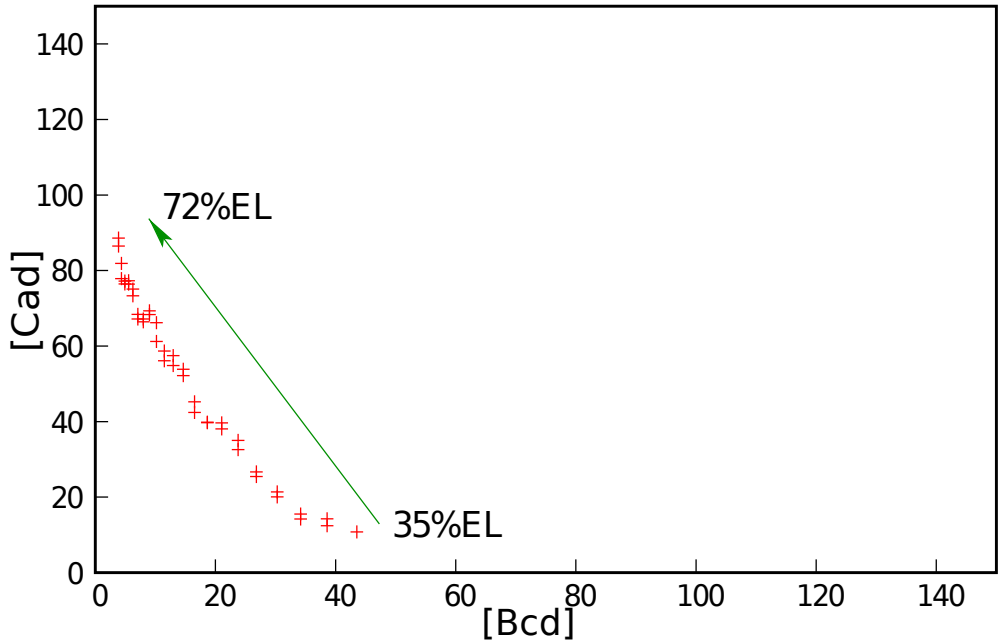


Figure 3.1.2: Combination of Bcd and Cad concentrations define % EL.

harder. Tll is important in the posterior formation of the gap gene expression hence we must confine our model to 35-72% EL region of the embryo. The position in the embryo (% EL) in the simplified model is defined by the Bcd and Cad concentrations (see figure 3.1.2). At each point on the A-P axis we can assign a pair (v^{Bcd}, v^{Cad}) to every nucleus. Then the A-P position x is varied by changing $v^{\mathrm{Bcd}}(x)$ and $v^{\mathrm{Cad}}(x)$ accordingly. The concentrations of Bcd and Cad used here correspond to the averaged data from wild type flies whose profiles were smoothed.

Mitosis occurs in an interval extending from 16 to 21 minutes after the onset of cycle 13. As transcription shuts down during mitosis only protein decay governs the dynamics during this period.

The Bcd gradient remains approximately constant during cycles 13 and 14 with the exception of time classes T7 and T8 (end of cycle 14), during which there is a considerable

decrease in Bcd concentration. Thus to keep our stationary approximation valid we consider

$$\frac{dv^a}{dt} = \begin{cases} R^a g(\sum_{b=1}^{N} T^{ab} v^b + \delta \sum_{b,c}^{N} T^{abc} v_i^b v_i^c + m^a v^{\mathrm{Bcd}}(x) + \\ \qquad E^{a \leftarrow \mathrm{Cad}} v^{\mathrm{Cad}}(x,t) + h^a) - \lambda^a v^a, & \text{if } t < t_6 \\ R^a g(\sum_{b=1}^{N} T^{ab} v^b + \delta \sum_{b,c}^{N} T^{abc} v_i^b v_i^c + m^a v^{\mathrm{Bcd}}(x) + \\ \qquad E^{a \leftarrow \mathrm{Cad}} v^{\mathrm{Cad}}(x,t_6) + h^a) - \lambda^a v^a, & \text{otherwise,} \end{cases} \tag{3.1.7}$$

where t_6 stands for the midpoint of T6. δ=0 for model 1, and δ=1 when model 2 is considered.

The initial conditions in our four dimensional Hb-Kr-Gt-Kni phase space are represented by specifying one state variable—namely Hb. *Kr, gt* and *kni* start to appear in cycle 13, thus the initial condition for all of them is 0, while for *hb* we use the expression from the maternal *hb* from cycle 12.

3.2 Dynamical Systems

This section introduces the use of dynamical systems as a way to understand the qualitative properties of the system without finding solutions to the system of ODEs. Methods for analyzing solutions of linear ODEs as well as identification of bifurcations are introduced.

3.2.1 Equilibrium points

Consider a system of ODEs

$$\dot{x} = f(x, \alpha), \tag{3.2.1}$$

where $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}^m$.

A *Dynamical system* is a manifold M in m-dimensional space, where for a given parameter α we have a map f (also called the *flow* of an ODE) which maps points in M into M. M is a *phase space*, which is all the possible states that the system (3.2.1) can take, where f acts as a rule of how exactly the points will be mapped. The *trajectory* of a point is the set of states that it will take in M under the rule f. This is the curve in the state space that is parametrized by time (t). Another way we can think of *phase space* is a collection of all

trajectories in the state space.

The simplest type of a trajectory is one that stays fixed for all time under f. An equilibrium point is a point which remains stationary under the flow of the ODE. These points have a great deal of influence on the structure of nearby solutions.

In a linear system $\dot{x} = Ax$ we can classify the behavior of solutions based on the eigenvalues of the matrix A. As eigenvalues can be complex and the stability of an equilibrium point depends on the sign of the real part, define the *signature* of the Jacobian to be the signs of the real parts of its eigenvalues. To make this more relevant to our model we will take $x \in \mathbb{R}^4$. If all the eigenvalues of a Jacobian at an equilibrium point have nonzero real part, this equilibrium point is called hyperbolic. The Jacobian of any hyperbolic equilibrium point in $\mathbb{R}^4$ will have four eigenvalues, thus the signatures can be:

1. four positive
2. four negative
3. mixture of positive and negative.

In the case of (1) the solutions are going to be exponentially repelled from the equilibrium point, thus any trajectory starting at an equilibrium point will remain there, but any starting point near x_0 will be carried away from it by the flow of the ODE (for a two-dimensional example see figure 3.2.1(B)). This point is called a *repeller* or a *source*. In the case of (2) the trajectories will be exponentially attracted to x_0, thus with any starting point (other than x_0, which remains invariant under the flow) the flow will carry it closer to x_0 (for a two-dimensional example see figure 3.2.1(A)). This equilibrium point is called a *sink* or an *attractor*. If we are in case (2) and the complex part of the eigenvalue is not zero, the point is called a $focus$ (for a two-dimensional example see figure 3.2.1(D)). While in both instances any small perturbation would bring us back to the equilibrium, for an attractor it is accomplished monotonically while for the focus it is oscillatory. For (3) it is a combination of (1) and (2) (for a two-dimensional example see figure 3.2.1(C)). Locally, the trajectories

will be attracted along the space of eigenvectors that correspond to the eigenvalues where the real part is negative, and at the same time trajectories will be repelled along the space of eigenvectors that correspond to the eigenvalues where the real part is positive. This kind of equilibrium point is called a *saddle point*.

In four dimensions we can have saddles with one eigenvalue with positive real part and three eigenvalues having negative real part are denoted by $S^{1,3}$ (see figure 3.2.1(I)). Saddles with two eigenvalues having positive real part and two eigenvalues having negative real part are denoted by $S^{2,2}$ (see figure 3.2.1(G)). Finally, saddles with three eigenvalues having positive real part and one eigenvalue having negative real part are $S^{3,1}$ (see figure 3.2.1(J)).

Most of the problems in biology, including the one presented in this work are non-linear. We can still use the theory from linear systems to find out local dynamics—the Hartman-Grobman theorem (see below) allows us to treat our non-linear system as being linear around equilibrium points and thus be able to classify the behavior of solutions as was done above.

As our model has two inputs v^{Bcd} and v^{Cad}, we denote the pair as α_0. We find the equilibrium points of (3.2.1) by setting the right hand side of the ODE to zero. Let x_0 be a *hyperbolic* equilibrium point. Expanding $f(x, \alpha_0)$ as a Taylor series around (x_0, α_0), we find

$$\dot{x} = \cancelto{0}{f(x_0, \alpha_0)} + Df|_{x_0,\alpha_0}(x - x_0) + O((x - x_0)^2), \tag{3.2.2}$$

where we have used the fact that $f(x_0, \alpha_0) = 0$. The Hartman-Grobman theorem states that the structure of a dynamical system in a neighborhood of x_0 is topologically equivalent to the linearized dynamical system,

$$\dot{x} = Df|_{x_0,\alpha_0}(x - x_0). \tag{3.2.3}$$

Theorem 1 *[53]* (The Hartman-Grobman Theorem) *Let E be an open subset of $\mathbb{R}^n$ containing the origin, let $f \in \mathbb{C}^1(E)$, and let ϕ_t be the flow of the nonlinear system (3.2.1). Suppose*

that $f(0) = 0$ and that the matrix $A = Df(0)$ has no eigenvalue with zero real part. Then there exists a homeomorphism H of an open set U containing the origin onto an open set V containing the origin such that for each $x_0 \in U$, there is an open interval $I_0 \subset \mathbb{R}$ containing zero such that for all $x_0 \in U$ and $t \in I_0$

$H \circ \phi_t(x_0) = e^{At} H(x_0)$;

i.e. H maps trajectories of (3.2.1) near the origin onto trajectories of (3.2.3) near the origin and preserves the parametrization by time.

A homeomorphism is a continuous function whose inverse is also continuous. Two spaces are said to be topologically equivalent, if there is a homeomorphism between them, that is to say that one space can be continuously deformed into the other. Topologically equivalent objects have the same properties, and thus it is beneficial, instead of working with a complex object of interest, to find the simplest topologically equivalent object to study. Topological equivalence of the two spaces has an important consequence in that the number and types of equilibria are the same. Hence the theorem says that qualitatively the dynamics near equilibria for the non-linear system are the same as for its linearization. Thus the behavior of solutions near the equilibrium point x_0 is determined by the linear term $Df|_{x_0,\alpha_0}$.

Let n_+ denote the number of eigenvalues of $Df|_{x_0,\alpha_0}$ with real part greater than 0 and let n_- denote the number of eigenvalues of $Df|_{x_0,\alpha_0}$ with real part less than 0 [32]. Let T^s be the linear generalized eigenspace of $Df|_{x_0,\alpha_0}$ corresponding to the n_- eigenvalues with real part less than 0. Finally let T^u denote the linear generalized eigenspace of $Df|_{x_0,\alpha_0}$ corresponding to the n_+ eigenvalues with real part greater than 0.

Any equilibrium point has two invariant sets associated with it:

$$W^s(x_0) = (x|\phi_t \to x_0 \text{ as } t \to +\infty) \text{ and } W^u(x_0) = (x|\phi_t \to x_0 \text{ as } t \to -\infty)$$

These sets are called the *stable invariant manifold* and the *unstable invariant manifold* respectively [32]. A connection between the invariant stable and unstable manifolds for an equilibrium point of system (3.2.1) and the stable and unstable spaces of the linearized version (3.2.3) is given by Local Stable Manifold Theorem.

Theorem 2 *[32]* (The Local Stable Manifold Theorem) *Let x_0 be a hyperbolic equilibrium point. Then the intersections of $W^s(x_0)$ and $W^u(x_0)$ with a sufficiently small neighborhood of x_0 contain smooth submanifolds $W^s_{\text{loc}}(x_0)$ and $W^u_{\text{loc}}(x_0)$ of dimension n_- and n_+ respectively. Moreover, $W^s_{\text{loc}}(x_0)$ ($W^u_{\text{loc}}(x_0)$) is tangent at x_0 to T^s (T^u).*

The theorem says that in the neighborhood of x_0 W^s and W^u have T^s and T^u as tangent spaces. Hence T^s and T^u serve as good approximations to W^s and W^u locally. Furthermore, if we are looking in the neighborhood of an equilibrium point and our unstable manifold of interest happens to be of dimension one, we can approximate it by starting the trajectories at two points ϵ away from our equilibrium along the direction of the eigenvector associated with the eigenvalue having positive real part.

If at the hyperbolic equilibrium point we perturb one of our parameters a bit (so that the signature doesn't change), the location of the point might change, but its type will not. Furthermore, the hyperbolic equilibrium point will not disappear or split into more fixed points; the dynamical behavior around it will be preserved. This is not true for a *nonhyperbolic* equilibrium point, which is an equilibrium point where one or more eigenvalues have zero real part. For these points any perturbation will result in a completely different behavior of the trajectories as any change would result in the change in the signature. The points where the signature changes, that is we go through an equilibrium point where the Jacobian has one or more eigenvalues with zero real part is called a *bifurcation point*.

If the eigenvalues that we are dealing with are real, the best way to think about the above ideas is to treat trajectories as a flow of water over the landscape. An attractor is a depression into which the water flows, a source is a mountain peak, and a saddle is a pass. As we change our parameters the landscape changes and the flow of the water automatically adjusts to the change. If no bifurcation occurs as the parameters are varied, then the landscape will look topographically the same, with just the location of mountaintops, passes and depressions shifted. If a bifurcation does occur, then you can see qualitative changes: depressions, saddles or summits may merge together, they can disappear, new ones can appear, etc.

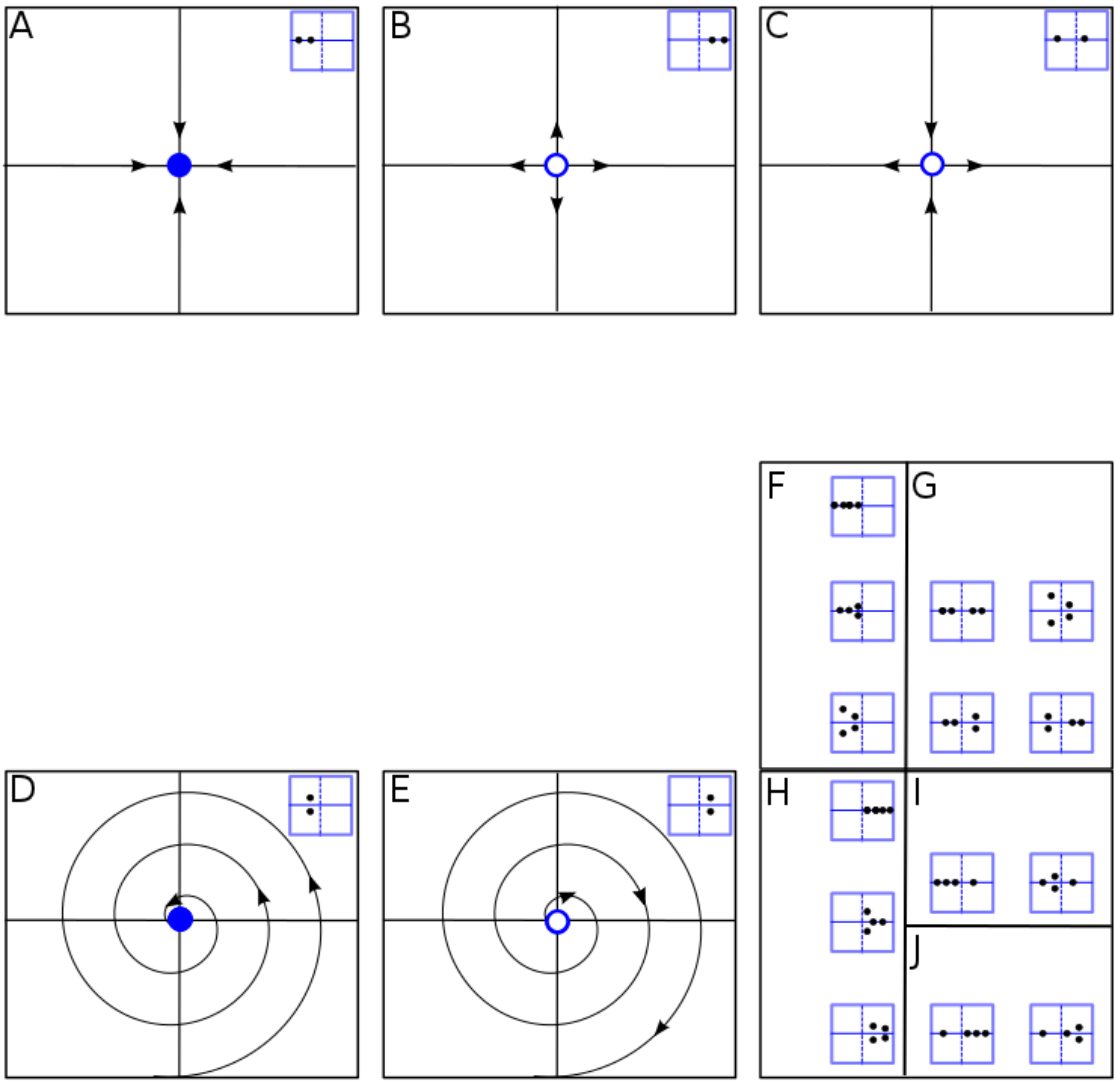


Figure 3.2.1: **(A)-(E) Phase portraits in the neighborhood of an equilibrium point in two dimensions. (F)-(J) All the possible combinations of different equilibria in four dimensions** The blue inset in the top right corner of each panel (A)-(E) is a representation of the eigenvalues corresponding to their associated equilibrium point. The solid line is the real axis and the dotted line is the imaginary axis. A blue circle is a stable equilibrium point, while an empty circle is an unstable equilibrium. **(A)** Attractor. **(B)** Repeller. **(C)** Saddle. **(D)** Stable focus. **(E)** Unstable focus. **(F)** In four dimensions attractors can have three different types of eigenvalues: all real, two with nonzero complex part or all with nonzero complex part (from top to bottom). **(G)** $S^{2,2}$ can have four different types of eigenvalues: all real, all complex, the stable part real and the unstable part complex and vice versa (left to right, top to bottom). **(H)** Repellers can have the same types as attractors in (F). **(I)** $S^{1,3}$ can have only two different types of eigenvalues: all real or a pair of complex (along the stable direction). **(J)**$S^{3,1}$ can have only two different types of eigenvalues: all real or a pair of complex (along the unstable direction).

As we are dealing with a high dimensional dynamical system we cannot expect the solutions to be completely dependent on equilibria and stable or unstable manifolds. Attracting sets could in fact be higher dimensional, for example an n-dimensional torus. The system could posses strange attractors, causing dynamics to be unpredictable due to great sensitivity on the initial conditions. Fortunately, in our case all of the trajectories (with one exception) that corresponded to relevant biological initial conditions (see section 3.1.2, page 34) lay in the basins of fixed point attractors and ended up at the attractor before or after gastrulation. The only exception to the above was one initial condition whose trajectory ended up on the limit cycle (to be discussed in section 5.2.1). Furthermore, our system did not display extreme sensitivity to initial conditions. When Hb concentration, that differed from the correct initial conditions by several units was considered, the resulting trajectory ended up at the same state in Hb-Kr-Gt-Kni space. Thus in this work chaotic behavior was not encountered, and the only attracting sets were the fixed point attractors and a limit cycle.

The study of hyperbolic equilibrium points is more or less straightforward and as long as we don't go through any bifurcations the dynamical system would remain topologically the same. Therefore, we will focus on the analysis of nonhyperbolic equilibrium points and their associated bifurcations. We will first show that our equations at the bifurcation point are reduced from 4 dimensions to 1 or 2 dimensions on a special manifold called the *Center manifold* (W^c). We will then focus our attention only on the W^c and look at the behavior of 1 or 2 dimensional systems that undergo bifurcations.

3.2.2 Reduction to the Center Manifold

The aim is to study bifurcations of equilibria in a 1 or 2-dimensional phase plane and apply these ideas to the analysis of n-dimensional systems. For our system, the bifurcations that will be encountered occur in a 1 or 2 dimensional invariant manifold. The behavior of the system off this invariant manifold is not of great interest as the bifurcation does not affect it.

Therefore we are justified in considering the systems with a 1 or 2 dimensional phase plane and studying the bifurcations in them. The rest of this subsection deals with the justification of this reduction.

Let α_0 be a parameter such that there exists an x_0 for which $f(x_0, \alpha_0) = 0$ and $Df|_{(x_0,\alpha_0)}$ has an least one eigenvalue whose real part is 0. Keeping α_0 fixed we can define $f(x, \alpha_0) = \tilde{f}(x)$. This constitutes a continuous-time system

$$\dot{x} = \tilde{f}(x) \tag{3.2.4}$$

which is not parameter-dependent, where $x \in \mathbb{R}^n$. $\tilde{f}(x_0) = 0$, with f sufficiently smooth. Also let $\lambda_1 ... \lambda_n$ be the eigenvalues of $D\tilde{f}|_{x_0}$. Consider x_0 to be a nonhyperbolic equilibrium point so that there are n_0 eigenvalues lying on the imaginary axis, n_- eigenvalues lying to the left of the imaginary axis and n_+ lying to the right of the imaginary axis. Let T^c be the linear generalized eigenspace of $D\tilde{f}|_{x_0}$ that is associated with n_0 eigenvalues that have the real part equal to 0. Let ϕ^t be the flow associated with (3.2.4).

Theorem 3 *[32]* (Center Manifold Theorem) *There is a locally defined smooth n_0-dimensional invariant manifold $W^c_{\text{loc}}(x_0)$ of (3.2.4) that is tangent to T^c at $x = x_0$. Moreover, there is a neighborhood U of $x = x_0$, such that if $\phi^t x \in U \; \forall t \geq 0 (t \leq 0)$, then $\phi^t x \to W^c_{\text{loc}}(x_0)$ for $t \to \infty (t \to -\infty)$*

Thus W^c_{loc} is invariant under the flow, and it is called a *center manifold* [32]. We should also note that while it can be proven that such a manifold exists it is not unique [32]. Expanding (3.2.4) in a Taylor series around our nonhyperbolic point x_0 we get

$$\dot{x} = \cancelto{0}{\tilde{f}(x_0)} + D\tilde{f}|_{x_0}(x - x_0) + O((x - x_0)^2). \tag{3.2.5}$$

Consider a basis for system (3.2.4) formed by eigenvectors corresponding to the eigenvalues of $D\tilde{f}|_{x_0}$. This is the eigenbasis of (3.2.4). Let v_i be the eigenvectors corresponding to the

λ_i such that the $Re\ \lambda_i = 0$ and u_i to be the eigenvectors corresponding to the λ_i such that the $Re\ \lambda_i \neq 0$. Thus we can rewrite (3.2.4) as

$$\begin{cases} \dot{v} = Bv + h(u, v) \\ \dot{u} = Au + g(u, v) \end{cases} \tag{3.2.6}$$

Here $u \in \mathbb{R}^{n_+ + n_-}$ and $v \in \mathbb{R}^{n_0}$, h and g have Taylor expansions with at least quadratic terms. Comparing (3.2.5) and (3.2.6) we see that A has all the eigenvalues of $D\tilde{f}|_{x_0}$ with nonzero real part and the eigenvalues of B are all the eigenvalues of $D\tilde{f}|_{x_0}$ with zero real part.

Locally we can represent the center manifold as

$$W^c_{x_0} = ((u, v) : u = U(v)), \tag{3.2.7}$$

where $U : \mathbb{R}^{n_0} \to \mathbb{R}^{n_+ + n_-}$.

Theorem 4 *[32]* (Reduction Principle) *System (3.2.6) is locally topologically equivalent near x_0 to the system:*

$$\begin{cases} \dot{v} = Bv + h(U(v), v) \\ \dot{u} = Au \end{cases}$$

While the behavior off the center manifold is determined by the linear terms, the dynamics on center manifold are governed by both linear and nonlinear terms of 3.2.6.

We can be even more explicit and break down u into the $u_- \in \mathbb{R}^{n_-}$ and $u_+ \in \mathbb{R}^{n_+}$,

$$\begin{cases} \dot{v} = Bv + h(U(v), v) \\ \dot{u_-} = -u_- \\ \dot{u_+} = u_+. \end{cases} \tag{3.2.8}$$

Looking at (3.2.8) we see why the center manifold is an important object; the behavior in a neighborhood of the bifurcation point x_0 can be predicted from just looking at how the system behaves on the center manifold, since the equations off the manifold are trivial (they are attracting along u_- and repelling along u_+).

If we want to work with the parameter-dependent system, it can be shown that the

behavior near the value of the parameter α where the bifurcation occurs is completely defined by the n_0 dimensional center manifold that we have already considered [32].

3.2.3 Codimension 1 Bifurcations

The *codimension* (codim) of a bifurcation is the number of parameters needed for that particular bifurcation to occur. In this work only two local bifurcations of codimension 1 are encountered—saddle-node and Hopf bifurcations.

Consider equations (3.2.1). In order for the system to undergo a bifurcation, the Jacobian would have to have at least one eigenvalue where the real part vanishes. If we start with an eigenvalue λ that has a nonzero real part, by varying α we can arrive at λ=0 (this is a saddle-node bifurcation, discussed below) or we can arrive at λ=$\pm$ iω (this is a Hopf bifurcation, discussed below).

Both of these bifurcations can be found by varying only one parameter. As there are two parameters in our system that can be varied, v^{Bcd} and v^{Cad}, the manifold on which they occur in the two-dimensional parameter space is a curve, called the bifurcation curve (β). From now on by Theorem 4 we will only focus on dynamics on the center manifold and deal with the reduced system.

