MATHEMATICAL MODELS FOR UNDERSTANDING DYNAMIC CELLULAR SYSTEMS

Henry Hughes Mattingly

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF CHEMICAL AND BIOLOGICAL ENGINEERING
Advisers: Stanislav Y. Shvartsman and Ioannis G. Kevrekidis

November 2017
Acknowledgements

Thank you to my parents, Bill and Cecile. A lifetime of effort on your parts made this possible, and I am so lucky to have you. Thank you to my advisors, Stas and Yannis—I couldn’t have asked for better. From you I learned how to be comfortable in the perpetual position of intellectual uncertainty that is science. Thank you to Ben—you are an incredible friend and mentor. And thank you to Natalia for all of your love and support—I am excited for our future.
Abstract

Cellular processes, such as fate decisions and mitotic divisions, are dynamic. Complex behaviors emerge from networks of interactions among numerous components: time-varying extracellular signals induce expression of specific genes, and periodic changes in intracellular molecule concentrations coordinate mitotic entry and exit. Quantitative understanding of these processes requires mathematical models. Models can make predictions about system dynamics in conditions that are difficult to probe experimentally, explain how systems-level behaviors emerge from a network of interactions, and convert observed data into constraints on future behaviors. This thesis uses mathematical modeling in each of these ways to learn about various dynamical biological phenomena. In the first chapter, models are used to estimate time-varying signals controlling meiosis that are difficult to measure experimentally. In the second, a simple model of the embryonic cell cycle is used to understand how robust oscillations arise in that system. And in the third and ongoing chapter, a model is used to explore how much we can expect learn about a biochemical mechanism from planned experiments.
# Table of Contents

Acknowledgements .................................................................................................................................................. i
Abstract ........................................................................................................................................................................ ii
List of Figures ............................................................................................................................................................... iv
List of Tables .................................................................................................................................................................... v
Introduction ................................................................................................................................................................... 1
Chapter 1: A Transport Model for Estimating the Time Course of ERK Activation in the *C. elegans* Germline .......................................................................................................................................................... 8
Chapter 2: The Design Space of the Embryonic Cell Cycle Oscillator ................................................................. 32
Chapter 3: Ongoing Work ................................................................................................................................................. 55
Appendix A: Supplement to Chapter 1 ................................................................................................................................. 68
Appendix B: Supplement to Chapter 2 ................................................................................................................................. 82
Appendix C: Supplement to Chapter 3 ................................................................................................................................. 94
References ...................................................................................................................................................................... 97
List of figures

Figure 1: The *C. elegans* germline ........................................................................................................ 10
Figure 2: Image analysis workflow ........................................................................................................ 16
Figure 3: Estimating distribution of germ cell positions ........................................................................ 22
Figure 4: Signaling time courses collapse ................................................................................................. 27
Figure 5: Consistent values of lumped physiological parameters ....................................................... 28
Figure 6: Minimal model of the embryonic cell cycle ............................................................................. 36
Figure 7: Phase plane analysis ................................................................................................................ 37
Figure 8: Analytical and numerical two-parameter bifurcation diagrams ............................................ 41
Figure 9: Bivariate distributions of oscillatory parameter vectors ....................................................... 44
Figure 10: Robustness of oscillations to variations in rate of cyclin synthesis ..................................... 47
Figure 11: Minimal model of ERK regulation ......................................................................................... 56
Figure 12: What can we learn about parameter values from typical experiments? ............................ 65
Figure A1: Evaluation of automated nuclei segmentation algorithm .................................................. 69
Figure A2: Evaluation of automated nuclei segmentation algorithm (zoom) .................................... 70
Figure A3: Detecting apoptotic cells in the germline .............................................................................. 73
Figure A4: Estimating the shape of the apoptosis sink function ............................................................ 76
Figure A5: Consistent physiological parameters, raw ......................................................................... 79
Figure A6: Representative uncertainty in $t(x)$ ....................................................................................... 80
Figure A7: dpERK signal versus normalized arc length ....................................................................... 81
Figure A8: Fold change in dpERK signal verus time ............................................................................ 81
Figure B1: Analytical and numerical domains of oscillations, non-small $\epsilon$ .................................. 91
Figure B2: Marginal distributions of oscillatory parameter vectors ..................................................... 91
Figure B3: Bivariate distributions of oscillatory parameter vectors .................................................... 92
Figure B4: Analytical versus exact oscillation amplitudes .................................................................... 92
Figure B5: Derivation of analytical oscillation amplitudes ................................................................... 93
Figure C1: Marginal distributions of parameter vectors consistent with synthetic data .... 96
List of Tables

Table I: Dimensionless parameter definitions.................................................................................................. 52
Table II: Ranges and scales used for parameter sampling .............................................................................. 52
Table III: Parameter values used to generate Figs. 7 and 8............................................................................ 54
Table B1: Parameter set with the largest value of $\delta$................................................................................... 90
Table CI: Definition of model parameters........................................................................................................ 95
Introduction

Cell biology is dynamic (1-3). In response to various environmental stresses, the transcription factor NF-κB periodically translocates into and out of the cell nucleus, and the resulting gene expression depends the number, period, and amplitude of oscillations (4, 5). The circadian clock in cyanobacteria, a three-protein posttranslational oscillator with 24-hour period, allows the organisms to anticipate the onset of morning and dusk and change their metabolic profiles accordingly (6-10). And among the most fundamental cellular processes is the cell cycle, with periodic growth, DNA synthesis, and mitosis (11).

Some of the best-studied examples of dynamic cell-biological processes involve signaling pathways, the biochemical machinery that cells use to sense their environments and respond with changes in behavior. The Ras/ERK signaling cascade, in particular, has served as a model system of cell signaling because of its role in regulating multiple cellular behaviors, such as proliferation, differentiation, and apoptosis, in various biological contexts (12, 13). Furthermore, abnormal pathway function has been identified in human diseases, including cancers (14-16) and developmental disorders (17-19), and an active research direction is the development of drug inhibitors to target pathway components and modify signaling dynamics (20-22). The pathway is activated when extracellular ligands bind to a variety of cell surface-bound receptors (e.g. (23)). Receptor-ligand binding activates the GTPase Ras at the cell surface, which then recruits and activates the kinase Raf (13). Raf phosphorylates the MAPK/ERK kinase (MEK) on two amino acid residues, making it catalytically active (24), and MEK activates the extracellular signal-regulated kinase (ERK) by phosphorylation, also on two amino acid residues (25-27). Mammalian cells have two isoforms of both MEK and ERK (MEK1/2 and ERK1/2), but the isoforms are
functionally redundant (28, 29). In its active, dually-phosphorylated form, ERK has numerous substrates in the cytoplasm and nucleus that interpret the signal and carry out responses by altering the cell’s state via, for example, changes in gene expression, cell shape, or cell cycle progression (13). Finally, multiple phosphatases act on the pathway; some are specific—such as MKP3, which specifically desphosphorylates ERK—while others target multiple pathway components (30-33).

The dynamics of phosphorylation of ERK and its substrates are modulated in unicellular organisms and developing embryos to elicit differential responses. Exposing cultured PC12 cells to epidermal growth factor elicits a transient pulse of intracellular ERK activity and a proliferative response, while nerve growth factor leads to sustained activation of ERK and cell differentiation (34). At the level of ERK substrates, it was recently found that ERK both activates and deactivates the transcription factor Elk-1 on different time scales (35). On short time scales (tens of minutes), ERK phosphorylates Elk-1 on a set of amino acids that allow it to interact with a Mediator complex and promote transcription; however on longer time scales (hours) ERK phosphorylates additional sites that disrupt this interaction and inhibit transcription. In the early *Drosophila melanogaster* embryo, a valuable model system for quantitative studies of dynamic developmental signals, an hour-long pulse of ERK activity at the poles of the embryo induces expression of genes necessary for the formation of head and tail structures (36-39), and a 30-minute pulse of ERK activation in the third hour of *Drosophila* embryonic development specifies two narrow bands of cells in the presumptive neuroectoderm that take on neural fates (40-42). These and numerous other examples attest to the importance of dynamics in cell and developmental biology.

Quantitative understanding of these dynamic processes requires mathematical models. They provide concise summaries of our current state of knowledge by abstracting the relevant components. They can make predictions in contexts that may be experimentally inaccessible and
that guide new experiments, leading to refined models and deeper understanding. And mathematical models allow us to understand the conditions in which certain behaviors emerge and can use observations to constrain which behaviors are realizable. This thesis uses mathematical models in each of these ways to learn about various dynamic biological phenomena.

In the first chapter of this thesis, a mathematical model was used to extract the most from available data by estimating the dynamics of ERK activity in the *Caenorhabditis elegans* germline from images, where time-resolved measurements are difficult to attain (43). The C elegans germline is a well-studied system for cell signaling and fate decisions (44-46). The germline consists of two U-shaped tubes full of germ cells and oocytes. At the distal tip, cells respond to a signal that induces mitotic divisions that populate the germline (47, 48). Cells then move away over time and, once out of range of the distal signal, differentiate and begin meiosis (48-50). At the loop, where the gonad tube bends, most cells undergo apoptosis, and it is thought that this donates materials to the surviving cells that become oocytes (51).

In the germline, ERK phosphorylates numerous downstream substrates with diverse biological functions. The MAPK pathway and, ultimately, ERK are activated by an external signal that is present when the animal is well-fed (52). During normal germ cell development, cells traveling through the germline see an extended pulse of ERK activity that shuts off at the loop (43, 45). If the animal is starving, the signal is lost, and germ cells enter a state of meiotic arrest (52). Mutations that cause over- or under-activation of ERK in the germline disrupt normal germ cell development and render the animal sterile (45).

One of the challenges in building a quantitative understanding of ERK-dependent processes in the germline is that live reporters for the active, dually-phosphorylated form of ERK (dpERK),
currently are not working in the germline. Germlines dissected from fixed animals can be stained with an antibody specific for dpERK, which allows measurement of ERK activity at a snapshot in time (45). Even then, germlines are of different sizes and shapes, and spatial profiles of dpERK are not directly comparable. In order to study signaling in the germline quantitatively, we needed to be able to compare across germlines. That is, we need a common coordinate against which dpERK intensity could compared in multiple germlines.

To address this, a transport model was derived in Chapter 1 that allowed the positions of germ cells in an image to be mapped to cell ages (43). This then provided a way of transforming spatial profiles of ERK activity into dynamics of ERK activity in developing germ cells that could be compared across germlines. This analysis revealed that germ cells in different animals experience pulses of ERK activity roughly at the same time and for the same duration during their development, despite moving through germlines of different geometries. This raised the question of whether germ cell transport and ERK activity are coupled to ensure that the time course of signaling seen by germ cells is invariant across animals. The method can be applied to estimate the dynamics of molecules besides ERK for which live reporters are lacking, to study multiple molecules in the germline simultaneously, and to explore the quantitative effects of mutations on signaling dynamics. A version of Chapter 1 appears in the published article Mattingly et al. (43).

From signaling dynamics in the germline of an adult organism, Chapter 2 moves to the dynamics of the cell cycle in a developing embryo. There, a mathematical model of the essential components of the cell cycle was used to understand how the interactions of cell cycle components give rise to nontrivial dynamical behaviors, particularly oscillations. A combination of analytical, numerical, and statistical methods were used to delineate the parameter regime in which stable oscillatory
solutions exist and identify design principles that make those solutions robust to parameter variations (53).

The cell cycle of a somatic cell is complex, consisting of a period of growth, followed by DNA replication, further growth, and then cell division, all of which are coordinated by phase-shifted periodic changes in several molecules called cyclins (11). However, cells in the early frog embryo, *Xenopus laevis*, a model system of the embryonic cell cycle, undergo rapid divisions by skipping the growth phases and only replicating DNA between divisions (11). In this context, the cell cycle is driven by the periodic accumulation and degradation of cyclin B, which forms a complex with cyclin-dependent kinase 1 (CDK1). The kinase activity of the cyclin-CDK complex depends on its phosphorylation state, which is modified by various kinases and phosphatases. Active cyclin-CDK, in turn, phosphorylates and modifies the activity of its own regulators, creating complex positive and negative feedback loops (54-60). Finally, the active cyclin-CDK complex also leads to activation of the anaphase-promoting complex (APC) (61) which marks mitotic cyclins for rapid degradation by ubiquitin-mediated proteolysis (62) and whose other degradation targets initiate chromosome segregation and cell division (61).

Despite having many components and, in some cases, poorly-understood interactions (61), the autonomous oscillatory behavior of the embryonic cell cycle can be captured by a simple model with two dependent variables: the amounts of active and inactive cyclin-CDK complex (63). In this model, the regulatory positive and feedback loops are lumped into Hill functions that depend on the level of active cyclin-CDK.

Even with a mathematical model written down, though, it is unclear under what conditions (i.e. for what values of the model parameters) the system generates sustained oscillations. The goal in
Chapter 2 was to fully characterize the region of parameter space that produces oscillations (the “design space”) in a simple model of the embryonic cell cycle using a combination of analytical, numerical, and sampling techniques. The model lumped both the positive and double-negative feedback loops involved in cyclin-CDK regulation into a single Hill function, making the model amenable to analytical analysis. Taking advantage of a separation of time scales in the model that is borne out by measurements in the Xenopus system, we analytically derived the boundaries of the region of oscillations and the amplitude of oscillations. Numerical continuation methods verified the analytical results, and sampling from the parameter space was used to explore the design space in regions where the assumptions in the analytical derivation broke down. This study revealed that the experimentally-measured parameter values for the Xenopus system have features found in the widest regions of the oscillatory domain, partially explaining the robustness of its oscillations to parameter variations. A version of Chapter 2 appears in the published article Mattingly et al. (53).

The next step is to connect systems-level studies like the ones above to the underlying biochemistry in cells. The third and final chapter of this thesis discusses an ongoing project to build a predictive model of the minimal network regulating the phosphorylation state of ERK using experimental biochemical data and to constrain this network’s possible dynamic behaviors. Our recent work showed that a system containing only ERK2, MEK1, and the phosphatase MKP3 can in principle exhibit rich long-term dynamics, including bistability, oscillations, and coexistence of limit cycles, if the parameters are allowed to take on any values (64). An effort has begun in the group to directly monitor the (de)phosphorylation reactions in test tubes using mass spectrometry in order to constrain the parameters in that model. In preparation for the incoming experimental data, we generated synthetic data from a parameter set that fits data reported in the literature monitoring
phosphorylation of ERK by MEK (65) and asked how much we can expect to learn about the parameter values from this experiment. This analysis identified a major source of uncertainty in the parameters and lead to suggestions for additional experiments.
Chapter 1: A Transport Model for Estimating the Time Course of ERK Activation in the C. elegans Germline

In collaboration with Jessica J. Chen\textsuperscript{1} and Swathi Arur\textsuperscript{1}

\textsuperscript{1}UT Graduate School of Biomedical Sciences and Department of Genetics, University of Texas M.D. Anderson Cancer Center, Houston, TX 77030

ABSTRACT

The Caenorhabditis elegans germline is a well-studied model system for investigating the control of cell fate by signaling pathways. Cell signals at the distal tip of the germline promote cell proliferation; just before the loop, signals couple cell maturation to organism-level nutrient status; at the proximal end of the germline, signals coordinate oocyte maturation and fertilization in the presence of sperm. The latter two events require dual phosphorylation and activation of ERK, the effector molecule of the Ras/MAPK cascade. In C. elegans, ERK is known as MPK-1. At this point, none of today’s methods for real-time monitoring of dually phosphorylated MPK-1 are working in the germline. Consequently, quantitative understanding of the MPK-1-dependent processes during germline development is limited. Here, we make a step toward advancing this understanding using a model-based framework that reconstructs the time course of MPK-1 activation from a snapshot of a fixed germline. Our approach builds on a number of recent studies for estimating temporal dynamics from fixed organisms, but takes advantage of the anatomy of the germline to simplify the analysis. Our model predicts that the MPK-1 signal turns on \textasciitilde30 h into germ cell progression and peaks \textasciitilde7 h later.
INTRODUCTION

The *Caenorhabditis elegans* germline is a well-studied model system for genetic studies of cell signaling (Fig. 1) (44-46). The adult hermaphrodite germline consists of two U-shaped tubes that meet at a common uterus. Each tube is filled with germ cells spatially arranged according to their maturity (66). Stem cell divisions at the distal tip of each tube (distal to the common uterus) maintain the germ cell population as the more mature cells move away from the distal tip and transition into meiosis (47, 67). After the loop, the bend in the U, a single-file line of oocytes prepare for fertilization and ovulation. Here, we focus on the anterior region of the germline, from the distal tip cell until the loop (Fig. 1), where approximately 1000 germ cells line the periphery of the gonad tube and are connected to a common cytoplasmic core called the rachis (68, 69). Nuclei are separated by incomplete cell membranes that are open to the rachis, so the terms “nucleus” and “cell” in this region are essentially interchangeable. Near the loop, a large portion of healthy germ cells undergo apoptosis, which may leave more cytoplasmic material for the surviving cells that become oocytes (51). In an adult hermaphroditic nematode, the germline reaches steady state, maintaining a roughly constant number of germ cells and ovulation rate for most of the organism’s reproductive life (68, 70).

