RICE UNIVERSITY

Design Principles of Cellular Differentiation Regulatory Networks

by

Jatin Narula

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Oleg A. Igoshin, PhD, Chair
Associate Professor of Bioengineering

Jeffrey J. Tabor, PhD
Assistant Professor of Bioengineering

Matthew R. Bennett, PhD
Assistant Professor of BioSciences

HOUSTON, TEXAS

November __, 2015
ABSTRACT

Design Principles of Cellular Differentiation Regulatory Networks

by

Jatin Narula

To understand cellular differentiation programs is to understand the often large and complex gene regulatory networks (GRNs) that control and orchestrate these programs. The work presented here aims to exploit the wealth of newly available experimental information and methods to identify design principles that relate how GRN structures relate to the functional requirements in three model differentiation/stress-response programs: embryonic hematopoiesis, sporulation and σ^B general stress response. First we used a statistical thermodynamic approach to characterize the biophysical mechanisms of combinatorial regulation by distant enhancers in eukaryotes and demonstrate how the GRN controlling embryonic hematopoiesis acts as an irreversible bistable switch with low-pass noise filtering properties. We further used our model of the hematopoiesis network to reconcile discrepant experimental observations about the regulator Runx1 and explained how it limits HSC emergence in vitro. In the second project we investigated the Bacillus subtilis sporulation network and showed how a cascade of feed-forward loops downstream of the master regulatory Spo0A~P control cell-fate during starvation. We also identified a rate-responsive network module in the Spo0A regulon to explain why accelerated accumulation of Spo0A~P leads to a dramatic reduction in sporulation efficiency. Further we found that the arrangement of two sporulation network genes on opposite ends of the chromosome ties Spo0A~P activation to the DNA replication status. We were also able to show that the slowdown of cell growth is the primary starvation signal that determines sporulation cell-fates by controlling Spo0A~P activation. For the third project we built a detailed model of the σ^B network in Bacillus subtilis to
mechanistically explain the experimentally observed pulsatile response of this network under stress. We further showed that the same network architecture that enables this pulsatile response insulates the $\sigma^B$ network from the effects of competition for cellular resources like RNA polymerase. The design principles identified in the studies of these networks are related to their topological structure and function rather than the specific genes and proteins that comprise them. As a result, we expect them to be widely applicable to and help in the study of a diverse array of other differentiation GRNs.
Acknowledgments

First and foremost I want to acknowledge my advisor and mentor, Oleg Igoshin. None of the work presented here would have been possible without his generous encouragement, unwavering support and astute guidance. Perhaps even more crucially he has shaped every aspect of my approach to biological research including the ways that I think, solve, write and talk about scientific problems.

I would like to thank all of the wonderful Igoshin lab members and experimental collaborators that I’ve had the great fortune of working with. They have played a critical role in shaping the work in this thesis. And without their efforts my work would have nowhere near the same import or significance. I would also like to thank Drs. Masaya Fujita, Gabor Balazsi, John Pimanda and my committee members Drs. Jeff Tabor and Matthew Bennett. As mentors and collaborators they have all been very generous with time and their critical feedback on my projects has helped refine my perspective and approach.

Next I am grateful to my friends, Abhinav, Sayantan, Richa, Ramdas and Achal for all the amazing years filled with good memories. I would especially like to acknowledge Abhinav for being my chief collaborator, willing to indulge in endless discussions about far-fetched ideas and crack-pot schemes.

Finally, I would like to thank my family, mom and Abhey. I miss them everyday and without their sacrifices none of this would be possible.
# Contents

ABSTRACT ........................................................................................................................................... iii  

Acknowledgments .............................................................................................................................. v  

List of Figures .................................................................................................................................... xi  

List of Tables ...................................................................................................................................... xvi  

Introduction ......................................................................................................................................... 1  

1.1. Structure of Gene Regulatory Networks: Parts, Assemblies, Modules and Systems  
.................................................................................................................................................................. 2  

1.2. Functional Requirements of Cellular Differentiation ................................................................. 5  

1.3. Melding experimental and systems modeling approaches to understand cellular differentiation  
.................................................................................................................................................................. 11  

Organization of the Thesis ................................................................................................................ 13  

Regulation of embryonic hematopoiesis  

Bistability and low-pass filtering in the network module determining blood stem cell fate  
................................................................................................................................................................ 18  