3.2.4 Saddle-node Bifurcation

A saddle-node bifurcation occurs when a stable equilibrium point and an unstable equilibrium point collide, annihilating each other, the end result being no fixed point at all. If we have a saddle-node bifurcation at α_0 then our system near the bifurcation point is topologically equivalent to

$$\dot{x} = x^2 + (\alpha - \alpha_0). \tag{3.2.9}$$

From now on we can use 3.2.9 to discuss the saddle-node bifurcation. For $\alpha < \alpha_0$ we have two equilibria (one stable and one unstable); for $\alpha = \alpha_0$ the two equilibria come together and

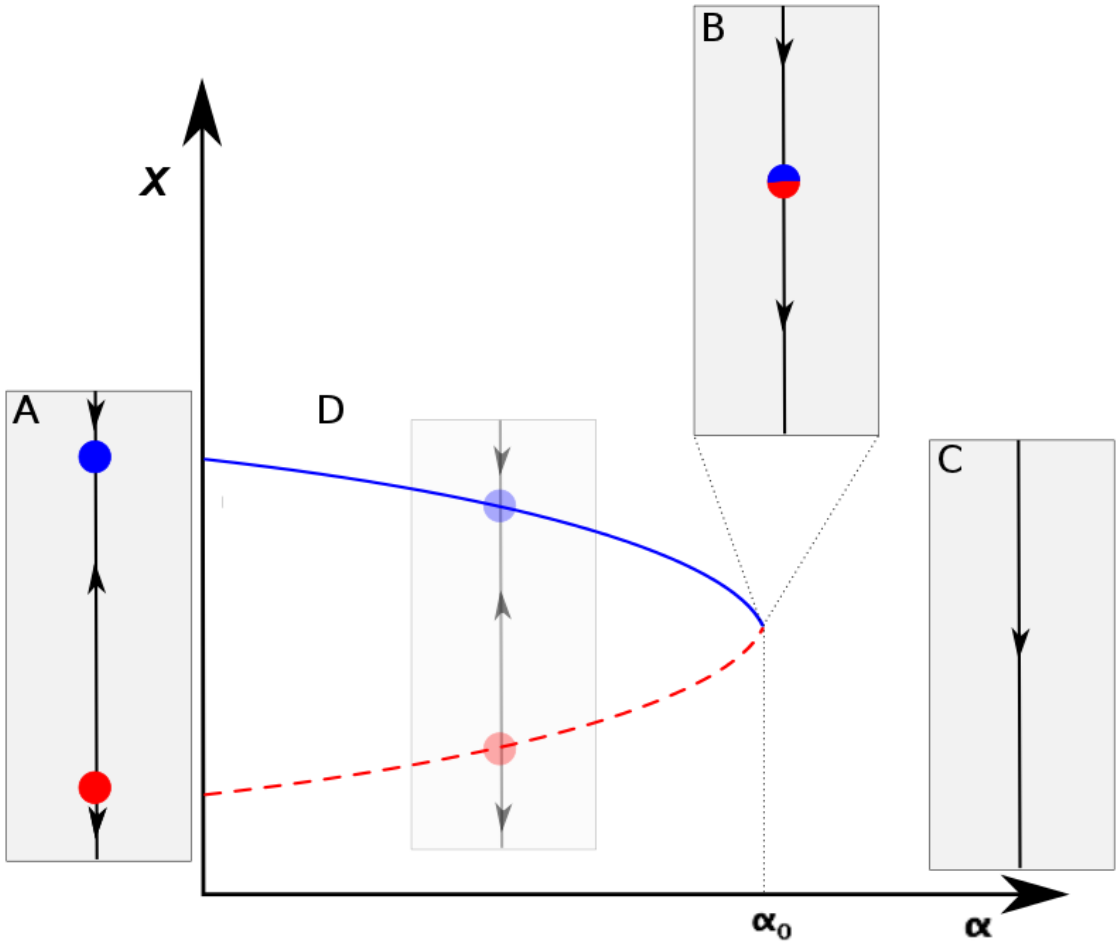


Figure 3.2.2: **(A)-(C)**Trajectories of solutions in phase space for saddle-node bifurcations. Bifurcation occurs at an equilibrium point for $\alpha = \alpha_0$. **(A)** $\alpha < \alpha_0$; **(B)** $\alpha = \alpha_0$ **(C)** $\alpha > \alpha_0$ **(D)**Corresponding bifurcation diagram. Dotted line (red) indicates unstable equilibrium points (red circle), solid line (blue) indicates stable equilibrium points (blue circle).

annihilate each other; finally, for $\alpha > \alpha_0$ there are no equilibria present (see figure 3.2.2).

3.2.5 Hopf Bifurcation

When the Jacobian at an equilibrium point has two complex eigenvalues (complex eigenvalues occur in conjugate pairs) that are purely imaginary, a Hopf bifurcation occurs. In the phase space this phenomenon corresponds to having one equilibrium point switch from attracting trajectories to repelling them and at the time of the switch there is a creation of a limit cycle to which the trajectories would now be attracted to or be repelled from. A limit cycle is a

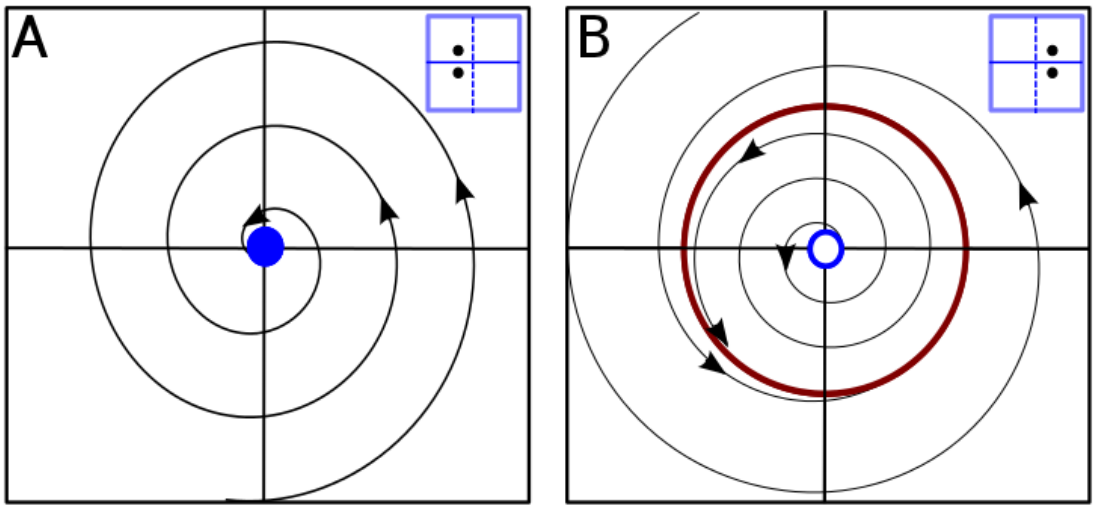


Figure 3.2.3: Trajectories of solutions in phase space for Hopf bifurcation. Bifurcation occurs at an equilibrium point for $\alpha = \alpha_0$. **(A)** $\alpha > \alpha_0$, the equilibrium point is an attractor; **(B)** $\alpha \leqq \alpha_0$, the equilibrium point switches to being a repeller and it also gives rise to a limit cycle, which is an invariant two dimensional manifold.

closed trajectory such that every point on the cycle is rotated around it by the flow of the ODE, and the trajectories in the neighborhood of the limit cycle spiral into it in forward or backward time. As the parameter is varied, the radius of the limit cycle is varied, and as we go back through the Hopf bifurcation point the radius becomes 0, in other words it merges into the original equilibrium point.

A Hopf bifurcation is best visualized in polar coordinates. Consider the system:

$$\begin{cases} \dot{\rho} = \rho((\alpha - \alpha_0) - \rho^2) \\ \dot{\phi} = 1. \end{cases} \tag{3.2.10}$$

where ρ is the distance from the origin and ϕ is the angle with the positive x-axis. For $\alpha < \alpha_0$ we have one equilibrium at ρ=0 and it is stable. For $\alpha > \alpha_0$ the equilibrium point above becomes unstable, and further there is a stable set called a limit cycle $\rho=\sqrt{(\alpha - \alpha_0)}$, having radius $\sqrt{(\alpha - \alpha_0)}$. The radius becomes 0 at $\alpha = \alpha_0$ and the limit cycle disappears (see figure 3.2.3).

3.2.6 Codimension 2 bifurcations

In order to find a bifurcation of codimension 2, both parameters need to be varied simultaneously. In the 2-dimensional parameter space these bifurcations will correspond to points (see figure 3.2.4 (*)).

There are five codimension 2 bifurcation in the continuous-time system. In this section only one will be mentioned. The location of that bifurcation in the parameter plane will be crucial in helping to identify gene circuits that display canalization (see Chapter 4). This would provide yet another connection between properties of the dynamical system and biological property of canalization of gap gene expression of the organism.

A cusp bifurcation occurs when one eigenvalue is zero and dim $W^c = 1$ but the quadratic coefficient of f is also zero. Two branches of saddle-node bifurcation curves come together at this point (see figure 3.2.4 (SN1 and SN2)). In the neighborhood of the cusp three equilibria are affected. These three equilibrium points annihilate each other pairwise along each of the two saddle-node bifurcation curves.

A hyperbolic equilibrium point has the leading term of its Taylor series expansion vanish and it does not have a center manifold. For a nonhyperbolic equilibria we can consider how the system behaves near the bifurcation on the center manifold. As we saw in the previous sections, for a saddle-node bifurcation the first order term vanishes as well as the leading term. When higher order bifurcations such as the cusp are considered even more leading terms vanish, and therefore higher order terms take precedence.

These ideas can be applied to experimental data from real biological systems as I will demonstrate in the next 3 chapters.

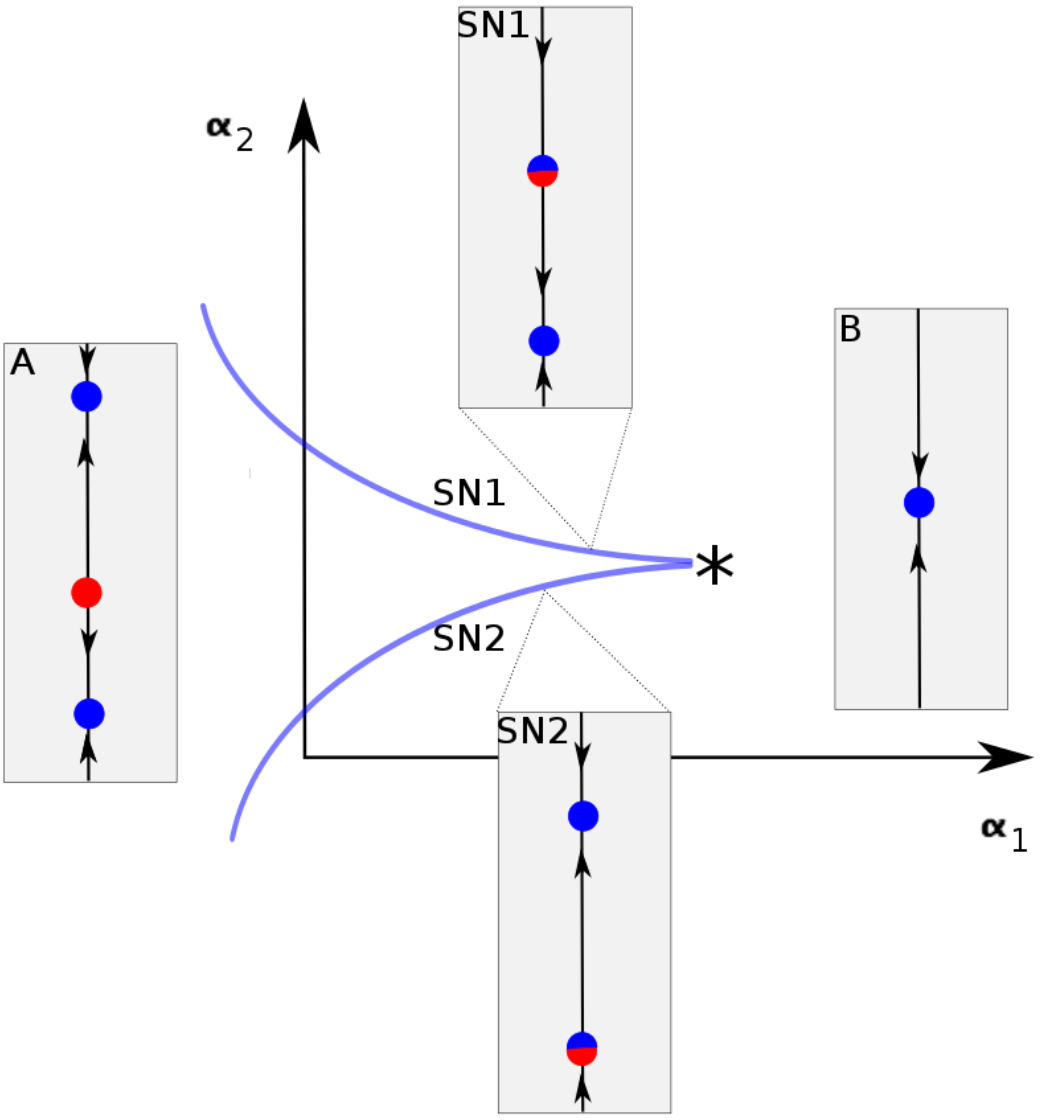


Figure 3.2.4: A cusp bifurcation. (*) denotes the location of a cusp bifurcation. The blue curves SN1 and SN2 denote different saddle-node bifurcation curves.
(A) Two attractors (blue circles) and one saddle (red circle) are present. **(SN1)** Attractor and saddle are annihilated. **(SN2)** Another attractor and saddle are annihilated. **(B)** Only one attractor is left. Going through a cusp bifurcation (*) from **(A)** to **(B)** all three equilibria merge together.

Chapter 4

Bifurcation Analysis of the Original Model

The goal of bifurcation analysis is to have a global understanding of the behavior of a system with respect to its parameters. In this chapter the model's (3.1.7) dependence on v^{Bcd} and v^{Cad} will be explored. The notation for concentration will be used when talking about the parameter space—[Bcd]-[Cad] plane. I will discuss the dependence of important equilibria on the parameters as well as locate and analyze relevant bifurcations.

To study the parameter space we will subdivide the [Bcd]-[Cad] plane into regions where the gap gene system has qualitatively the same behavior. To do this, we will start from any equilibrium point of our system and perform parameter continuation. Parameter continuation is a numerical method that will tell us the location of the equilibria that we are considering in the Hb-Kr-Gt-Kni phase space for different parameter values. The new location of the equilibrium point is calculated while varying one of the two parameters until we find a saddle-node or Hopf bifurcation. The curve in the [Bcd]-[Cad] plane is found such that all the points on that curve exhibit the same bifurcation. This will separate the plane into 2 regions with different dynamical behaviors, and the curve that separates the two is the locus where that particular bifurcation happens (see figure 4.0.1). If we can find all

the possible types of bifurcations and continue each in the [Bcd]-[Cad] plane we will derive the *parametric portrait* for our model. This will be a subdivision of the [Bcd]-[Cad] plane into regions, where in each region no bifurcation occurs—therefore the qualitative behavior of the solutions of the system does not change. As a result we can classify all the possible behaviors for different values of Bcd and Cad $(v^{\text{Bcd}}, v^{\text{Cad}})$.

4.1 Continuation of Equilibria and Detection of Bifurcations using AUTO

Recall that our model (3.1.7) consists of a system of 4 ODEs for each nucleus. Let us rewrite (3.1.7) as

$$\dot{x} = f(x, \alpha), \tag{4.1.1}$$

where $x = (v^{\text{Hb}}, v^{\text{Kr}}, v^{\text{Gt}}, v^{\text{Kni}}) \in \mathbb{R}^4$, and $\alpha = (v^{\text{Bcd}}, v^{\text{Cad}}) \in \mathbb{R}^2$.

For this chapter only model 1 (see equations 3.1.1 and 3.1.2) will be considered.

We start off by choosing a nucleus. For every nucleus there corresponds a pair of values $(v^{\text{Bcd}}, v^{\text{Cad}})$ that describe the concentration of Bcd and Cad in that nucleus. For the starting nucleus such a pair is $\alpha_0 = (v_0^{\text{Bcd}}, v_0^{\text{Cad}})$. All the equilibrium points of the starting nucleus are found by setting the left hand side of our equation to zero, and using the Newton-Raphson iteration method to find all the combinations of $x = (v^{\text{Hb}}, v^{\text{Kr}}, v^{\text{Gt}}, v^{\text{Kni}})$ that satisfy $f(x, \alpha_0) = 0$. Next we choose one of the hyperbolic equilibrium points and continue it as one of the parameters changes (either v^{Bcd} or v^{Cad} but not both of them together). This will give the dependence of an equilibrium on α. Take $x = x_0$ to be such that for α_0, $f(x_0, \alpha_0) = 0$ (i.e. x_0 is our first solution). To see what happens to our equilibrium point as one of the parameters is varied, we need to compute the solution x_1 at $\alpha_1 = \alpha_0 + \Delta\alpha$. In order to do that AUTO [8] is used.

AUTO is a continuation software, originally developed by Eusebius Doedel; it also does

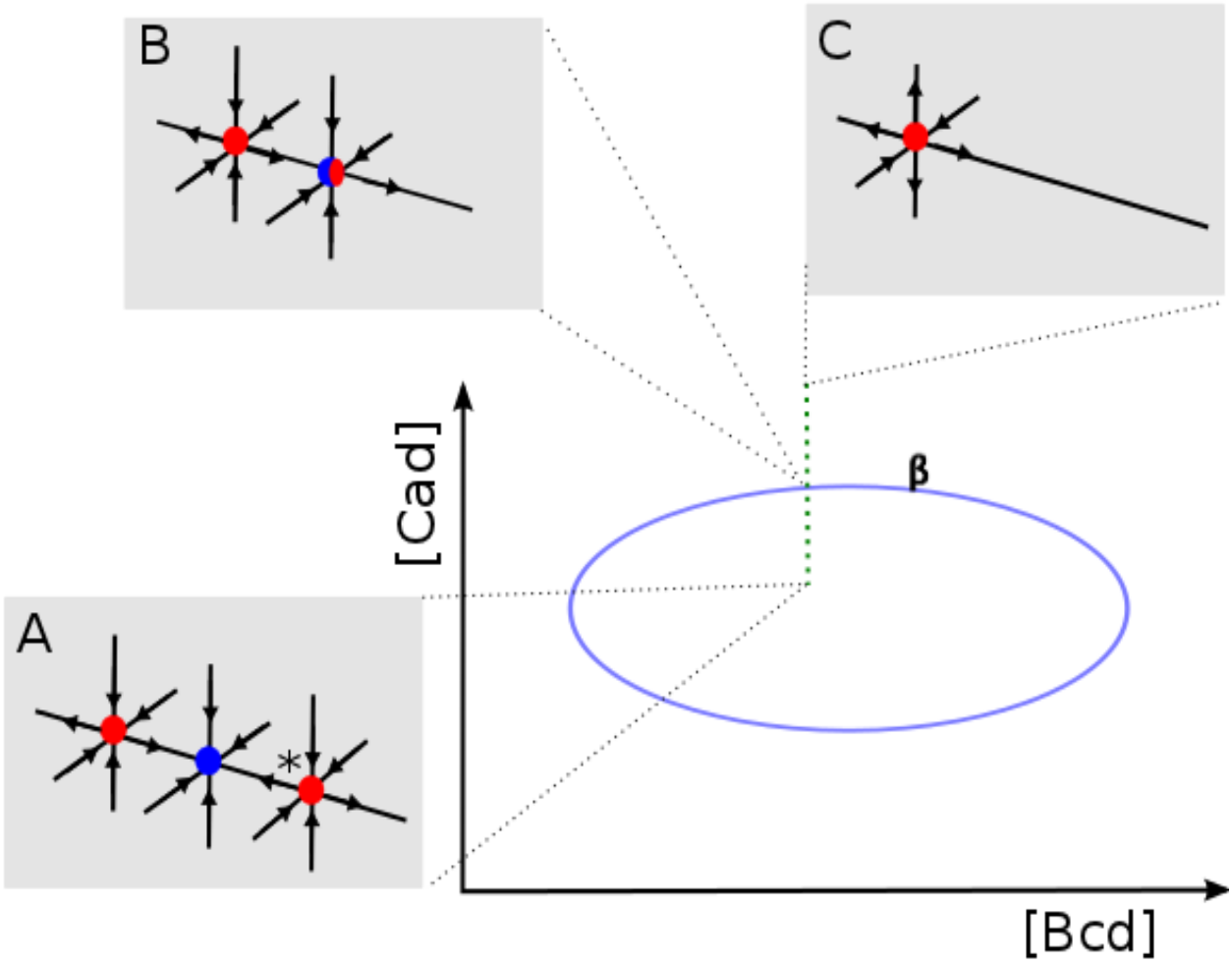


Figure 4.0.1: Finding a bifurcation curve in the [Bcd]-[Cad] plane. Each point in the plane corresponds to a 4-dimensional phase space Hb-Kr-Gt-Kni. The grey inserts represent the projection of the phase space onto 3 dimensions. Red circles are the saddles, blue are the attractors, and black arrows are the direction of the flow in the vicinity of the equilibrium. **(A)** The starting point is the right saddle equilibrium (with an asterisk next to it). The dotted green line is the continuation with respect to [Cad]. **(B)** The right hand equilibrium formed by a merger of the middle equilibrium in (A) with that marked with an asterisk is nonhyperbolic and is undergoing a saddle-node bifurcation. The bifurcation curve β is the light blue curve. It comprises all ([Bcd], [Cad]) such that the bifurcation conditions hold, and the phase space looks like (B) all along this curve. **(C)** If we go past the bifurcation only the left saddle remains.

limited bifurcation analysis. AUTO uses *parameter continuation*, to continue an equilibrium solution. The continuation algorithm is essentially a predictor-corrector method.

1. Start with x_0.

2. Predict $\tilde{x_1}$, such that $f(\tilde{x_1}, \alpha_1) = 0$ holds.

   This is done using tangent prediction,

   $\tilde{x_1}$=$x_0 + \Delta\alpha\dot{x}|_{x_0,\alpha_0}$,

   where $\Delta\alpha$ is the step size and $\dot{x}$=$\frac{dx}{d\alpha}$.

3. Perform a correction to $\tilde{x_1}$ by Newton iterations, such that the iteration converges to x_1, which is a point that is within a specified accuracy of the true equilibrium,

   $\tilde{x}_1^{n+1} = \tilde{x}_1^n - \frac{f(\tilde{x}_1^n,\alpha_1)}{Df|_{\tilde{x}_1^n,\alpha_1}}$.

As long as Df is nonsingular and the step size is sufficiently small Newton iterations converge to x_1 from predicted point $\tilde{x}_1$. It should be noted that the step size is controlled by the following: if at our step size $\Delta\alpha$ no convergence occurs the step size is increased; if on the other hand convergence occurs after only a few iterations, the step size is decreased.

The goal is to find solutions x_i, as α is incremented further using the previous values x_{i-1}. The method above fails when Df becomes singular, i.e. one of the eigenvalues of Df is 0. This is the point where a saddle-node bifurcation occurs and the solution curve encounters a fold, by having a branch of solution turn around (see figure 3.2.2). In this case AUTO executes *Keller's Pseudo-Archlength Continuation* [8, 32]. The best way to parametrize a curve is through its arclength, thus instead of making a step in α, AUTO makes a step in the pseudo-arclength, which we denote as s. Here we reparametrize our system by introducing s such that both x and α are functions of s. By letting $v(s) = (x(s), \alpha(s))$, we find equilibrium points for the extended system $\Phi(v(s), s) = 0$.

For every equilibrium point AUTO calculates the Jacobian at that point and also keeps track of the Fold (this is to detect saddle-node bifurcation) and Hopf functions. These are

test functions that depend on the eigenvalues of the Jacobian. They are defined to have regular zeros at the bifurcation points, such that

$$\text{Fold} = \prod \lambda_i(x, \alpha) \tag{4.1.2}$$

and

$$\text{Hopf} = \prod_{i>j} (\lambda_i(x, \alpha) + \lambda_j(x, \alpha)), \tag{4.1.3}$$

where $\lambda_i(x, \alpha)$ are the eigenvalues of the $Df|_{x,\alpha}$.

At the point where a bifurcation occurs we can start a 2 parameter continuation of the bifurcation, to find the locus of the saddle-node or Hopf bifurcation points in the [Bcd]-[Cad] plane. The continuation problem consists of finding all points in the (x-α)-plane that satisfy

$$f(x, \alpha) = 0 \tag{4.1.4}$$

and

$$\det Df|_{x_f,\alpha_f} = 0. \tag{4.1.5}$$

Since $x \in \mathbb{R}^4$ and $\alpha \in \mathbb{R}^2$, we have a six dimensional space and five constraints, meaning that there is a 1-dimensional curve that satisfies both (4.1.4) and (4.1.5). When we project this curve into our parameter plane we get the bifurcation curve β.

4.2 Parametric Portrait for Gap Gene Circuit

The first circuit analyzed was from [44, 43]. Manu [44, 43] looked for the bifurcations along the A-P axis of the embryo. A saddle-node bifurcation located around 53% EL was found to be responsible for separating the dynamics in an embryo. Anterior to this bifurcation Bcd was crucial for pattern formation. This bifurcation has important biological meaning and will be called a biologically significant bifurcation (BSB). *bcd* mutants are missing all

the segments in the head and thorax (see bottom of figure 4.2.1), which using the fate map (defined in the introduction) translates back to the embryo missing all the structures up to 53% EL (fig 4.2.1) We can see that something fundamentally important happens in the fly around 53% EL, which has not been biologically explained, and Manu's analysis showed that indeed this behavior is captured in the dynamics of the model.