Within the rachis are spatial gradients of effector molecules that regulate germ cell development as they progress through the germline (44, 45). Among the regulatory molecules in the rachis is the extracellular-signal regulated kinase (ERK), the terminal kinase of the Ras/MAPK signaling cascade. The *C. elegans* homolog of ERK is known as MPK-1 (45, 71). In the pachytene region of the germline, DAF-2 insulin-like signaling activates the MAPK cascade when the organism is in a nutritionally replete environment, resulting in MPK-1 to become dually-phosphorylated (dpMPK-1) and catalytically active (52); activation of MPK-1 drives meiotic progression and
oocyte production. The ligand to the DAF-2 receptor that results in activation of MPK-1 in the germline is unknown as of this writing. In the proximal part of the germline, MPK-1 activation by signals from sperm couples oocyte maturation to sperm availability (70).

Figure 1: An adult, hermaphrodite *C. elegans* germline. A) Shows the anterior region of the germline, from the distal tip (left in the image) through the first few cellularized oocytes (right in the image). B) Focuses on the germline from A, but zoomed in only on the region from the distal tip to the loop. Blue: Nuclei stained with DAPI. Mitotic cells in the Mitotic Region undergo divisions that maintain the germ cell population. As the cells divide and expand the mitotic region, more mature cells are pushed towards the loop. Green: Germ cells differentiate in the Transition Zone, entering meiosis, marked by the HIM-3 protein in green. HIM-3 labels the synaptonemal complex axis of the meiotic nuclei. Red: Activation of MPK-1, which occurs in the final two-thirds of the anterior germline. Near the loop, a large fraction of arriving germ cells undergo apoptosis (apoptotic cells in each color channel marked by white arrowheads). Surviving cells pass the loop and become oocytes.
Active MPK-1 has numerous substrates that control multiple biological functions (72). Mutations that result in complete loss of active MPK-1 cause germ cells to arrest in early pachytene stage of meiosis I. Reduction of MPK-1 activation results in multiple phenotypes, such as delayed progression of pachytene stage germ cells and the formation of excessively large oocytes (45). Alternatively, mutations that over-activate MPK-1 pathway result in higher and ectopic activation of MPK-1 in the loop region, causing defects in oocyte growth and maturation (45). All of these mutations render the animal infertile.

Clearly, the amplitude of the MPK-1 signal in the germline is important for normal germ cell development. The duration of MPK-1 signaling has also been shown to be important in other cell decision-making contexts, such as the proliferation versus differentiation decision in PC12 cells (34, 73-76), and may also be a relevant factor in the germline. However, since existing methods for real-time monitoring of MPK-1 signaling in living organisms are not yet working in the germline, quantitative information about the dynamics of this signal are not available as of this writing (77-86). In the early *Drosophila melanogaster* embryo, ERK activation dynamics were estimated from fixed samples by matching morphological features to developmental time (41, 42, 87). Here, we provide an alternative method to quantitatively estimate MPK-1/ERK activation dynamics in the *C. elegans* germline. Our approach builds on recent work in cultured cells estimating dynamic information from static data (88, 89). However, these studies required molecular markers to order cells according to their progress in the cell cycle; in the germline, cells are naturally ordered in space. We take advantage of this anatomical feature and develop a model-based computational approach for the reconstruction of active MPK-1 dynamics in the distal germline (which we define here as the region from the distal tip cell to the loop). While multiple mathematical models for germ cell dynamics have been proposed (90-92), all of these simulate the
stem cell programs in the distal region of the germline. To our knowledge, this is the first model aimed at the quantitative estimation of signaling dynamics in the germline from fixed samples and that can be extended to multiple signaling conditions and environmental contexts.

MATERIALS AND METHODS

Nematode fixation and staining

Wild-type N2 hermaphroditic animals were placed on NGM (Normal Growth Media, (93)) plates seeded with OP50 *Escherichia coli* bacteria at fourth larval stage of development (L4) at 20°C for 24 h, and germlines were extruded at the end of 24 h. Dissections were performed as described in Arur et al. (72, 94). Briefly, all dissections were performed under 5 minutes (immediately after adding levamisole to render the animals immobile) to achieve optimal dpMPK-1 staining. The dissected germlines were then fixed in 3% paraformaldehyde for 10 min at room temperature, followed by fixation in 100% methanol at -20°C for 1 h. The fixed germlines were then processed for immuno-fluorescence staining via blocking in 30% Normal Goat Serum (in 1x PBST with 0.1% Tween-20) for 1 h, followed by incubation with anti-MAP Kinase antibody used at 1:200 (Clone MAPK-YT; Sigma-Aldrich, St. Louis, MO) overnight in 30% Normal Goat Serum at 4°C. Secondary antibodies were donkey anti-mouse Alexa Fluor 594. The secondary antibodies were used at 1:400. All germlines were treated with the same staining solution in the same vial to mitigate antibody-staining variation, as described in Lee et al. (45).

dpMPK-1 staining was standardized using wild type germlines and germlines from an *mpk-1* null allele, *gal7*, processed in the same tube (45, 52). *gal7* was described in Lackner and Kim (95) and is an early stop in the *mpk-1* gene, resulting in no expression of the MPK-1 protein. The
MAPK-YT antibody was rigorously tested for specificity by multiple laboratories against \textit{mpk-1(ga117)} null mutant germlines (45, 96). The staining conditions have been standardized such that the antibody does not detect any signal in the \textit{mpk-1} null germline, but reproducibly detects the two peaks of MPK-1 activation in wild-type N2 germlines.

**Imaging**

An A1 confocal microscope with Plan-Apo VC 60x Oil objective was used for imaging germlines (Nikon, Melville, NY). Z-stacks, with 0.5 \( \mu \)m spacing between slices, were taken of the distal part of each germline. A single z-stack did not fit the entire distal germline in the field of view. Multiple z-stacks taken along the length of a single germline were stitched together using a custom MATLAB program (The MathWorks, Natick, MA). The overlap position between a pair of overlapping z-stacks was found by finding the position with the highest the normalized cross-correlation between the DAPI images in the stacks (MATLAB \textit{normxcorr2}).

**Image segmentation and extracting fluorescence data**

Individual germ cells were detected in each DAPI image slice of the stitched z-stacks using the two-dimensional (2D) circular Hough Transform. The built-in MATLAB function \textit{imfindcircles} was used to implement this. This approach was used because the germ cells took on irregular shapes in 3D when pressed by the cover slip. However, within a slice they appeared circular. As a result, the structure imposed by the 2D circular Hough transform produced more accurate detection rates than other, less structured methods, such as watershed-based methods. Before applying the Hough transform, images were smoothed with a 3-pixel-wide Gaussian filter (\textit{imfilter}), eroded (\textit{imerode}) with a 4-pixel disk structuring element (\textit{strel}), and reconstructed (\textit{imreconstruct}), using
the Gaussian-smoothed image as the mask. The purpose of these preprocessing steps was to blur out intracellular features, while leaving the circular cell shapes intact.

Within a slice, circles with projected area overlap above a threshold were merged by keeping the circle containing the highest average DAPI fluorescence intensity among those overlapping and removing the others. Because germ cells span multiple slices, a single germ cell would almost always be identified in multiple slices. Therefore, multiple detections of the same germ cell across slices had to be registered. This was done by starting from the bottom (or top) slice, detecting circles in that slice, and then assigning the mutually closest circle in the next slice to the same cell. This was repeated until the last slice, with the maximum allowed extent of a single cell being four slices. To determine the effectiveness of the image segmentation code, three germlines were manually segmented in ImageJ (National Institutes of Health, Bethesda, MD) using the oval selection tool to circle cells in each slice. The result of this process was a list of cells, the slices they appear in, their 2D positions in each slice, and their radii in each slice. The automated detections then had to be matched to corresponding manual detections. For each manually detected cell, all automatically detected cells in the same slice range were considered. It was visually determined that sometimes the automated segmentation code found that cells extended one slice above or below the same cell in the manual segmentation, so the slice range for each manually detected cell was increased above and below by one slice.

Among the automatically detected cells in the same slice range as a given manually detected cell, the one with the mutually closest 2D position (averaged over slices) was found. A metric was then needed to decide whether this closest cell was actually the same cell as the manually detected one. Because we knew the two cells were in the same slice range, we only considered their projected areas. The metric used was the area of intersection of the 2D projections of the two cells divided
by the area of their union. This metric is common in the object detection community and accounts for both the locations and scales of the objects being compared (97, 98). If the metric was above a preset threshold (0.3), the automatically detected cell was considered to be the same cell as the manually detected one. Otherwise, the manually detected cell was considered to be undetected by the segmentation code. An automatically-detected cell that was matched to a manually-detected one could not be matched to any other manually detected ones. After cycling through all of the manually detected cells, any cells that were automatically detected but not assigned to a manually detected one were considered false detections. Among the three manually segmented germlines, on average the detector found 90% of the germ cells present, and of the ones detected, 97% were true positives. Images comparing the automated and manual detections are provided in Section A1 in Appendix A.

With the three-dimensional (3D) positions and radii of each germ cell, fluorescence intensities could be extracted from z-stacks (Fig. 2). In some cases, the segmentation code could detect faint DAPI signal even when only the tip of a cell was present in a slice. The dpMPK-1 signal from the cells in these slices was very weak and not representative of the actual dpMPK-1 content of the cell. For this reason, the slice of maximum DAPI intensity was determined for each cell, and the average nuclear dpMPK-1 intensity was measured only from that slice. If the dpMPK-1 signal in a germline significantly decreased with slice depth due to photobleaching during imaging, then only the signal from cells in the first few slices were used to measure the dpMPK-1 spatial profile. Additionally, background fluorescence was subtracted from the dpMPK-1 profiles. In each germline, background fluorescence was calculated by averaging the dpMPK-1 intensity in the mitotic and early meiotic region, where there should be much less active MPK-1 than in the pachytene region. For each experiment, ~6-10 intact germlines were imaged with the exact same
**Figure 2:** Workflow of image analysis for a given germline. A) Confocal fluorescence microscope image of a DAPI-stained germline, extruded from a wild-type adult. White circles denote detected nuclei. Only the distal part of the germline is shown. Scale bar is 21 μm. B) Nuclei detected from z-stack of the germline in A. The center line estimated for this germline is shown in black. Nuclei are colored by their arc length position along the center line. At the distal tip $x = 0$, and at the beginning of the loop $x = x_{\text{max}}$. Germ cells are move through the germline with average velocity $v(x)$. C) The red channel image of the germline in A, corresponding to the dpMPK-1 antibody staining. D) The nuclear mask in B was used to extract nuclear fluorescence intensity of dpMPK-1 for each germ cell. Plotted is dpMPK-1 fluorescence intensity versus arc length position. The black dots are the measurements of dpMPK-1 intensity from each nucleus. The dark red line is the smoothed average, while the shaded region shows plus/minus one standard deviation.
microscope settings, and each experiment was repeated at least three times. The data was internally compared for consistency.

**Finding the center line of the gonad tube**

The scatter plot of individual cell positions in a single germline viewed from a distance appears to be a 1D curve. Diffusion Maps, a nonlinear dimensionality reduction technique (99), essentially takes advantage of this coarse perspective to order the germ cells according to their positions along the 1D curve, which in this case corresponds to the center line of the gonad tube. More details about the Diffusion Maps algorithm are given in Appendix A, Section A2. Diffusion Maps is not necessarily required to find the center line; we used it because it is quickly and robustly sorts cells by arc length. A user needed to manually specify which end of the 1D curve corresponded to the distal tip. Once the cells had been ordered along the center line and the distal tip had been chosen, the positions of germ cells within a local neighborhood of each other were averaged, producing a noisy estimate of the center line. Following the approach of Kafri et al (88), the ordered set of points was converted to spherical coordinates, and each of the three ordered sets of coordinates was smoothed by taking moving averages using MATLAB’s *smooth* function. Converting back to Cartesian coordinates resulted in a smooth estimate of the center line. Due to neighborhood averaging, the center line estimated this way did not extend all the way to the two ends of the distal germline. Each end of the center line was extended in a straight line in the direction of the tangent to the center line at that end. The center line was extended at each end until it intersected the plane containing the distal- (or proximal-) most cell and with normal vector parallel to the direction in which the center line was being extended.
**Estimating probability distribution functions**

Germ cells positions were projected from 3D space onto the closest point on the center line of the gonad tube to calculate their arc length positions. Smooth probability density functions and cumulative distribution functions were estimated in MATLAB using Z. Botev’s code *kde*, available on the MathWorks website at:


**RESULTS**

**Transport model for the germline**

In this section, we formulate a mathematical model of germ cell transport along the germline. The prediction of this model will be the average time it takes a germ cell to travel from the distal tip to any given position in the germline. The approach, much like models used to describe unidirectional traffic flow (101), will be to estimate the local velocity as a function of spatial position. If we were given the average velocity of germ cells at each position, we could immediately calculate the average transit times. Because we do not get this information from fixed germlines, we will use our model to express velocity in terms of quantities that are measurable from fixed samples.

We will only consider distal-to-proximal motion and neglect perpendicular motion because the length of the gonad tube is much longer than its diameter. Position in the germline will be measured by arc length: if one were to measure the total distance driven down a winding road, that distance would be the arc length of the road. Arc length position, $x$, will be measured as distance in microns from the distal tip, with $x = 0$ located at the distal tip. We aim to derive an equation $t(x)$ that
describes the time it takes germ cells to travel from the distal tip to each arc length position along the germline.

In this analysis, we focus only on the distal germline. The approach described here for estimating time as a function of arc length is statistical in nature and therefore benefits from having as many cells as possible in the sample. In the proximal germline, where oocytes are in single-file, there are too few cells to make any meaningful estimates using this approach.

We will only consider adult hermaphroditic organisms (defined as 24 h past the fourth larval molt (L4)), with fully-developed germlines because these germlines have reached steady state (68, 70). Specifically, we assume that the number of germ cells in the germline and the rate of oocyte ovulation are constant over time (70). Young adults are in a state of germline expansion, increasing their germ cell numbers. By 24 h past the fourth larval molt, the germline switches to a homeostatic mode, maintaining the number of germ cells more or less constant (70, 102). In normal laboratory conditions at 20°C, this state lasts until ~72-90 h past L4 (52, 103). Time course analysis of hermaphroditic germlines assayed at 18, 20, 24, 36, and 48 h past the L4 stage of development for dpMPK-1 activation suggests that the concentration gradients inside the germline are also time-invariant in homeostatic adult germlines (52).

Next, we assume that germ cells move unidirectionally through the germline, from which it follows that germ cells are perfectly arranged according to their maturity. This is not strictly true, but there is clear directionality to germ cells’ motion on the time scale of hours, the time scale over which germ cells develop. This assumption is weakest in the mitotic region, where it is unclear whether there is a mechanism that would prevent some cells from moving unidirectionally down the germline, such as asymmetric cell divisions (68), trapping by distal tip cell (DTC) processes (68,
104), or diffusive motion created by random orientation of the axes of cell divisions (92). We acknowledge that our model does not account for affects like these. However, while an individual cell could get trapped in the mitotic region for an extended period of time before beginning unidirectional motion towards the loop, this would only add a time delay to the beginning of the time course of MPK-1 activation, without otherwise changing the dynamics.

Finally, we assume that in a wild type germline, germ cells undergo stereotypical dynamics, so that the population-level dynamics are representative of an individual cell. Furthermore, steady state physiological parameters, such as the total rate of cell production and the total rate of cell apoptosis, are assumed to be similar across animals.

The intuition behind the derivation is that the rate at which cells arrive at each position must be balanced by the rate at which cells leave that position for there to be no accumulation of cells at any position over time. We can construct a balance on the number of cells arriving and leaving an arbitrary arc length position:

\[ 0 = -\frac{dJ}{dx} + S(x) - R(x). \]  

(1)

Here, \( J(x) \) describes the number of cells per time that pass through each position \( x \) and is a convective flux of cells. In this formulation, cell flux and cell flow rate are equivalent. \( S(x) \) is a spatially-dependent source term that accounts for the production of cells by mitotic divisions, and \( R(x) \) is a sink term that accounts for the removal of apoptotic cells from the germline before they reach the loop. The total rate at which cells are removed via apoptosis is \( r \equiv \int_0^{x_{\text{max}}} R(x) \, dx \) cells per hour.

We assumed that all new cells are introduced into the germline at \( x = 0 \), reducing Eq. 1 to
\[
0 = -\frac{dJ}{dx} - R(x),
\]  
(2)

with cell production included as a boundary condition on the flux at \(x = 0\), \(J(x = 0) = s\) cells per hour. The left-hand sides of Eqs. 1 and 2 correspond to the rate of accumulation of cells over time at each position \(x\), which is equal to zero under the steady state assumption.

The convective flux term can be expressed as \(J(x) = C(x)v(x)\), where \(C(x)\) is the concentration of cells at \(x\), and \(v(x)\) is the average velocity of cells at \(x\). We want to solve for \(v(x)\). The concentration of cells can be expressed as \(C(x) = N_{tot}f_X(x)\), where \(N_{tot}\) is the average number of germ cells in the distal germline, and \(f_X(x)\) is the probability density function (PDF) of germ cell arc length positions \(x\) (Fig. 3). The concentration \(C(x)\) is a smoothed histogram of germ cell positions, and \(f_X(x)\) is the same but normalized so that the area under the curve equals one. We estimated the function \(f_X(x)\) in each germline from images as follows: germ cells were automatically detected in a \(z\)-stack of an extruded germline using custom code written in MATLAB; the center line of the germline was constructed; arc length positions of cells in that germline were calculated; then \(f_X(x)\) for that germline was estimated from the collection of cell positions using kernel density estimation, a standard technique for empirically estimating probability distributions (100).