2.1. Introduction .................................................................................................................................... 18  

2.2. Results ......................................................................................................................................... 21  

2.2.1. Thermodynamic model for enhancement of gene expression ................................................ 21  

2.2.2. Mathematical model for the Scl-Gata2-Fli1 Network triad module ........................................ 24  

2.2.3. Steady state response of the triad module ................................................................................. 27  

2.2.4. Enhancer mutations change the steady-state response of the triad ........................................ 31  

2.2.5. Dynamical response of the triad module architecture ............................................................... 34  

2.3. Discussion ..................................................................................................................................... 39  

2.4. Methods ....................................................................................................................................... 46  

Thermodynamic models of combinatorial gene regulation by distant enhancers  
................................................................................................................................................................ 59  

3.1. Introduction .................................................................................................................................... 59  

3.2. Results ......................................................................................................................................... 63  

3.2.1. Thermodynamic formalism to model combinatorial gene regulation via chromatin mechanism  
............................................................................................................................................................... 63
3.2.2. Implementation of cis-regulatory logic gate functions with chromatin mechanism .................................................................................................................. 66
3.2.3. Sensitivities of logic input functions to free energy values ......................... 71
3.2.4. Parameter estimation for chromatin model from experiments .................... 73
3.2.5. Comparison of stochastic kinetics of gene regulation by direct contact and chromatin mechanisms .................................................................................................................. 79
3.3. Discussion ....................................................................................................... 87
Effects of genetic perturbations on the dynamics of hematopoietic stem cell emergence .............................................................................................................. 92
4.1. Introduction ..................................................................................................... 92
4.2. Results ........................................................................................................... 95
  4.2.1. Notch1 is necessary for irreversible activation of the triad ....................... 95
  4.2.2. The model explains accelerated emergence of HSCs in Runx1 heterozygotes 99
  4.2.3. Runx1 overexpression can compensate for Notch1 deletion .................. 102
  4.2.4. Timing of Runx1 conditional deletion or Smad1 overexpression determines triad activation dynamics ................................................................. 103
  4.2.5. Novel predictions about the roles of Runx1 and Smad6 in HSC emergence 107
4.3. Discussion ...................................................................................................... 107
4.4. Methods ......................................................................................................... 111
Regulation of the Bacillus subtilis sporulation decision
Ultrasonic dependence of sporulation decision on Spo0A~P...... 123
5.1. Introduction ..................................................................................................... 123
5.2. Results ........................................................................................................... 126
  5.2.1. Sporulation in ASI is induced in ultrasonic manner ............................... 126
  5.2.2. Activation of Spo0A is not ultrasonic ................................................... 128
  5.2.3. Activation of $\sigma^F$ in the forespore is ultrasonic but occurs below the sporulation threshold ................................................................. 131
  5.2.4. Ultrasonic $\sigma^E$ activation in the mother cell determines cell-fate ...... 135
  5.2.5. Activation of $\sigma^F$ does not ensure successful sporulation in wild-type B. subtilis cells .................................................................................................................. 139
5.3. Discussion ..................................................................................................... 143
5.4. Materials and Methods .................................................................................. 146
The importance of proper Spo0A dynamics in sporulation .......... 151

6.1. Introduction .............................................................................................................. 151
6.2. Results ......................................................................................................................... 154
   6.2.1. KinC-Spo0A artificial two-component system triggers entry into sporulation .......................................................... 154
   6.2.2. The role of 0A~P accumulation kinetics in efficient sporulation ...................... 158
   6.2.3. Accelerated accumulation of 0A~P perturbs the temporal order of 0A regulon expression .............................................................................................................. 162
   6.2.4. Accelerated 0A~P accumulation leads to premature repression of DivIVA which impairs chromosome segregation .......................................................... 168
   6.2.5. Autoregulation of 0A ensures proper temporal coordination of high and low threshold gene expression .......................................................... 170
6.3. Discussion ..................................................................................................................... 173
6.4. Experimental Procedures ............................................................................................ 176

Chromosomal locations of phosphorelay genes controls the Bacillus subtilis response to starvation ................................................................. 181