Anterior to the BSB, the gap gene pattern forms by the point attractors dependence on Bcd concentration. Posterior to BSB, the pattern arises by attraction to a 1-dimensional manifold and the dependence of the final state of the system by gastrulation on the initial Hb concentration.

While Manu looked at a one dimensional subset of the [Bcd]-[Cad] plane for bifurcations, I extended his analysis of the same circuit by looking at the whole [Bcd]-[Cad] plane. This chapter will show how a feature that is directly related to the BSB can help us determine which circuits are the best representatives for our system by looking at the parametric portrait alone.

To find the parametric portrait for the previously published circuit [44, table S3 Column 1] I continued all of the solutions to the equilibrium equation $f(x, \alpha) = 0$ from every nucleus by first varying v^{Bcd} and then varying v^{Cad}. Upon encountering a Hopf or a saddle-node bifurcation AUTO continued each of them in two parameters. A parametric portrait was constructed by plotting all the bifurcation curves in the [Bcd]-[Cad] plane. (see figure 4.2.2). From the parametric portrait we note that the most unstable bifurcations are the ones of codimension two (these correspond to points on the diagram), as any perturbation in either of the two parameters will result in a different behavior. For a codimension one bifurcation (these correspond to curves) there are some perturbations of both parameters simultaneously that result in similar dynamics. At any point in a region we can look at phase space and calculate how many attractors and saddles there are. If this is done for all regions explicit change in dynamics can be discerned from the parametric portrait as we go from one region to another.

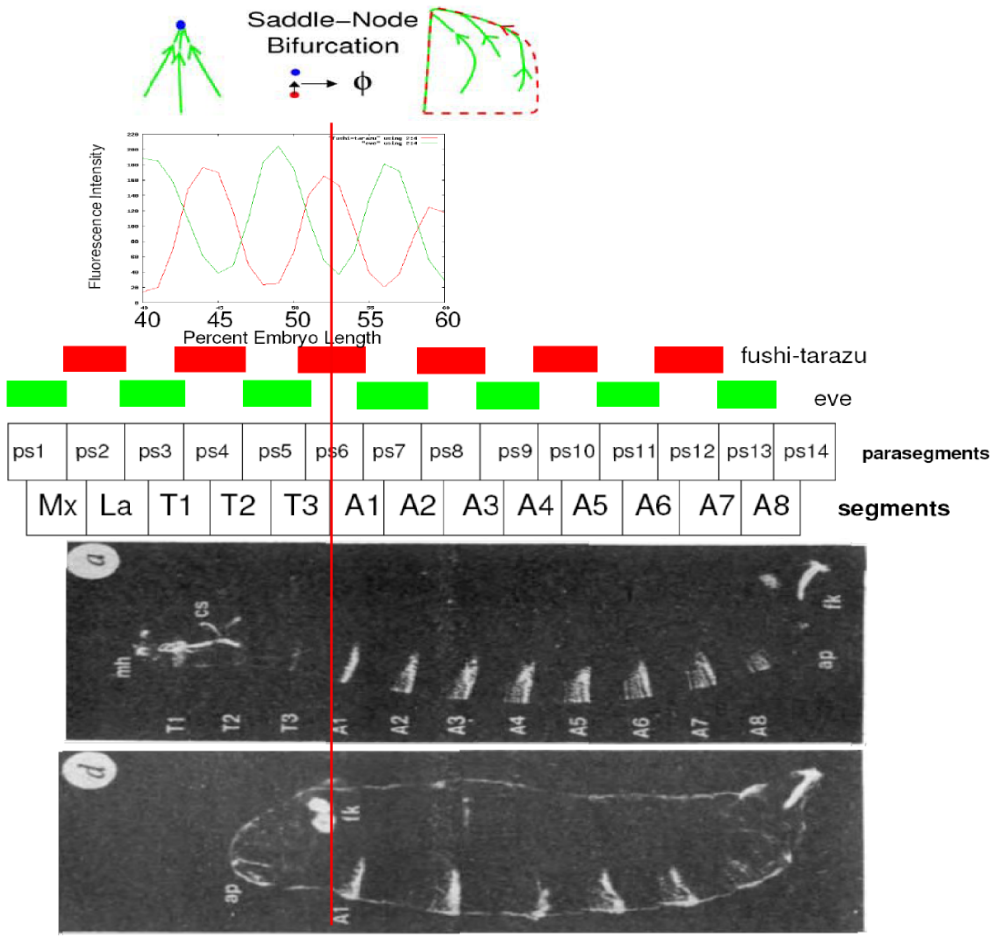


Figure 4.2.1: **Saddle-node bifurcation at 53% EL corresponds to the location of morphological anomaly in *bicoid* embryo.** Mapping cuticular segments (bottom) to percent embryo length (top) by locating segments and parasegments with respect to fushi-tarazu and eve positions(middle) which in turn map to % EL in our model [7, figure 9]. Cuticular images are at the bottom [13]. Upper panel is wild type. Lower panel is strong *bcd* phenotype (bcd^{E1}/bcd^{E1}) head and thorax missing. The top picture representing different ways of pattern formation in the anterior and posterior is taken from [43, figure 7]. Solid red line corresponds to the location of the saddle-node bifurcation.

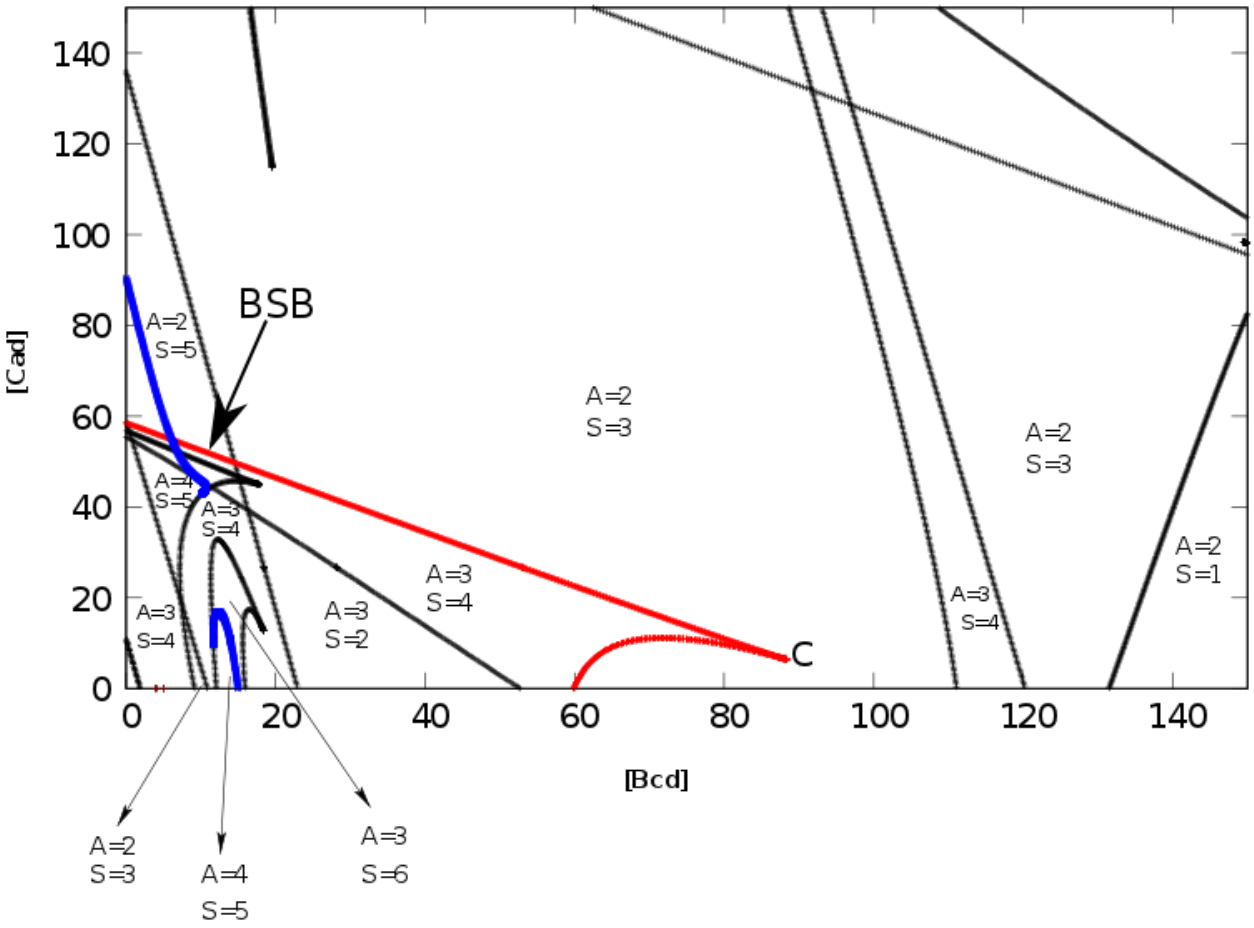


Figure 4.2.2: Parametric portrait of the [Bcd]-[Cad] plane. The solid lines divide this plane into regions with different dynamics. The behavior of the system is qualitatively the same in each domain. A is the number of attractors, S is the number of saddles. The location of the BSB in [44, 43] is indicated by a black arrow, and its continuation is shown in red. C indicates the cusp from which the BSB arises for appropriate values of Bcd and Cad. Blue curve indicates a locus where a Hopf bifurcation occurs.

One important feature of our parametric portrait should be noted—it is the most prominent cusp (C) located on the bottom of the figure (4.2.2). This cusp is generated from continuation of the BSB in the [Bcd]-[Cad] plane. At a cusp bifurcation, three equilibrium points collide together and result in one equilibrium point. This equilibrium point is unstable at the cusp, and a perturbation in our parameters would result in unfolding those three equilibrium points from the one equilibrium point.

The importance of the cusp will become evident once more circuits are considered and their parametric portraits generated.

4.3 Other Circuits and their Parametric Portraits

Manu [42] found other circuits for the model as well. The candidate circuits were found by first fitting them to integrated gap gene data using PLSA (Parallel Lam Simulated Annealing), such that the pattern of gap gene expression produced by the model was biologically correct within a certain tolerance. To judge the quality of the fit root mean squared difference (RMS) between data and model was used. In [42] the following three conditions were required to be fulfilled:

1. The squared difference between the model and the data had to be less than 12. RMS < 12 was chosen to weed out the models whose fits were unequivocally unacceptable.

2. The circuit had to produce the correct gap gene pattern, those with serious defects were not considered.

3. According to the experimental [6] and the theoretical [52] investigations Kr represses *hb*. The circuits needed to reproduce this phenomenon.

23 circuits from a total of 65 optimizations passed the above three criteria. These were called consistent circuits, and they had the same network topology as in Jaeger [24]. Namely, Bcd and Cad were activators of *hb, Kr, gt* and *kni*; Tll was an activator of *hb*, and a repressor

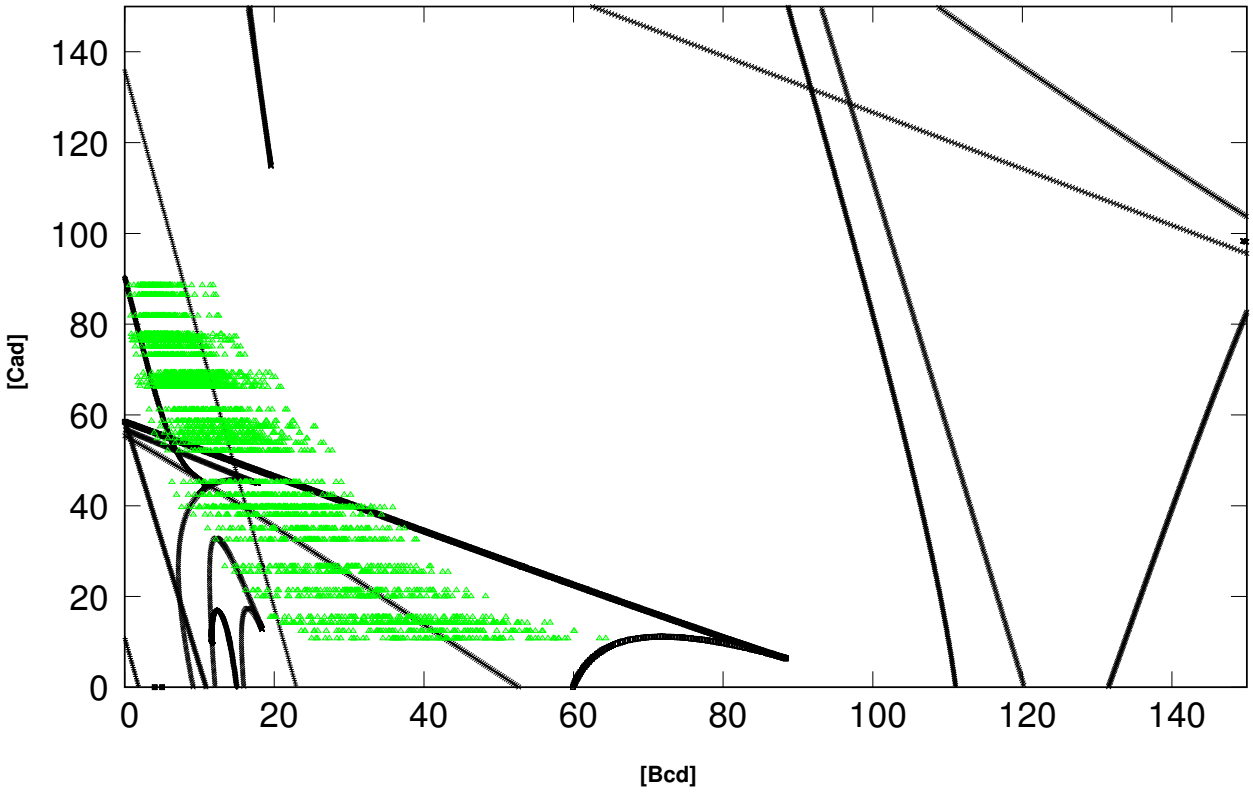


Figure 4.3.1: Parametric portrait of Bcd-Cad plane, along with (Bcd, Cad) concentrations from 88 simulations.

of the other gap genes; interaction between hb, Kr, gt, kni were of mutual repression except that Gt was an activator of hb in all circuits obtained; kni was an activator of gt in about half the circuits and a repressor in the other half. These 23 circuits were further subdivided into canalizing and noncanalizing circuits.

Canalizing circuits were the circuits that showed positional accuracy (i.e. positional variance for the six borders described below was comparable to variation of these borders in the data) for all of the following six 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. There were a total of 15 canalizing and 8 noncanalizing circuits.

In my analysis, these circuits were further clustered into groups. The canalizing circuits were subdivided into three groups and the noncanalizing circuits into two groups, by inspect-

ing the location and the types of bifurcations that their parametric portraits admitted. This was done by visual inspection of parametric portraits, where members of the same group had a very similar parametric portrait, and members of different groups had drastically different portraits. Thus one property of a circuit was its associated parametric portrait while the other property was an ensemble of the gap gene expression under Bcd perturbations, described below.

Manu simulated how circuits would reproduce the pattern of gap genes under the effects of variation between different Bcd profiles. Embryos in a population do not have the same Bcd profile but they produce gap gene pattern that canalizes over time. It is important that circuits were robust to different Bcd input, and reduced the variance of the gap gene pattern in accordance with the experimental data. Thus to each circuit a graph of how the gap gene pattern looked under 88 different Bcd profiles could be associated (for examples see figure 4.3.2). I found that there was a direct correspondence between the parametric portraits and the pattern of gap genes under the different Bcd profiles.

This correspondence was evident by focusing on the location of the cusp C described in the previous section. All the canalizing groups possessed such a cusp, and the saddle-node bifurcation boundary that contained the BSB intersected with the range of 88 ([Bcd], [Cad]) values that were present in the embryo. On the other hand, half of the noncanalizing groups did not possess such a cusp. Moreover, those that did have the cusp had it in locations such that the biologically realistic ([Bcd], [Cad]) values did not lie on the appropriate bifurcation boundary. Members of the same group were located in a cluster (see figure 4.3.2) when the location of the cusp in the [Bcd]-[Cad] plane was plotted. The only exceptions were several circuits that could not be identified clearly as members of one group and they possessed distinct patterns under Bcd variation. Each cluster could be associated with an ensemble of the gap gene patterns under Bcd variation (see figure 4.3.2).

Hence it is reasonable to suppose that the patterns of gap gene variability under different Bcd profiles for a circuit depend on the location of the cusp on the [Bcd]-[Cad] plane of

the parametric portrait for that circuit. Furthermore it can be conjectured that the correct canalization of a circuit, which is a biological property that has been experimentally confirmed and correctly modeled, is directly related to its bifurcation structure, which is embedded in the dynamics of the ODE describing the system.

To put the idea, that a bifurcation structure is responsible for the canalization of gap gene pattern, to the test we need to consider the case when the pattern decanalizes. As was previously discussed, the variability of the pattern in mutants is greater than in wild type, therefore considering mutants would be the next logical step. Unfortunately, model 1 does not reproduce the mutants' pattern correctly [42]. In order to study how to relate a property associated with a dynamical system to the phenomenon of canalization we need to consider a model that can pattern both wild type and mutants. The next two chapters will discuss how model 2 allows us to investigate pattern formation as well as to identify the structure responsible for canalization.

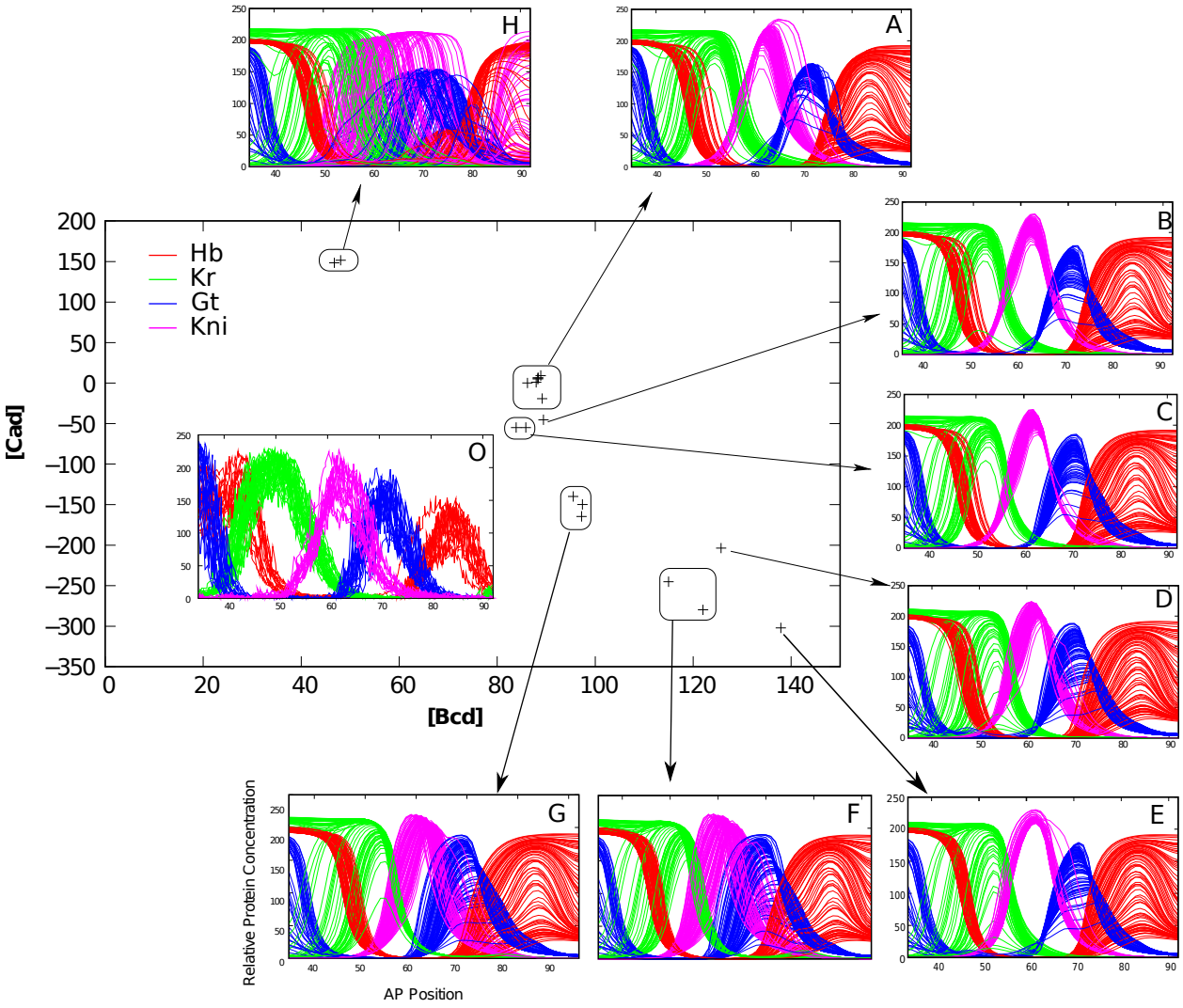


Figure 4.3.2: Locations of the cusp bifurcations on [Bcd]-[Cad] plane. Each group corresponds to a circled cluster, where each cluster has an associated characteristic gap gene pattern under Bcd variation. **(O)** Reproduced from Manu [42]. Gap gene expression data for late cleavage C14 wild type embryos. There are 18 profiles for *hb*, 33 for *Kr*, 20 for *gt* and 17 for *kni*. **(A)** This group contains the circuit analyzed in this chapter and in [42]. **(B)**-**(H)** Other groups with varying degree of canalization (these circuits were taken from [42]).

Chapter 5

Pattern Formation

In chapter 4 we demonstrated a connection between the degree of canalization in a circuit for model 1 and the location of the cusp bifurcation. Chapters 5 and 6 will deal with circuits from model 2 and use dynamical analysis to explain pattern formation. The degree of canalization or the lack thereof will once again be tied to particular features of the dynamical system.

The goal of this chapter will be to understand the dynamics of pattern formation in Kr^- as it was described in section 2.6.1. First, I will describe how to find a suitable model that captures both wild type and Kr^- gap gene expression. Then I will compare the model output to the data. Lastly, by analyzing equilibria and the dependence of solutions on the maternal input, I will provide a reason for the significantly lowered levels of expression in Kr^-.

5.1 Selection of the circuit

Since previous efforts to model Kr^- using model 1 have not been successful [42, 22], model 2 will be used from now on. As mentioned previously, the only difference between the two models is the way gap gene products interact with each other. Unlike model 1, model 2 contains extra terms $T^{abc}v^bv^c$ which describe a multiplicative contribution of the product of two genes b and c on the target gene a (see equation 3.1.3). The T^{abc} matrix is symmetric. In

order to talk about specific terms we will use the convention $T^{a\leftarrow(b,c)}$, where $T^{a\leftarrow(b,c)}$ describes how the combination of gene products b and c affects the expression of a.

Model 2 was fitted to integrated gap gene data for both wild type and Kr^- [68, 67] using Parallel Lam Simulated Annealing [5, 4]. To determine which circuit is best we needed to make sure that the pattern produced for wild type and mutant was qualitatively comparable to data. Furthermore, the wild type system needed to canalize Bcd variability at the level of gap gene expression.

The following two criteria were used:

1. The first measure of the quality of a fit was the RMS. There were separate RMS for wild type and Kr^-. To eliminate circuits with major patterning defects I considered only the circuits whose RMS was low. It is generally the case that circuits with RMS higher than 12 contain serious anomalies in the pattern [24]. Those circuits with RMS less than 12 were further visually inspected for anomalies in patterning.

2. The other property that the circuit had to have was correct canalizing behavior of the gap genes in Kr^- and wild type. The aim was to find a circuit where wild type would canalize Bcd perturbations while Kr^- would not.

In an effort to keep the number of parameters in equation 3.1.3 small the goal was to find the minimal number of nonzero T^{abc} elements that capture the correct patterning as well as canalizing behavior in both systems. K. Kozlov undertook the task above by starting with a large combination of nonzero T^{abc} terms, then gradually narrowing it down to as few terms as possible. The different combinations of nonzero T^{abc} elements that he found are depicted in table 5.1. All of the combination in the table produced correct patterns other than numbers 1, 2, 3, 10 and 14. I replicated Kozlov's trials 1-7 as well as tried 10 new combinations of nonzero T^{abc} elements (see table 5.2). In my case only combination numbers 7-10 produced correct patterns for both wild type and mutants. The only combination of T^{abc} terms that produced correct canalization behavior was nonzero $T^{a\leftarrow(Hb,Bcd)}$, $T^{a\leftarrow(Kr,Kr)}$ and $T^{a\leftarrow(Kr,Kni)}$ (number 5 from table 5.1).