**Solving the model**

In this section we solve the transport model for \(v(x)\), the average local velocity, and \(t(x)\), the average transit time to each position along the gonad tube. First, we substitute \(J(x) = N_{tot}f_X(x)v(x)\) in for the convective flux term of Eq. 2. Moving the convective flux term to the left-hand side and integrating once gives:
\[ N_{\text{tot}} f_X(x) v(x) = s - \int_0^x R(u) du, \]  

where \( s \), the rate of cell production, is the constant of integration that results from applying the boundary condition \( f(x = 0) = s \). Solving for \( v(x) \) gives:

\[ v(x) = \frac{s - \int_0^x R(u) du}{N_{\text{tot}} f_X(x)}. \]  

We can draw an analogy here to unidirectional, steady state fluid flow in a pipe. In the pipe flow problem, \( v(x) = Q(x)/A(x) \)—the average local velocity of fluid elements equals the local volumetric flow rate divided by the local cross-sectional area. The numerator of the expression on the right-hand side of Eq. 4 is the local cell flow rate (cells per unit time), analogous to the local volumetric flow rate in a pipe (volume per unit time). The cell flow rate at \( x = 0 \), where cells are produced by mitotic divisions, equals \( s \), and the second term in the numerator reduces the local
cell flow rate as cells undergo apoptosis. In the absence of cell death, the cell flow rate everywhere in the germline would be constant and equal to the production rate, $s$. The denominator of the expression is analogous to the local cross-sectional area in the pipe flow case.

The velocity $v(x)$ is the time rate of change of a cell’s position $x$, or $v(x) = dx/dt$. Separating the variables $x$ and $t$, giving $dt = dx/v(x)$, and then integrating gives:

$$t(x) = \int_0^x \frac{N_{\text{tot}} f_X(w)}{s - \int_0^w \tilde{R}(u) du} \, dw. \quad (5)$$

The variables $u$ and $w$ are dummy variables of integration. The lower bound of the outer integral on the right-hand side is zero because at $x = 0$, $t = 0$. The average time cells take to transit from the distal tip at $x = 0$ to the loop at $x = x_{\text{max}}$ is denoted $\tau$.

Factoring out $N_{\text{tot}}/s$ from Eq. 5 shows that there are essentially two lumped parameter combinations that govern the behavior of $t(x)$: the ratios $N_{\text{tot}}/s$ and $r/s$. After factoring, the equation becomes

$$t(x) = \frac{N_{\text{tot}}}{s} \int_0^x \frac{f_X(w)}{1 - \int_0^w \tilde{R}(u) du} \, dw, \quad (6)$$

where $\tilde{R}(x)$ is a dimensionless function with the same shape as $R(x)$ and, as mentioned in the previous section, $r$ is the average rate at which cells undergo apoptosis. The first ratio, $N_{\text{tot}}/s$, is the total number of cells in the germline divided by the cell production rate and has units of time. In the absence of apoptosis, the average transit time from the distal tip to the loop would be equal to $N_{\text{tot}}/s$. The second ratio, $r/s$, is the fraction of germ cells that undergo apoptosis and is dimensionless. $N_{\text{tot}}/s$ gives the scale of $t(x)$ but does not affect its shape, which is determined by the expression inside the integral.
In practice, \( f_X(x) \) was estimated for each germline and \( R(x) \) was calculated by pooling data from multiple germlines. Estimation of \( R(x) \), which accounts for germ cell apoptosis and removal from the germline, required pooling data from multiple germlines and is described in the Supplemental Material (Appendix A, Sections A3 and A4). The same values of the parameters \( N_{tot} \), \( s \), and \( r \), were used for all germlines. In particular, \( N_{tot} \) and \( s \) have strong opposing effects on the model’s predictions; as a result, using a value of \( N_{tot} \) from an individual germline and the average value for \( s \) would give biased results. During the error analysis (Appendix A, Section A5), the parameters \( N_{tot} \), \( s \), and \( r \) were sampled uniformly within ranges consistent with the literature to propagate parameter uncertainties through the model, which will be discussed in the next section. The trapezoidal rule was used to approximate all integrals.

**Exploring the feasible space of physiological parameters**

The model contains one equation and four physiological parameters: \( N_{tot} \), the average number of cells in a germline at any given time once adulthood is reached; \( s \), the average rate of cell production; \( r \), the average rate of cell removal by apoptosis; and \( \tau \), the average time cells take to travel from the distal tip to the loop. The steady state assumption provides an equation relating these physiological parameters, meaning their values are no longer entirely independent of each other. Evaluating the indefinite integral in Eq. 6 at \( x = x_{max} \), where \( t(x_{max}) = \tau \) gives:

\[
\tau = \frac{N_{tot}}{s} \int_0^{x_{max}} \frac{f_X(w)}{1 - \frac{r}{s} \int_0^w R(u)du} dw.
\]

The density function \( f_X(x) \) is measurable for each individual germline, and \( \hat{R}(x) \) can be estimated by pooling data from multiple germlines. This leaves only the physiological parameters as
unknowns in the equation. Additionally, choosing values for three of the parameters constrains the value of the fourth.

Previous studies have estimated ranges for these physiological parameters from data on wild type animals. The total number of cells in an adult hermaphroditic germline is approximately 800-1000 cells (68). The cell production rate by mitotic divisions, $s$, has been reported to be between 16 and 24 cells per hour (68, 105). An engulfment marker to show that in wild type animals each gonad arm contains 2-4 apoptotic cells at a given time (106). However, the time rate of cell death, per se, has not been precisely measured. Under the steady state assumption, the rate of cell removal via apoptosis, $r$, must equal the difference between the rate of cell production, $s$, and the rate of cell removal via ovulation. The latter has been estimated to be between 2 and 5 cells per hour per gonad arm (70). Therefore, a consistent estimate for the rate of cell death, $r$, lies between 11 and 22 cells per hour. Finally, pulse-chase experiments suggest that $\tau$ lies between 48 and 54 h (107).

Although there are only three independent physiological parameters after applying Eq. 7, we have information about all four parameters. To capitalize on this, we sampled the parameters $N_{tot}$, $s$, and $r$ uniformly within the literature-reported ranges, then discarded all parameter combinations for which the value of $\tau$, calculated using Eq. 7, was outside of the range consistent with the literature. This procedure, inspired by the Data Collaboration methodology, carved out a sliver in the 3-dimensional parameter space consisting of the parameter combinations that are consistent with our current knowledge (108). This collection of consistent parameters is called the feasible set.
Model predictions

With Eq. 6 and the feasible set of parameters, we were able to calculate, for each germline, a prediction of the transit time to each position along that germline, $t(x)$. These maps from position to time were used to transform the spatial profiles of dpMPK-1 fluorescence intensity to dynamic profiles for each germline. Essentially, dpMPK-1 intensity at each position in a germline was plotted against the time at which cells arrive at that position. The result was a picture of the time course of MPK-1 signaling that germ cells experience/produce as they move through that germline (Fig. 4). Doing this for multiple germlines allowed us to compare signaling dynamics across germlines.

Comparing the estimated time courses of dpMPK-1 from multiple germlines shows that they approximately collapse. It was not obvious a priori that this would be the case from the raw spatial profiles of dpMPK-1. Additionally, naively plotting dpMPK-1 intensity against normalized arc length for each germline does not cause the profiles to collapse (Fig. A5). The collapse of these curves suggests that, despite varying geometries and dpMPK-1 gradients among germlines, germ cells see stereotypical dynamics of MPK-1 activation during meiotic progression.

The average time at which germ cells in different germlines first begin to be exposed to active MPK-1 (30.5 h, standard deviation 1.5 h) and the time of peak dpMPK-1 concentration (mean 37.7 h, standard deviation 1.8 h) are very similar across germlines.
These averages were taken over all germlines and all parameter combinations in the feasible set. The uncertainties are dominated by the uncertainty in the rate of apoptosis $r$. Future experiments that constrain the range of possible values for the parameter $r$ will dramatically reduce the uncertainty of this model’s estimates.

The model also makes quantitative predictions about the relationships between physiological parameters in an animal. From Eqs. 6 and 7, one can immediately see that because ratios of physiological parameters govern the behavior of $t(x)$, numerous values of the individual parameters $N_{tot}, s$, and $r$ can give rise to the same dynamics in the germline. Additionally, looking at the dependence of $\tau = t(x_{max})$, the average transit time from the distal tip to the loop, on the lumped parameter ratios, is also informative (Fig. 5). Consider a situation in which the rate of

**Figure 4:** A) Raw spatial profiles of dpMPK-1 fluorescence intensity in six germlines from a single experiment, imaged under identical conditions. B) For each germline, the function $t(x)$ relating arc length positions to the transit times to reach those positions was estimated using Eq. 5. Plotted are the results of transforming the spatial dpMPK-1 profiles in A into average time courses of dpMPK-1 seen by traveling cells.
ovulation decreases, causing germ cells to spend more time in the region distal to the loop. In order to maintain steady state, the animal’s physiological parameters must change. According to the model, steady state can be achieved by: 1) decreasing the rate of cell production, \( s \), 2) increasing the number of cells in the distal region \( N_{tot} \), or 3) increasing the rate of apoptosis, \( r \). The quantitative model predictions are therefore in line with qualitative intuitions.

Thus, during the first wave of MPK-1 activation in the germline, each of the developing germ cells is exposed to high levels of MPK-1 signaling for several hours, during which MPK-1 activation is translated into changes in a large number of processes, including apoptosis, meiotic progression and membrane biogenesis.

**DISCUSSION**

Cell fate decision processes rely on the joint dynamics of multiple molecular players. Some of these species may be difficult to follow in real time, calling for approaches for reconstructing dynamics based on snapshots from fixed samples. In the Drosophila melanogaster embryo, cell signaling and gene induction dynamics were estimated from fixed samples by matching embryo

---

**Figure 5:** Dependence of \( \tau \), the average transit time from the distal tip to the loop, on the two lumped parameter ratios in the model. The first, \( N_{tot}/s \), is the ratio of the number of cells in the germline to the cell production rate. The second, \( r/s \), is the ratio of the rate of apoptosis to the cell production rate, or the fraction of cells that undergo apoptosis.
morphology to time (41, 42, 87). Additionally, a number of recent articles demonstrated that for a stationary process, in which cells progress unidirectionally through a sequence of states, average dynamics can be reconstructed from fixed cells. For instance, Kafri and colleagues (88) used data from a population of unsynchronized, fixed vertebrate cultured cells to reconstruct the average growth dynamics of an individual cell. In this work, each cell was labeled for DNA and Geminin (a protein degraded during mitosis), and this pair of signals provided a quantitative measure of cell cycle progression for each cell. The large numbers of cells available to them in each experiment enabled estimation of cell size distributions throughout the cell cycle and revealed a feedback mechanism that reduces the variability in cell sizes. A similar approach was used by Akopyan and colleagues to reconstruct the dynamics of the mitotic entry network (89). In both of these studies, reconstruction of dynamics from snapshots relied on the introduction of some measure of progress, which was critical for temporal ordering of data from multiple individual cells. While this step is essential for a dataset extracted from cells that are independent of each other, it is not necessary for data from cells within the germline, where the arc length provides a natural ordering coordinate.

Assuming that the number and spatial arrangement of cells are invariant over time, we derived a transport model, in which cells move similar to cars on a packed highway, to relate arc length, rather than the cell state, to time. In doing so, we used literature estimates of the physiological parameters in the germline (68, 70, 105, 107). Finally, we pooled data from multiple germlines to estimate the average local rates of cell death. This led to the average time course of MPK-1 activation within a developing germ cell. Looking at this pulse, one can ask, what controls the onset of MPK-1 activation, its amplitude, and duration? It is known that this phase of MPK-1 signaling relies on a spatially uniform insulin receptor, but the relevant ligand and the mechanisms responsible for the intracellular control of the duration of MPK-1 activation are still unknown. As
the molecular candidates for controlling these processes are proposed, such as a localized extracellular ligand, our approach can be used to probe their quantitative contributions to multiple aspects of MPK-1 signaling.

The estimates derived from our model-based approach of cell transit times and MPK-1 activation dynamics remain to be verified by data from studies with live reporters of MPK-1 activation. At this point, these estimates provide the only available insight into the dynamics of a key regulatory signal in one of the most advanced experimental systems for studies of cell fate decisions. As measurements of the physiological parameters in the model become more precise, the model may be updated and its uncertainties reduced. Applying this approach to germlines at other steady states under altered conditions is relatively straightforward. Changes in the rate of cell division, rate of cell death, total number of cells, or time cells take to reach the loop can cause germlines to operate at different steady states. These alternative steady states can be achieved in the lab setting by changing the environmental conditions (temperature, crowding, or presence of a noxious substance) or by introducing genetic perturbations. However, for analysis of mutant backgrounds, the assumptions that the germline is at steady state, that the dpMPK-1 gradient is time invariant, and that the distribution of germ cells is not changing over time must be verified. Additionally, at least three of the four physiological parameters for that mutant background must be measured: the average rate of cell production, the average number of cells in the germline, the average rate of apoptosis, and the average time cells take to travel from the distal tip to the loop.

Since this chapter was published in article form in 2015, a few articles have reported live imaging of the mitotic region of the germline (102, 109). These results suggest that it may be possible in the near future to verify the predictions made by our transport model, particularly those about the local, time-averaged velocities of cells in the germline. The main experimental challenge is
imaging multiple parts of the germline for long enough periods of time to get reliable time-averaged estimates of the local velocities, without harming in the animal in some way that disrupts germ cell development.
Chapter 2: The Design Space of the Embryonic Cell Cycle Oscillator

In collaboration with Moshe Sheintuch

1Department of Chemical Engineering, Technion - Israel Institute of Technology, Haifa, Israel

ABSTRACT

One of the main tasks in the analysis of models of biomolecular networks is to characterize the domain of the parameter space that corresponds to a specific behavior. Given the large number of parameters in most models, this is no trivial task. We use a model of the embryonic cell cycle to illustrate the approaches that can be used to characterize the domain of parameter space corresponding to limit-cycle oscillations, a regime that coordinates periodic entry into and exit from mitosis. Our approach relies on geometric construction of bifurcation sets, numerical continuation, and random sampling of parameters. We delineate the multidimensional oscillatory domain and use it to quantify the robustness of periodic trajectories. While some of our techniques explore the specific features of the chosen system, the general approach can be extended to other models of the cell cycle engine and other biomolecular networks.

INTRODUCTION

Starting from the early 1990s, computational models of the embryonic cell cycle have led the way in the systems-level analysis of biomolecular networks (110-112). The modeling approach to the cell cycle and other biochemical networks starts by converting the information about the established components and interactions into a dynamical system. Usually, at this stage, the aim is to build a model that qualitatively predicts the desired behavior(s). This is followed by analytical or computational analysis of the model, with the minimal requirement to find a vector of model
parameter values that yields the desired behavior, such as periodic dynamics with the right period and waveform. As a rule, this vector is not unique and belongs to a set of parameter vectors that are equally successful in describing the data (113, 114). Ideally, one would like to characterize this entire set, which would make it possible to explore functional capabilities of the model only on the basis of the included components and processes.

Since even the simplest models of biochemical networks contain a large number of parameters, characterizing their design spaces is a nontrivial task and most early modeling studies were limited to providing one or several parameter sets consistent with the desired dynamics. Note that even this step might be quite challenging, although more and more algorithms are being proposed for this purpose (114, 115). At the same time, rapid improvements in computational power and numerical methods make it possible to systematically sample the design space of a model and enable the types of analyses that had previously been very difficult or impossible (64, 116, 117). For instance, one might be interested in determining the parameter vectors that can withstand the maximal possible perturbations without disrupting the desired dynamics. In this paper, we try to accomplish these tasks by probing the design space of a mathematical model of the embryonic cell cycle, aiming to highlight the issues that might be relevant for a broad class of models.

Our model is a simplified version of earlier descriptions of cell cycles in Xenopus egg extracts, one of the leading experimental systems for studies of cell cycle regulation (118). The first model of this system was proposed in 1993 by John Tyson and Bela Novak, shortly after the elucidation of the core processes responsible for the periodic dynamics in the activity of the cyclin-dependent kinase (CDK) (119, 120). Their model accounts for steady synthesis of cyclin, association with CDK, reversible activation of the cyclin-CDK complex by phosphorylation, and proteolytic degradation. The model has 10 variables, which could be reduced to two under certain
assumptions, and about 20 parameters that correspond to the rates of protein synthesis, strengths of protein-protein interactions, and rate constants of enzymatic reactions. The authors found a set of parameter values that successfully described the periodic activity of CDK and made several predictions that have been confirmed by subsequent experiments (121). Starting from 2003, James Ferrell and his group initiated a systematic effort aimed at quantifying the main regulatory interactions in this model. These studies led to updated models with two variables and about 10 parameters (63, 122-124). Note that none of the current models can be viewed as a “first-principles” description of the real system, which is still only partially understood. For example, a recent study investigated the role of the dynamic control of cyclin synthesis, which was assumed to be constant in the first generation of models (125).