7.1. Introduction .............................................................................................................. 181
7.2. Results ......................................................................................................................... 183
   7.2.1. 0A activity pulses follow the completion of DNA replication ......................... 183
   7.2.2. A hidden negative feedback loop in the phosphorelay ...................................... 184
   7.2.3. Chromosomal arrangement of phosphorelay genes provides a pulse triggering perturbation .......................................................... 187
   7.2.4. Experimental tests confirm the role of gene-dose imbalance and negative feedback in 0A~P pulsing .............................................................................. 190
   7.2.5. Lack of 0A~P pulsing leads to sporulation defects ........................................... 194
7.3. Discussion ..................................................................................................................... 197
7.4. Experimental Procedures ............................................................................................ 199

Slowdown of growth controls cellular differentiation ......................... 207

8.1. Introduction .............................................................................................................. 207
8.2. Results ......................................................................................................................... 210
   8.2.1. 0A~P pulse amplitudes are correlated with cell growth rate ......................... 210
   8.2.2. Decrease in growth rate leads to accumulation of stable proteins. .......... 212
8.2.3. Accumulation of phosphorelay proteins with growth slowdown is sufficient to explain observed increase in $\sigma^A\sim P$ levels ......................................................... 213
8.2.4. Growth slowdown dynamics control sporulation deferral ...................... 217
8.2.5. KinA activity does not depend on the growth rate ................................. 223
8.3. Discussion ................................................................................................. 227
8.4. Materials and Methods ............................................................................ 229

Regulation of the Bacillus subtilis general stress response

Design and functional properties of the general stress response network in Bacillus subtilis ................................................................. 239

9.1. Introduction .................................................................................................. 239
9.2. Results ......................................................................................................... 243
  9.2.1. Biochemically accurate model of $\sigma^B$ pulsing mechanism ................ 243
  9.2.2. $\sigma^B$ response dynamics depend on operon relative stoichiometry ........ 246
  9.2.3. $\sigma^B$ network encodes phosphatase burst size into pulse amplitudes ...... 248
  9.2.4. $\sigma^B$ network encodes phosphatase ramp rates into pulse amplitudes .... 250
  9.2.5. The design of the $\sigma^B$ network enables it to compete with $\sigma^A$ for RNA polymerase ..................................................................................................................... 252
  9.2.6. Negative feedback designs of stress-response sigma factor networks minimizes interference .................................................................................................................. 255
9.3. Discussion .................................................................................................. 259
9.4. Methods ..................................................................................................... 261

Appendix 1. Bistability and low-pass filtering in the network module determining blood stem cell fate ................................................................. 265
Appendix 2. Thermodynamic models of combinatorial gene regulation by distant enhancers .......................................................................................... 279
Appendix 3. Reconciling Effects of Genetic Perturbations on HSC Emergence ................................................................................................. 291
Appendix 4. Understanding the ultrasensitive dependence of sporulation decision on $\text{Spo0A} \sim P$ .................................................................................. 300
Appendix 5. The importance of proper $\text{Spo0A}$ dynamics in sporulation ................................................................................................................................. 327
Appendix 6. Chromosomal locations of phosphorelay genes controls the Bacillus subtilis response to starvation ........................................... 347
Appendix 7. Slowdown of growth controls cellular differentiation ... 363
Appendix 8. Design and functional properties of the general stress response network in Bacillus subtilis ............................................. 379
References ........................................................................................................................................ 394
List of Figures

Figure 2.1 Regulation of gene expression in the Scl-Gata2-Fli1 triad................. 19

Figure 2.2. Steady state signal-response analysis of the triad module to Notch, Bmp4 and Gata1 signals demonstrates irreversible bistability. ......................... 28

Figure 2.3. Selective deletion of enhancer binding sites can change the steady state response characteristics. ................................................................. 33

Figure 2.4. Comparison of dynamical responses of the triad and the reduced module to Notch, Bmp4 and Gata1 signals. .................................................. 36

Figure 3.1. Two mechanisms of combinatorial gene regulation by distant enhancers................................................................. 62

Figure 3.2. Designs of distant enhancers that exhibit a logic gate response. ..... 67

Figure 3.3. Gene expression response functions for the logic gates of the chromatin mechanism. ................................................................. 69

Figure 3.4. Sensitivity of OR gate response to variations of free energies values. ......................................................................................... 73

Figure 3.5. Application of the parameter estimation method to the Gata2-3 enhancer. ......................................................................................... 76