	Nonzero T^{abc} terms
1	$T^{a\leftarrow(\text{Hb,Bcd})}$
2	$T^{a\leftarrow(\text{Kr,Kni})}$
3	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Kr,Kr})}$
4*	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Kr,Kni})}$
5* C	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Kr,Kr})}$, $T^{a\leftarrow(\text{Kr,Kni})}$
6*	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Kr,Kr})}$, $T^{a\leftarrow(\text{Kr,Kni})}$, $T^{a\leftarrow(\text{Hb,Gt})}$
7*	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Kr,Kr})}$, $T^{a\leftarrow(\text{Kr,Kni})}$, $T^{a\leftarrow(\text{Hb,Kr})}$
8*	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Kr,Kr})}$, $T^{a\leftarrow(\text{Kr,Kni})}$, $T^{a\leftarrow(\text{Hb,Kr})}$, $T^{a\leftarrow(\text{Hb,Gt})}$
9*	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Kr,Kr})}$, $T^{a\leftarrow(\text{Kr,Kni})}$, $T^{a\leftarrow(\text{Hb,Kr})}$, $T^{a\leftarrow(\text{Hb,Gt})}$, $T^{a\leftarrow(\text{Gt,Bcd})}$
10	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Hb,Kr})}$, $T^{a\leftarrow(\text{Hb,Gt})}$, $T^{a\leftarrow(\text{Gt,Bcd})}$, $T^{a\leftarrow(\text{Kr,Bcd})}$, $T^{a\leftarrow(\text{Kni,Bcd})}$
11*	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Hb,Kr})}$, $T^{a\leftarrow(\text{Hb,Gt})}$, $T^{a\leftarrow(\text{Gt,Bcd})}$, $T^{a\leftarrow(\text{Kr,Bcd})}$, $T^{a\leftarrow(\text{Kni,Bcd})}$, $T^{a\leftarrow(\text{Hb,Cad})}$, $T^{a\leftarrow(\text{Kr,Kni})}$
12*	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Kr,Kr})}$, $T^{a\leftarrow(\text{Kr,Kni})}$, $T^{a\leftarrow(\text{Hb,Kr})}$, $T^{a\leftarrow(\text{Hb,Gt})}$, $T^{a\leftarrow(\text{Gt,Bcd})}$, $T^{a\leftarrow(\text{Hb,Cad})}$, $T^{a\leftarrow(\text{Kr,Bcd})}$, $T^{a\leftarrow(\text{Kni,Bcd})}$
13*	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Hb,Kr})}$, $T^{a\leftarrow(\text{Hb,Gt})}$, $T^{a\leftarrow(\text{Gt,Bcd})}$, $T^{a\leftarrow(\text{Kni,Bcd})}$, $T^{a\leftarrow(\text{Kr,Gt})}$, $T^{a\leftarrow(\text{Hb, Cad})}$, $T^{a\leftarrow(\text{Kr,Kni})}$, $T^{a\leftarrow(\text{Kr,Bcd})}$
14	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Hb,Kr})}$, $T^{a\leftarrow(\text{Hb,Gt})}$, $T^{a\leftarrow(\text{Gt,Bcd})}$, $T^{a\leftarrow(\text{Kni,Bcd})}$, $T^{a\leftarrow(\text{Hb, Tll})}$, $T^{a\leftarrow(\text{Kr,Kni})}$, $T^{a\leftarrow(\text{Kr,Tll})}$, $T^{a\leftarrow(\text{Gt,Tll})}$, $T^{a\leftarrow(\text{Bcd,Cad})}$, $T^{a\leftarrow(\text{Tll,Bcd})}$
15*	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Hb,Kr})}$, $T^{a\leftarrow(\text{Hb,Gt})}$, $T^{a\leftarrow(\text{Gt,Bcd})}$, $T^{a\leftarrow(\text{Kni,Bcd})}$, $T^{a\leftarrow(\text{Kr,Gt})}$, $T^{a\leftarrow(\text{Kr,Kr})}$, $T^{a\leftarrow(\text{Hb,Cad})}$, $T^{a\leftarrow(\text{Kr,Kni})}$, $T^{a\leftarrow(\text{Kr,Bcd})}$, $T^{a\leftarrow(\text{Gt,Kni})}$, $T^{a\leftarrow(\text{Tll,Bcd})}$

Table 5.1: Combinations of nonzero T^{abc} terms tested by K. Kozlov. The notation $T^{a\leftarrow(b,c)}$ describes how combination of gene products b and c affects the expression of a. Asterisk (*) marks combinations that produced correct patterning behavior and C marks correct canalizing behavior.

	Nonzero T^{abc} terms
1	$T^{a\leftarrow(\text{Kr,Kr})}$, $T^{a\leftarrow(\text{Kr,Kni})}$
2	$T^{a\leftarrow(\text{Kr,Kr})}$, $T^{a\leftarrow(\text{Gt,Kr})}$
3	$T^{a\leftarrow(\text{Kr,Kr})}$, $T^{a\leftarrow(\text{Hb,Kr})}$
4	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Gt,Kni})}$
5	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Gt,Kr})}$
6	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Gt,Gt})}$
7*	$T^{a\leftarrow(\text{Kr,Kr})}$, $T^{a\leftarrow(\text{Kr,Kni})}$, $T^{a\leftarrow(\text{Hb,Kr})}$
8*	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Kr,Kni})}$,$T^{a\leftarrow(\text{Hb,Kni})}$
9*	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Kr,Kni})}$, $T^{a\leftarrow(\text{Hb,Kni})}$, $T^{a\leftarrow(\text{Kr,Gt})}$, $T^{a\leftarrow(\text{Hb,Kni})}$
10*	$T^{a\leftarrow(\text{Hb,Bcd})}$, $T^{a\leftarrow(\text{Kr,Kr})}$, $T^{a\leftarrow(\text{Kr,Gt})}$, $T^{a\leftarrow(\text{Kr,Kni})}$, $T^{a\leftarrow(\text{Kr,Bcd})}$, $T^{a\leftarrow(\text{Kr,Cad})}$

Table 5.2: 10 combinations of nonzero T^{abc} terms tested in this thesis. Asterisk (*) marks combinations that produced correct patterning behavior.

Model 2 with nonzero $T^{a\leftarrow(Hb,Bcd)}$, $T^{a\leftarrow(Kr,Kr)}$ and $T^{a\leftarrow(Kr,Kni)}$ was used for further study in this thesis. Using the above mentioned model 39 circuits that correctly patterned both wild type and mutant were found. Out of those 39 only two circuits showed the correct canalization for both wild type and Kr^-. These circuits produced parameters that were strikingly similar (see table 5.3). Both circuits were analyzed and the results discussed in chapters 5 and 6 are consistent between the two circuits. For clarity we chose one of them (`h_-3.5_wt_kr_11`) and from now on the analysis of the model will refer to this circuit. The RMS for wild type and Kr^- were 11.49 and 11.30 respectively.

5.1.1 Comparison of data to model predictions

We now discuss how well our model reproduced the major findings from the Kr^- data described in Chapter 2. One of the most striking features of the Kr^- data was the drastically reduced levels of gap gene expression. Our model correctly represented the lowered Hb, lowered Gt and greatly reduced Kni expression (see figure 5.1.1). Moreover, the model

Table 5.3: Parameters for full model `h_-3.5_wt_kr.11 and h_-3.5_wt_kr.13` with diffusion and *tll*. Asterisk (*) marks the parameters that were not annealed on.

Parameter	Model 11	Model 13
R^{hb}	15.0000	15.0000
R^{Kr}	15.0000	15.0000
R^{gt}	15.0000	15.0000
R^{kni}	15.0000	15.0000
D^{hb}	0.2000	0.2000
D^{Kr}	0.2000	0.2000
D^{gt}	0.180303	0.180610
D^{kni}	0.2000	0.2000
$t_{1/2}^{hb}$	16.0665	16.0678
$t_{1/2}^{Kr}$	11.7515	11.7378
$t_{1/2}^{gt}$	8.5051	8.5008
$t_{1/2}^{kni}$	10.3035	10.3026
$h^{hb}*$	−3.5000	−3.5000
$h^{Kr}*$	−3.5000	−3.5000
$h^{gt}*$	−3.5000	−3.5000
$h^{kni}*$	−3.5000	−3.5000
m^{hb}	0.080075	0.080120
m^{Kr}	0.053878	0.054241
m^{gt}	0.033992	0.034083
m^{kni}	0.058476	0.058769
$E^{hb\leftarrow\mathrm{Cad}}$	0.019791	0.019783
$E^{Kr\leftarrow\mathrm{Cad}}$	0.025042	0.024954
$E^{gt\leftarrow\mathrm{Cad}}$	0.017928	0.017931
$E^{kni\leftarrow\mathrm{Cad}}$	0.018573	0.018567
$E^{hb\leftarrow\mathrm{Tll}}$	−0.004557	−0.004547
$E^{Kri\leftarrow\mathrm{Tll}}$	−0.021468	−0.021213
$E^{gt\leftarrow\mathrm{Tll}}$	0.002626	0.002631
$E^{kni\leftarrow\mathrm{Tll}}$	−0.175158	−0.174912

Parameter	Model 11	Model 13
$T^{hb\leftarrow\text{Hb}}$	0.021351	0.021344
$T^{Kr\leftarrow\text{Hb}}$	−0.016405	−0.016651
$T^{gt\leftarrow\text{Hb}}$	−0.033599	−0.033619
$T^{kni\leftarrow\text{Hb}}$	−0.007471	−0.007593
$T^{hb\leftarrow\text{Kr}}$	0.014623	0.014643
$T^{Kr\leftarrow\text{Kr}}$	0.042268	0.042698
$T^{gt\leftarrow\text{Kr}}$	0.046056	0.046008
$T^{kni\leftarrow\text{Kr}}$	0.019732	0.019756
$T^{hb\leftarrow\text{Gt}}$	0.002661	0.002672
$T^{Kr\leftarrow\text{Gt}}$	−0.065261	−0.064835
$T^{gt\leftarrow\text{Gt}}$	0.020197	0.020224
$T^{kni\leftarrow\text{Gt}}$	−0.008831	−0.008857
$T^{hb\leftarrow\text{Kni}}$	−0.174888	−0.174833
$T^{Kr\leftarrow\text{Kni}}$	−0.017079	−0.017177
$T^{gt\leftarrow\text{Kni}}$	0.009825	0.009774
$T^{kni\leftarrow\text{Kni}}$	0.015671	0.015675
$T^{hb\leftarrow\text{(Hb,Bcd)}}$	−0.000560	−0.000561
$T^{Kr\leftarrow\text{(Hb,Bcd)}}$	0.000414	0.000417
$T^{gt\leftarrow\text{(Hb,Bcd)}}$	0.000929	0.000928
$T^{kni\leftarrow\text{(Hb,Bcd)}}$	−0.000742	-0.000742
$T^{hb\leftarrow\text{(Kr,Kr)}}$	−0.000106	−0.000106
$T^{Kr\leftarrow\text{(Kr,Kr)}}$	−0.000145	−0.000147
$T^{gt\leftarrow\text{(Kr,Kr)}}$	−0.000683	−0.000684
$T^{kni\leftarrow\text{(Kr,Kr)}}$	−0.000351	−0.000352
$T^{hb\leftarrow\text{(Kr,Kni)}}$	−0.000305	−0.000304
$T^{Kr\leftarrow\text{(Kr,Kni)}}$	−0.000036	−0.000037
$T^{gt\leftarrow\text{(Kr,Kni)}}$	−0.000800	−0.000798
$T^{kni\leftarrow\text{(Kr,Kni)}}$	0.000268	0.000269

captured that the expression of *kni* in Kr^- is on par with the background level by the time of gastrulation [67].

Another behavior our model represented correctly was the earlier time at which the Kr^- gap gene expression peaked. Just like the data, our model showed that the expression reached maximum around time class 5 and declined to very low levels thereafter, while the wild type took longer to reach its maximum.

On the other hand, our model did not capture the full magnitude of the anterior shift of the *gt* domain into the central part of the embryo as documented in the data. Our circuit predicted the anterior shift, but not to the same degree as seen in data. Another shortcoming of our model was that in Kr^-, the posterior domain of *gt* in the data is much broader than in the model. While the model predicted that the *gt* domain in Kr^- was more variable than its wild type counterpart, the variability was exaggerated.

Overall, the circuit captured the gap gene expression of wild type correctly. In addition it reproduced the key features of Kr^-, such as a lowered expression of all gap genes in addition to the anterior shift of Gt.

5.2 Pattern formation in wild type

5.2.1 Anterior regime

In chapter 3 we discussed how framing the gap gene circuits as a dynamical system allows us to use the techniques of dynamical analysis in order to explain processes such as pattern formation and canalization. This chapter will use the techniques from chapter 3 with simplifications that were introduced in chapter 4 to show how pattern formation arises from our equations. We began our analysis by using a simplified version of `h_-3.5_wt_kr_11` without diffusion or Tll input and considering only the region 35-72% EL. Our analysis involved gathering information about the equilibria and the behavior of the trajectories for each nucleus. We first analyzed wild type—considering R^{Kr}=15. We calculated the equilibria using the

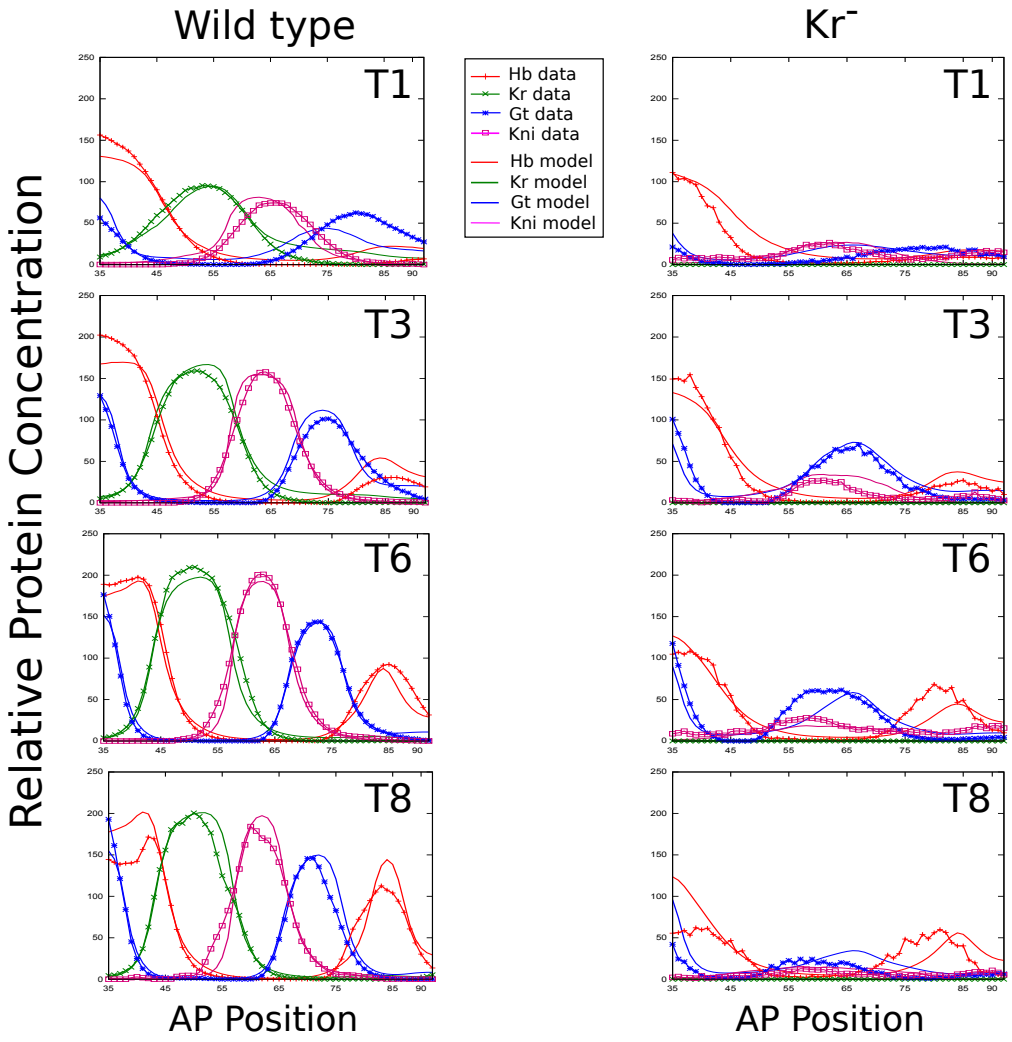


Figure 5.1.1: **Comparison between the dynamics of gap gene expression in the full model `h_-3.5_wt_kr_11` to data for wild type and Kr^-.** Each panel on the left provides a comparison of full model of wild type to wild type data for all four gap genes at selected times. Each panel on the right provides a comparison of full model of Kr^- for the three relevant gap genes at selected times.

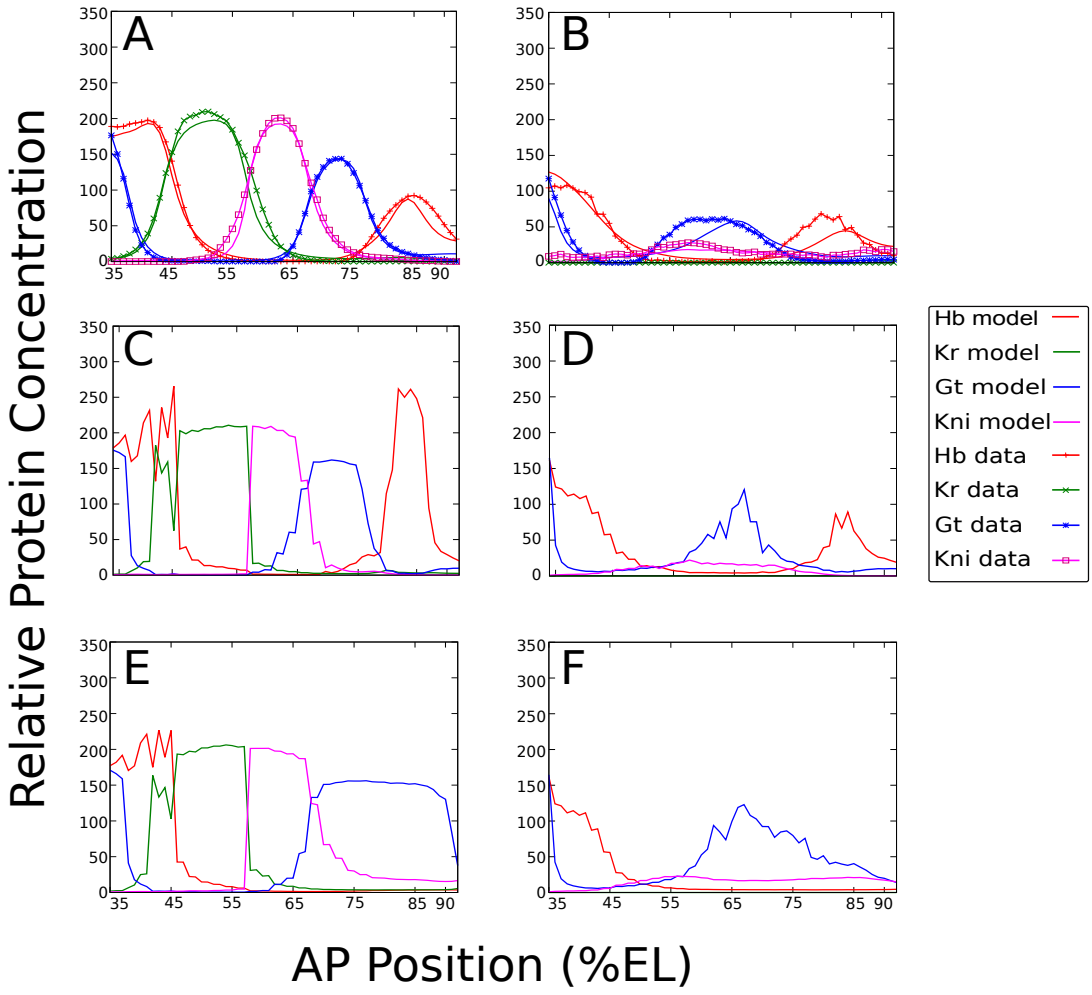


Figure 5.1.2: **Comparison between the wild type and Kr^- gap gene patterns.** Time class T6 is represented. **(A)**Gap gene pattern for wild type model and data with diffusion and Tll. **(B)**Gap gene pattern for Kr^- model and data with diffusion and Tll. **(C)**Gap gene pattern for wild type model without diffusion. **(D)**Gap gene pattern for Kr^- model without diffusion. **(E)** Gap gene pattern for wild type model without diffusion and without Tll. **(F)** Gap gene pattern for Kr^- model without diffusion and without Tll.

Newton-Raphson method starting from initial conditions that were uniformly distributed inside the hypercube (0,350)×(0,350)×(0,350)×(0,350). Since the synthesis term is bounded and decay term is bounded and nonzero, solutions lie in a bounded region. More specifically solutions v^a to the equations $R^a g(u^a) - \lambda^a v^a = 0$, where $g(u^a)$ is between 0 and 1, are such that $v^a \leq R^a/\lambda^a$. Hence, the maximum v^a will be is slightly less than 350. The equilibrium points are the possible end states of the system, and the trajectories are the dynamics of our ODE. Each nucleus had a specific initial condition associated to it—Hb forms a gradient that decreases towards the posterior (see last column of table 5.4) and the other gap genes are all 0.

% EL	Bcd	Cad	Initial Conditions of Hb
35	43.55	10.80	57.43
37	38.58	14.27	53.63
39	34.18	15.50	50.05
41	30.28	21.38	47.73
43	26.82	26.65	42.27
45	23.76	35.01	39.12
47	21.05	38.07	32.69
49	18.65	39.70	29.40
51	16.52	45.25	23.39
53	14.64	53.84	20.67
55	12.97	54.83	15.57
57	11.49	58.68	13.20
59	10.18	66.19	9.45
61	9.02	69.31	8.06
63	7.99	66.44	5.72
65	7.08	68.41	4.77
67	6.27	75.08	3.38
69	5.56	76.40	2.95
71	4.92	77.19	2.24

Table 5.4: Description of nuclei in terms of Bcd, Cad and Hb. The dynamics of the nucleus are determined by the pair (Bcd, Cad) but the specific trajectory and the final state of the system at T6 are determined by the initial conditions of Hb.

We use the following notation in the analysis below. Firstly, attractors are denoted as A_i where i is a reference to a specific attractor. Secondly, the saddles are identified as $S_i^{j,k}$, where j is the number of positive and k is the number of negative eigenvalues of the Jacobian at the saddle equilibrium. i is a reference to a specific saddle. Furthermore, to add a quantitative description to the saddles and attractors of interest, the notation a, b-on will represent an equilibrium point with high level of protein a and b (in this work that means above 50) while the other proteins are at a low level [44, 43]. The notation *all*-off will represent an equilibrium point where all the proteins are at a low level (in this work that means below 25); see table 5.6 for a description of all the equilibrium points encountered in our system. Lastly, there will be certain cases in this thesis where we will be interested in one branch of a manifold emerging from saddles of type $S^{1,3}$ and ending in an attractor A_n. In those specific instances the branch of interest ending in A_n will be denoted as U_n^+.

The rest of this subsection will describe how the underlying dynamics explain wild type pattern formation in the anterior by looking at nuclei in the range of 35-57% EL. Table 5.5 contains a summary of this discussion.

At nucleus 35 there are two attractors: A_1 and A_2. There are also three saddles: $S_1^{1,3}$, $S_2^{1,3}$ and $S_3^{2,2}$. Starting from the initial conditions on the Hb-axis that correspond to nucleus 35 the trajectory is attracted to A_1, getting there by gastrulation, hence the state of the system is hb, gt-on.

Moving posterior to nucleus 37 the system undergoes a bifurcation that creates two new saddles: $S_4^{1,3}$ and $S_5^{2,2}$. This change does not affect the trajectory that starts at initial conditions for nucleus 37, hence by gastrulation the state of the system ends up at the hb, gt-on attractor once again.

Moving to nucleus 39 affects the overall dynamics significantly, as the newly created $S_4^{1,3}$ and A_1 annihilate each other in a saddle-node bifurcation. The saddle-node bifurcation between nuclei 37 and 39 is responsible for the formation of the posterior border of the anterior gt as it annihilates the attractor A_1. The trajectories are attracted to a limit cycle

% EL	Bifurcations anterior to this nucleus	Attractors	Saddles	Asymptotic state from biological initial conditions	State of the system at T6
35		A_1	$S_1^{1,3}$, $S_2^{1,3}$		
		A_2	$S_3^{2,2}$	A_1	*hb,gt*-on
37	$S_4^{1,3}$ and $S_5^{2,2}$ are created (saddle-node)				
		A_1	$S_1^{1,3}$, $S_2^{1,3}$, $S_4^{1,3}$		
		A_2	$S_3^{2,2}$, $S_5^{2,2*}$	A_1	*hb,gt*-on
39	A_1 and $S_4^{1,3}$ are annihilated (saddle-node)				
		A_2	$S_1^{1,3}$, $S_2^{1,3}$	Limit cycle	*hb*-on
			$S_3^{2,2}$, $S_5^{2,2*}$		
41	$S_5^{2,2*} \mapsto A_4^*$ (Hopf). A_3 and $S_6^{1,3}$ are created (saddle-node)				
		A_2	$S_1^{1,3}$, $S_2^{1,3}$, $S_6^{1,3}$		
		A_3	$S_3^{2,2}$	A_3	*hb*-on
		A_4^*			
43-45		A_2	$S_1^{1,3}$, $S_2^{1,3}$, $S_6^{1,3}$		
		A_3	$S_3^{2,2}$	A_2 via U_2^+	Kr, hb-on
		A_4^*			
47-57		A_2	$S_1^{1,3}$, $S_2^{1,3}$, $S_6^{1,3}$		
		A_3	$S_3^{2,2}$	A_2	Kr-on
		A_4^*			

Table 5.5: Summary of equilibria for each anterior nucleus for wild type (see table 5.6 for biological characterization of these equilibria). The bifurcations which occur anterior to each nucleus, the asymptotic behavior of the trajectories for each nucleus and the location of the system in phase space by T6. Creation and annihilation are in the direction from anterior to posterior. Equilibria marked with an asterisk have complex eigenvalues.

Equilibrium point	Description
A_1	hb, gt-on
A_2	Kr-on
A_3	hb-on
A_4	all-off
$S_1^{1,3}$	Kr, kni-on
$S_2^{1,3}$	hb, Kr-on
$S_3^{2,2}$	all-off
$S_4^{1,3}$	all-off
$S_5^{2,2}$	all-off
$S_6^{2,2}$	hb-on

Table 5.6: Biological characterization of the equilibrium points at the location in phase space when the point first appears in table 5.5.