One of the key insights of the original Tyson-Novak model and its descendants is that the observed CDK dynamics can be viewed as relaxation oscillations, a periodic trajectory with fast and slow motions (126). Furthermore, quantitative changes in the values of model parameters can lead to qualitative changes in the long-term dynamics, from oscillations to either unique steady states or bistability. Here, we have coarse-grained interactions considered in previous models of the cell cycle, producing a model that is amenable to analytical treatment. We show how the design space of this model can be delineated using a combination of analytical and computational tools and used to explore the robustness of the oscillatory regime. Our results provide new insights into the design principles leading to robust oscillations and can be extended to more complex models of cell regulation systems.
RESULTS

Model description and nondimensionalization

We consider a model with two species: $I$ and $A$, corresponding to the inactive and active forms of the cyclin-CDK1 complex, respectively (Fig. 6A). Similar to previous models (63, 119, 123), the complex appears at a constant rate in the active form, reflecting synthesis of cyclins, their rapid association (127) with a large pool of CDK1 (128, 129), and rapid phosphorylation of the cyclin-CDK complex by the CDK-activating kinase (CAK). The inactive form is converted into the active form in a process that is promoted by $A$, reflecting cyclin-CDK’s double negative feedback loop with Wee1/Myt1 (54-56, 130-133) and positive feedback loop with Cdc25 (57-60, 134-136). $A$ is converted back to $I$ with first-order kinetics. Both forms are degraded in a process that is also promoted by $A$, reflecting activation of the anaphase-promoting complex (APC) by active cyclin-CDK and the subsequent degradation of cyclins by the APC (61, 62, 137-140).

These processes are modelled by the following system of differential equations:

\[
\frac{dI}{dt} = -k_d(A)I - (k_a(A)I - k_iA), \tag{1}
\]

\[
\frac{dA}{dt} = s - k_d(A)A + (k_a(A)I - k_iA), \tag{2}
\]

where $s$ is the rate of synthesis of $A$, $k_d(A)$ is the function describing the rate constant for complex degradation, $k_a(A)$ is the corresponding function for complex activation, and $k_i$ is the rate constant for the complex deactivation. Following experimental studies, the functional forms of $k_d(A)$ and $k_a(A)$ are modeled as Hill nonlinearities (63, 122, 123), each of which is characterized by a basal value, threshold, sharpness, and gain (Fig. 6B):
To begin the analysis, we first make problem dimensionless. Rescaling $A$ and $I$ by $s/k_d$ gives the dimensionless variables $a$ and $i$, and rescaling time such that $\tau = k_d^{-1} t$ gives the dimensionless time. The dimensionless equations take the following form:

$$\frac{di}{d\tau} = -(1 + \beta d f(a))i - \frac{1}{\epsilon} \left( (\mu + \beta_a g(a))i - a \right)$$  \hspace{1cm} \text{(5)}$$

$$\frac{da}{d\tau} = 1 - (1 + \beta d f(a))a + \frac{1}{\epsilon} \left( (\mu + \beta_a g(a))i - a \right).$$  \hspace{1cm} \text{(6)}$$

$\textbf{Figure 6:}$ Model reaction diagram and functional forms of regulatory rate constants. A) Reaction diagram depicting the synthesis of active cyclin-CDK, $A$, conversion into the inactive form, $I$, inactivation of the active form, and degradation of both forms. The rate constants of activation and degradation depend nonlinearly on the concentration of the active form. B) Schematic showing the dependence of the activation and degradation rate constants $k_a(A)$ and $k_d(A)$ on $A$. The parameters $k_a^-$ and $k_d^-$ are the basal values for activation and degradation rate constants when the concentration of $A$ is low. The rate constants reach their half-maximal values when $A = \theta_a$ or $A = \theta_d$, respectively, and the maximal increases are denoted by $\Delta_a$ and $\Delta_d$ (Eqs. 3 and 4).
Figure 7: Phase plane analysis and the emergence of oscillations. A) Nullcline for $a$ in the limit that $\varepsilon$ goes to zero and $n_a$ is large. Arrows depict the direction of change in $a$ over time on each side of the nullcline. B) Exact nullcline for $a$, with the same parameter values as in A). The limit points of the $a$-nullcline are marked with arrows. C) Nullcline for $w$ in the limit that $\varepsilon$ goes to zero and $n_d$ is large. Arrows depict the direction of change in $w$ over time on each side of the nullcline. D) Exact nullcline for $w$, with the same parameter values as in C). E) When the nullclines for $a$ and $w$ intersect in the transition regions of their respective Hill functions, the system has a unique, unstable steady state and exhibits sustained oscillations. The thick red line is the $a$-nullcline, and the thick black line is the $w$-nullcline. The thin black lines are trajectories started from different initial conditions, which all approach a limit cycle at long times. F) The time course of oscillations on the limit cycle in E). Here, $\varepsilon$ is very small, generating a large separation of time scales and saw tooth-shaped oscillations.
In these equations, \( f(a) \equiv \frac{a^{n_d}}{y_d^{n_d} + a^{n_d}} \) and \( g(a) \equiv \frac{a^{n_a}}{y_a^{n_a} + a^{n_a}} \) are the rescaled Hill functions. After rescaling, 8 dimensionless parameters remain, which are defined in Table I. Adding Eqs. 5 and 6 gives the equation for the dynamics of the total amount of complexes, \( w = a + i \):

\[
\frac{dw}{d\tau} = 1 - (1 + \beta_d f(a))w. \tag{7}
\]

Going forward, we will analyze the \((a, w)\) system, since \(i\) is determined by mass conservation.

**Phase plane analysis in the limit of strong time scale separation**

The first insights into the dynamics are provided by phase plane analysis in the limit of strong time scale separation, when the interconversion between \(i\) and \(a\) occurs much faster than changes in the total amount of complexes. In this limit, \(\varepsilon \ll 1\), \(a\) becomes a fast variable, and Eq. 6 becomes:

\[
\varepsilon \frac{da}{d\tau} = (\mu + \beta_a g(a))(w - a) - a. \tag{8}
\]

The nullcline for \(a\), found by setting the time derivative to zero, is:

\[
a = \frac{\mu + \beta_a g(a)}{1 + \mu + \beta_a g(a)} w. \tag{9}
\]

When the shape of the Hill nonlinearity in \(g(a)\) approaches a step function, this nullcline can be approximated by a piecewise linear function (Fig. 7A). When \(a < \gamma_a\), \(g(a) \approx 0\), and \(a \approx \frac{\mu}{1+\mu} w\).

On the other hand, when \(a > \gamma_a\), \(g(a) \approx 1\), and \(a \approx \frac{\mu + \beta_a}{1 + \mu + \beta_a} w\). The two straight lines are separated by a discontinuity at \(a = \gamma_a\). A similar analysis leads to a piecewise linear approximation for the second nullcline: \(w \approx 1\), when \(a < \gamma_d\), and \(w \approx \frac{1}{1 + \beta_d}\) when \(a > \gamma_d\) (Fig. 7C).
For certain parameter values, the $\alpha$-nullcline takes the form of an S-shaped curve, where three distinct steady values of activity ($\alpha$) are possible for a particular range of total complex concentrations ($\omega$). Phase plane analysis shows that all steady states located between the extrema of this curve are unstable. At the same time, the nullcline for the slow variable $\omega$ always remains a single-valued function. In the limit of strong time scale separation, oscillations in our model exist when the two nullclines intersect only once and when this intersection is located between the limit points of the $\alpha$-nullcline (Fig. 7E,F, Table 3) (126, 141).

**Delineating the oscillatory domain**

As mentioned above, oscillations exist when the nullclines intersect only once and between the limit points of the $\alpha$-nullcline. This criterion can be expressed as two inequalities: the steady state value of $\alpha$ ($\alpha_{ss}$) must be less than the value of $\alpha$ at the upper limit point ($\alpha_+$) and greater than that at the lower one ($\alpha_-$) (see Fig. 7B), or $\alpha_- < \alpha_{ss} < \alpha_+$. We derived analytical expressions for these inequalities in terms of parameters only for the limit of $\epsilon \ll 1$.

While the full derivation can be found in the supplementary information, the outline of the derivation is as follows. We need to translate the above conditions on the state variables to conditions on the parameters. First, we derive the locations of the extrema in the $\alpha$-nullcline in terms of parameters and derive conditions for when these extrema exist. Then, we want the steady state of the system to lie between these extrema. When the steady state lies exactly on an extremum, the system is at the boundary between oscillatory and non-oscillatory behavior. We find these boundaries by making the $\omega$-nullcline intersect the $\alpha$-nullcline at a limit point, giving one equation for each limit point, each of which can be expressed in terms of parameters only.
These equalities define two hypersurfaces that bound the domain of oscillations (i.e. $a_{ss} = a_-$ and $a_{ss} = a_+$).

In terms of parameters, $a_- < a_{ss} < a_+$ is satisfied when $\gamma_{d,-} < \gamma_d < \gamma_{d,+}$, where

$$\gamma_{d,\pm} = \gamma_a \left( \frac{g_\pm}{1-g_\pm} \right)^{1/n_a} \left( \frac{(1+\beta_d)n_a\beta_a\gamma_a g_\pm(1-g_\pm)}{(\mu+\beta_a g_\pm)^2-n_a\beta_a\gamma_a g_\pm(1-g_\pm)} \right)^{1/n_a} \left( \frac{g_\pm}{1-g_\pm} \right)^{1/n_d}. \quad (10a,b)$$

In this expression, $g_\pm = g(a_\pm)$, and

$$g_\pm = -\frac{(1-\eta\pm2\mu\pm n\pm\beta_a(\eta\pm1-n\pm2)\pm(\mu\pm\beta_a+1))}{2(\beta_a+n\pm)}. \quad (11a,b)$$

The expressions in Eqs. 10a,b define two 7-dimensional hypersurfaces in the 8-dimensional parameter space that enclose the domain of oscillations, and any $\gamma_d$ between them will produce oscillations. These equations fully characterize the oscillatory domain in the small-$\epsilon$ regime.

What does the region of oscillations in the small-$\epsilon$ regime look like? We are limited to cross-sections of the domain to gain insights by visualization. Holding the values of all parameters but $\gamma_a$ and $\gamma_d$ fixed, we used Eqs. 10a,b to trace out the boundaries of the oscillatory region in a two-parameter diagram (Fig. 8A). In this cross-section, we can see that the region of oscillations is a single simply-connected domain. The accuracy of the above analytical expression, valid in the limit $\epsilon \ll 1$, can be evaluated by comparing with the results of numerical bifurcation analysis, which traces the domains of oscillatory and other types of solutions. The numerically-calculated and analytically-derived boundaries show good agreement when $\epsilon$ is small (Fig. 8A) and deviate as $\epsilon$ increases above 0.01 (Fig. B1). Furthermore, numerical continuation reveals that in this cross-section there are two disjoint regions in which the system is bistable (Fig. 8B).
Figure 8: Analytical and numerical two-parameter bifurcation diagrams. A) Comparison of the numerically-computed and analytically-derived regions of oscillations in the \((\gamma_a, \gamma_d)\) plane, for the same parameters as in Fig. 7. The light gray patch is the region of oscillations determined by numerical continuation, and the green and purple lines are the analytically-derived boundaries in the limit that \(\varepsilon \to 0\). At this small value of \(\varepsilon\), the results are indistinguishable. The upper boundary (green line) comes from the constraint that the \(w\)-nullcline intersects the \(a\)-nullcline at its upper extremum, and similarly for the lower boundary (purple line). These conditions are shown graphically in the insets. See also Fig. B1. B) Numerically-calculated two-parameter bifurcation diagram showing the regions of oscillations and bistability in the \((\gamma_a, \gamma_d)\) plane for the same parameter set. The light gray patch is again the region of oscillations, the dark gray lines are Hopf bifurcation points, and the black lines are saddle-node bifurcation points. The insets without borders show nullclines and steady states for selected parameter sets in the 2-parameter diagram. In these insets, solid red lines are the nullclines for \(a\), solid black lines are the nullclines for \(w\), filled dots are stable steady states, and open dots are unstable steady states. The inset bordered by a dashed black line zooms in on the bottom-left portion of the diagram where there is a small region of bistability.
Exploring the full parameter space by sampling

In this section we numerically explore the parameter space by randomly sampling and show how this analysis can produce parameter relationships that improve the probability of generating oscillations. In the two-parameter cross-section of the domain of oscillations (Fig. 8), we noticed that the two threshold parameters bounding the oscillatory domain, $\gamma_a$ and $\gamma_d$, appeared to be highly correlated—oscillations were unlikely when $\gamma_a/\gamma_d$ was far from 1. To find other correlations among model parameters relevant to oscillations, we sampled from the 8-dimensional parameter space of the full model (Eqs. 6 and 7). Details about how the sampling was performed can be found in the Methods section. This screen found 1 oscillatory parameter set in every 1000 samples, ultimately producing ~75,000 oscillatory parameter sets.

Inspecting the one- and two-dimensional marginal distributions of the parameter samples (Figs. B2 and B3), we first verified features we expected from theory. First, oscillations were more likely when $\varepsilon$ was small, meaning there is a strong separation of time scales between the rates of cyclin activation/deactivation and the rate of cyclin degradation. We also verified that the correlation between $\gamma_a$ and $\gamma_d$ along the boundaries, observed in the two-parameter cross-section, was a general feature of the model’s oscillatory domain. From the bivariate distribution of $\gamma_a$ and $\gamma_d$, oscillations are likely to emerge when $\gamma_d \approx \gamma_a$ and disappear when the two threshold values differ significantly (Fig. 9A). This should come as no surprise: when $\gamma_a$ and $\gamma_d$ are similar, the extrema of the $a$-nullcline and the transition of the $w$-nullcline occur at roughly the same value of $a$, almost guaranteeing that the $w$-nullcline will intersect the $a$-nullcline between the extrema. When these are equal, the remaining parameters can vary considerably while still maintaining the oscillatory function.
The marginal distributions also provided insights into how to choose the remaining parameters in a way that would tend to generate oscillations (Figs. 9, B2, and B3). For example, \( n_d \), the exponent for the degradation rate constant, could be arbitrarily large, corresponding to a sharp, horizontal transition in the \( w \)-nullcline. However, the univariate distribution of \( n_a \), the exponent in the activation rate constant, peaked at about \( n_a = 3 \). The bivariate distribution showed dependence between \( n_a \) and \( n_d \) (Fig. 9B): larger values of \( n_d \) were necessary for larger values of \( n_a \) to generate oscillations, whereas smaller values of \( n_a \) could generate oscillations for most values of \( n_d \). We also noticed correlations between \( \mu \) and the two threshold parameters, \( \gamma_a \) and \( \gamma_d \) (Fig. 9C). If \( \mu \) is small relative to \( \gamma_a \), then the \( a \)-nullcline will intersect the \( w \)-nullcline below the first extremum of the \( a \)-nullcline, producing a stable steady state. If \( \mu \) is large relative to \( \gamma_a \), the extrema of the \( a \)-nullcline occur at smaller values of \( w \), and fewer values of the remaining parameters cause the nullclines to intersect between the extrema. The correlation between \( \mu \) and \( \gamma_d \) is the result of both parameters being correlated with \( \gamma_a \). Finally, a nonlinear relationship was found between \( \mu \) and \( n_a \) (Fig. 9D). This relationship could be explained by the conditions under which the \( a \)-nullcline has extrema, which are necessary for oscillations and were derived in the process of deriving Eqs. 10a,b. Overlaying the boundary between the regions where the \( a \)-nullcline does and does not have extrema (in the limits that \( \beta_a \to \infty \) and \( \epsilon \to 0 \)) (Eqs. B1.18 and B1.19), it becomes clear that this is the source of the relationship between \( \mu \) and \( n_a \).
Figure 9: Selected bivariate distributions of the sampled oscillatory parameter sets. The color denotes the probability density at each point, with lighter colors being higher in value. At pairs of parameter values where the density is higher, a larger space of the remaining 6 parameters generates oscillations. See also Figs. B2 and B3. A) The joint distribution of $\log_{10}(\gamma_a)$ and $\log_{10}(\gamma_d)$ shows that oscillations are more likely when the two are very close in value. B) The bivariate distribution of $n_a$ and $n_d$ shows that increasing $n_d$ makes larger values of $n_a$ capable of generating oscillations. C) The joint distribution of $\log_{10}(\gamma_a)$ and $\log_{10}(\mu)$ shows that oscillations are also more likely when the threshold parameter $\gamma_a$ is close in value to the basal activation rate $\mu$. The same trend was observed between $\gamma_a$ and $\mu$. D) The joint distribution of $n_a$ and $\log_{10}(\mu)$. The white dashed line is the boundary separating the regions where the $\alpha$-nullcline can and cannot have extrema, in the limits that $\beta_a \to \infty$ and $\epsilon \to 0$ (Eq. B1.19). Although this boundary was derived for small $\epsilon$, no sampled parameter vectors crossed it, even those with arbitrary values of $\epsilon$. 

The high density of oscillatory parameter sets at small values of \( \varepsilon \) suggested that our analytical approximations made in the previous section would be good predictors of oscillations. Of the 75,000 samples, 96.7% satisfied the analytical conditions that the nullclines intersect between the extrema of the \( \alpha \)-nullcline, which was derived for small \( \varepsilon \) (Eqs. 10a,b). This meant that a large portion of the oscillatory domain lay inside the analytically-derived domain. However, containment does not imply equivalence. It could be that a large portion of the analytical domain contains parameter sets that do not generate oscillations. To determine how well the analytically-derived domain approximates the true oscillatory domain, we sampled from the analytically-derived domain and determined whether each sample actually generated oscillations or not. Of about 13,000 samples, 38.4% actually generated oscillations. This demonstrated that the analytical domain delineated by Eqs. 10a,b contains almost all of the true oscillatory domain but overestimates its volume by a factor of \(~2.5\) over the range of parameter values considered here.