Figure 3.6. Waiting time distributions for transcriptional machinery bound and unbound states.................................................................................. 82

Figure 4.1. GRN responsible for regulating HSC specification. ......................... 95

Figure 4.2. Notch1 is necessary for triad activation. ......................................... 97

Figure 4.3. The role of Runx1 in triad activation............................................. 100

Figure 4.4. Runx1 overexpression compensates for Notch1 deletion.......... 103

Figure 4.5. Timing of Runx1 conditional deletion or Smad1 overexpression determines whether Triad activation is affected or not. .......................... 105

Figure 5.1. Regulation of sporulation in B. subtilis........................................ 124
Figure 5.2. The phosphorelay response is not ultrasensitive, due to response time requirements. ................................................................. 127

Figure 5.3. $\sigma^F$ activation overestimates the fraction of cells that sporulate. ...... 132

Figure 5.4. $\sigma^E$ activation is the ultrasensitive switch that controls cell fate. ...... 136

Figure 5.5. Ultrasensitivity and cell-fate decision in wild-type cells.................. 141

Figure 6.1. The sporulation network phosphorelay........................................ 151

Figure 6.2. Simultaneous overproduction of KinC and Spo0A reduces sporulation efficiency. ............................................................................. 156

Figure 6.3. Dynamic Threshold model explains the dependence of sporulation on IPTG and xylose. ................................................................. 159

Figure 6.4. Rapid Spo0A~P accumulation reverses the temporal order of low and high threshold genes. ......................................................... 163

Figure 6.5. Dynamic Threshold model explains the existence of optimal IPTG and xylose concentrations for sporulation. ............................... 167

Figure 6.6. Overproduction of KinC and Spo0A results in early repression of DivIVA and improper chromosome segregation. ............................ 169

Figure 6.7. $\sigma^H$-mediated feedback prevents decrease in sporulation efficiency at high KinC levels............................................................... 171

Figure 7.1. Coordination of the sporulation response with the cell-cycle in B. subtilis. .......................................................... 183

Figure 7.2. Substrate inhibition of KinA by 0F produces a negative feedback in the phosphorelay. ................................................................. 186

Figure 7.3. Mathematical model identifies the mechanism of 0A~P pulsing and its necessary conditions. ............................................................... 189

Figure 7.4. 0A~P pulsing depend on $\text{kinA} \cdot 0\text{F}$ chromosomal arrangement and transcriptional feedback from 0A~P to 0F negative feedback............. 192

Figure 7.5. Loss of coordination of sporulation program with DNA replication in non-pulsing strains could lead to sporulation defects.................. 196
Figure 8.1. Decision making during *B. subtilis* sporulation. ........................................ 209

Figure 8.2. Decrease in growth rate affects 0A−P levels and controls cell-fate. 215

Figure 8.3. Growth slowdown controls sporulation deferral. ................................. 218

Figure 8.4. Selective exposure to slow growth eliminates sporulation deferral. 221

Figure 8.5. Decrease in growth rate affects 0A−P levels and controls cell-fate. 225

Figure 9.1. σ^B general stress response network. ......................................................... 241

Figure 9.2. Pulsatile response of the σB network to stochastic phosphatase bursts ................................................................. 249

Figure 9.3. Rate sensitivity of the σ^B pulsatile response .............................................. 251

Figure 9.4. Negative feedback insulates the σ^B response from competition with housekeeping sigma factor σ^A. ................................................. 254

Figure 9.5. Negative feedback minimizes competition between stress σ factors for RNA polymerase. ......................................................... 257

Figure A1.1. Scl+19 distant enhancer controls the probability not rate of gene expression. ........................................................................ 265

Figure A1.2. Schematic diagram of ratchet model of distal enhancer action using Scl+19 enhancer as an example. ......................................................... 267

Figure A1.3. Robustness of switchable bistable response to variability in chromatin equilibrium constants. ................................................. 267

Figure A1.4. Dose-response to Notch and Bmp4 in Gata2, Scl and Fli1 heterozygotes. ........................................................................ 268

Figure A2.1. Gene expression response functions for contact model logic gates. ........................................................................ 279

Figure A2.2. Comparison of the sensitivities to free energy values of the AND, NAND and XOR logic gate responses. ................................................. 280