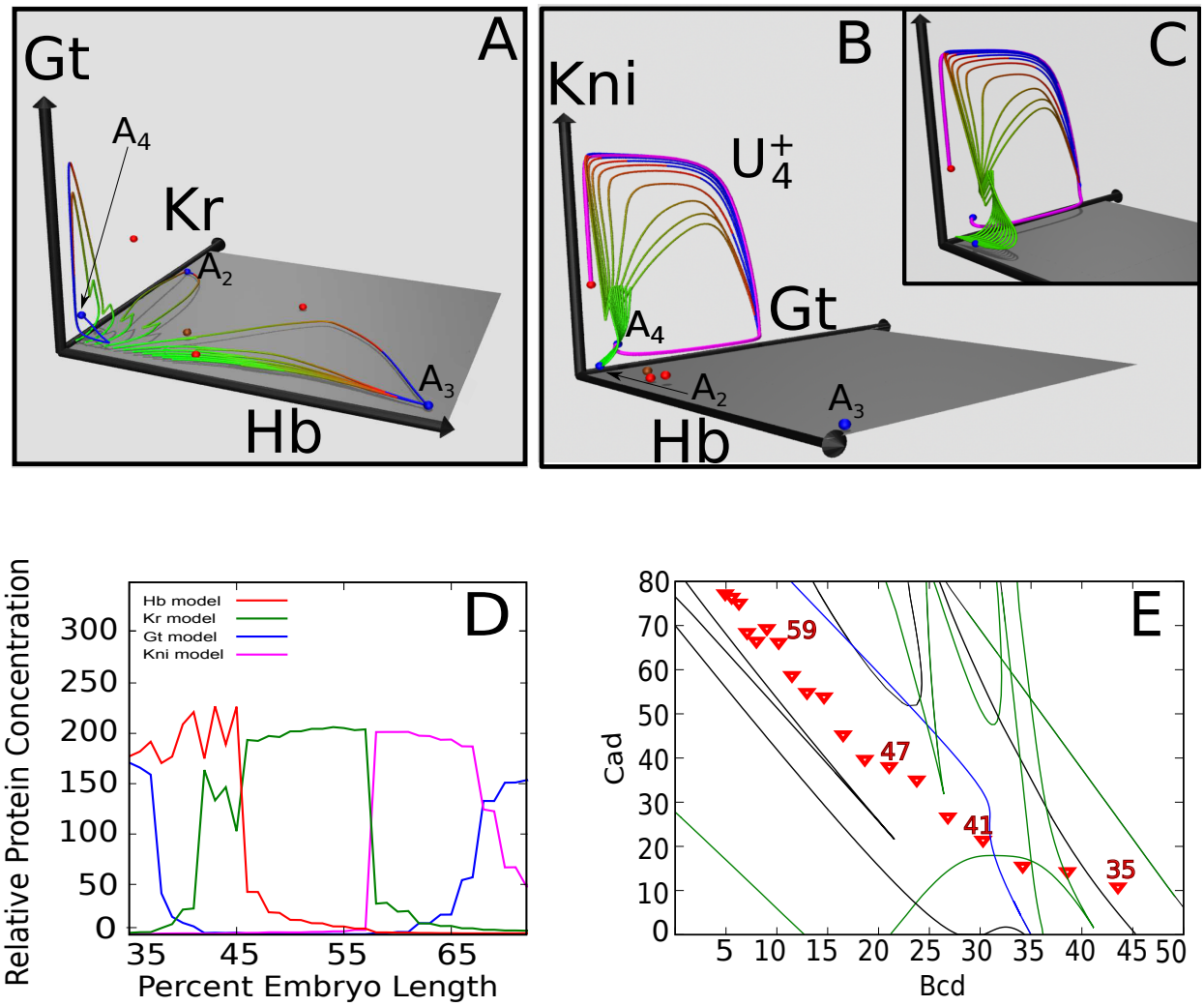


Figure 5.2.1: Pattern formation for wild type from 35% EL to 72% EL. Blue spheres are attractors. Red spheres are saddles of type $S^{1,3}$. Brown spheres are saddles of type $S^{2,2}$. Time is represented by the color of the trajectories. Trajectories are green from cycle 13 to beginning of C14A. Throughout C14A the color changes from green to red. Red corresponds to gastrulation. There is an abrupt transition to blue after gastrulation. **(A)** 47% EL. 10 trajectories are shown with starting points uniformly distributed on the Hb axis between 0-100. **(B)**59% EL. 10 trajectories are shown with starting points uniformly distributed on the Hb axis between 0-10. The unstable magenta manifold is generated by $S_1^{1,3}$. **(C)** Panel (B) with Hb scale increased by a factor of 3. **(D)** Gap gene pattern for wild type (without diffusion and without tailless). **(E)** Parametric portrait of the circuit for wild type. Black lines are bifurcations involving only saddles. Green lines are bifurcations involving saddles and attractors. The dark blue line is the Hopf bifurcation. Red triangles correspond to nuclei 35-71, with 35% EL on the right bottom corner.

that is associated with a Hopf bifurcation discussed in the paragraph below. The limit cycle doesn't have biological significance as it is transient and it does not allow for any special dynamical behavior, it only serves as an attracting object to place the system at *hb*-on by gastrulation.

Moving more to the posterior the system undergoes the last two bifurcations, a Hopf and a saddle-node. The Hopf bifurcation turns $S_5^{2,2*}$ which is an *all*-off saddle with complex eigenvalues into the attractor A_4^* with complex eigenvalues. As we approach the Hopf bifurcation the radius of the limit cycle decreases and this causes the space where the trajectories were moving away from $S_5^{2,2*}$ to shrink. Eventually the limit cycle's radius becomes 0 and disappears altogether, leaving only the trajectories that were spiraling into it. This is how A_4^* is formed. This attractor will pay an important part in posterior pattern formation. At this anterior point on the A-P axis it does not play a part at all as its basin of attraction consists of low values of Hb, which are only accessible from the initial conditions existing in the posterior part of the embryo.

The saddle-node bifurcation creates $S_6^{1,3}$ and A_3. Hence at 41% EL we have a total of three attractors: A_2, which has been present from the most anterior point considered together with the newly created A_3 and A_4^*.

As no more bifurcations occur at more posterior points, the patterning in this region is a result of how changes in Bcd and Cad affect the position of different equilibria as well as to which basin of attraction the initial condition on Hb belongs to and what structure the trajectory gets attracted to. At 41% EL the nucleus is in the basin of A_3.

Going from nucleus 41 to 43 the phase space undergoes two major changes in terms of the location of equilibria. The $S_6^{1,3}$ saddle goes from being *hb*-on to hb, kr-on and A_3 moves to an even higher location along the Hb-axis. Whereas in nucleus 41 the majority of the trajectories tended to A_3, now only the trajectories with high Hb initial conditions end up at A_3. Trajectories with a starting point at the intermediate Hb values go to A_2 via an unstable manifold U_2^+ from $S_2^{1,3}$ to A_2. By the onset of gastrulation they wind up at the

hb, Kr-on state. The spike in the Kr expression at its anterior border is solely an artifact of the diffusionless model. In the model with diffusion there are no such spikes, but rather a smooth transition from Kr-off to Kr-on. This is also the only place in the anterior where the trajectories are attracted to the unstable manifold rather than point attractors.

More posteriorly, the saddle $S_2^{1,3}$ moves in the direction of lower hb and higher Kr. This movement of $S_2^{1,3}$ is the mechanism by which the posterior border of anterior hb is formed. Nucleus 47 is in the basin of A_2 and the trajectories get to Kr-on by gastrulation (see figure 5.2.1 panel (A)). Hence the anterior border of the central Kr domain is formed by switching basins into A_2 from A_3; the switch is facilitated by the movement of $S_2^{1,3}$. From 47% EL to 57% EL there are no significant changes in the phase space, and the nuclei are persistently located in the basin of A_2.

5.2.2 Posterior regime

From 59% EL until 72% EL the dynamical structure of the phase space remains invariant and the initial Hb concentration places the nuclei in the basin of the *all*-off attractor A_4^*. Trajectories approach A_4^* by converging to the unstable manifold arising from $S_1^{1,3}$ (see magenta curve in figure 5.2.1 panel (B)). The unstable manifold traverses all of the states in the embryo posterior to 57% EL. The trajectories approach the unstable manifold by gastrulation and the final state in the phase space depends continuously on the initial Hb position (see figure 5.2.1 panel (C)). The higher initial conditions of Hb, correspond to more anterior nuclei which end up at *kni*-on, while the lower values correspond to more posterior nuclei which end up at *gt*-on by gastrulation (see table 5.4 and figure 5.2.1 panel (C)). Thus the posterior border of the central Kr domain is formed by crossing into the basin of A_4^*, while both the anterior and posterior borders of *kni* and the anterior border of posterior *gt* are formed by the continuous dependence of the final state on the initial values of Hb as facilitated by the attraction to the unstable manifold.

These results support a previous analysis of wild type [43]. In the anterior part of the

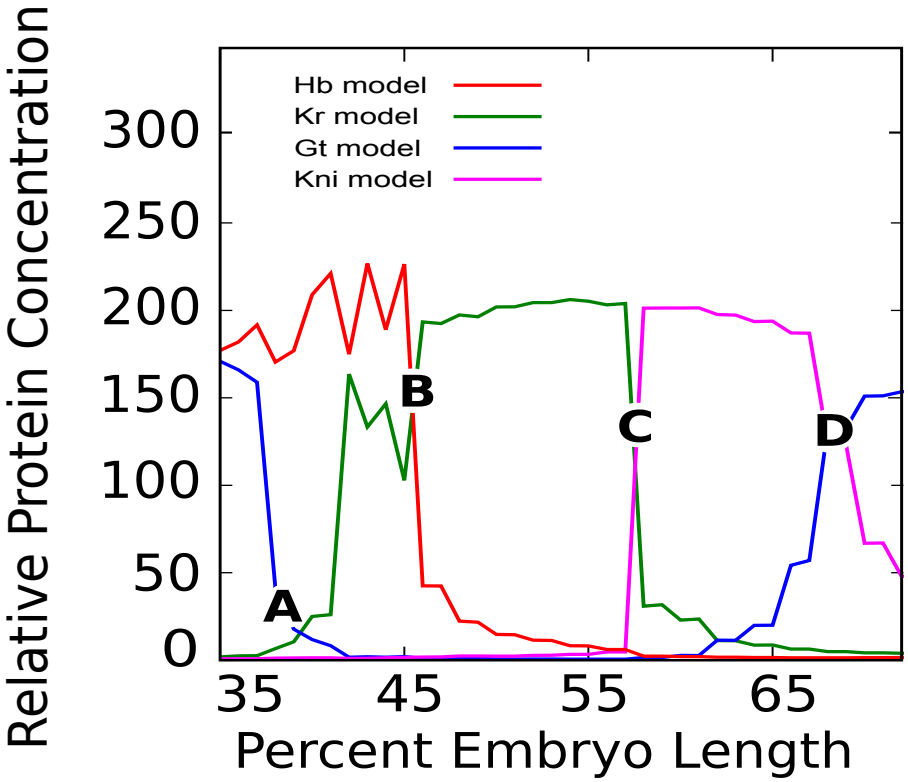


Figure 5.2.2: **Mechanisms of border formation for wild type.** This is an annotated version of panel (D) from figure 5.2.1. Borders are marked by letters **A**-**D** **(A)** Bifurcation annihilates A_1. **(B)** Trajectory switches from being in A_3 basin to A_2 basin. **(C)**Trajectory switches from being in A_2 basin to A_4^* basin. **(D)** Trajectory approaches the attractor via an unstable manifold, with the state of the system by gastrulation dependent on the initial value of Hb.

embryo the main mechanism of pattern formation was the attraction to point attractors. The change in concentration in Bcd and Cad was responsible for the change in the dynamics that allowed for correct pattern formation.

In the posterior the object of attraction changed to an unstable manifold, and the states of trajectories at gastrulation were sensitive to changes in the initial conditions of Hb. The position of the manifold in Hb-Kr-Gt-Kni space remained fairly invariant for all nuclei posterior to 57% EL. In fact, as we shall see in Chapter 6, this particular unstable manifold will be fairly invariant to perturbations in Bcd alone with Cad fixed, and will play a major role in canalizing the gap gene pattern. In the case of [42], the switch between two different mechanisms for pattern formation came about through a saddle-node bifurcation, but in this study it was through switching into the basin of A_4^*.

5.3 Pattern formation in Kr^-

The model for Kr^- consists of the same parameters as the one for wild type with the exception of R^{Kr} which is set to 0. Our analysis of mutants proceeds similarly to the wild type. We start off by finding the equilibria for each nucleus and calculating the dynamics that different combinations of Bcd and Cad bring about. There are two different ways of finding equilibria. We can either think of our mutant model as a separate entity from wild type and find the zeros using the Newton-Raphson method, or we can think of Kr^- as a numerically mutated wild type version. By continuously changing one parameter in our model, the maximum synthesis rate of Kr, we can smoothly turn down the function of Kr in such a way as to model an allelic series running from homozygous wild type levels to heterozygous, and then to hypomorphic and finally to functional null. These two methods were used to cross-validate each other.

At every nucleus each equilibrium found for wild type was used as the starting point for continuation in the R^{Kr} parameter. AUTO (see section 4.2) allowed us to trace what

happened to each of the equilibrium points as the transition between wild type and Kr^- took place (see figure 5.3.1).

The first aspect to notice is the annihilation of A_2. A_2 gets annihilated with $S_1^{1,3}$ as R^{Kr} is lowered (see figure 5.3.1 panel (A)). This saddle-node bifurcation occurs around R^{Kr}=6 for 35% EL, but the position moves continuously to higher values of R^{Kr} as we consider more posterior nuclei. For 45% EL it is around 10 and stays so until 72% EL.

Another observation to note is that A_4 appears sooner in mutants—by nucleus 39, while in wild type it is only present from nucleus 41 onward (see table 5.7). This is the only time where the stability of an equilibrium point is changed when we transition from wild type to Kr^-. All the attractors in the Kr^- dynamics come from wild type through continuation in the R^{Kr} parameter. In fact an even stronger statement can be made, which is that the attractors remain fairly invariant throughout the transformation (see table 5.7). Furthermore, with the exception of 35% EL all the saddles that appear in the Kr^- dynamics are continuations of saddles from the wild type (see figure 5.3.1).

Considering the dynamics of 35% EL with parameter R^{Kr} being lowered, the only two equilibria that have not originated from wild type are $S_4^{1,3}$ and $S_5^{2,2}$, which were created through a saddle-node bifurcation at R^{Kr}=1.

The number of equilibria for each nucleus decreased from between 3 and 7 in wild type to between 1 and 3 in Kr^-.

At 35% EL the system possesses one attractor, A_1 and two saddles, $S_4^{1,3}$ and $S_5^{2,2}$. The trajectory reaches A_1 by gastrulation.

At 37% EL $S_4^{1,3}$ has a dramatic gain in Hb concentration, thus guiding the trajectories along higher Hb coordinates en route to A_1. A_1 is reached long after the gastrulation; at the onset of gastrulation the system is in *hb*-on state.

There are two bifurcations between nucleus 37 and 39. The first is a Hopf bifurcation which turns $S_5^{2,2}$ with complex eigenvalues into A_4 with complex eigenvalues. The second bifurcation occurs when $S_4^{1,3}$ and A_1 come together and annihilate each other in a saddle-

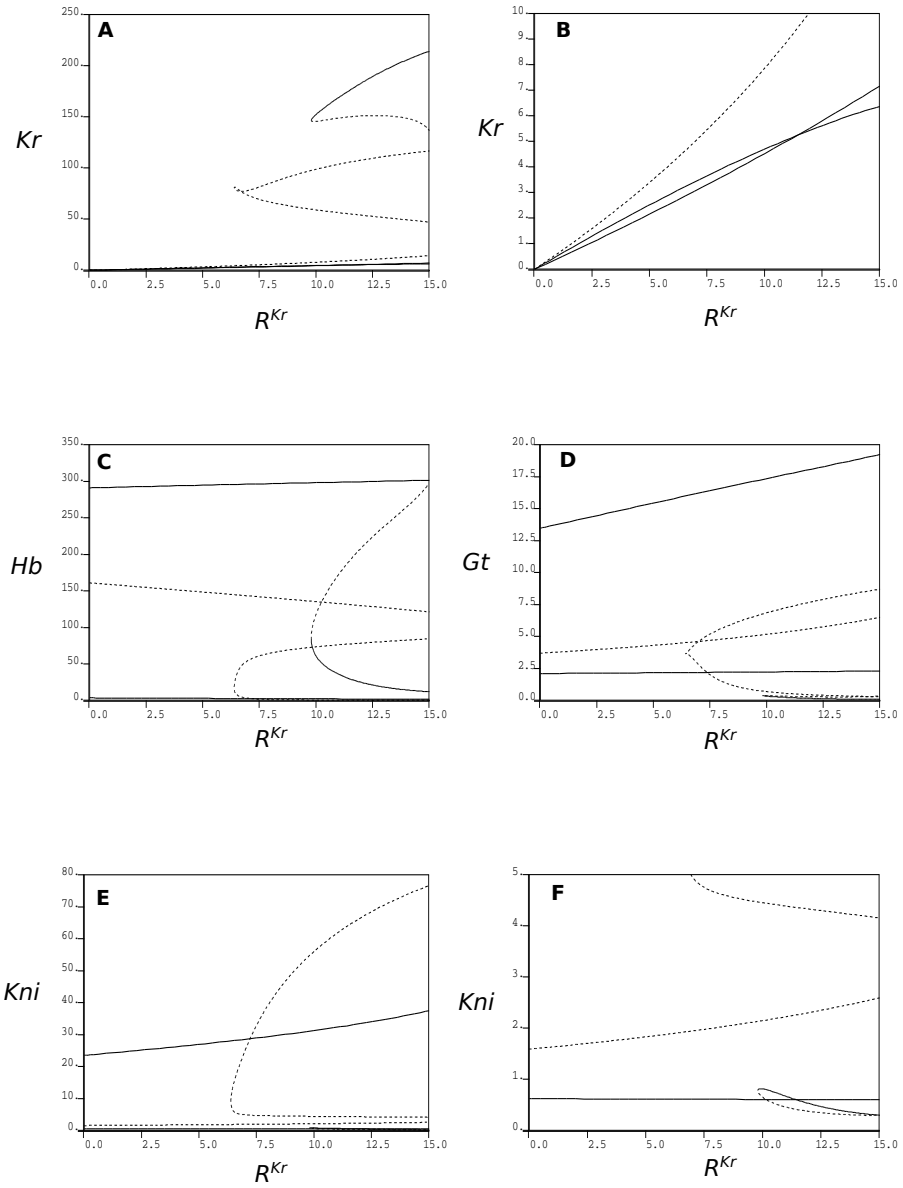


Figure 5.3.1: **Continuation of equilibria at 43% EL from wild type (R^{Kr}=15) to Kr^- (R^{Kr}=0).** Solid lines represent locations of attractors. Dotted lines represent locations of saddles. The X-axis is R^{Kr}. The equilibria are continued with respect to R^{Kr}. The Y-axis is the particular gap gene component of the equilibrium. As the equilibria lie in a four-dimensional space, each has four components to it. There are 7 zeros for wild type and only 3 zeros for mutants **(A)** Kr component **(B)** Kr component at finer scale, showing that at R^{Kr}=0, all the Kr components are 0. **(C)** Hb component; **(D)** Gt component; **(E)** Kni component. **(F)** Finer scale of Kni component for small values of Kni.

% EL	R^{Kr}=15 (wild type)	Dynamics for Intermediate values of R^{Kr}	R^{Kr}=0 (Kr^-)
35	A_1, A_2 $S_1^{1,3}$, $S_2^{1,3}$, $S_3^{2,2}$	A_2 and $S_2^{1,3}$ annihilate each other; $S_1^{1,3}$ and $S_3^{2,2}$ annihilate each other; $S_4^{1,3}$ $S_5^{2,2}$ are created	A_1 $S_4^{1,3}$, $S_5^{2,2}$
37	A_1, A_2 $S_1^{1,3}$, $S_2^{1,3}$, $S_3^{2,2}$, $S_4^{1,3}$, $S_5^{2,2}$	A_2 and $S_2^{1,3}$ annihilate each other; $S_1^{1,3}$ and $S_3^{2,2}$ annihilate each other;	A_1 $S_4^{1,3}$, $S_5^{2,2}$
39	A_2 $S_1^{1,3}$, $S_2^{1,3}$, $S_3^{2,2}$, $S_5^{2,2}$	A_2 and $S_2^{1,3}$ annihilate each other; $S_1^{1,3}$ and $S_3^{2,2}$ annihilate each other; $S_5^{1,3}$ turns into A_4	A_4
41	A_2, A_3, A_4 $S_1^{1,3}$, $S_2^{1,3}$, $S_3^{2,2}$, $S_6^{1,3}$	A_2 and $S_2^{1,3}$ annihilate each other; $S_1^{1,3}$ and $S_3^{2,2}$ annihilate each other; A_3 and $S_6^{1,3}$ annihilate each other	A_4
43-72	A_2, A_3, A_4 $S_1^{1,3}$, $S_2^{1,3}$, $S_3^{2,2}$, $S_6^{1,3}$	A_2 and $S_2^{1,3}$ annihilate each other; $S_1^{1,3}$ and $S_3^{2,2}$ annihilate each other;	A_3, A_4 $S_6^{1,3}$

Table 5.7: Equilibria of Kr^- obtained by continuous reduction of R^{Kr} from the equilibria of wild type. This table summarizes the dynamics that occur as R^{Kr} is lowered, as well as the final equilibria of the Kr^- system. If two equilibria are either annihilated or created it is done through a saddle-node bifurcation. If a saddle becomes an attractor or vice-versa it is through a Hopf bifurcation.

% EL	Bifurcations anterior to this nucleus	Attractors	Saddles	Asymptotic state from biological initial conditions	State of the system at T6
35		A_1	$S_4^{1,3}$ $S_5^{2,2}$	A_1	hb, gt-on
37		A_1	$S_4^{1,3}$ $S_5^{2,2*}$	A_1	hb-on
39-41	$S_5^{2,2*} \mapsto A_4^*$ (Hopf). A_1 and $S_4^{1,3}$ are annihilated (saddle-node)	A_4^*		A_4^*	hb-on
43-51	A_3 and $S_6^{1,3}$ are created (saddle-node)	A_3 A_4	$S_6^{1,3}$	A_4	Nuclei 43-45 hb-on, Nuclei 47-51 all-off
53-71		A_3 A_4^*	$S_6^{1,3}$	A_4^*	Nuclei 53-57 all-off Nuclei 57-59 medium Gt, Nuclei 59-71 gt-on

Table 5.8: Summary of equilibria for each nucleus for Kr^- (see table 5.6 for biological characterization of these equilibria). The bifurcations which occur anterior to each nucleus, the asymptotic behavior of the trajectories for each nucleus and the location of the system in phase space at T6 are presented. Creation and annihilation are in the direction from anterior to posterior. Equilibria marked with an asterisk have complex eigenvalues.

node bifurcation. Thus at 39% EL and 41% EL the system possesses only one attractor, namely the *all*-off A_4^*. The initial conditions of nuclei 39 and 41 place them too far away from this attractor to reach it before the gastrulation, by the onset of which they are in *hb*-on state (see figure 5.3.2 panel (A)).

The final bifurcation occurs between 41% EL and 43% EL. It is a saddle-node bifurcation which creates $S_6^{1,3}$ and A_3. Note that even though the system now possesses A_3, the basin of its attraction consists of initial Hb values that are greater than 50, while by 43% EL the Hb value is only 42.27 (see table 5.4). Hence this attractor does not play a role in the pattern formation for Kr^-. By nucleus 49 A_4 is no longer complex. As we move to the posterior, the initial conditions are closer to A_4, which results in posterior nuclei having lower Hb final state by the time of gastrulation. This is how the posterior border of the anterior *hb* domain forms.

For nuclei posterior to 63% EL, A_4^* becomes complex again, with trajectories approaching it via a spiral. The spiral lies in the Gt-Kni plane, with very a low Kni component (see figure 5.3.2 panels (C) and (D)). The *gt* domain is formed by attraction to a stable focus.

We now explain the reasons behind two main features of Kr^- data: the reduced levels of gap gene expression and the anterior shift of the *gt* domain. In the anterior the Hb levels are reduced in mutants due to the loss of A_1 via a saddle-node bifurcation (trajectories are attracted to *all*-off A_4). In the posterior the gap gene expression levels are drastically lowered by attraction to an *all*-off attractor—A_4^*. In particular Kni expression is especially low due to trajectories traversing only the low values of Kni as they spiral into A_4^*.

In the wild type system nuclei posterior to 57% EL belonged to the basin of A_4^*. Of those, only the most posterior nuclei (posterior to 65% EL) ended up in a *gt*-on state due to the way that trajectories were attracted to U_4^+. In Kr^- case A_4 alone was responsible for formation of *gt* domain. All nuclei posterior to 37% EL belonged to the basin of A_4^*, and nuclei posterior to 57% EL had initial conditions close enough to the *all*-off attractor to end up in a *gt*-on state by gastrulation (as trajectories spiral into A_4^* along high values on Gt

axis). Thus the anterior shift in the *gt* domain of Kr^- is due to more anterior nuclei being in the basin of A_4^*.

There are many similarities between the bifurcation portraits of Kr^- and the wild type system (see figures 5.2.1 panel (E) and 5.3.2 panel (E)). Within the scope of 35% EL-72% EL the same bifurcations are undergone by both systems. In fact even their equilibrium points are similar in the sense that Kr^- equilibria can be though of as derived from the wild type without major changes (see figure 5.3.1 and table 5.7). However, the final expression pattern of the gap genes of interest are drastically different.