**Exploring the robustness of oscillatory solutions**

With many samples from the domain of oscillations in hand, we were then able to ask questions about functions over this set. In particular, to which perturbations is the oscillatory behavior most robust? As mentioned above, the domain of oscillations appears to be narrow and extended along the \((\gamma_a, \gamma_d)\) direction, suggesting that the oscillatory behavior may withstand considerable perturbations along this direction. Recall also that \( \gamma_a \equiv \frac{\theta_a}{s/k_d} \) and \( \gamma_d \equiv \frac{\theta_d}{s/k_d} \) are the only dimensionless parameters that depend on the rate of cyclin synthesis \( s \), so perturbations in the \((\gamma_a, \gamma_d)\) direction are equivalent to perturbations in the synthesis rate. Robustness of oscillations to variations in cyclin synthesis would be desirable, given that the assumption of constant synthesis rate has been shown to be a simplification (125).
Dividing the rate of synthesis by a perturbation factor $f$, we asked what are the largest and smallest values of $f$ for which each parameter set produces oscillations. Note that dividing $s$ by $f$ is equivalent to multiplying both thresholds by $f$. Taking a subsample of 1000 parameter sets from the point cloud generated above, we calculated the upper and lower bounds of $f$ for which oscillations still existed via the bisection method. We defined a measure of robustness to this perturbation as:

$$\delta \equiv \frac{s_{\text{max}}}{s_{\text{min}}} = \frac{f_{\text{max}}}{f_{\text{min}}}$$  \hspace{1cm} (12)

where $s_0$ was the “initial” value of $s$ for the sampled parameter set, $s_{\text{max}} \equiv s_0/f_{\text{min}}$ was the largest value of $s$ that generated oscillations, and $s_{\text{min}} \equiv s_0/f_{\text{max}}$ was the smallest. This measure of robustness does not depend on the particular value of $s_0$; rather, it measures the “width” in log space of the oscillatory domain along the $(\gamma_a, \gamma_d)$ direction at the sampled point. Additionally, although $f_{\text{min}}$ and $f_{\text{max}}$ depend on the parameter vector from which the bisection method was started, $\delta$ does not. Therefore, $f_{\text{min}}$ and $f_{\text{max}}$ can be rescaled such that $f_{\text{min}} = 1$, without changing the value of $\delta$. Fig. 10A shows the histogram of $\delta$ values among the subsampled points.
Figure 10: Robustness of oscillations to perturbations in the rate of cyclin synthesis, and analysis of the most robust parameter vector. 

A) Histogram of the measure of robustness \( \delta \) to perturbations of the synthesis rate \( s \) for 1,000 points in the domain of oscillations. \( \delta \equiv \frac{s_{\text{max}}}{s_{\text{min}}} = \frac{f_{\text{max}}}{f_{\text{min}}} \) is the ratio of the largest and smallest values of \( s \) that generate oscillations for each parameter set. 

B) One-parameter continuation diagram in \( f \) for the parameter vector with the largest value of \( \delta \) (Table B1). Note that the synthesis rate \( s \) is inversely proportional to \( f \). The solid lines denote stable steady states, the dashed line denotes unstable steady states, and black dots denote stable limit cycles. The amplitude of oscillations in \( a \) are roughly constant (see also Fig. B4).

C) Period, \( T \), versus the continuation parameter \( f \) for the parameter vector with the largest value of \( \delta \). At smaller \( f \), the period is relatively insensitive to the value of \( f \), but increases as \( f \) increases. Additionally, at the bifurcation points, the period diverges, corresponding to infinite-period bifurcations. Note that in Figs. 10B,C, \( f \) was scaled such that \( f_{\text{min}} = 1 \).
We then asked what is the most “robust” set of parameter values according to this measure (Table B1), and how do the period and amplitude of oscillations of the active cyclin-CDK complex depend on $f$? For this parameter set, $\delta \sim 52.8$, indicating that the largest value of the synthesis rate $s$ that generates oscillations is about 53 times larger than the smallest one. The parameter vector also exhibited many of the heuristic properties that, from sampling, were found to increase the likelihood of oscillations: $\beta_d$ was large, $n_a$ was close to 3, $n_d$ was large, $\mu$ and the thresholds were of similar value, and $\varepsilon$ was small. This suggested that the parameter vector would be robust to other parameter variations, as well.

Performing numerical continuation in $f$ for this parameter vector, we found that the amplitude of oscillations in the active component, $a$, remained roughly constant over the range of viable values of $f$ (Fig. 10B). Since this parameter vector fell in the small-$\varepsilon$ regime, we were able to use an analytical argument to explain this invariance to $s$ (see supplementary information). In this regime, the limit cycle nearly traces the shape of the $\alpha$-nullcline, with jumps at the limit points (Figs. 7E and B5). Using this observation and our expressions for the positions of the limit points (Eqs. B2.4a,b), we derived an approximate expression for the amplitude of oscillations in $\alpha$:

$$\Delta a = \gamma a n_a \beta a \left( \frac{\mu + \beta a}{1 + \mu + \beta a} \frac{g^- (1 - g^-) (g^-)^{\frac{1}{n_a}}}{(\mu + \beta a g^-)^2} - \frac{\mu}{1 + \mu} \frac{g^+ (1 - g^+) (g^+)^{\frac{1}{n_a}}}{(\mu + \beta a g^+)^2} \right),$$  \hspace{1cm} (13)

where $\Delta a$ is the amplitude of oscillations in $\alpha$ and $g_{\pm}$ are given by Eqs. 11a,b. Equation 13 is linear in $\gamma a$, which along with $a$ is the only term scaled by $s$ that appears in the expression. As a result, this equation predicts that variations in $s$ should cancel out, leaving the amplitude insensitive to the rate of cyclin synthesis. Fig. B4 shows that Eq. 13 is a good predictor of the amplitude of oscillations in $\alpha$ for most of the oscillatory parameter vectors found by sampling.
Finally, the period of oscillations, $T$, for the most robust parameter vector was relatively insensitive to the value of $f$ at smaller $f$, but increased at larger values of $f$. At the bifurcation points, the period diverged, corresponding to saddle-node infinite period bifurcations (142), which have been previously observed in cell cycle models (63, 119) (Fig. 10C). The loss of limit cycles at these bifurcations is not due to a change in stability of the unique steady state (a Hopf bifurcation), but rather due to the appearance of a new, stable steady state elsewhere in the phase space. These calculations indicate that oscillations of robust period and amplitude can be achieved at higher synthesis rates (smaller $f$) for this parameter set.

**DISCUSSION**

Studies of the embryonic cell cycle are prime examples of how mathematical models of biochemical networks are valuable for summarizing current understanding, testing assumptions and guiding experiments, and exploring networks’ functional capabilities (63, 119, 120, 122-124). Typically, specifying the interactions in the model is not sufficient to generate a particular functional behavior, which also depends on the values of the system parameters. Ideally, one could characterize the parameter domains with desired functions in order to explore a model’s functional behaviors and understand under what conditions they arise.

Here, we analyzed a minimal set of interactions, based on biochemical experiments, that describe the embryonic cell cycle and is capable of all of the relevant qualitative behaviors—monostability, bistability, and oscillations. The model’s simplicity allowed us to derive analytical expressions for the boundaries of the domain of oscillations in the model’s parameter space, which are valid in the limit of strong separation of time scales. We then explored the design space of oscillations by
sampling the parameter space and screening for oscillations. The marginal distributions of the resulting samples revealed heuristics for choosing parameter vectors that were highly likely to generate oscillations. The bivariate distributions, in particular, effectively acted as two-parameter diagrams averaged over the entire oscillatory domain, as opposed to the cross-sectional view provided by conventional numerical continuation. Sampling also allowed us to check the global validity of our approximate boundaries for the domain of oscillations. Finally, with a large collection of samples from the oscillatory domain in hand, we investigated the robustness of oscillations to variations in the rate of cyclin synthesis. The most robust parameter set satisfied many of the heuristics found from the marginal distributions of oscillatory parameter sets, suggesting it lay deep in the domain of oscillations. Observing that the amplitude of oscillations in the active component was roughly insensitive to changes in the rate of cyclin synthesis led us to derive an approximate analytical expression for the amplitude that explained this finding. Hence, through a combination of analytical and computational techniques, we characterized the design space of this model and a measure of robustness over this region of parameter space.

The fact that our model contained only two dependent variables allowed geometric analysis of the nullclines and an analytical derivation of the domain of oscillations in the limit of strong separation of time scales. We expect the methods demonstrated here to be applicable to any system reducible to two dependent variables (143, 144). Although these features do not generalize to other models, numerical continuation and sampling of the parameter space are viable methods for exploring a model’s functional behaviors and characterizing their parameter domains (64, 116, 117).

Experimental studies have identified several features contributing to robust oscillations in Xenopus embryos. These include bistability in the steady state response of active cyclin-CDK to the amount of total cyclin (122, 145), ultrasensitivity in the positive feedback loop (146, 147), similar
threshold parameters in the Hill functions of the positive and negative feedback loops (63), and a large Hill exponent for cyclin-CDK degradation (63). Our analysis reveals that these features lead to robust oscillations in the Xenopus system, which is a point in the parameter space, because they are also features of the widest parts of the oscillatory region, as measured by the high fraction of oscillatory points that share these features. For example, having similar thresholds, $\gamma_a$ and $\gamma_d$, in the Hill functions describing the positive and negative feedback loops and a large Hill exponent for cyclin-CDK degradation, $n_d$, leaves the remaining parameters with a large degree of flexibility. It is only by considering the entire oscillatory region that this becomes clear.

**METHODS**

**Numerical continuation of bifurcation points**

All numerical continuations (Figs. 8 and 10) were performed using Matcont (148), a numerical continuation software for MATLAB (MathWorks, Natick, MA).

**Parameter sampling**

The model contains eight dimensionless parameters, defined in Table 1. Table 2 summarizes the ranges and scales used to generate samples from the parameter space. Since we were interested in the location and shape of the region of oscillations, ranges were chosen to exclude as much of the non-oscillatory parameter space as possible. In some cases, taking a parameter to a limiting value of zero or infinity would preserve oscillations. In these cases, parameter ranges were chosen such that the model outputs were still somewhat sensitive to the parameters. Finally, whereas all other
dimensionless parameters were sampled on a logarithmic scale, Hill coefficients were sampled on a linear scale since their experimentally-measured values are typically of order 1.

### Table I: Dimensionless parameter definitions

<table>
<thead>
<tr>
<th>Dimensionless Parameter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>( k_a / k_i )</td>
</tr>
<tr>
<td>( \beta_a )</td>
<td>( \Delta_a / k_i )</td>
</tr>
<tr>
<td>( \beta_d )</td>
<td>( \Delta_d / k_d )</td>
</tr>
<tr>
<td>( \gamma_a )</td>
<td>( \theta_a / (s / k_a) )</td>
</tr>
<tr>
<td>( \gamma_d )</td>
<td>( \theta_d / (s / k_d) )</td>
</tr>
<tr>
<td>( n_a )</td>
<td>( n_a )</td>
</tr>
<tr>
<td>( n_d )</td>
<td>( n_d )</td>
</tr>
<tr>
<td>( \varepsilon )</td>
<td>( k_d / k_i )</td>
</tr>
</tbody>
</table>

### Table II: Ranges and scales used for parameter sampling

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Scale</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>log</td>
<td>0.01</td>
<td>10</td>
</tr>
<tr>
<td>( \beta_a )</td>
<td>log</td>
<td>0.01</td>
<td>100</td>
</tr>
<tr>
<td>( \beta_d )</td>
<td>log</td>
<td>0.1</td>
<td>100</td>
</tr>
<tr>
<td>( \gamma_a )</td>
<td>log</td>
<td>0.01</td>
<td>10</td>
</tr>
<tr>
<td>( \gamma_d )</td>
<td>log</td>
<td>0.01</td>
<td>10</td>
</tr>
<tr>
<td>( n_a )</td>
<td>linear</td>
<td>0.01</td>
<td>15</td>
</tr>
<tr>
<td>( n_d )</td>
<td>linear</td>
<td>0.01</td>
<td>15</td>
</tr>
<tr>
<td>( \varepsilon )</td>
<td>log</td>
<td>0.0001</td>
<td>15</td>
</tr>
</tbody>
</table>

Parameter ranges used for sampling and the scale on which sampling was performed uniformly for each parameter (i.e. \( \mu \) was sampled uniformly in logarithm between 0.01 and 10, while \( n_a \) was sampled uniformly on a linear scale between 0.01 and 15).

For each sample, the first step was to find a steady state by a local method and check the linear stability of that steady state. Steady states were found by first solving for \( w \) in \( \frac{dw}{dt} = 0 \) and then plugging that expression into the equation \( \frac{da}{dt} = 0 \). The value of \( \log(a) \) (to ensure positive solutions) that satisfied this expression was found using MATLAB’s `fzero` function, with an initial guess of \( a = 0.1 \).
The linear stability of the steady state was checked by finding the two eigenvalues of the Jacobian matrix, $J_{ij} = \frac{\partial \dot{X}_i}{\partial X_j}$, evaluated at the steady state, where $i, j = 1, 2$, $X_1 = a$ and $X_2 = w$, and the dot indicates a time derivative.

If the found steady state was stable, corresponding to the real parts of both eigenvalues being negative, the parameter set was immediately thrown out, because it could not produce oscillations. If an unstable steady state was found, further tests were performed to check whether the parameter set generated oscillations.

If the steady state was unstable, the system was integrated in time using MATLAB’s ode15s until either 1), the system time reached 1000 units or 2) the time course had exhibited 3 peaks. If the system time reached 1000 units before producing 3 peaks, then it reached a stable steady state. This happened when there were three steady states—either two unstable and one stable or two stable and one unstable—and the local solver happened to find an unstable one. Alternatively, if the system generated 3 peaks before reaching 1000 time units, it could either be due to sustained oscillations or damped oscillations. To check which, the local search for a steady state was repeated from the last point of time-integration. If the resulting steady state was stable, then the oscillations were damped, and the parameter set was discarded. If the steady state was unstable, then the oscillations were sustained.

Finally, if the parameter set corresponded to sustained oscillations, we found the converged limit cycle by numerically solving:

$$\tilde{F}(a_0, w_0, T) = \left[ a_0 - a(T) \right] \left[ w_0 - w(T) \right] = 0, \quad (14)$$
where $a_0$ and $w_0$ were initial points on the limit cycle, $T$ was the period, $a_0$ was fixed, and $w_0$ and $T$ were varied.

<table>
<thead>
<tr>
<th>Dimensionless parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>1/3</td>
</tr>
<tr>
<td>$\beta_a$</td>
<td>2/3</td>
</tr>
<tr>
<td>$\beta_d$</td>
<td>7</td>
</tr>
<tr>
<td>$\gamma_a$</td>
<td>1/8, varied</td>
</tr>
<tr>
<td>$\gamma_d$</td>
<td>1/8, varied</td>
</tr>
<tr>
<td>$n_a$</td>
<td>10</td>
</tr>
<tr>
<td>$n_d$</td>
<td>10</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>varied</td>
</tr>
</tbody>
</table>

**Density Estimation**

Univariate and bivariate marginal distributions were estimated in MATLAB using Z. Botev’s code *kde* and *kde2d*, both available on the MathWorks website:


and


(accessed 12/16/16) (100).
Chapter 3: Ongoing Work

The preceding chapters used mathematical modeling to understand biological dynamics at a systems level. The first chapter looked at Ras/ERK signaling in the distal part of the *C. elegans* germline, where a gradient in the extracellular signal regulates germ cell maturation as the cells move down the gonad tube. Studies of the role of ERK in germ cell development have been limited to fixed samples, which provide snapshots in time. Here, modeling was used to get the most from available data by estimating dynamics from images. The model was built on two key features of the germline: adult germlines are approximately at steady state, and germ cells are arranged in order of maturity. Conservation of mass (or rather cells) lead to derivation of an expression for the transit time for cells to reach a given position and hence the time course of active ERK they see as they moved through the germline. This revealed that, although germlines are of different shapes and sizes across animals, the time course of ERK activation seen by cells in different animals is remarkably similar: after 30 hours of travel and development, cells see a pulse of ERK activation that lasts 15 hours. This raised questions about whether cell motion and the spatial profile of ERK signal are somehow coupled to give rise to this invariance.

In the second chapter, a host of analytical and computational analyses were used to learn as much as possible about how oscillations emerge from the network of interactions composing the *Xenopus* embryonic cell cycle. In that system, periodic changes in the level of the catalytically-active complex of cyclin B and CDK1 control the timing of cell divisions. Active cyclin-CDK regulates its own activity via positive and negative feedback loops, but this does not guarantee sustained oscillations. A simple model of the system was constructed to understand under what conditions on the model parameters oscillations arise. Analysis of the nullclines identified geometric conditions that would produce oscillations, and in the limit that a parameter controlling a
separation of time scales went to zero, these conditions could be expressed analytically. In this limit, the 8-dimensional oscillatory domain in parameter space was enclosed by two 7-dimensional nonlinear hypersurfaces, and the accuracy of the analytical expression was verified in a two-parameter cross section via numerical continuation. To explore the full oscillatory domain, random samples were drawn from the parameter space, and the bivariate distributions of oscillatory parameter vectors were used to find relationships that enrich for oscillations and confer robustness to parameter variations. In particular, robustness of the system to variation in the rate of cyclin synthesis was explored. Finally, parameters previously measured experimentally for the *Xenopus* system obey many of the relationships found to enrich oscillations, helping explain the robustness of this system to parameter variations.

**BRIDGING SYSTEMS AND MECHANISMS**

Going forward, we need to bridge the gap between systems-level studies and biochemistry, the individual reactions that occur in a cell. This requires experiments in which most species can be measured and their interactions are well-understood. One approach is to tackle small modules at a time, such as the multisite phosphorylation system regulating ERK kinase activity, via *in vitro* experiments combined with dynamical models.

**Figure 11:** Minimal model of ERK regulation. Each arrow is an enzyme-mediated reaction, modeled using mass-action kinetics. A version of this figure appears in Rubinstein et al. (64).
Our recent work explored the dynamical behaviors that are possible in the simplest model of ERK regulation consistent with biochemical studies (64). Sampling the parameter space, we found a surprisingly rich set of dynamical behaviors, including bistability, oscillations, and co-existence of limit cycles. This network consisted of only three mammalian proteins: ERK2; MEK1, the kinase that phosphorylates ERK; and MKP3, a phosphatase that specifically dephosphorylates ERK. Active MEK1 phosphorylates ERK2 in by an ordered mechanism: first it transfers a phosphoryl group from a molecule of ATP to the tyrosine residue Y183 and then to the threonine T185 (149). If phosphothreonine ERK is present in solution, MEK can phosphorylate its tyrosine, but a mixture of only ERK, MEK, and ATP produces very little phosphothreonine ERK. MKP3 dephosphorylation of ERK2 is ordered as well: first MKP3 removes the phosphoryl group from the tyrosine and then from the threonine (150). Similarly, MKP3 is capable of dephosphorylating phosphotyrosine ERK. Hence, this network has a cyclic structure with two additional reaction paths (Fig. 11).

It was found that this simple network is capable of complex behaviors, but which dynamical regime the real system operates in depends on the values of the kinetic parameters in the model. To constrain the model parameters, experimental observations of the kinetics of ERK phosphorylation and dephosphorylation are needed. The Shvarstman laboratory has begun addressing this for the kinase branch of the reaction through a combination of biochemical experiments, mathematical modeling, and parameter estimation techniques. Wild-type mammalian ERK2 is being purified from E. coli transfected with an inducible promoter; the isolated ERK2 is mixed with dually-phosphorylated MEK1 and ATP; and the phosphorylation state of ERK is monitored at discrete times via mass spectrometry. This assay is capable of discriminating all four phosphoforms of ERK. In parallel, methods are being developed to learn model parameters from this data.
The experimental protocol is quite involved and has been optimized to minimize experimental variability. Aliquots are taken from the reaction mixture and diluted with urea to stop the reaction. Each time point is then processed to label the ERK molecules (of all phosphoforms) with a unique barcode peptide. Then, the samples from all time points are mixed together to be injected into the mass spectrometry. However, this only gives the relative amount of each phosphoform of ERK over time, not the relative abundance between phosphoforms, nor their absolute amounts. In order to determine these, a known quantity of labeled standard ERK peptide is also added for each phosphoform, which allows internal calibration of the total amounts. These experiments produce measurements of the absolute concentrations of each ERK phosphostate over time.

PARAMETER ESTIMATION: CHALLENGES AND METHODS

Estimating model parameters from time course data faces several challenges. Kinetic ODE models are typically sloppy, a phenomenon where the contributions of individual reaction rates to model behavior are obscured in the collective system dynamics (113, 115, 151-154). For example, the effect of changing the value of one parameter can often be compensated by changing the value of another, and some processes in the model may occur too fast or slow to be constrained. Sloppiness is exacerbated when some model species cannot be observed experimentally, as is the case in our experiments, which can’t distinguish between free and MEK-bound ERK. Despite large uncertainties in individual parameters, model predictions are often well-constrained. While this may seem paradoxical, it arises because the observed system behavior only depends on a small number of parameter combinations, and the combinations that govern predictions are well-constrained. Which parameter combinations are important is a local property: it depends on where
in the parameter space the system operates. Due to model sloppiness, many parameter vectors make the model predictions fit the observed data, up to measurement uncertainty. As a result, the parameter vector that best-fits the data is not informative for making predictions, and one must consider an ensemble of parameter vectors that fit the data (113, 151).

A first step in parameter estimation is often to construct a cost function that measures the deviation of the model predictions from data, which is then minimized by varying the values of the parameters. A common cost function is the sum of squared differences between model predictions and measurements, or least-squares cost:

\[ C(\hat{\theta}) = \sum_i \left( y_i - f_i(\hat{\theta}) \right)^2, \]  

(1)

where \( y_i \) are the measurements, \( \hat{\theta} \) are the parameters, and \( f_i(\hat{\theta}) \) are the model predictions for each measurement. Minimizing this cost function is a special case of maximum likelihood estimation. The likelihood of a parameter vector \( \hat{\theta} \) is the probability of observing the data if the “true” parameter values were \( \hat{\theta} \), or \( p(\{y_i\}|\hat{\theta}) \). This requires a model for the measurement uncertainty, which is usually assumed to be Gaussian and uncorrelated between measurements. The least-squares cost function above corresponds to a Gaussian noise model, or likelihood function, in which all measurements have the same standard deviation, and \( p(\{y_i\}|\hat{\theta}) \propto e^{-\frac{1}{2}C(\hat{\theta})} \). The next level of generality, which is usually sufficient for fitting time course data, is to allow different levels of uncertainty for each data point, giving rise to the cost function:

\[ C(\hat{\theta}) = \sum_i \left( \frac{y_i - f_i(\hat{\theta})}{\sigma_i} \right)^2. \]  

(2)
Typically, nonlinear local methods, which take into account the derivatives of the cost with respect to parameters, are used to minimize the cost, such as the Levenberg-Marquardt algorithm (115, 152, 155, 156). These require an initial guess of the parameter values, from which incremental changes are made. These methods fail when the derivative of the cost with respect to parameters at the initial guess is nearly zero, so oftentimes one has to try multiple starting points or choose the starting point in an informed way. Although these methods only find locally-optimal solutions, for the cost functions associated with fitting physical models, many local optima of nearly equal cost lie in a single deep, extended valley in the cost function (157). It is uncommon to find multiple deep minima that aren’t connect by a continuous region of similar cost value (excluding the trivial case of when there is symmetry in the parameters) (115).

Minimizing the cost function gives a best-fit parameter set, but really one is interested in an ensemble of parameter vectors consistent with the data, and several methods exist for constructing ensembles. One simple approach is to run a local minimization algorithm from many starting points in the parameter space and keep those with cost value below some threshold (157). This approach can be restrictive for large models, but for models with order 10 parameters it is computationally tractable. The drawback of this algorithm is that it does not fully explore the uncertainty in the parameter values.

Alternatively, a popular method for constructing ensembles is the Markov Chain Monte Carlo (MCMC) algorithm, which produces many parameter vectors of varying qualities of fit in proportion to their likelihood or posterior probability (151, 158). MCMC can be applied to large systems and does not require derivatives of the model predictions with respect to the parameters. Samples are generated by randomly jumping through parameter space and either accepting or rejecting proposed jumps with probability based on the relative likelihoods of the current and
proposed parameter vectors. Variants of this method exist, broadly called simulated annealing, that first execute a wide search of parameter space and gradually narrow it down by proposing smaller jumps and making the acceptance criteria for proposed jumps more restrictive (159, 160). MCMC fully explores the uncertainty in the parameter values, producing fits in proportion to their likelihood of having generated the data.

**PRELIMINARY RESULTS**

In preparation for the incoming experimental data, it would useful to know how many and which parameters we can expect to constrain from a typical time course. This question can be addressed computationally, using the phosphorylation branch of the model in (64). This model denotes the four phosphoforms of ERK as $S_{00}, S_{01}, S_{10},$ and $S_{11},$ where the subscripts refer to the tyrosine and threonine amino acid residues of ERK: 0 denotes unphosphorylated, and 1 denotes phosphorylated. It is assumed that $S_{10}$ cannot be formed, since MEK phosphorylates the tyrosine on ERK first. Additionally, in the model dually-phosphorylated ERK can form a complex with MEK, blocking binding with other ERK phosphoforms. The model contains 8 parameters describing the rates of binding, unbinding, and catalysis. Typical data sets will consist of measurements of the total amount of each phosphoform (free and MEK-bound) at about 10 time points, i.e. $y_i = y(t_i), i = 1, \ldots, 10.$ It was assumed that the measurement uncertainty is Gaussian-distributed and uncorrelated, with constant standard deviation equal to 10% of the total amount of ERK. A reference, “true” parameter vector was found by least-squares fitting the model to observations of ERK phosphorylation assayed with Western blots, as reported by Aoki et al. (65). After corrupting
the “true” model’s predictions with synthetic measurement noise, what can we recover about the parameter values?

A ball of measurement uncertainty surrounds the “true” trajectories, which induces uncertainty in the parameter values. To get at this parameter uncertainty, the following was done: 1) added a random realization of Gaussian measurement noise to the noise-free model predictions, 2) found the best fit to the data using a least-squares solver initiated from a random point in parameter space (drawn uniformly in log), and 3) repeated 1) and 2) with new realizations of noise and new initial guesses 1000 times. This produced an ensemble of best-fit parameter vectors drawn from the joint distribution of possible data and initial parameter guesses. Using this procedure, the parameter vectors are generated roughly in proportion to their likelihood of generating data, but there are systematic biases. The ensemble of points in the 8-dimensional parameter space was then analyzed to see what we can expect to learn about the parameter values from this type of experiment.

As might be expected from model sloppiness, the histograms of individual parameters showed that none of them was well-constrained (Fig. C1). The bivariate plots of the logarithms of parameters showed that the catalytic and unbinding rates associated with the second phosphorylation step were correlated, as were some binding and unbinding rates, suggesting that ratios of those parameters might be constrained (Fig 12A). The two catalytic constants also exhibited a nonlinear relationship, where either one could be very large but usually not both.

A standard technique to assess the uncertainty in parameter estimates is to compute the Fisher information matrix (FIM), a local quantity that lower bounds uncertainties in the parameter values (161-163). The eigenvalues and eigenvectors of the FIM contain information about the uncertainty in linear combinations of parameters: the standard deviation of the parameter estimates along each
eigenvector in parameter space is $1/\sqrt{\lambda}$, where $\lambda$ is the corresponding eigenvalue. The threshold $\sqrt{\lambda} = 1$ approximately separates constrained (identifiable) and unconstrained (practically unidentifiable) parameter combinations, with constrained combinations having $\sqrt{\lambda} > 1$ (164). See Appendix C for more information about the FIM. The FIM was calculated with respect to the logarithm of parameters for each parameter set in the ensemble in order to assess uncertainties in products and ratios of the parameters. Among the samples, between 2 and 3 (out of 8) parameter combinations were constrained, as measured by the number of eigenvalues greater than 1 (Fig. 12B).

Together, these results suggest that probing this model with typical time course data generated by our experimental assay is not sufficient to resolve all of the model’s parameters. This statement rests on the assumption that the model used is adequate for describing the real experimental system.

What is the source of the large degree of parameter uncertainty? Since our experiment cannot distinguish between free and MEK-bound ERK, the dynamics of those species and the parameters controlling them may not have been constrained by the data. To check this, the model predictions for free MEK and the complexes of ERK and MEK over the ensemble were inspected and found to be poorly constrained, as expected. Furthermore, at least two mutually-exclusive limiting regimes were consistent with the data: the binding-limited regime, where the enzyme is always free, and any complex that is formed rapidly disappears; and the enzyme-limited regime, where the enzyme is always fully-bound, and any new enzyme that gets freed up is immediately bound again. The ensemble contained the union of these two regimes with different governing parameters, as well as the space between the limits, which in part contributed to the large spread in the ensemble of consistent parameter vectors.
To get a sense of which parameters controlled the unmeasured dynamics, principal component analysis was applied to the dynamics of the complexes and free enzyme. The first principal component, which accounted for 64% of the variance in the dynamics, corresponded to the amount of free enzyme over the time course, as demonstrated by coloring the dynamics by their projection onto this mode (Fig 12A). Coloring the corresponding parameter vectors in the same way showed that the binding and unbinding rates controlled the transition from binding-limited to enzyme-limited (Fig 12C).

These observations suggest that the inability to distinguish free and bound enzymes and substrates is a major limitation to constraining parameters in our model. Since both the binding- and enzyme-limited regimes, whose dynamics are determined by different governing parameter combinations, are consistent with the synthetic data, we can’t determine which parameter combinations are constrained. An experiment that rules out one of the limits would make it possible to answer this. For example, we could measure the time course of ERK phosphorylation dynamics at varying enzyme concentrations, since the dynamics of binding- and enzyme-limited regimes should scale differently with the amount of enzyme. These experiments would also reduce uncertainty in the parameters controlling the unmeasured species. However, in doing model-based experimental design, it is important that the assumptions that went into the original model still hold in the new experiments (164).
Figure 12: Analysis of parameter uncertainty. A) Two-dimensional projections of the ensemble of parameter vectors consistent with synthetic data. Colors denote the projection of the parameters’ predictions for the dynamics of free enzyme and complexes onto the first principal mode described in the text. B) Leading eigenvalues of the Fisher Information Matrix over the ensemble. The threshold of $\sqrt{\lambda} = 1$ separating “constrained” and “unconstrained” parameters is marked by the black dashed line. C) Predicted dynamics of free enzyme over the ensemble colored by their projection onto the first principal mode described in the text. The first principal mode corresponds to the amount of free enzyme over the time course.
The fact that multiple incompatible limiting regimes can be consistent with data has implications for model reduction. A powerful model reduction algorithm has been proposed recently that automatically finds limiting parameter regimes (165). This method starts from an initial parameter vector and numerically solves a set of differential equations that locally evolves the parameters towards an interpretable limit, producing an approximate model with one less parameter. It was reported that this method is insensitive to the initial choice of parameter values, however the results described here suggest that this will not be the case if multiple mutually-exclusive limits are consistent with the data. In this case, the final reduced model will depend on which regime the initial parameter vector was closest to. This scenario can be expected whenever some species in the model are not observed experimentally, which is common in biology. At the same time, model reduction techniques can formalize the statements made here by explicitly finding the various limiting regimes and determining their connectivity to see which ones are mutually exclusive (166).

Returning to the experiments, the data collected so far have already led to unexpected results. Most striking is the consistent depletion of the pool of ATP, despite there being much more than necessary to phosphorylate all of the ERK in the samples by stoichiometry. This suggests that the model in Rubinstein et al. (64) is inadequate for describing this experiment. The ATPase activity of MEK and ERK, which was lumped into catalysis, must be modeled explicitly. Going forward will involve a combination of literature review, modeling and parameter estimation, and in vitro experiments to learn as much as possible about the quantitative mechanisms of ERK phosphorylation and ATP consumption. The methods outlined above should be useful for determining which parameters are constrained by data and for proposing new experiments. Finally, to constrain which dynamical behaviors can be achieved by the real futile cycle of ERK regulation
studied in Rubinstein et al., we need to experimentally monitor and mathematically model the dephosphorylation of ERK by MKP3. Until then, we are fortunate that these very reactions have been monitored and the measurements published before (150, 167).

All of this is building up to make predictions about the behaviors of real networks inside cells. The approach is to monitor individual biochemical reactions and use what is learned about the parameters of these steps to construct predictive models of the larger systems. Ultimately, one would like to be able to predict the effects of disease-associated mutations in individual components on network behavior for the purpose of designing therapeutic treatments. Already we know that mutations can have nontrivial effects \textit{in vivo} (39), and having a systems-level model based on detailed biochemistry would be a significant step in predicting and understanding these effects.
Appendix A: Supplement to Chapter 1

Section A1: Comparison of Manual and Automated Image Segmentation
Figure A1: Detection of germ cells in z-stacks of germlines. Each row is an image of the same germline taken at a different depth. The distal tip is at the bottom-left of each image. The green circles in the left column are the cells that were correctly identified by the automatic segmentation algorithm, and the orange circles are the corresponding manually-segmented cells. Green circles without an orange partner correspond to cases in which the cell was first detected in that slice by the automatic segmentation algorithm, but first detected in the slice above or below that one by the manual segmentation. The red circles in the right column are the objects that were incorrectly identified as cells by the automatic segmentation, while the orange circles are the cells that were identified by the manual segmentation but not by the automatic one.
Figure A2: Zoomed-in regions of the images in Fig. A1. Circle colors have the same meaning as in Fig. A1. The second row is zoomed in on the distal tip. The third row is zoomed in on the pachytene region, just before the loop.

Section A2: Diffusion Maps Algorithm for Ordering Cells along the Center Line

In the Diffusion Maps algorithm (99), a random walk is constructed over a set of data points, with hopping probabilities between pairs of points determined by their pairwise distances and a kernel function. Here, the data points were the positions of the germ cells in a single germline. When the data points lie on a lower-dimensional manifold, the algorithm produces a robust ordering of the data along its principal nonlinear axis (or axes) on the manifold. In the case of the germline, the germ cells essentially lie on a 1D manifold, the center line of the germline. The algorithm was implemented using custom MATLAB code.

First, weights between data points (germ cell positions) were calculated by passing their pairwise distances through a Gaussian kernel function. Weights between data points are related to the probability of a random walker jumping from one of those data points to the other, with higher weights corresponding to higher hopping probabilities. The width of the Gaussian kernel determines the relevant scale of hopping. If the kernel width is much smaller than even the smallest
distance between cells, then all weights between data points will be near zero, and a random walker cannot jump between any pair of data points. As the width is increased from zero, there is a scale at which the data appear 3D, then 2D, then 1D. At the 2D scale, a hopper can jump across the entire depth of the flattened germline in one jump, but not the diameter or length of the germline. At the 1D scale, the hopper can jump across the entire diameter of the germline in one jump, but not the length of the germline. If the kernel width is larger than the entire germline, all weights between data points will be close to one, and a random walker can jump between any two data points, no matter how far apart they are in space. At this scale, the data is essentially zero-dimensional from the point of view of the hopper, all collapsing to a single point. Previous work has developed an automated way of choosing the kernel width (168, 169). In practice, Diffusion Maps is not sensitive to the precise value of the kernel width, as long as it is in the correct dimensionality regime. We chose the kernel width for each germline so that the data “appeared” one-dimensional to a random walker.