The next chapter will show the importance of the extra saddles that the wild type system has and the Kr^- system lost during mutation of the R^{Kr} parameter. It will demonstrate that those saddles aid in producing higher levels of expression in wild type. Most importantly one of those saddles will be responsible for canalization of the Bcd perturbations, which is visible in the wild type but is completely missing in Kr^-.

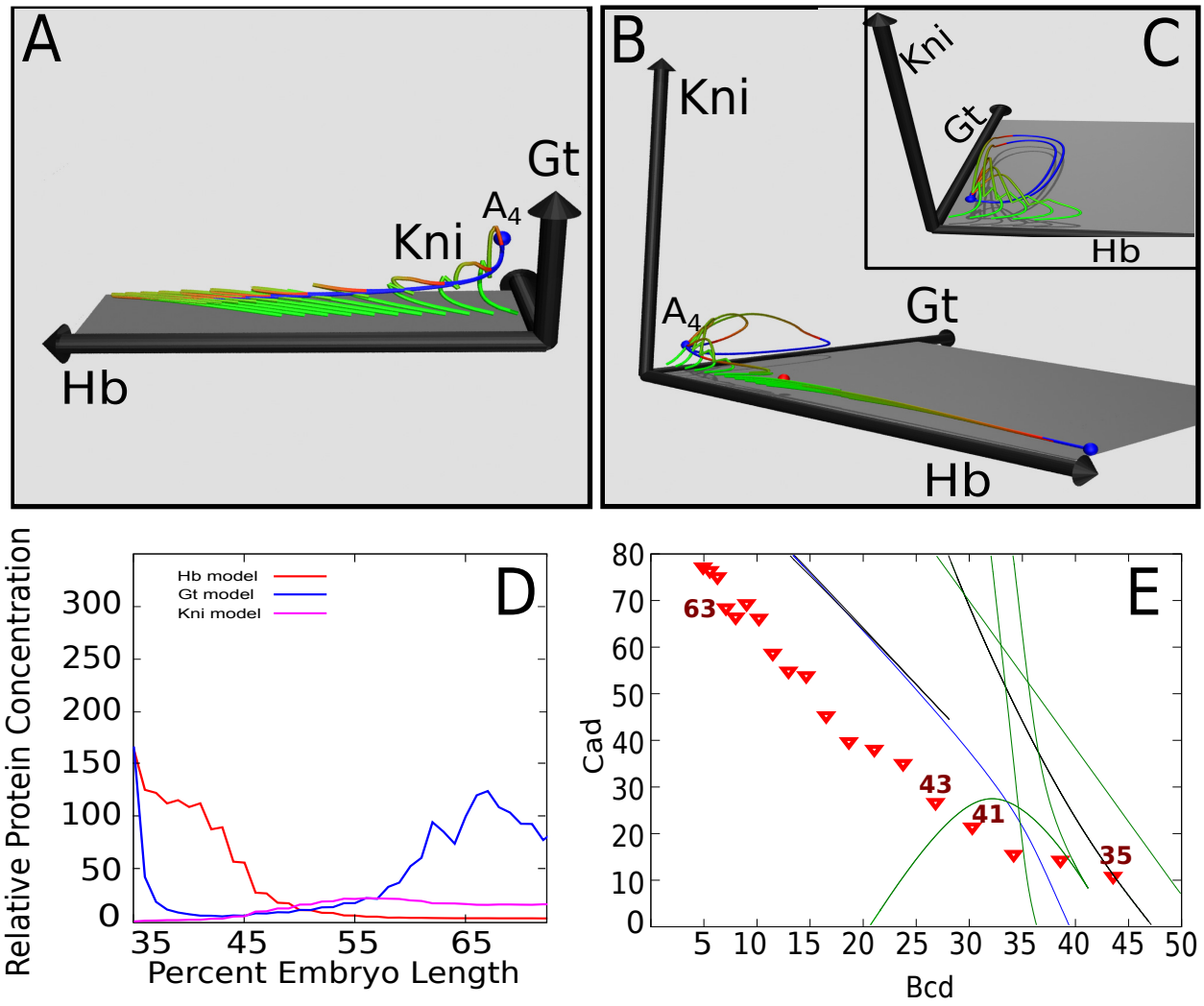


Figure 5.3.2: **Pattern formation for Kr^- from 35% EL to 72% EL.** Red spheres are saddles of type $S^{1,3}$. Brown spheres are saddles of type $S^{2,2}$. Time is represented by the color of the trajectories. Trajectories are green from cycle 13 to the beginning of C14A. Throughout C14A the color changes from green to red. Red corresponds to gastrulation. There is an abrupt transition to blue after gastrulation. Orientation of the axes was chosen to best illustrate the dynamics of each nucleus.
(A) Phase portrait for 39% EL. 10 trajectories are shown with starting points uniformly distributed on the Hb axis between 0-100. Length of Hb axis is 205, length of Gt and Kni axes is 65. **(B)** Phase portrait for 67% EL. 10 trajectories are shown with starting points uniformly distributed on the Hb axis between 0-100. Length of axis is 355 for Hb and Gt, 255 for Kni. **(C)** Dynamics from panel(B) near A_4. Scale of Hb is increased by a factor of 3. 5 trajectories are shown with starting points uniformly distributed on the Hb axis between 0-20. **(D)** Gap gene pattern for Kr^- (without diffusion and without tailless). Red is Hb, blue is Gt and magenta is Kni. **(E)** Parametric portrait of the circuit for Kr^-. Red triangles correspond to nuclei 35-72, with 35% EL on the right bottom corner.

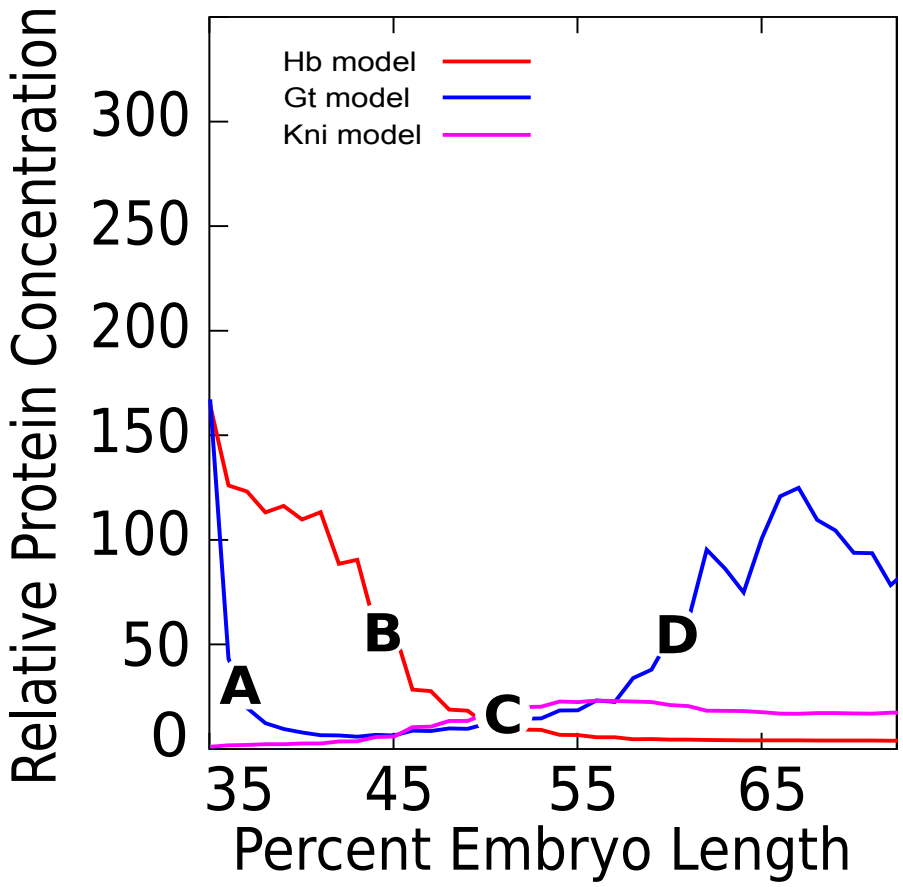


Figure 5.3.3: Mechanisms of border formation for Kr^{-}. Borders are marked by letters **A-D** **(A)** Movement of $S_4^{1,3}$ followed by annihilation of A_1 through saddle-node bifurcation. **(B)** Attraction to A_4. **(C)**-**(D)** Attraction to A_4 via a spiral on Gt-Kni plane, with very low Kni component.

Chapter 6

Mechanism of Decanalization in Kr^- Embryos

It has been noted by Waddington [75] and Rendel [58], among others that the variability between individuals within a population of wild type stock is much less than between those of mutant stock. This difference in variance is especially clear in *D.melanogaster*. In Chapter 2, data showed that Kr^- serves as an example of the above statement. The difference in embryo-to-embryo variability in the gap gene pattern of Kr^- is greater then in wild type [67].

As it was mentioned in Chapter 4, in order to simulate the embryo-to-embryo variability of gap gene expression *in silico* for both wild type and Kr^- we used 88 different Bcd profiles [42]. Fixing all the parameters other than v^{Bcd}, we are able to see what is the gap gene expression under different biologically admissible Bcd inputs. This is done by repeatedly running the diffusionless model without *Tll* using all 88 possible Bcd profiles separately. Figure 6.0.1 demonstrates that there is a characteristic ensemble of gap gene expression for wild type and one for Kr^-. The model predicts canalization in wild type and decanalization of the gap gene expression in mutants.

The most clear example of complete decanalization of Kr^- in comparison to wild type

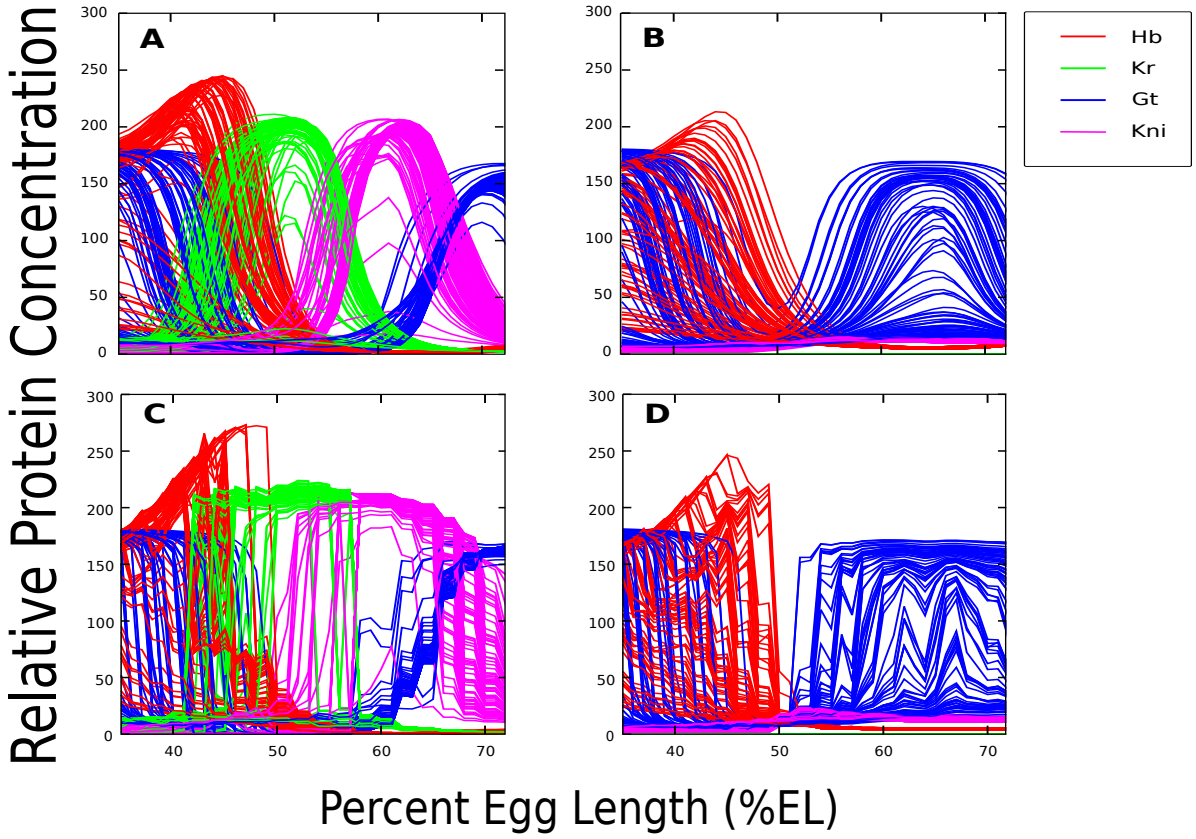


Figure 6.0.1: **Wild type and Kr^- gap gene pattern ensembles under different Bcd profiles in the region 35-72% EL.** The model was run on the range 35-92%EL of which only 35-72% EL is shown. The time class displayed is T6. Hb is red, Kr is green, Gt is blue and Kni is magenta. **(A)** An ensemble of 88 wild type gap gene patterns, using the full model, resulting from the 88 Bcd profiles (each Bcd profile corresponds to a gap gene pattern). **(B)** An ensemble of 88 Kr^- gap gene patterns using the full model. **(C)** An ensemble of 88 wild type gap gene patterns, using a simplified model with no Tll and no diffusion. **(D)** An ensemble of 88 Kr^- gap gene patterns using the simplified model.

is illustrated by variability in the expression of the posterior *gt* domain in mutants. In fact the focus of the analysis in this work will be on the posterior part of the embryo, namely 59-72% EL and more specifically on the *gt* expression. For each nucleus in the posterior of the wild type system, different inputs of Bcd bring about small changes in the relative protein concentration of Gt. Thus for a specific nucleus *gt* is either on or off for all possible 88 Bcd concentrations. Over on the mutant side we have a much more uncertain picture. For every nucleus in the posterior, the *gt* expression is highly dependent on the Bcd input, with the state of *gt* having all the possibilities from completely on to completely off depending on what Bcd concentration is chosen. Figure 6.0.2 demonstrates how extreme values of Bcd chosen from the 88 profiles result in extreme expression in *gt*. Thus we can see that while wild type takes the variability in maternal input and minimizes it at the *gt* expression level, Kr^- exaggerates the same variability at the *gt* level.

In this section we first investigate the mechanism by which wild type canalizes Bcd input. We then show why the same mechanism cannot be extended to a Kr^- circuit and under which conditions it breaks down. The focus will be on the Gt coordinate in phase space (figure 6.0.2) and this in turn will translate to the gap gene expression pattern (figure 6.0.1).

In the analysis done in Chapter 5 we have seen that in the wild type system all the posterior nuclei are attracted to the unstable manifold, with the state of the system by gastrulation dependent on the initial conditions of Hb. Thus it makes sense that changes in Bcd result in similar states of the trajectories by gastrulation. It was also shown in Chapter 5 that wild type and Kr^- possess different mechanisms of patterning the posterior. While A_4 is the attracting state for both Kr^- and wild type trajectories in the posterior, the way the trajectories approach the attractor are different. The wild type system possesses an unstable manifold U_4^+ stemming from $S_1^{1,3}$ and ending up at A_4 that the trajectories get attracted to by gastrulation (see figure 5.2.1 panel (B)). For Kr^- the trajectories approach A_4 directly in a manner that is dependent on the A-P position.

In order to ascertain how the wild type canalizes and Kr^- fails to canalize Bcd pertur-

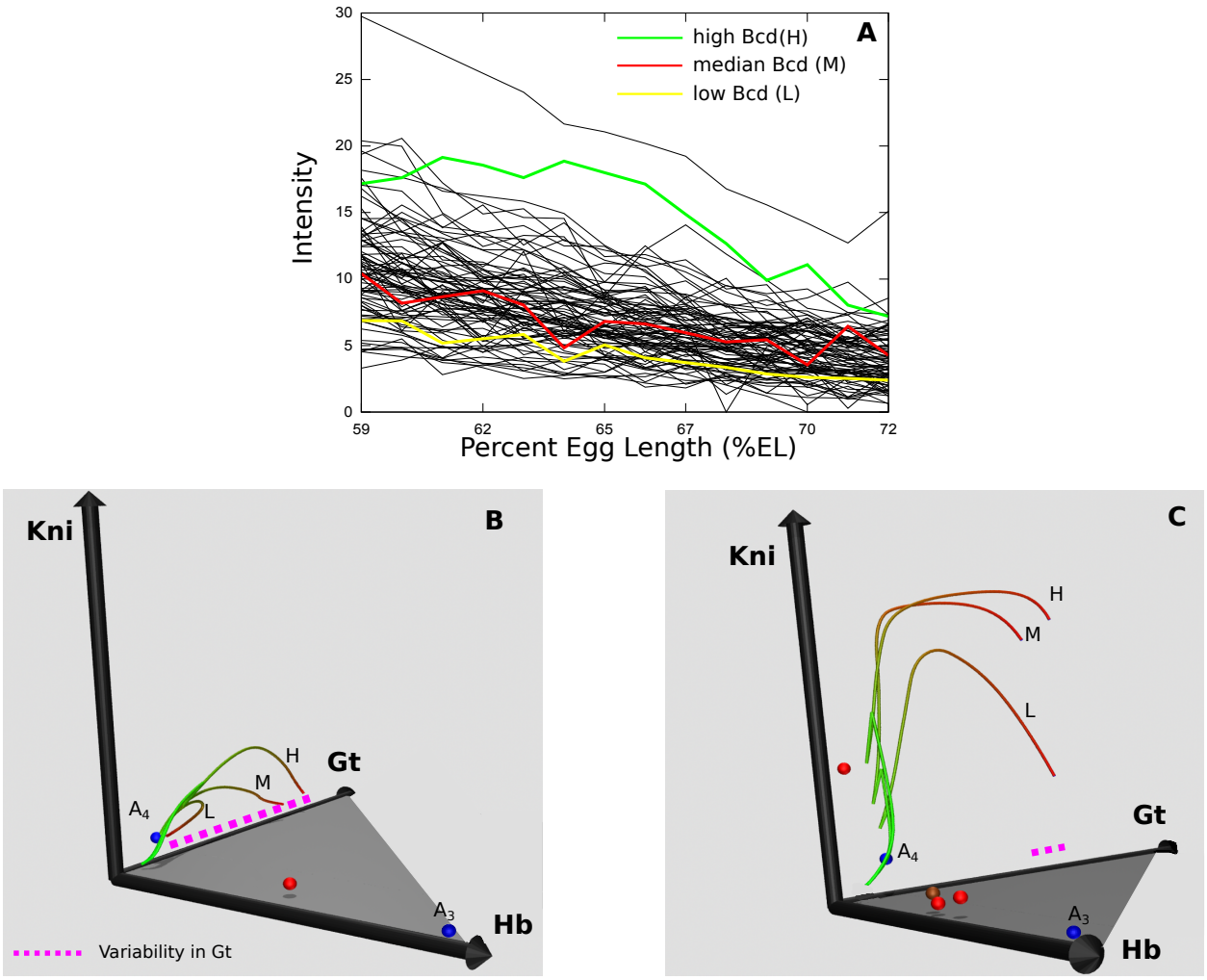


Figure 6.0.2: **Wild type and Kr^- phase space under three different Bcd profiles for nucleus 67. (A)** 88 different Bcd profiles in the region 59-72% EL. Red is the median Bcd profile, green is the representative for a higher than median Bcd profile and yellow is the lower then median Bcd profile. The scale in panel B and C is the same. **(B)** Phase space for Kr^-. For each of the 3 colored Bcd profiles there is a corresponding trajectory. H corresponds to the trajectory with higher than median Bcd input, M is the trajectory with median Bcd input and L is for lower than median Bcd input. The trajectories are stopped at gastrulation. The magenta dotted line represents the range in Gt coordinate of the three trajectories at gastrulation. **(C)** Phase space for wild type.

bations we need to see what happens to the mechanism of patterning in the posterior as the levels of Bcd are changed.

6.1 Dependence of the unstable manifold U_4^+ on Bcd concentration

For wild type U_4^+ dependence on Bcd concentration needs to be examined. U_4^+ depends on the starting and ending point, namely $S_1^{1,3}$ and A_4. The behavior of the trajectories in the neighborhood of U_4^+ is influenced by the different equilibria that are nearby. Firstly, varying the Bcd concentration in the biological range of interest produces no significant local changes to the equilibria present. Secondly, we will use AUTO's continuation package to track the location and changes in type of $S_1^{1,3}$ and A_4. We will continue $S_1^{1,3}$ and A_4 as we vary Bcd in the range of all the possible concentrations found in the 88 profiles. These two equilibria were continued separately from each other.

Considering $S_1^{1,3}$ first, figure 6.1.1 shows that not only does $S_1^{1,3}$ remain of the same type, without undergoing any bifurcations, as Bcd concentration is varied but it also remains in the same state, namely Kr, kni-on, for almost all Bcd values. The location of the saddle-node bifurcation that eventually annihilates the saddle (black arrow in figure 6.1.1) varies slightly between the posterior nuclei with 63% EL being at Bcd concentration of 4.07 (here the Bcd inputs are larger as it is more anterior), 69% EL at 0.15 and the rest of them between Bcd concentrations of 2.62 and 3.29.

Next we consider continuation of A_4. As we can see from figure 6.1.2, the attractor persists only for Bcd concentrations less than 13.49 for nucleus 67. The attractor changes into $S_4^{2,2}$ as a result of a Hopf bifurcation[1] (figure 6.1.2). As soon as A_4 is transformed to a saddle point $S_4^{2,2}$, the unstable manifold switches its end point to the other attractor—A_3

[1] As Bcd concentration increases further $S_4^{2,2}$ in essence becomes $S_7^{2,2}$ through the creation and annihilation of the "transient" saddle $S_t^{2,2}$.

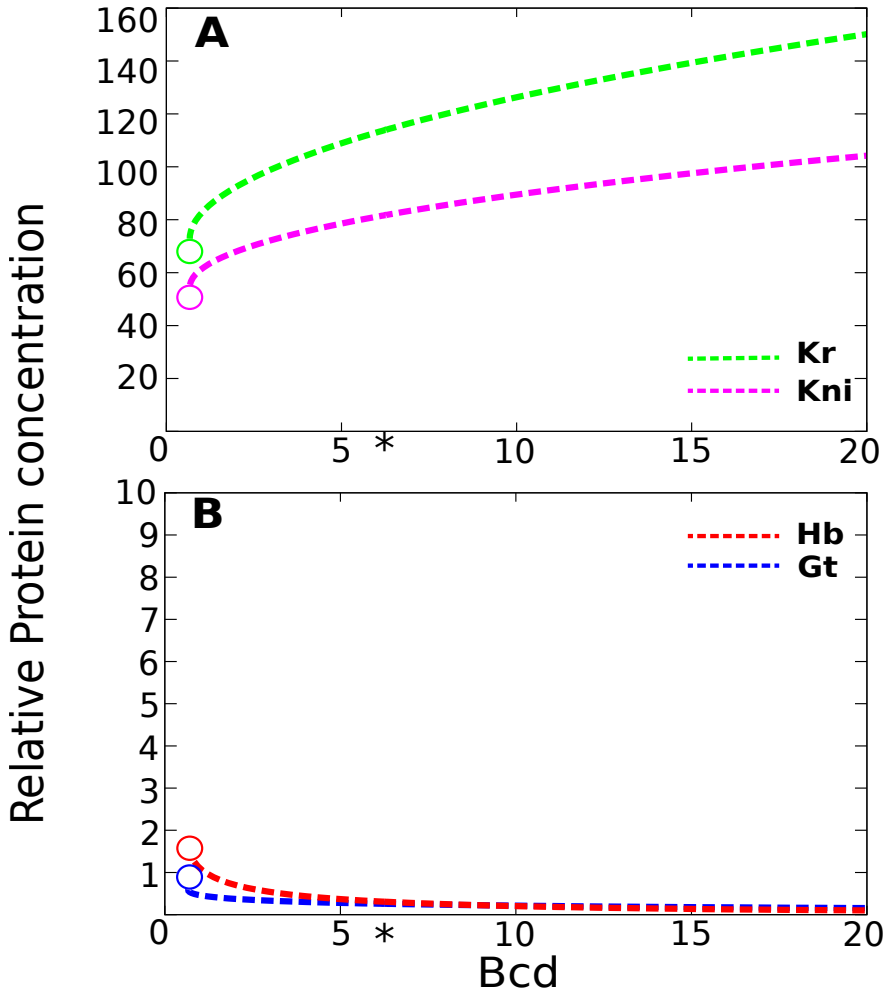


Figure 6.1.1: **Location of $S_1^{1,3}$ under Bcd perturbation for nucleus 67.** **(A)** Kr and Kni coordinates of the saddle. **(B)** Hb and Gt coordinates of the saddle. As Bcd is varied between 0 and 20 (this range includes all the possible concentrations of Bcd from the family of 88 profiles), the location of $S_1^{1,3}$ in four dimensional phase space is tracked using AUTO. Note that $S_1^{1,3}$ is present for all of the possible Bcd concentrations that we are considering. The colored circles point to the Bcd concentration at which the saddle-node bifurcation occurs. The asterisk indicates the concentration of Bcd from the median profile corresponding to 67% EL (this was used as a starting point for continuation by AUTO).

(see figure 6.1.3 panel (B)). Hence for Bcd concentrations below 13.49 we have U_4^+ which ends in A_4 and above 13.49 we have U_3^+ which ends in A_3. A_3 has very low components of Gt, Kr and Kni, just like A_4. The location of U_4^+ and U_3^+ on the Kr-Kni plane remains practically the same.

The presence of $S_1^{1,3}$ and its unstable manifold buffers the maternal perturbations. The Bcd range among the 88 profiles can be split into two parts—below and above 13.49. For Bcd concentration less than 13.49, U_4^+ remains in the same location and for Bcd concentration greater than 13.49 we have U_3^+ which attracts the trajectories to *gt*-on state by gastrulation. U_3^+ and U_4^+ persist due to the existence of $S_1^{1,3}$. As this saddle remains invariant in location and type under the perturbations in Bcd, the unstable manifolds are not affected by different levels of Bcd and the wild type system canalizes *gt* expression in the posterior.

Another finding from the wild type system must be noted, as it will be of importance in the Kr^- case. The Gt component of A_4 increases to high level as Bcd increases, while the Hb, Kr and Kni components remain mostly the same under Bcd changes (see figure 6.1.2). With a decrease in Bcd concentration A_4 becomes a stronger attractor as the absolute value of the eigenvalues corresponding to the attractor increases (see table 6.1 last column). The dependence of position and strength of A_4 on Bcd is of no consequence in the wild type case, as all these changes do not influence the location of the unstable manifold nor its ability to attract trajectories to *gt*-on state by T6, but it will be crucial for Kr^-.

6.2 Dependence of A_4 in Kr^- on Bcd concentration

For Kr^- the dependence of A_4 on Bcd parallels that of wild type. First, the Gt component of the attractor rises with increasing Bcd concentration, as shown in figure 6.2.1. As is the case with wild type, the absolute value of the real part of eigenvalues of A_4 tend to increase as Bcd concentration falls (see table 6.1 first column).

The difference between the wild type and Kr^- continuations of A_4 with respect to Bcd

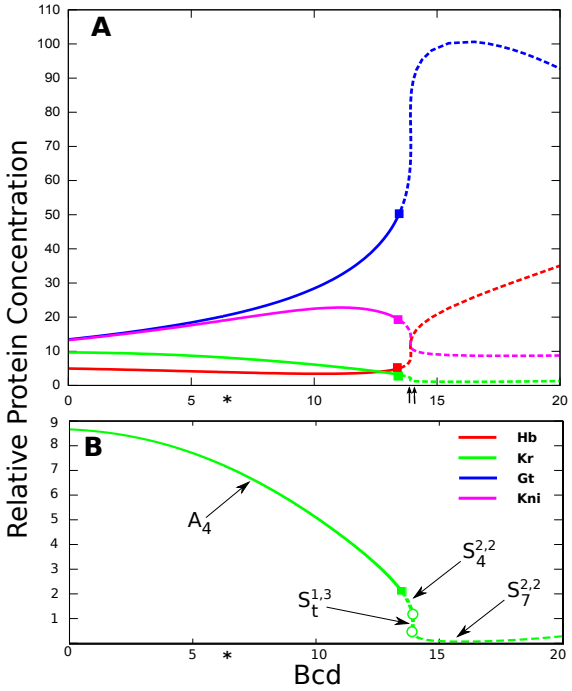


Figure 6.1.2: **Location of A_4 under Bcd perturbation for 67% EL in wild type.** Starting with the median Bcd concentration for nucleus 67 (*) we continued A_4 in the direction of increasing and then in the direction of decreasing Bcd. Solid lines signify that the equilibrium is an attractor; dotted lines mean that it is of a saddle type. Red is the Hb coordinate, green is Kr, blue is Gt and magenta is Kni. The colored squares are the location of the Hopf bifurcation on a coordinate of corresponding color. **(A)** All four coordinates are represented with their corresponding colors. For all Bcd concentrations less than 13.49 A_4 remains in the state *all*-off and of type attractor. At a Bcd concentration of 13.49 there is a Hopf bifurcation turning the attractor into an $S_4^{2,2}$ saddle. At a Bcd concentration of 13.92 there is a saddle-node bifurcation creating $S_t^{1,3}$ and $S_7^{2,2}$ and at 13.93 another saddle-node bifurcation annihilates $S_4^{2,2}$ along with $S_t^{1,3}$. $S_7^{2,2}$ persists for all higher Bcd concentrations. $S_7^{2,2}$ is *gt*-on. Two arrows point to the saddle-node bifurcations. **(B)** Only the Kr component is shown. Circle shows the position of the bifurcation on the Kr axis.

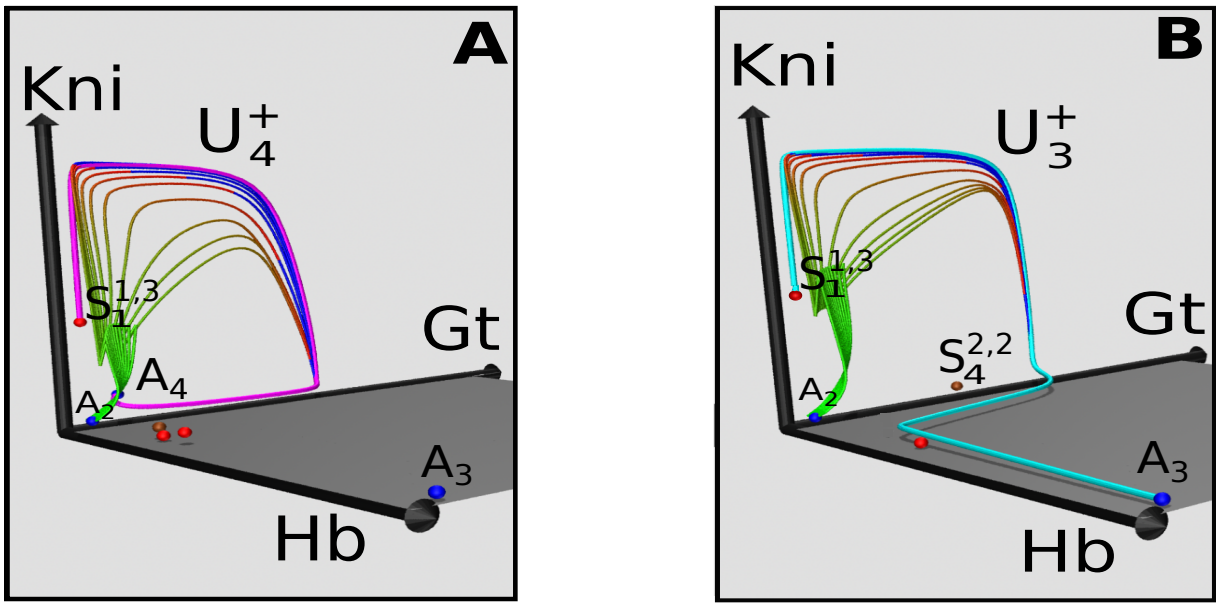


Figure 6.1.3: **Comparison of U^+ before and after Hopf bifurcation.** Phase portraits at 67% EL with different Bcd concentrations. 10 trajectories are shown with starting points equally distributed on the Hb axis between 0 and 10. **(A)** U_4^+ before Hopf bifurcation. Bcd concentration is 6.269 (from median Bcd profile). **(B)** U_3^+ after the Hopf bifurcation. Bcd concentration is 15. As A_4 was turned into $S_4^{2,2}$, U_3^+ connects $S_1^{1,3}$ and A_3.

is that for Kr^- the Hopf bifurcation occurs at larger concentrations of Bcd. Therefore for sufficiently large concentrations of Bcd the A_4 of the wild type system is a saddle, while in the Kr^- system it is still an attractor. Furthermore A_4 for the Kr^- has two complex eigenvalues, whereas A_4 for wild type has four complex eigenvalues (see table 6.1 middle row first and last columns). Furthermore, as Bcd concentration decreases the complex part of the conjugate eigenvalues gets smaller and for sufficiently low Bcd A_4 in the Kr^- system has four eigenvalues real. On the other hand all of the wild type attractor's eigenvalues maintain nonzero complex part throughout the range of interest for Bcd.

These findings about dependence of A_4 on Bcd are very important in the Kr^- case as there are no other equilibria in the vicinity of the initial conditions to govern the dynamics. Trajectories are dependent on the location and type of eigenvalues of A_4. High Bcd concentration results in a trajectory that spirals along large values on the Gt axis without reaching the attractor by gastrulation(see figure 6.0.2 (H)). This is a consequence of the large Gt

Bcd concentration	R^{Kr}=0	R^{Kr}=4	R^{Kr}=7.5	R^{Kr}=10	R^{Kr}=15
13.11	-0.059	-0.057-0.006i	-0.057-0.008i	-0.057-0.008i	-0.056-0.008i
	-0.059	-0.057+0.006i	-0.057+0.008i	-0.057+0.008i	-0.056+0.008i
	-0.025-0.022i	-0.020-0.026i	-0.015-0.028i	-0.012-0.028i	-0.005-0.027i
	-0.025+0.022i	-0.020+0.026i	-0.015+0.028i	-0.012-0.028i	-0.005+0.027i
10.02	-0.063	-0.057-0.009i	-0.055-0.012i	-0.052-0.013i	-0.052-0.014i
	-0.059	-0.057+0.009i	-0.055+0.012i	-0.052+0.013i	-0.052+0.014i
	-0.036-0.013i	-0.034-0.019i	-0.032-0.024i	-0.030-0.026i	-0.026-0.030i
	-0.036+0.013i	-0.034+0.019i	-0.032+0.024i	-0.030+0.026i	-0.026+0.030i
6.27	-0.068	-0.059-0.011i	-0.056-0.015i	-0.054-0.016i	-0.051-0.018i
	-0.059	-0.059-0.011i	-0.056+0.015i	-0.054+0.016i	-0.051+0.018i
	-0.043-0.005i	-0.042-0.012i	-0.041-0.017i	-0.040-0.019i	-0.037-0.024i
	-0.043+0.005i	-0.042+0.012i	-0.041+0.017i	-0.040-0.019i	-0.037+0.024i
4.02	-0.070	-0.061-0.011i	-0.058-0.016i	-0.056-0.018i	-0.052-0.019i
	-0.059	-0.061-0.011i	-0.058+0.016i	-0.056+0.018i	-0.052+0.019i
	-0.049	-0.043-0.009i	-0.043-0.013i	-0.041-0.016i	-0.041-0.021i
	-0.041	-0.043+0.009i	-0.043+0.013i	-0.041+0.016i	-0.041+0.021i
2.26	-0.071	-0.062-0.011i	-0.059-0.015i	-0.057-0.017i	-0.053-0.020i
	-0.059	-0.062-0.011i	-0.059+0.015i	-0.057+0.017i	-0.053+0.020i
	-0.052	-0.045-0.007i	-0.044-0.011i	-0.043-0.014i	-0.040-0.018i
	-0.040	-0.045+0.007i	-0.044+0.011i	-0.043+0.014i	-0.040-0.018i

Table 6.1: Eigenvalues of A_4 for selected values of R^{Kr} and Bcd for nucleus 67. Bcd concentrations are taken from 88 Bcd profiles.

component of A_4 combined with the smaller absolute value of the negative real part of the eigenvalues as well as two of the eigenvalues being complex conjugates. Low Bcd concentration results in the trajectories that go straight to A_4 (see figure 6.0.2(L)). This behavior is due to A_4 having a small Gt component and all of its eigenvalues having a larger absolute value of the negative real part as well as all eigenvalues being real. Thus the Kr^- system amplifies the variability in the Bcd input by admitting all the possible responses from high Gt to very low Gt concentration for a relatively small range of Bcd concentrations.

So far it has been shown that in the wild type system, $S_1^{1,3}$ and its associated unstable manifold are responsible for patterning the posterior of the embryo. Moreover this system canalizes Bcd perturbations at the level of posterior *gt* expression. On the other hand in Kr^- posterior pattern formation is governed by the system's dependence on A_4 and there is no mechanism to buffer the effect that variation in Bcd concentration has on A_4. As a result Kr^- has a decanalized posterior *gt* expression. What happens to unstable manifold in between wild type and mutants?

6.3 Loss of canalization by annihilation of $S_1^{1,3}$

In the previous section we have seen that the unstable manifold is crucial to canalization. In this section we investigate why Kr^- does not allow for the unstable manifold that wild type possesses and what happens to $S_1^{1,3}$ as we transition from wild type to the mutant system. This transition is undertaken by continuously changing R^{Kr}, the parameter for the maximum synthesis rate of Kr. Through this change we can smoothly turn down the function of Kr to get to Kr^-. Thus we have a way to mutate our system all the way from wild type (R^{Kr}=15) to Kr^- (R^{Kr}=0).

All of the bifurcations that the system undergoes when continuing equilibria from wild type to Kr^- were briefly mentioned in Chapter 5, but the discussion below will go into more detail with respect to two particular equilibria and one specific bifurcation that is important

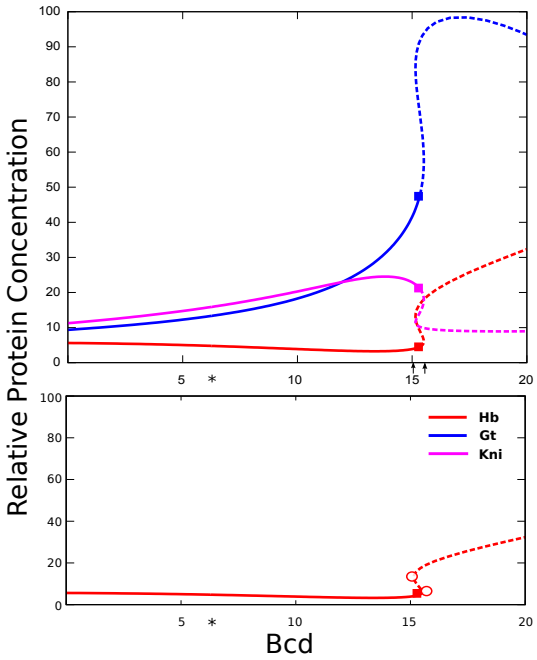


Figure 6.2.1: **Location of A_4 under Bcd perturbation for 67% EL in Kr^-.** Starting with the median Bcd concentration for nucleus 67 (*) we continued A_4 in the direction of increasing and then in the direction of decreasing Bcd. Solid lines signify that the equilibrium is an attractor; dotted lines mean that it is of a saddle type. Red is the Hb coordinate, blue is Gt and magenta is Kni. The colored squares are the location of the Hopf bifurcation on a coordinate of corresponding color. Arrows point to the saddle-node bifurcations. **(A)** All three coordinates are represented with their corresponding colors. For all Bcd concentrations less than 15.34 A_4 remains in the state *all*-off and of type attractor. At Bcd 15.34 there is a Hopf bifurcation turning the attractor into an $S_4^{2,2}$ saddle. At Bcd concentration of 15.15 there is a saddle-node bifurcation creating $S_t^{1,3}$ and $S_7^{2,2}$ and at 15.51 another saddle-node bifurcation annihilates $S_4^{2,2}$ along with $S_t^{1,3}$. $S_7^{2,2}$ persists for all higher Bcd concentrations. $S_7^{2,2}$ is *gt*-on. **(B)** Only the Hb component is shown. Circles shows the position of the bifurcations on the Hb axis.

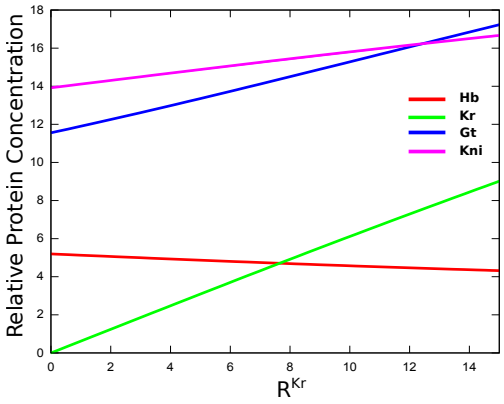


Figure 6.3.1: **Continuation of A_4 with respect to R^{Kr} at 67% EL.** This figure shows how A_4 responds to different values of R^{Kr}. Not only does A_4 persists throughout all the possible values of R^{Kr} from wild type to mutant, but its location in Hb-Kr-Gt space remains practically invariant.

in order to understand decanalization.

In the posterior region of the *gt* domain, two equilibria play a major part in pattern formation and canalization in wild type: A_4 and $S_1^{1,3}$. Tracking these two equilibria while turning R^{Kr} down, we note that A_4 remains in the same location in the Hb-Gt-Kni space (see figure 6.3.1). However, the eigenvalues corresponding to the extreme values of R^{Kr} are quite different (see table 6.1 first and last columns). Whereas the attractor in the wild type system has all complex eigenvalues, the attractor for the Kr^- system has two real and two complex ones. When we lower R^{Kr} from 15 to 0, the complex part of one of the two pairs of conjugate eigenvalues becomes smaller and by R^{Kr}=0 it goes to zero, hence the two eigenvalues become real.

The saddle $S_1^{1,3}$ starts at high Kr and Kni concentrations when R^{Kr}=15, and as R^{Kr} is lowered these two concentrations decrease. There is a saddle-node bifurcation that annihilates the saddle before R^{Kr} gets to 0 (see figure 6.3.2). Therefore this saddle has no analog for Kr^-. This bifurcation is of major importance as the saddle that is destroyed is responsible for canalization in the wild type. Without this saddle the system is not able to buffer the perturbations in Bcd input that result in changes to A_4. Low Bcd input creates strong *all*-off attractor, causing trajectories to go straight to A_4, resulting in low Gt by gas-

trulation. High Bcd input creates weak *all*-off attractor, leading trajectories along high Gt en-route to A_4, which is reached long past gastrulation, resulting in high Gt by gastrulation. Continuation of $S_1^{1,3}$ was calculated for each nucleus separately, hence the location of the saddle-node bifurcation is slightly different for each nucleus. It is largest for nuclei 63-65 at R^{Kr} between 6.57 and 6.58 and for the other posterior nuclei it is between 6.24 and 6.32. The equilibrium point that is annihilated along with the above saddle started off as $S_6^{2,2}$. $S_6^{2,2}$ underwent a Hopf bifurcation which converted it to an attractor A_t. This attractor annihilates $S_1^{1,3}$ and for all the values of R^{Kr} below 6.24 there is no $S_1^{1,3}$ for any of the nuclei.

If we look at the gap gene pattern for each R^{Kr} (see figure 6.3.3) we see a sharp change in canalization of the posterior *gt* domain as soon as the saddle-node bifurcation annihilates $S_1^{1,3}$. Before the saddle-node bifurcation (see figure 6.3.3 (A)-(D)) posterior *gt* expression was canalized. Once the saddle disappears and with it the associated unstable manifold, expression becomes completely decanalized taking all possible values from high Gt concentration to low Gt concentration (see figure 6.3.3 (E)-(F)).

Three more points need to be made. Firstly, canalization is more pronounced for more posterior nuclei. Recalling U_4^+ from figure 5.2.1 panel (B) we see that the slope with respect to Gt is negligible for more posterior nuclei, hence different trajectories project onto the same Gt coordinates by gastrulation. On the other hand, the trajectories that start with initial conditions corresponding to more anterior nuclei end up at the part of U_4^+ that has a large slope with respect to Gt. For these nuclei, even though the Bcd fluctuations are buffered by the manifold, the differences in Gt coordinates by gastrulation are larger than for their more posterior counterparts.

Secondly, *kni* expression becomes smaller with decreasing R^{Kr} values. The reason for this is the change in $S_1^{1,3}$ location as R^{Kr} changes. For the wild type $S_1^{1,3}$ has high Kr and Kni values but as R^{Kr} is lowered both Kr and the Kni coordinates decrease in value (see figure 6.3.2). This reduces the distance between $S_1^{1,3}$ and A_4, the start and end point for unstable manifold, respectively. This produces an unstable manifold that spans lower values

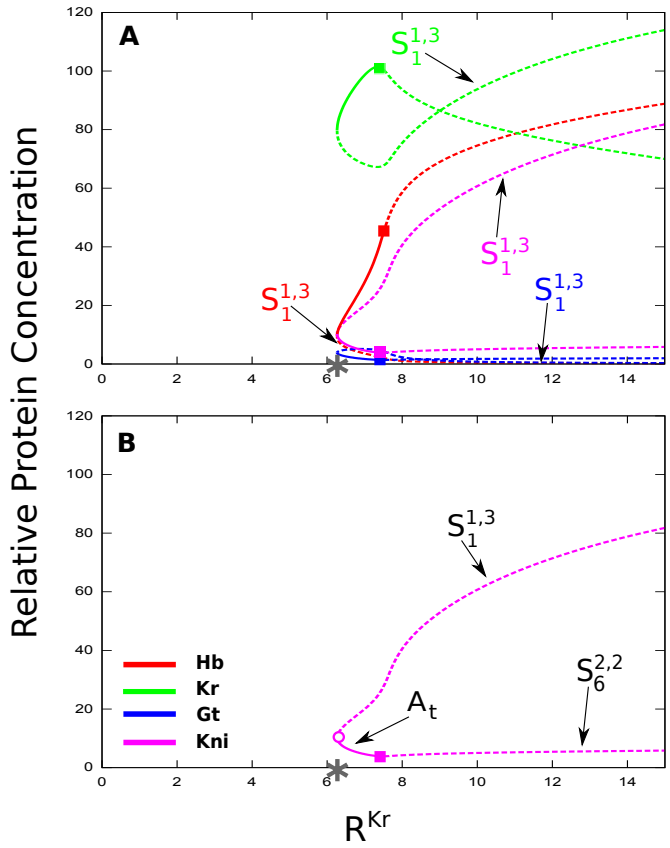


Figure 6.3.2: **Continuation of $S^{1,3}$ with respect to R^{Kr} at 67% EL.** By continuously changing R^{Kr} (maximum synthesis rate of Kr) in our model, we smoothly turn down the function of Kr all the way down to null. Continuing $S_1^{1,3}$, which is associated with the unstable manifold using AUTO allows us to find a point at which unstable manifold no longer exists. The starting point of the continuation is wild type (R^{Kr}=15). Kr^- corresponds to R^{Kr}=0. The colored squares are the location of the Hopf bifurcation on a coordinate of corresponding color. This bifurcation converts $S_6^{2,2}$ into A_t as R^{Kr} is turned down. Asterisk indicates the position of the saddle-node bifurcation that annihilates $S_1^{1,3}$ at R^{Kr}=6.26. **(A)**All four coordinates are represented with their corresponding colors. Arrows point to the branch to which $S_1^{1,3}$ belongs to for each of the four coordinates. **(B)** Kni coordinate is illustrated. Circle shows the position of the saddle-node bifurcation on the Kni axis.

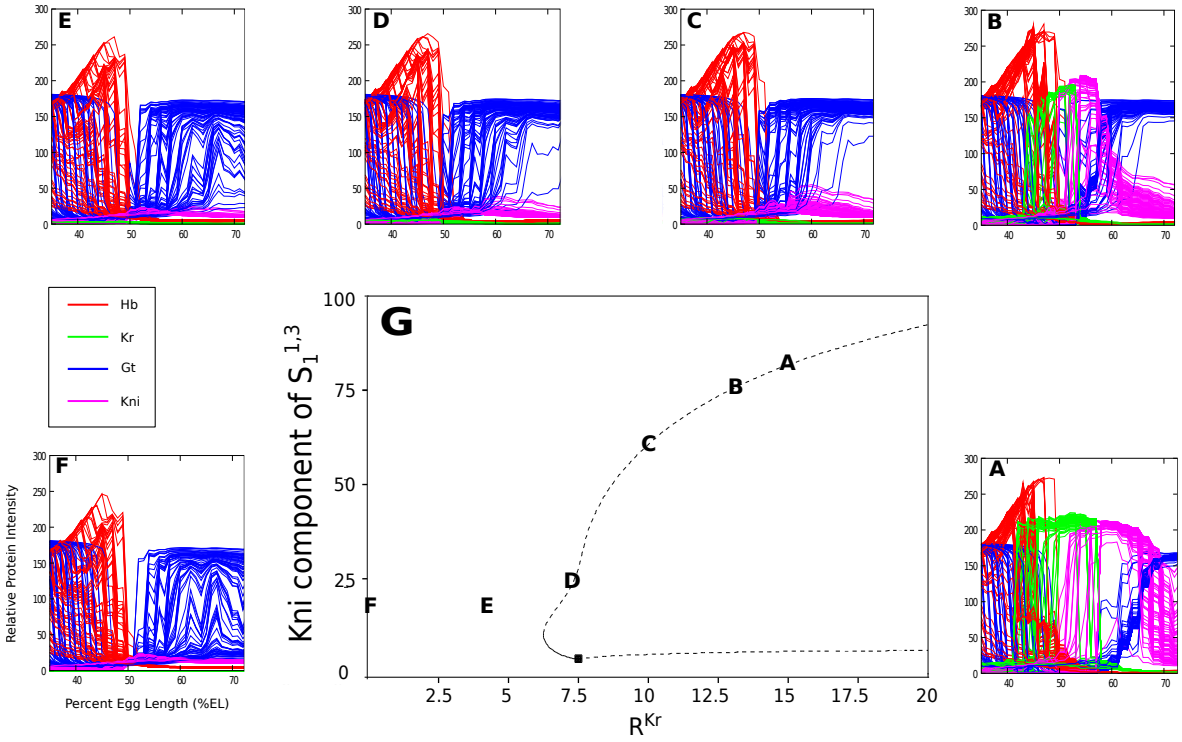


Figure 6.3.3: **Continuation of $S_1^{1,3}$ with respect to R^{Kr} corresponding to 67% EL and characteristic ensembles of gap gene expression corresponding to different values of R^{Kr}.** For each R^{Kr} between wild type (R^{Kr}=15) and Kr^-(R^{Kr}=0) we can associate an ensemble of 88 gap gene patterns, corresponding to different Bcd inputs. These ensembles were generated from the model with no diffusion and no Tll. The ensembles A-F are grouped around panel G, and points in G corresponding to the ensembles in A-F are labeled with the corresponding letter. The values are **(A)** R^{Kr}=15, which is the wild type, **(B)** R^{Kr}=13, **(C)** R^{Kr}=10, **(D)** R^{Kr}=7.5, **(E)** R^{Kr}=4, **(F)** R^{Kr}=0, which is Kr^-. **(G)** Panel (B) from figure 6.3.2.

on Kni axis.