The weights were then assembled into a symmetric matrix, with entry $(i,j)$ containing the weight between germ cell $i$ and germ cell $j$. The rows were normalized so that the sum of each row equaled one. This normalized matrix can be interpreted as a Markov transition matrix, with entry $(i,j)$ containing the probability of a hopper located at data point $i$ jumping to data point $j$ in one time step. As the number of data points approaches infinity, the eigenvectors of this Markov matrix approach the eigenfunctions of the Laplacian (diffusion) operator with Neumann (reflecting) boundary conditions (168). The first eigenvector of this matrix is a vector of ones, and contains no information. The first nontrivial eigenvector is one-to-one with and parameterizes the principal nonlinear axis of the data, the center line of the gonad tube. Element $i$ of this eigenvector is associated with germ cell $i$; therefore, the monotonic ordering of the elements gives the ordering
of the cells according to their positions along the center line. Diffusion Maps does not give the arc length positions of the cells, only their ordering.

**Section A3: Estimating the Apoptosis Term, $R(x)$**

Evaluation of Eq. 5 of Chapter 1 requires an expression for $R(x)$, which accounts for cell death in the germline. It is common to observe several germ cells undergoing apoptosis in a given fixed germline. Apoptotic germ cells are recognizable because their chromatin condenses and they undergo cellularization (51, 170). The former causes the cells to exhibit strong fluorescence when stained for DNA, while the latter causes the cells to exhibit essentially no fluorescence when stained for dpMPK-1 (Fig. A1). From the time that a germ cell first shows symptoms of apoptosis to the time that the cell is removed from the germline by sheath cells is about 1 hour (51). In this time, a dying germ cell can only travel about one cell diameter before being cleared from the germline (68). As a result, the frequency at which cell corpses are observed at a given position is essentially the same as the frequency at which cells undergo apoptosis at that position.
Figure A3: Image of the loop region of a *C. elegans* germline. Apoptotic cells (denoted by white arrowheads) are recognizable by their strong DAPI signal and lack of dpMPK-1 signal.

By this argument, we assumed that the death/clearance rate $R(x)$ is proportional to the number of cell corpses observed at $x$. We can decompose $R(x)$ into a shape function times a constant of proportionality that determines its scale. Under our assumptions, the shape function is given by the probability distribution of cell corpse locations (Fig. A2). The constant of proportionality is $r$, the total rate of cell death, which can be estimated from data in the literature, as discussed in the main text. Note that the source term $S(x)$ could be estimated in a similar manner by looking at the relative frequency of cell divisions as a function of arc length across germlines, but this was not explored here.

However, in a given germline there are too few corpses to estimate the shape of $R(x)$ accurately. To address this, we pooled corpse counts from multiple fixed germlines to estimate the average shape of $R(x)$ over multiple germlines. The pooling process, itself, requires aligning spatial positions across different germlines. Here we will describe a method for registering arc length positions across different germlines. This is the only part of our approach that requires averaging across germlines.
Section A4: Registering Arc Length Positions across Germlines

Since there are so few cell corpses in a given germline, we need to estimate the shape of \( R(x) \) from multiple germlines. Germlines come in different sizes, so pooling corpse positions across germlines requires that arc length positions in different germlines be registered or transformed to a common axis. Registering positions between two germlines is equivalent to determining an invertible function that maps positions in one germline to corresponding positions in the other. We assume that germ cells at “corresponding positions” are at the same developmental maturity, are the same age, and have spent the same amount of time in their respective germlines. Under the assumption that germ cells are arranged according to their maturity, this invertible function exists. The mapping will locally stretch or compress positions in one germline, like an accordion, to match the corresponding positions in the other. There will be a different mapping between each pair of germlines.

If \( x_1 \) refers to arc length positions in the first germline and \( x_2 \) refers to those in the second. The goal is to estimate an invertible function that maps \( x_1 \) to \( x_2 \), i.e. \( x_2 = g(x_1) \) and \( x_1 = g^{-1}(x_2) \). In the main text, we introduced the probability density of germ cell arc length positions, \( f_X(x) \), which quantifies the local “concentration” of germ cells in a particular germline. A related quantity is the cumulative distribution of arc length positions, \( F_X(x) \), which quantifies the cumulative fraction of germ cells located at or before \( x \).

Since \( x_1 \) and \( x_2 \) are related by an invertible function, their cumulative distributions must satisfy:

\[
F_{X_1}(x_1) = F_{X_2}(x_2). \tag{A1}
\]
Proof of Eq. A1:

1. \( F_{X_1}(x_1) = P(X_1 \leq x_1) \)  
   (definition of a cumulative distribution function)

2. \( = P(g(X_1) \leq g(x_1)) \)  
   (applying \( g(\cdot) \) to both sides, and noting that \( g(\cdot) \) is invertible, one-to-one and onto, and monotonically increasing)

3. \( = P(X_2 \leq x_2) \)  
   (using \( X_2 = g(X_1) \) and \( x_2 = g(x_1) \))

4. \( = F_{X_2}(x_2) \).  
   (definition of a cumulative distribution function)

Therefore, the invertible function we are seeking is \( x_2 = g(x_1) = F_{X_2}^{-1}(F_{X_1}(x_1)) \). For each germline, \( F_X(x) \) and its inverse are both measurable from data, meaning the mapping between any pair of germlines is measurable.

Using this approach, germ cell corpse positions from all germlines were transformed to their corresponding positions in a single germline. The particular germline used does not affect the result. The shape of \( R(x) \) was estimated from these corpse positions the same way as \( f_X(x) \), by kernel density estimation. This shape function for \( R(x) \) was then transformed back to the arc length axis of each germline.

Note that the local source term \( S(x) \) in Eq. 1 of Chapter 1 can also be estimated in a similar fashion by identifying the positions of mitotically dividing cells in multiple germlines and registering the positions across germlines. Recent work suggests, though, that production throughout the mitotic region is roughly uniform (105).
Figure A4: Estimation of $R(x)$. Assuming that the time rate at which cells undergo apoptosis at a given location is proportional to the frequency with which corpses are observed at that location, the shape of the spatially-dependent sink term $R(x)$ is the probability density function of cell corpse locations. A) Shows a schematic of a germline, where red cells are undergoing apoptosis. B) The histogram of cell corpse positions. With enough corpses, normalizing this histogram would give a good approximation of the shape of $R(x)$. Since there are not enough corpses in a single germline to estimate $R(x)$, corpse positions from multiple germlines must be aligned and pooled. C) Plotted in red is $R(x)$, estimated from 63 cell corpses pooled from 6 germlines and plotted against arc length position in a particular germline. The spatial dpMPK-1 profile from the same germline is shown in black. This shows that the peak rate of cell death occurs spatially (and temporally) after the peak of the dpMPK-1 pulse.

Section A5: Error Analysis

Uncertainty in $f_X(x)$

Assuming that we can measure germ cell arc length positions accurately, uncertainty in the probability density functions $f_X(x)$ for each germline can be approximated by their root mean squared error (RMSE) in the asymptotic limit of many samples. The mean squared error (MSE) of the estimate of $f_X(x)$ is the squared bias of the estimate (introduced by oversmoothing the true function) plus the variance of the estimate (introduced by estimating the function from a finite set of observations). When kernel density estimation is used to estimate probability density functions, the expression for the MSE is (171):
\[ \delta f_X(x)^2 = \frac{h^4(f_X''(x))^2}{4} + \frac{R f_X(x)}{nh}. \]  

(A2)

The first term in the sum is the squared bias of the estimate of \( f_X(x) \), and the second term is the variance of the estimate. \( h \) is the bandwidth of the smoothing kernel used in the density estimation, and \( n \) is the number of observations (here, the number of cells in a germline). \( f_X''(x) \) is the second derivative of the density, meaning that regions of the density function with larger curvature are more difficult to estimate accurately. This quantity was calculated by fitting the estimates of \( f_X(x) \) with splines (MATLAB \texttt{csapi}) and taking the second derivative of the spline (MATLAB \texttt{fnder}). Finally, \( R \) is a property of the kernel function used in density estimation; for a kernel function \( g(u) \), \( R = \int_{-\infty}^{\infty} g(u)^2 du \). Here, an approximately Gaussian kernel was used, for which \( R = 1/2\sqrt{\pi} \). Technically, density estimation was done via solving a diffusion equation, which acts much like a Gaussian smoothing kernel, but with better estimates of the density near the boundaries of the domain. This expression for the MSE should overestimate the error of the estimate near the boundaries.

**Uncertainty in \( t(x) \)**

Uncertainty in the estimates of \( t(x) \) propagate from: uncertainty in \( f_X(x) \), uncertainty in the values of the parameters in the model, and uncertainty in the shape of the apoptosis function \( R(x) \).

Uncertainty in the model parameters was accounted for by uniformly sampling the literature ranges for \( N_{tot}, s \), and the rate of ovulation (used to calculate \( r \)). Sampling was performed using a Latin Hypercube design (MATLAB’s \texttt{lhsdesign}) to generate 100,000 samples. The value of \( \tau \) produced by each parameter combination was calculated; if the value of \( \tau \) was outside of the literature
reported range for $\tau$ (48-54 hrs (107)), then the parameter set was discarded. After this pruning, 31,552 parameter sets remained. This collection of acceptable parameter sets was sampled from during the next step.

To estimate the effect of uncertainty in the shape of the apoptosis function $R(x)$, 5,000 samples were bootstrapped per germline from the collection of corpse observations. The shape of $R(x)$ was calculated for each randomly-sampled set of corpses. Then, for each sample, a parameter set was drawn at random from the collection of acceptable parameter sets, with replacement. Finally, $t(x)$ was calculated for that set of corpses, that parameter set, and that germline. The result was 5,000 estimates of $t(x)$ for each germline, the distribution of which accounted for uncertainty in the model parameters and the shape of $R(x)$. We denote the standard deviation of this distribution, as a function of $x$, $\delta_{t_{boot}}(x)$.

The total uncertainty in $t(x)$, for each germline, is given by:

$$
\delta t(x)^2 = \delta t_{boot}(x)^2 + \left( \frac{dt}{df_X} \right)^2 \delta f_X(x)^2,
$$

(A3)

where $\frac{dt}{df_X}$ is, from Eq. 5 in Chapter 1, $\frac{dt}{df_X} = \int_0^\infty s \frac{N_{tot}}{R(u)dw} dw$, and $\delta f_X(x)$ is the RMSE of $f_X(x)$. This quantity was calculated separately for each germline.
**Figure A5:** The collection of acceptable samples of $N_{tot}$, $s$, and $r$. Colors correspond to the value of $\tau$, in hours, calculated from that parameter set, using Eq. 7 of Chapter 1. Axes limits are the ranges consistent with the literature. Using knowledge from the literature of all four parameters significantly reduces the volume of acceptable parameter combinations. The resulting region is called the feasible set (108).

*Uncertainty in MPK-1 activation dynamics*

Uncertainty in the dynamics of dpMPK-1 estimated from fixed samples arose from measurement uncertainty of the antibody staining and propagation of uncertainty from the time estimates. The measurement uncertainty was taken to be the standard deviation of the nuclear dpMPK-1 intensity measurements around the smoothed dynamics for that germline. If $y(t)$ is the fluorescence intensity of dpMPK-1 with respect to time and $\delta y_{\text{measure}}$ is the measurement uncertainty, then the total uncertainty in $y(t)$ is:

$$\delta y(t)^2 = \delta y(t)^2_{\text{measure}} + \left(\frac{dy}{dt}\right)^2 \delta t(x)^2.$$  (A4)
where $dy/dt$ is the derivative of $y(t)$ with respect to $t$, and $\delta t$ is the uncertainty in $t$. This derivative was calculated by fitting $y(t)$ with splines (MATLAB `csapi`) and taking the derivative (MATLAB `fnder`). This calculation was performed for each germline.

Figure A6: Sample $t(x)$ from a germline, plus and minus one standard deviation $\delta t(x)$, which accounts for uncertainties propagated from errors in estimating $f_X(x)$, uncertainty in the parameter values, and uncertainty in the shape of $R(x)$. 
Section A6: Supplementary Figures

**Figure A7:** dpMPK-1 fluorescence intensity versus normalized arc length (arc length divided by the total distance from the distal tip to the loop) for multiple germlines. The images were acquired in the same experiment, at a set microscope condition. Plotting this way does not cause the spatial dpMPK-1 profiles from multiple germlines to collapse.

**Figure A8:** The fold change in dpMPK-1 fluorescence intensity, relative to background levels, as a function of time. Background fluorescence was calculated by averaging the dpMPK-1 intensity in the mitotic and early meiotic region, where there should be much less active MPK-1 than in the pachytene region.
Appendix B: Supplement to Chapter 2

Section B1: Derivation of bistability condition

We suspected that oscillations in this system are built on bistability. Therefore, as a first step to finding an analytical approximation to this region in the limit that $\varepsilon$ is small, we derived the conditions under which bistability can occur. In order for bistability to be possible, the nullcline for $a$ must have an “S” shape with two extrema.

In dimensionless form, the system can be described by the following two differential equations (Eqs. 5 and 6 in Chapter 2):

$$\frac{da}{d\tau} = -(1 + \beta_d f(a))a + \frac{1}{\varepsilon} \left( (\mu + \beta_a g(a))(w - a) - a \right), \quad (B1.1)$$

$$\frac{dw}{d\tau} = 1 - (1 + \beta_d f(a))w, \quad (B1.2)$$

where $a$ and $w$ are the dimensionless concentrations of active and total cyclins, respectively. Also note that:

$$f(a) \equiv \frac{a^{n_d}}{\gamma_d^{n_d} + a^{n_d}}, \quad (B1.3)$$

$$g(a) \equiv \frac{a^{n_a}}{\gamma_a^{n_a} + a^{n_a}}, \quad (B1.4)$$

When $\varepsilon$ is small, the differential equation for $a$ can be approximated by

$$\varepsilon \frac{da}{d\tau} = \left( (\mu + \beta_a g(a))(w - a) - a \right) \equiv G(a, w), \quad (B1.5)$$

and the concentration of $a$ changes rapidly compared to other time scales in the problem.
Setting \( G(a, w) = 0 \) gives an implicit equation for the \( a \) nullcline. Rearranging, we can make this equation explicit by solving for \( w \):

\[
w = a + \frac{a}{\mu + \beta a g(a)}.
\] (B1.6)

This curve has extrema when

\[
\frac{dw}{da} = 0 = 1 + \frac{1}{\mu + \beta a g(a)} - \frac{a \beta a g'(a)}{(\mu + \beta a g(a))^2}
\] (B1.7)

has real solutions, where \( g'(a) = \frac{dg}{da} \). Multiplying through by \((\mu + \beta a g(a))^2\) gives

\[
0 = (\mu + \beta a g(a))(1 + \mu + \beta a g(a)) - a \beta a g'(a).
\] (B1.8)

Note that

\[
a g'(a) = a \left( \frac{n a^{n-1}}{y_a^n + a^n} - \frac{n a^{n-1} a^n}{(y_a^n + a^n)^2} \right) = \frac{n a^n y_a^n}{(y_a^n + a^n)^2}.
\] (B1.9)

where we have dropped the subscript on \( n_a \) to reduce clutter. In this and the following equation, \( n \) will refer to \( n_a \). Using the definition of \( g \), Eq. B1.9 can be written in terms of just \( g \) and \( n_a \):

\[
a g'(a) = \frac{n a^n y_a^n}{(y_a^n + a^n)^2} = n \frac{a^n}{(y_a^n + a^n)(y_a^n + a^n)} = n g(1 - g).
\] (B1.10)

Plugging this expression for \( a g'(a) \) into Eq. B1.8 gives the condition for bistability in terms of just \( g \) and parameters:

\[
0 = (\mu + \beta_a g)(1 + \mu + \beta_a g) - \beta_a n_a g(1 - g).
\] (B1.11)

in which we have returned the subscript to \( n_a \). This expression is a quadratic function in \( g \):

\[
0 = (\beta_a + n_a)\beta_a g^2 + (1 - n_a + 2\mu)\beta_a g + \mu(1 + \mu).
\] (B1.12)
Equation B12 has two real solutions when the discriminant is greater than zero:

$$\Delta = \left((1 - n_a + 2\mu)\beta_a\right)^2 - 4((\beta_a + n_a)\beta_a)(\mu(1 + \mu)) > 0. \quad (B1.13)$$

This can be written

$$\Delta = \beta_a n_a \left(\beta_a \left(n_a + \frac{1}{n_a} - 2\right) - 4\mu(\mu + \beta_a + 1)\right) > 0, \quad (B1.14)$$

which simplifies to the following condition:

$$\beta_a (n_a + \frac{1}{n_a} - 2) > 4\mu(\mu + \beta_a + 1). \quad (B1.15)$$

Here, we can see that if $n = 1$, the left-hand side equals 0, and the inequality can’t be satisfied. However, it can be satisfied by $0 < n_a < 1$ or $n_a > 1$. Multiplying through by $n_a$ and rearranging gives:

$$n_a^2 - \left(2 + 4\frac{\mu}{\beta_a}(\mu + \beta_a + 1)\right)n_a + 1 > 0. \quad (B1.16)$$

We can find the boundary of the domain specified by this inequality by making it an equality. Then, solving for $n_a$ via the quadratic equation gives two solutions:

$$n_{a,\pm} = 1 + 2\frac{\mu}{\beta_a}(\mu + \beta_a + 1) \pm \sqrt{\left(1 + 2\frac{\mu}{\beta_a}(\mu + \beta_a + 1)\right)^2 - 1}. \quad (B1.17a,b)$$

The expression in Eq. B1.17 defines two surfaces in the space of $(\mu, \beta_a, n_a)$. The inequality in Eq. B1.14 is satisfied when $n_a > n_{a,+}$, corresponding to $n_a > 1$, or $n_a < n_{a,-}$, corresponding to $0 < n_a < 1$. These are two disconnected regions in parameter space, separated by the surface $n_a = 1$.
in each of these regions, which we will do in the next section. Performing this calculation, we
would find that points in the region \( n_a < n_a^* \) give negative or imaginary values for the value of
\( \alpha \) at the extrema. Therefore, these are not valid solutions, and the only valid condition for the \( \alpha \)-nullcline to have extrema is:

\[
n_a > 1 + 2 \frac{\mu}{\beta_a} (\mu + \beta_a + 1) + \sqrt{\left( 1 + 2 \frac{\mu}{\beta_a} (\mu + \beta_a + 1) \right)^2 - 1}. \tag{B1.18}
\]

This last condition is the necessary criterion for the system to display bistability and bistability-based oscillations in the limit that \( \varepsilon \) approaches zero. The domain that satisfies this condition grows
as \( \beta_a \) increases, approaching

\[
n_a > 1 + 2\mu + \sqrt{(1 + 2\mu)^2 - 1}. \tag{B1.19}
\]
in the limit that \( \beta_a \to \infty \).

**Section B2: Derivation of the boundaries of the oscillatory region**

When the condition in Eq. B1.18 is satisfied, Eq. B1.12 has two real solutions corresponding to
the values of \( g \) at the extreme points in the \( \alpha \)-nullcline. We expect the system to oscillate when
there is a unique steady state that lies between these two extreme points. The boundaries of the
oscillatory region then correspond to parameter sets for which the steady state lies exactly on one
of the two extreme points. First, we will derive expressions for the values of \( \alpha \) and \( w \) at the extreme
points in terms of the parameters in the problem. Then we will derive two equations, each one
corresponding to the steady state of the system lying exactly on one of the two extreme points.
Eliminating $a$ and $w$ from these expressions will produce two parametric curves, one for each extreme point, that define the boundaries of the oscillatory region.

From Eq. B1.12, the value of $g$ at the two extreme points of the nullcline for $a$ can be determined:

$$g_{\pm} = \frac{-(1-n_a+2\mu)\beta_a \pm \sqrt{\Delta}}{2(\beta_a+n_a)\beta_a} = \frac{-(1-n_a+2\mu)\pm \sqrt{n_a(\beta_a(n_a+1/\beta_a)-4(\mu+\beta_a+1)}}{2(\beta_a+n_a)}.$$

(B2.1a,b)

This expression maps one-to-one to the corresponding values of $a = a_{\pm}$ at the extreme points by the definition of $g$:

$$a_{\pm} = \gamma_a \left( \frac{g_{\pm}}{1-g_{\pm}} \right)^{1/n_a}.$$

(B2.2a,b)

One can find the expression for the value of $w$ at the extrema using Eqs. B1.6 (explicit equation for the $a$ nullcline) and B1.11 (the equation whose solutions are the values of $g$ at the extrema of the $a$-nullcline). Taking the reciprocal of Eq. B1.6, and multiplying numerator and denominator by $(\mu + \beta_ag_{\pm})$, gives:

$$\frac{1}{w_{\pm}} = \frac{(\mu+\beta_ag_{\pm})^2}{a_{\pm}(\mu+\beta_ag_{\pm})(1+\mu+\beta_ag_{\pm})}.$$

(B2.3a,b)

The denominator can be simplified and $a_{\pm}$ eliminated using Eqs. B1.11 and B2.2a,b:

$$\frac{1}{w_{\pm}} = \frac{(\mu+\beta_ag_{\pm})^2}{na\beta_ag_{\pm}(1-g_{\pm})(\frac{g_{\pm}}{1-g_{\pm}})^{1/n_a}}.$$

(B2.4a,b)

With Eqs. B2.1, B2.2, and B2.4, we have expressions for the extreme points of the nullcline for $a$ in terms of parameters only.
The missing piece in finding the boundaries of the region of oscillations is that the steady state of the system must occur at one of the two extreme points. Eqs. B2.2 and B2.4 are the coordinates for \(a\) and \(w\) at the extrema of the nullcline for \(a\); to be steady states, the nullcline for \(w\) must also intersect those points. First we set Eq. B1.2 equal to zero to get the nullcline for \(w\) and solve for \(w\), giving:

\[
w = \frac{1}{1 + \beta_d f(a)}. \tag{B2.5}
\]

To enforce that the nullcline for \(w\) intersects the nullcline for \(a\) at one of the extreme points, we set the reciprocal of Eq. B2.5 equal to the expression in Eq. B2.4:

\[
1 + \beta_d f(a_\pm) = \frac{(\mu + \beta_a g_\pm)^2}{n_a \beta_a g_\pm (1 - g_\pm) (\frac{g_\pm}{1 - g_\pm})^{1/n_a}}. \tag{B2.6a,b}
\]

In order to get an expression that only contains parameters, we must eliminate \(a_\pm\) from this expression. Doing so will give us two curves, one for each extremum, that define the boundaries of the region of oscillations in parameter space. To eliminate \(a_\pm\) we use Eqs. B1.3 and B2.2, giving:

\[
1 + \beta_d \gamma_a^{n_d (\frac{g_\pm}{1 - g_\pm})^{n_d/n_a}} = \frac{(\mu + \beta_a g_\pm)^2}{n_a \beta_a g_\pm (1 - g_\pm) (\frac{g_\pm}{1 - g_\pm})^{1/n_a}}. \tag{B2.7a,b}
\]

Solving the expression in Eq. B2.7 for \(\gamma_{d,\pm}\) allows analytical calculation of the two-parameter diagram in \(\gamma_a\) and \(\gamma_d\). To simplify the algebra, let us define

\[
B_{\pm} \equiv \frac{(\mu + \beta_a g_\pm)^2}{n_a \beta_a g_\pm (1 - g_\pm) (\frac{g_\pm}{1 - g_\pm})^{1/n_a}}. \tag{B2.8a,b}
\]
and

\[ C_\pm \equiv \gamma_a^{n_d} \left( \frac{g_\pm}{1-g_\pm} \right)^{n_d/n_a}. \]  \hspace{1cm} (B2.9a,b)

Then Eq. B2.7 can be written:

\[ 1 + \beta_d \left( \frac{C_\pm}{\gamma_d \pm \gamma_a} \right) = B_\pm, \]  \hspace{1cm} (B2.10a,b)

which can be solved explicitly for \( \gamma_{d,\pm} \):

\[ \gamma_{d,\pm} = C_\pm^{1/n_d} \left( \frac{\beta_d}{B_\pm - 1} - 1 \right)^{1/n_d}. \]  \hspace{1cm} (B2.11a a,b)

Substituting \( B_\pm \) and \( C_\pm \) back in and simplifying gives:

\[ \gamma_{d,\pm} = \gamma_a \left( \frac{g_\pm}{1-g_\pm} \right)^{1/n_a} \left( \frac{(1+\beta_d)n_\beta a \gamma a g_\pm (1-g_\pm) \left( \frac{g_\pm}{1-g_\pm} \right)^{1/n_a} - (\mu+\beta a g_\pm)^2}{(\mu+\beta a g_\pm)^2 - n_\beta a \gamma a g_\pm (1-g_\pm) \left( \frac{g_\pm}{1-g_\pm} \right)^{1/n_a}} \right)^{1/n_d} \]  \hspace{1cm} (B2.12a,b)

These equations define two 7-dimensional hypersurfaces that enclose the domain of oscillations at \( \epsilon = 0 \). By varying \( \gamma_a \) and keeping all other parameters constant, Eqs. B2.12a,b give \( \gamma_{d,+} \) and \( \gamma_{d,-} \), which are the boundaries of the oscillatory domain in the \( (\gamma_a, \gamma_d) \) plane.

**Section B3: Derivation of the amplitude of oscillations in \( \alpha \)**

When \( \epsilon \to 0 \), the nullcline for \( \alpha \) is:

\[ \alpha = \frac{\mu+\beta a g(a)}{1+\mu+\beta a g(a)} w \]  \hspace{1cm} (B3.1)
In this regime, the limit cycle traces the $a$-nullcline. Therefore, the maximal value of $a$ occurs when $w = w_-$ and the minimal value of $a$ at $w = w_+$, where $w_\pm$ are the values of $w$ at the limit points (see Fig. B5). Hence:

\[
a_{\text{min}, \text{max}} = \frac{\mu + \beta_a g(a)}{1 + \mu + \beta_a g(a)} w_\pm
\]

\[
a_{\text{min}, \text{max}} = \frac{\mu + \beta_a g(a)}{1 + \mu + \beta_a g(a)} \frac{n_a \beta_a g_\pm (1 - g_\pm)(1 - g_\pm)^{1/n_a}}{(\mu + \beta_a g_\pm)^2}
\]

\[
a_{\text{min}, \text{max}} = \gamma_a n_a \beta_a \frac{\mu + \beta_a g(a)}{1 + \mu + \beta_a g(a)} \frac{g_\pm(1 - g_\pm)(1 - g_\pm)^{1/n_a}}{(\mu + \beta_a g_\pm)^2}
\]

Note that since $w_-$ corresponds to $a_{\text{max}}$, so does $g_-$. Then, assuming $g(a) = 0$ at $a_{\text{min}}$ and $g(a) = 1$ at $a_{\text{max}}$:

\[
a_{\text{max}} = \gamma_a n_a \beta_a \frac{\mu + \beta_a}{1 + \mu + \beta_a} \frac{g_- (1 - g_-)(1 - g_-)^{1/n_a}}{(\mu + \beta a g_-)^2}
\]

\[
a_{\text{min}} = \gamma_a n_a \beta_a \frac{\mu}{1 + \mu} \frac{g_+ (1 - g_+)(1 - g_+)^{1/n_a}}{(\mu + \beta a g_+)^2}
\]

The amplitude of oscillations in $a$ is simply the difference between the maximal and minimal values of $a$ on the limit cycle:
\[ \Delta a = a_{\text{max}} - a_{\text{min}} \]

\[ = \gamma a n_a \beta_a \left( \frac{\mu + \beta_a}{1 + \mu + \beta_a} \frac{g_- (1 - g_-) \left( \frac{g_-}{1 - g_-} \right)^{\frac{1}{n_a}}}{\left( \mu + \beta_a g_- \right)^2} \right) \]

\[ - \frac{\mu}{1 + \mu} \frac{g_+ (1 - g_+)}{\left( \mu + \beta_a g_+ \right)^2} \left( \frac{g_+}{1 - g_+} \right)^{\frac{1}{n_a}} \]

(B3.7a,b)

From Eq. B2.1a,b, \( g_F \) depends only on \( \mu, \beta_a, \) and \( n_a \), none of which were scaled by \( s \). Since \( \beta_a \) and \( \gamma a \) are the only terms scaled by \( s \), and the Eq. B3.7a,b is linear in both, fold changes in \( s \) should cancel out, making the amplitude of oscillations in \( a \) invariant to variations in \( s \).

**Section B4: Supplementary Tables and Figures**

**Table B1: Parameter set with the largest value of \( \delta \)**

<table>
<thead>
<tr>
<th>Dimensionless parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>0.02897</td>
</tr>
<tr>
<td>( \beta_a )</td>
<td>1.2272</td>
</tr>
<tr>
<td>( \beta_d )</td>
<td>85.385</td>
</tr>
<tr>
<td>( \gamma a )</td>
<td>0.025933f</td>
</tr>
<tr>
<td>( \gamma d )</td>
<td>0.011456f</td>
</tr>
<tr>
<td>( n_a )</td>
<td>2.4569</td>
</tr>
<tr>
<td>( n_d )</td>
<td>14.633</td>
</tr>
<tr>
<td>( \epsilon )</td>
<td>0.00011613</td>
</tr>
</tbody>
</table>

Parameter set with the largest value of \( \delta \) (\( \delta_{\text{max}} = 52.8 \)), a measure of robustness of oscillations to perturbations in the synthesis rate. \( \gamma a \) and \( \gamma d \) are reported as the sampled values of the two parameters for which oscillations exist, times a factor \( f \), which can vary from 0.0048 to 9.69 and still preserve oscillations in the system. Continuation in \( f \) for this parameter set is shown in Fig. 10B.
Figure B1: Comparison of analytically-derived and numerical-computed domains of oscillations at a larger value of $\varepsilon$. Remaining parameter values are the same as in Fig. 8. Colors have the same meaning as in Fig. 8A of the main text.

Figure B2: Univariate marginal distributions of parameter samples that generated oscillations (from ~75,000 samples).
Figure B3: All bivariate marginal distributions of parameter sets that generated oscillations. Note that for the distributions shown in Fig. 9 of the main text, the axes are flipped here.

Figure B4: Density plot comparing the predicted amplitudes of $a$, using Eq. 13 in Chapter 2, to the true amplitudes. The predicted amplitudes overestimated the true amplitude for all parameter vectors.
Figure B5: Diagram accompanying the derivation of $\Delta \alpha$ in Section B3.
Appendix C: Supplement to Chapter 3

Section C1: Fisher Information Matrix

The Fisher Information Matrix (FIM) is defined as:

\[
I_{ij}(\theta) = -\int p(\tilde{y}|\tilde{\theta}) \frac{\partial^2 \log(p(\tilde{y}|\tilde{\theta}))}{\partial \theta_i \partial \theta_j} d\tilde{y} = \int p(\tilde{y}|\tilde{\theta}) \frac{\partial \log(p(\tilde{y}|\tilde{\theta}))}{\partial \theta_i} \frac{\partial \log(p(\tilde{y}|\tilde{\theta}))}{\partial \theta_j} d\tilde{y}. 
\] (C1)

For a Gaussian likelihood function, the FIM is approximately equal to the Hessian of the cost function (115). In this case, the FIM is also closely related to the sensitivity matrix, or matrix of derivatives of model predictions with respect to parameters:

\[
\text{FIM}(\tilde{\theta}) = S^T \Sigma S = \frac{1}{\sigma^2} S^T S, 
\] (C2)

where \( S_{ij} = \frac{\partial x_i}{\partial \theta_j} \) is the sensitivity matrix, which also depends on \( \tilde{\theta} \), and \( \Sigma \) is the covariance matrix of the measurement noise, which for uncorrelated noise with constant variance is just a constant times the identity matrix.

Since we assume Gaussian noise, rather than compute the eigenvalues and eigenvectors of the FIM, it is sufficient to compute the singular vectors and values of the matrix \( \frac{1}{\sigma^2} S \), the latter of which equal the square roots of the eigenvalues of the FIM. A set of differential equations for \( S \) can be derived by taking the derivatives of the model ODEs with respect to parameters:

\[
\frac{d}{dt} \begin{bmatrix} \frac{dx_i}{d\theta_j} \end{bmatrix} = \sum_k \frac{\partial f_i}{\partial x_k} \frac{dx_k}{d\theta_j} + \frac{\partial f_i}{\partial \theta_j} 
\] (C3)

where \( x_i(t) \) is a species in the model, \( \frac{dx_i}{d\theta_j} \) is the derivative of that species concentration with respect to parameter \( \theta \), and \( f_i(t, \bar{x}, \theta) \) is the function describing the time rate of change of species \( x_i \), i.e.
\[ \frac{dx_i}{dt} = f_i(t, \tilde{x}, \theta) \]. Equations C3 are called the sensitivity equations, and the model ODEs are solved simultaneously to get \[ \frac{dx_i}{d\theta_j} \]. Then, the experimentally-measured quantities \( y \) may be different from the species concentrations \( x \). If \( y_i(t) = h_i(\tilde{x}(t)) \), then

\[
\frac{\partial y_i}{\partial \theta_j} = \sum_k \frac{\partial h_i}{\partial x_k} \frac{dx_k}{d\theta_j} + \frac{\partial h_i}{\partial \theta_j}
\]

can be used to compute \( S \).

Section C2: Supplementary Tables and Figures

<table>
<thead>
<tr>
<th>Rate Constant</th>
<th>Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k_1 )</td>
<td>( ES_{00} \rightarrow ES_{01} )</td>
</tr>
<tr>
<td>( k_2 )</td>
<td>( ES_{01} \rightarrow ES_{11} )</td>
</tr>
<tr>
<td>( k_{on1} )</td>
<td>( E + S_{00} \rightarrow ES_{00} )</td>
</tr>
<tr>
<td>( k_{on2} )</td>
<td>( E + S_{01} \rightarrow ES_{01} )</td>
</tr>
<tr>
<td>( k_{on3} )</td>
<td>( E + S_{11} \rightarrow ES_{11} )</td>
</tr>
<tr>
<td>( k_{off1} )</td>
<td>( ES_{00} \rightarrow E + S_{00} )</td>
</tr>
<tr>
<td>( k_{off2} )</td>
<td>( ES_{01} \rightarrow E + S_{01} )</td>
</tr>
<tr>
<td>( k_{off3} )</td>
<td>( ES_{11} \rightarrow E + S_{11} )</td>
</tr>
</tbody>
</table>

Rate constants associated with each reaction step in the model of ERK phosphorylation by MEK.
Figure C1: Marginal histograms of individual parameters over the ensemble.
References