Lastly, different R^{Kr} result in different methods of patterning the posterior of the embryo. For high R^{Kr}, the pattern is a result of the trajectories choosing a state on unstable manifold based on the initial conditions. As R^{Kr} is lowered the patterning is more and more influenced by the effect Bcd input has on the eigenvalues of A_4.

It has been shown in this chapter that while high biologically relevant Bcd concentration always result in high *gt* expression, lower Bcd concentrations need to be buffered to produce relatively invariant posterior *gt* pattern by the U_3^+ or U_4^+. As the function of Kr is turned down Kr and Kni coordinates of $S_1^{1,3}$ are lowered causing *kni* domain to reduce significantly. As R^{Kr} is lowered further, $S_1^{1,3}$ is annihilated via a saddle-node bifurcation. With the loss of $S_1^{1,3}$, the low values of Bcd result in drastically different *gt* expression and therefore a complete loss of canalization is triggered by the abrupt absence of unstable manifold.

Chapter 7

Conclusions

Dynamical analysis has been employed in this dissertation in order to explain the distinct gap gene pattern in the Kr^- embryos. Of particular interest were the lowered levels of *hunchback*, *giant* and *knirps*, in light of the fact that it has been noted that *Krüppel* is a repressor of all of them [31, 30, 20, 40]. Furthermore, bifurcation analysis highlighted the mechanism and the structure responsible for canalization of posterior *gt* in wild type. This analysis has displayed what happens to the mechanism as the function of Kr is turned down. And finally it demonstrated why in the Kr^- system there is a complete decanalization of *gt* pattern.

We have employed a model that could be used for both wild type and well as null mutant studies simultaneously. This model accurately predicted the gap gene expression for both wild type and Kr null mutant phenotype in the range 35% EL-72% EL in the embryo. In addition the model predicted canalization of gap gene pattern for wild and decanalization in Kr^-. The latter was a property of the model, it was not built into it.

The mathematical modeling approach employed in this work was of an inverse optimization problem. We did not make any assumptions about the regulatory mechanisms instead we let the large quantitative data set of high temporal resolution determine the values for the parameters in the model. To this end the parameters were optimized such that they best

fit the data for both wild type and Kr^- gap gene expressions.

The model was simplified to allow us to use the techniques of dynamical analysis and to have the ability to perform numerical calculations in reasonable time. Although Tll and diffusion were set to 0 as a result of the simplifications, the patterns were reproduced correctly for the region between 35% EL and 72% EL. Furthermore canalization was present in wild type and absent in Kr^- (see figure 5.1.2 and figure 6.0.1).

The analysis in this work was focused more on the trajectories in the phase space and describing the system in a global qualitative way as well as identifying stable sets rather than approaching the problem from a completely quantitative direction and concentrating on the point attractors. This not only simplified our calculations but also allowed us to view the system in its entirety and see what states could be achieved and what conditions were needed to achieve them. Furthermore structures responsible for patterning and canalizing were readily visible and their evolution with respect to different inputs could be tracked.

7.1 Dynamics of the phase space

7.1.1 Wild type system

In Chapter 5 we found that there are two mechanisms responsible for patterning gap gene domains in wild type—as previously reported ([44, 43]). The mechanism for the anterior was the attraction to point attractors and in the posterior attraction was to the one-dimensional stable set. For the mutants, only one mechanism was present—attraction to the point attractor, in particular to A_4.

Attractors are an important feature of the phase space as they influence the nearby trajectories, but tracking only the movement of attractors or bifurcations involving attractors would not provide us with enough detail to understand patterning or canalization. Saddles as well as their associated unstable manifolds play a very important role. With regards to border formation for the wild type gap gene expression only one of them was formed as a

direct result of annihilation of an attractor—posterior part of anterior gt. The other anterior borders were formed by switching basins between two attractors as facilitated by a change in the location of a saddle in Hb-Kr-Gt-Kni plane. For example the posterior border of anterior hb and the anterior border of Kr was formed by the location of $S_2^{1,3}$ moving in the direction of higher Kr and lower Hb, as more posterior nuclei were considered. This combined with reduced initial conditions on Hb caused the trajectories to switch basins from the A_3 attractor to A_2 .

Another example was a saddle, $S_5^{2,2}$, that was present in anterior nuclei (35% EL-39% EL). That saddle turned out to be the most important feature of the posterior dynamics for Kr^- as well as for wild type. By undergoing a Hopf bifurcation, $S_5^{2,2}$ was transformed into A_4 which was the attractor for the trajectories that corresponded to posterior part of the embryo.

Not all the bifurcations were significant at the point that they occurred, but they had influence over the dynamics of the overall phase space. For example, there was a saddle-node bifurcation between nucleus 35 and nucleus 37 that created $S_5^{2,2}$ and $S_4^{1,3}$. Seemingly this did not affect the trajectories as they ended up at A_1 for 37% EL, and yet afterwards $S_4^{1,3}$ collided with A_1, thereby annihilating the relevant attracting state via the saddle-node bifurcation and forming the posterior border of anterior gt.

It should be noted that in the previous model of gap gene system [43] there were three mechanisms of boundary formation: movement of attractor, initial conditions switching from the basin of one attractor to another and initial conditions choosing different states on the attracting manifold. In this model we have seen only the last two mechanisms, though we have not explored the boundary formation for every individual Bcd profile from the 88 as it was done in the previous study [16].

7.1.2 Kr^- system

Kr^- system has a reduction in the number of attractors and saddles in its dynamics (see table 5.7). The exclusion of Kr from the interactions limited the number of equilibria. For all nuclei, with 35% EL and 39% EL being the exception, each of the equilibria for the Kr^- system had an analog from the wild type system. As we transitioned from wild type to Kr^- the equilibria that had a low Kr component for wild type translated into having no Kr component in Kr^- but with no drastic change in their location in Hb-Gt-Kni space. The equilibria with high Kr coordinate did not have an analog in Kr^- system, they were annihilated as R^{Kr} parameter was lowered (see table 5.7). This makes sense as intermediary systems as well as Kr^- do not have as much or any need for equilibria with higher Kr component as Kr axis loses its importance and dynamics transition into 3-dimensional phase space.

Whereas in the wild type system posterior patterning occurred by attraction to an unstable manifold, which stemmed from $S_1^{1,3}$ and ended up at an A_4, Kr^- had a different means of pattern formation. $S_1^{1,3}$ was one of the saddles that was annihilated in the continuation of equilibria from wild type to null mutant (see table 5.7), as the Kr component of this saddle was high. Hence in Kr^- we only had A_4. Therefore the trajectories were completely dependent on this attractor and patterns for all the nuclei after 37 were formed based on dependence of A_4 on Bcd as well as initial conditions on Hb.

The complexity of the eigenvalues of A_4 clarified the phenomenon that for Kr^- the gap gene expression reached maximum expression earlier than wild type and afterwords went down. This was a direct result of the fact that the trajectories approached A_4 via a spiral due to the attractor having complex eigenvalues. This was true for both wild type and mutant, but in wild type the presence of unstable manifold allowed the trajectories to take a longer route and reach maximum expression later in cycle 14.

7.2 Canalization

The wild type system is structurally stable, there are several bifurcations in the anterior part but only one that explicitly affects local dynamics of interest. This structural stability is not enough to ensure canalization as Kr^- system is structurally stable in the same way as the wild type system is. As the asymptotic state and the state of the system at gastrulation are not always the same, especially in the posterior part of the embryo, we need to examine the question of canalization in more detail. When Bcd perturbations are introduced in the posterior what changes the dynamics of relevant trajectories and the gastrulation state of the system is not a change in topology but a quantitative change in the eigenvalues of A_4. The real and complex part of the eigenvalues of A_4 are dependent on Bcd concentration (see table 6.1). In Kr^- case low Bcd concentration result in A_4 having all real eigenvalues, whereas in wild type eigenvalues of A_4 remain complex. The changes in eigenvalues dictate the way trajectories approach the equilibrium. Without any other attracting structures nearby, trajectories become extremely dependent on A_4. However, in wild type $S_1^{1,3}$ and its associated unstable manifold mitigate the effect changes in the eigenvalues of A_4 have on the trajectories.

It was shown in Chapter 6 that the unstable manifold (U_3^+ or U_4^+) is the reason for canalization in wild type with respect to Bcd variability. The unstable manifold is formed exclusively in wild type system. The only difference between the wild type and Kr^- system is lack of Kr in the latter. The formation of the unstable manifold is associated with the network of genes balancing each other out, and if one of the genes is missing the manifold is not able to form. This manifold facilitates the stability of the trajectories. The saddle associated with the unstable manifold is an extremely important feature for canalization. Support for the importance of $S_1^{1,3}$ for canalization is afforded by comparison with hypomorphic and null Kr mutants in which the saddle is not present. These systems are severely decanalized (figure 6.3.3).

Going back to the analogy of Waddington's epigenetic landscape, we can associate un-

stable manifold with some feature of the landscape (i.e. valley) that guides the flow to a particular fate, in our case a *gt*-on state. As the epigenetic landscape is shaped by the organism's genotype and the networks that the genes form, the Kr null mutant's network would be missing all the connections associated with Kr. These missing connections are responsible for a collapse of the structure that is associated with the unstable manifold. As a result the flow is unhindered in its ability to explore all the different possible fates (all the possible states from *gt*-off to *gt*-on) if it is perturbed, as the landscape is flat.

7.3 Other models

Another model of both wild type and Kr^- gap gene pattern was implemented by Kozlov et. al. [29], where he successfully modeled both wild type and Kr^- gap gene expressions in the posterior part of the embryo. Reasoning that the model was oversimplified as it had only one parameter representing the effects of a regulator on its target, Kozlov considered only the posterior part of the embryo (47% EL-92% EL), where the gap genes *hb*, Kr, *gt* and *kni* are expressed in one domain, thus justifying the use of a single parameter. Kozlov used model 1 (equation 3.1.2); whereas model 1 cannot be used to reproduce both wild type and Kr^- patterns correctly for the range considered in this work (35% EL-92% EL), it can be used to model the reduced range between 47% EL and 92% EL.

Kozlov postulated that the decreased gap gene expression as well as the large anterior shift of posterior *gt* expression were due to two events. Firstly, there is a decrease in the strength of Cad activation in Kr^-, which also leads to less auto-activation of all the gap genes (in particular this is the reason for the reduction in posterior *gt* expression). Secondly, as Kr is not longer part of the network in the mutants, new interactions, particularly those of *hb* and *gt*, become important. Anterior *gt* domain, which is formed by the repression from Kr and *hb* spreads anteriorly due to missing Kr expression. This spread of *gt* domain reduced *kni* expression by moving into the region where *kni* was, and it lowers *hb* expression

as *gt* is an activator of *hb* in this model.

While this model [29] fits both wild type and Kr^- gap gene pattern well in the posterior, the model doesn't provide any fits of the anterior pattern. Furthermore, it hasn't been checked for robustness in the face of fluctuations as per [44, 43] or this work. The model presented in this dissertation fits the data for wild type as well as Kr mutants expression comparable to that of Kozlov. Our model predicts broad *gt* domain in Kr^- more accurately at earlier times than [29], whereas [29] predicts the dynamic anterior shift in *gt* better. Both models miss the slight anterior shift of posterior *hb* in Kr^- in later time class.

Another approach to modeling pattern formation is logical modeling of regulatory gene networks [60]. This approach was also used to predict single and multiple mutant phenotype but provided only qualitative results. Logical variables represented the product of each gene in the network. Some variables had up to 4 levels to represent the situations where different levels of concentration were needed to regulate different products. Regulatory interactions were deduced from different mutants studies from the literature. Generalized logical equations were formulated based on the regulatory interactions, and a state table was constructed based on the generalized logical equations.

This approach does not provide high temporal resolution but rather divides the embryo into four areas. While the model predicts several mutant phenotypes without prior knowledge of mutant expression, the predictions are done with a very rough resolution and the level of gap gene expression is not quantified, thus the model fails to capture the dynamics of the mutants. For example the fact that the expression levels reach their maximum in Kr^- earlier than in wild type or that *gt* has a significantly reduced expression in Kr^- is missed in the logical analysis. This approach would not be suitable to study canalization.

Model 1 (3.1.2) with the same Bcd profiles as used in this thesis was employed by Gursky [16] for an in-depth study of canalization of the *hb* border. Gursky found that *hb* border formation was due to transition of the initial Hb profile from the basin of attractor that was *hb*-on to the basin of another attractor that was in *hb*-off state (or a *hb*-off position

on the unstable manifold). Thus the border formed when Hb profile crossed the boundary between the two basins. Canalization of the Hb border came about because of the geometrical fact that although the Hb profile was monotonically decreasing towards the posterior, the position of the basin boundary along the AP axis was increasing. Thus the variability of the intersection of the two was small when the effect of 88 different Bcd profiles on the basin boundaries was considered.

Formation of the *hb* border in this thesis is consistent with the previous result [16]. In Chapter 5 it was shown that the border forms by the trajectory switching from being in A_3 basin to A_2 basin. The mechanism for canalization discussed in this thesis is different from the one found by Gursky as *gt* domain lies in the posterior where the pattern formation is influenced by the unstable manifold and not point attractors (this observation holds true for [16] as well). Unfortunately we cannot compare the canalization mechanisms in the anterior to [16] as our model displays large variability in *hb* border which makes studying canalization not possible.

Up to now the other efforts to study canalization have been mostly focused on identifying proteins [38, 59] or microRNA [39] that buffer perturbation to the system. These molecular mechanisms for canalization consider the buffering against mutational effects and not canalization of intrinsic fluctuations that are a result of embryo-to-embryo variability. Furthermore they do not provide a model that explicitly shows what causes decanalization in mutants. On the other hand we have shown that transformation of the model from wild type to mutant causes changes to dynamical structures (i.e. equilibria, unstable manifold) of the deterministic system and leads to decanalization of the gap gene pattern.

7.4 Limitations of the current modeling approach

The model currently is not able to correctly predict the dynamic shift of *gt* fully. In our model *gt* expression levels rise and fall at the correct time and the *gt* domain is shifted to

the anterior in comparison to wild type but not as much as the data from Kr^- embryos suggests.

As we studied the diffusionless approximation to the model it is of utmost importance that this agrees with full model. There are only two discrepancies in the range of our interest (35% EL-72% EL). These is an erroneous spike in Kr expression at nucleus 43 in wild type system and a slightly heightened expression of gt after nucleus 65 in Kr^-.

While cooperative interactions (encoded in T^{abc} matrix) between transcription factors that are bound in the vicinity of each other have been suggested by various computational models [63, 17], the only one that was experimentally verified is Bcd cooperative binding [41, 36]. The latter has been tried as the only cooperative interaction in our model, but the fit was not as good as the model used in this work, and the wild type system didn't canalize the Bcd perturbation.

7.5 Future possibilities

We have used three cooperative interactions in this work. There were several other interactions tried but those did not produce the best fits to the gap gene pattern and did not show canalization for the wild type system. It would be instructive to find biologically viable combinations of nonzero T^{abc} elements and try to anneal on those. Furthermore, it would be interesting to test out different combinations of T^{abc} matrices that correctly patterns both wild type and mutants as well as canalize and see if there is a characteristic topology for the gap gene network.

A more ambitious endeavor would be to consider what type of bifurcations would occur if we perturb any combination of other parameters, thus generating multi-dimensional parametric portraits. These types of parametric portraits would provide us with full knowledge of the systems dependence on the parameters and what different states the system could possibly take. This, of course, would be a challenge, as of right now AUTO could only

handle continuation in two parameters.

This work only focused on simulations of one single null mutant. We have high resolution quantitative data for other single null mutants as well as for double mutants. Considering mutant phenotypes is a good test for the model, it would be instructive to find models that could simulate each of those mutant data. Then we can see if there is a consensus at the topology level and if there is one model that can predict more than one mutant gap gene pattern. Next we could see if the model allows for the unstable manifold in the posterior to provide canalization. An obvious step after that would be to see how having one or more missing gap gene proteins destroys the manifold.

The anterior part (35% EL-57% EL) of the wild type system is not very sensitive to the initial conditions on the Hb axis. As a result the basins of point attractors consist of a range of concentrations of Hb for each nucleus. These basins are a property of our model, and to test if they are biologically significant we could introduce different *hb* gradients. With a range of different Hb concentrations for each nucleus we could compare the state of the system in the Hb-Kr-Gt-Kni space in our model to the one from empirical data. One of the ways to generate different *hb* profiles is by having one, two or four maternal copies of the *hb* gene [65]. Furthermore lack of *hb* expression could be introduced by using mothers with *hb* mutant germ cells. We could also consider uniform levels of *hb* expression throughout the embryo by mutations in *oskar* gene [37].

Additionally, to test the claims that were made at the end of Chapter 6 about canalization of the posterior *gt* domain in intermediary systems, we could consider Kr heterozygotes. Analyzing each individual embryo separately one can get an insight into the variability of the *gt* expression pattern in data. Surkova [67] has looked at the *eve* pattern in Kr heterozygotes but not at the posterior *gt* domain.

In Chapter 5 we have seen that the dynamics of the trajectories is dominated by A_4 for both wild type and mutant systems. As the trajectories approached the attractor via a spiral it would be possible to see the level of *gt* go up and then down in the span of cycle 14 in

vivo. Additionally it would be possible to observe if the predicted trajectories were accurate for a single nucleus for an individual embryo. For this we would have to create *gt* as well as the rest of the gap gene protein fusions with GFP.

Bibliography

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

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

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

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

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

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

[7] W. Damen. Parasegmental organization of the spider embryo implies that the parasegment is an evolutionary conserved entity in arthropod embryogenesis. *Development*, 129:1239–1250, 2002.

[8] E. Doedel. AUTO: Lecture notes on numerical analysis of nonlinear equations. `http://indy.cs.concordia.ca/auto`, 2010.

[9] H. Driesch. The potency of the first two cleavage cells in echinoderm development. experimental production of partial and double formations. In B. H. Willern and J. M. Oppenheimer, editors, *Foundations of Experimental Embryology*. Hafner, New York, 1892.

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

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

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

[13] H. G. Frohnhöfer and C. Nüsslein-Volhard. Organization of anterior pattern in the *Drosophila* embryo by the maternal gene *bicoid*. *Nature*, 324:120–125, 1986.

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

[15] R. C. Gonzalez and R. E. Woods. *Digital image processing*. Prentice-Hall Inc, Upper Saddle River, NJ, USA, second edition, 2002.

[16] V. V. Gursky, L. Panok, E. M. Myasnikova, Manu, M. G. Samsonova, J. Reinitz, and A. M. Samsonov. Mechanisms of gap gene expression canalization in the *Drosophila* blastoderm. *BMC Systems Biology*, 5:118, 2011. doi:10.1186/1752-0509-5-118.

[17] X. He, M. A. H. Samee, C. Blatti, and S. Sinha. Thermodynamics-based models of transcriptional regulation by enhancers: The roles of synergistic activation, cooperative binding and short-range repression. *PLoS Computational Biology*, 6:e1000935, 2010.

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

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

[20] H. Jäckle, D. Tautz, R. Schuh, E. Seifert, and R. Lehmann. Cross-regulatory interactions among the gap genes of *Drosophila*. *Nature*, 324:668–670, 1986.

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

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

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

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

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

[26] G. Jürgens, E. Wieschaus, C. Nüsslein-Volhard, and H. Kluding. 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, 1984.

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

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

[29] K. Kozlov, S. Surkova, E. Myasnikova, J. Reinitz, and M. Samsonova. Modeling of gap gene expression in *Drosophila Kruppel* mutants. *PLoS Computational Biology*, 8:e1002635, 2012.

[30] R. Kraut and M. Levine. Mutually repressive interactions between the gap genes *giant* and *Krüppel* define middle body regions of the *Drosophila* embryo. *Development*, 111:611–621, 1991.

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

[32] Y. Kuznetsov. *Elements of Applied Bifurcation Theory*. Springer-Verlag, New York, 1998.

[33] J. Lam and J.-M. Delosme. An efficient simulated annealing schedule: Derivation. Technical Report 8816, Yale Electrical Engineering Department, New Haven, CT, Sept. 1988.

[34] J. Lam and J.-M. Delosme. An efficient simulated annealing schedule: Implementation and evaluation. Technical Report 8817, Yale Electrical Engineering Department, New Haven, CT, Sept. 1988.

[35] P. A. Lawrence. *The Making of a Fly*. Blackwell Scientific Publications, Oxford, UK, 1992.

[36] D. Lebrecht, M. Foehr, E. Smith, F. J. P. Lopes, C. E. Vanario-Alonso, J. Reinitz, D. S. Burz, and S. D. Hanes. Bicoid cooperative DNA binding is critical for embryonic patterning in *Drosophila*. *Proceedings of the National Academy of Sciences USA*, 102:13176–13181, 2005.

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

[38] S. F. Levy and M. L. Siegal. Network hubs buffer environmental variation in *Saccharomyces cerevisiae*. *PLoS Biology.*, 6:e264, 2008. doi:10.1371/journal.pbio.0060264.

[39] X. Li, J. Cassidy, C. A. Reinke, S. Fischboeck, and R. Carthew. A microrna imparts robustness against environmental fluctuation during development. *Cell*, 137:273–282, 2009.

[40] J. D. Licht, M. D. Grossel, J. Figge, and U. M. Hansen. *Drosophila* Krüppel protein is a transcriptional repressor. *Nature*, 346:76–79, 1990.

[41] X. Ma, D. Yuan, K. Diepold, T. Scarborough, and J. Ma. The *Drosophila* morphogenetic protein Bicoid binds DNA cooperatively. *Development*, 112:1195–1206, 1996.

[42] Manu. *Canalization of Gap Gene Expression During Early Development in Drosophila melanogaster*. PhD Thesis, Department of Applied Mathematics and Statistics, Stony Brook University, 2007.

[43] Manu, S. Surkova, A. V. Spirov, V. Gursky, H. Janssens, A. Kim, O. Radulescu, C. E. Vanario-Alonso, D. H. Sharp, M. Samsonova, and J. Reinitz. Canalization of gene expression and domain shifts in the *Drosophila* blastoderm by dynamical attractors. *PLoS Computational Biology*, 5:e1000303, 2008. doi:10.1371/journal.pcbi.1000303.

[44] Manu, S. Surkova, A. V. Spirov, V. Gursky, H. Janssens, A. Kim, O. Radulescu, C. E. Vanario-Alonso, D. H. Sharp, M. Samsonova, and J. Reinitz. Canalization of gene expression in the *Drosophila* blastoderm by gap gene cross regulation. *PLoS Biology*, 7:e1000049, 2009. doi:10.371/journal.pbio.1000049.

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

[46] M. Mlodzik and W. J. Gehring. 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, 1987.

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

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

[49] C. Nüsslein-Volhard, H. G. Frohnhofer, and R. Lehmann. Determination of anteroposterior polarity in *Drosophila*. *Science*, 238:1675–1687, 1987.

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

[51] C. Nüsslein-Volhard, E. Wieschaus, and H. Kluding. 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, 1984.

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

[53] L. Perko. *Differential Equations and Dynamical Systems*. Springer-Verlag, New York, 1996.

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

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

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

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

[58] J. M. Rendel. The canalization of the *scute* phenotype of *Drosophila*. *Evolution*, 13(4):425–439, June 1959.

[59] D. Samakovli, A. Thanou, C. Valmas, and P. Hatzopoulos. Hsp90 canalizes developmental perturbation. *Journal of Experimental Botany.*, 58:3513–3524, 2007. doi:10.1093/jxb/erm191.

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

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

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

[63] E. Segal, T. Raveh-Sadka, M. Schroeder, U. Unnerstall, and U. Gaul. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature*, 451:535–540, 2008.

[64] J. Shen and S. Castan. Further results on drf method for edge detection. In *9th I.C.P.R.*, Rome, 1988.

[65] G. Struhl, P. Johnston, and P. A. Lawrence. Control of *Drosophila* body pattern by the Hunchback morphogen gradient. *Cell*, 69:237–249, 1992.

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

[67] S. Surkova, E. Golubkova, Manu, L. Panok, L. Mamon, J. Reinitz, and M. Samsonova. Quantitative dynamics and increased variability of segmentation gene expression in the *Drosophila Krüppel* and *knirps* mutants. *Developmental Biology*, 376:99–112, 2013.

[68] S. Surkova, D. Kosman, K. Kozlov, Manu, E. Myasnikova, A. Samsonova, A. Spirov, C. E. Vanario-Alonso, M. Samsonova, and J. Reinitz. Characterization of the *Drosophila* segment determination morphome. *Developmental Biology*, 313(2):844–862, 2008.

[69] S. Surkova, E. Myasnikova, H. Janssens, K. N. Kozlov, A. Samsonova, J. Reinitz, and M. Samsonova. Pipeline for acquisition of quantitative data on segmentation gene expression from confocal images. *Fly*, 2:58–66, 2008.

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

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

[72] R. Thom. Topological models in biology. *Topology*, 8:313–335, 1969.

[73] R. Thom. *Structural stability and morphogenesis*. W. A. Benjamin, Reading, Massachusetts, USA, 1975.

[74] A. M. Turing. The chemical basis of morphogenesis. *Transactions of the Royal Society London, Series B*, 237:37–72, 1952.

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

[76] C. H. Waddington. *The Strategy of Genes*. George Allen & Unwin, London, 1957.

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

[78] E. Wieschaus, C. Nüsslein-Volhard, and G. Jürgens. 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, 1984.