Figure A2.3. Sensitivity of the parameter estimation method to the choice of chromatin equilibrium constant  \( \kappa \) . ........................................................................ 281
Figure A3.1. Enhancer and promoter binding sites for the cis-regulatory regions that control Smad1, Smad6 and Runx1 expression ................................................................. 291

Figure A3.2. Runx1 requirement for triad activation depends on α. ..................... 291

Figure A3.3. Robustness of WT-Runx1⁺⁻ activation time difference predictions. ................................................................. 292

Figure A3.4. Robustness of minimum Runx1 requirement time window ............. 292

Figure A3.5. WT-Runx1⁺⁻ activation time difference in the presence of Notch dynamics. ........................................................................ 293

Figure A4.1. KinA induction in ASI ........................................................................ 300

Figure A4.2. Ultrasensitivity of the phosphorelay increases response times. .... 301

Figure A4.3. Ultrasensitive activation of σF in the forespore at sub-threshold KinA concentrations. ........................................................................ 302

Figure A4.4. Ultrasensitive activation of σE in the mother cell. ......................... 304

Figure A4.5. Low spoIIG expression limits σE activation and sporulation below the KinA threshold. ................................................................. 306

Figure A4.6. σF activation does not ensure σE activation or completion of sporulation in starving wild-type cells. ................................................................. 307

Figure A5.1. Protein levels of KinC and Spo0A in the IPTG- and xylose-inducible strain ........................................................................ 327

Figure A5.2. Protein expression from IPTG- and xylose-inducible promoters in a single cell ........................................................................ 328

Figure A5.3. Increase in xylose concentration speeds up Spo0A~P accumulation. ........................................................................ 329

Figure A5.4. Early onset of divIVA repression at high IPTG and xylose concentrations. ........................................................................ 329

Figure A5.5. Fluorescence microscopy of KinC-Spo0A strain. ......................... 331

Figure A5.6. Overexpression of KinC and Spo0A leads to an increase in chromosome segregation defects ................................................................. 331
Figure A5.7. *spo0A* auto-regulation ensures proper target expression even at high KinC levels.

Figure A6.1. Pulsing from *P_{OF}* promoters.

Figure A6.2. Mathematical modeling of the post-translational interactions shows the ultrasensitive dependence of [0A\_P] on the [0F\_T]/[KinA\_T] ratio.

Figure A6.3. Modeling predictions for the effects of *OF* translocation and inducible *OF* expression.

Figure A6.4. *OF* and *kinA* chromosomal locations are evolutionarily conserved in spore forming bacteria.

Figure A6.5. Comparison of pulsing and non-pulsing responses.

Figure A7.1. Effect of growth slowdown on pulsing from *P_{OF}* promoter, cell volume and DNA replication.

Figure A7.2. Dependence of sporulation cell-fate on *P_{OA}* promoter activity and growth rate.

Figure A7.3. Effect of nutrient availability on growth slowdown and sporulation dynamics.

Figure A7.4. Model predictions for 0A\_P pulsing in an inducible KinA strain.

Figure A7.5. Experimental measurement of the sporulation cell-fate outcomes as a function of KinA level and growth rate.

Figure A8.1. *\sigma^B* does not pulse for networks that lack negative feedback.

Figure A8.2. Dependence of *\sigma^B* pulse amplitudes on post-translational parameters.

Figure A8.3. Pulsatile response of the *\sigma^B* network encodes stochastic phosphatase burst sizes not burst frequency.

Figure A8.4. Sensitivity of the *\sigma^B* target expression to *\sigma^A* and competition for RNA polymerase.

Figure A8.5. Pulsatile response of the multiple stress sigma factor model.
List of Tables

Table A1.1. Enhancer library reporter expression results ........................................... 269
Table A1.2. Notation for various symbols used in the main text and supplementary information ................................................................. 269
Table A1.3. Free energy values used in the model for Scl-Gata2-Fli1 triad ..... 272
Table A3.1. Notation for species concentrations .......................................................... 294
Table A3.2. Parameter values for the triad regulatory functions .................................. 294
Table A3.3. Parameter values for the Smad1 regulatory function .............................. 295
Table A3.4. Parameter values for the Smad6 regulatory function .............................. 295
Table A3.5. Parameter values for the Runx1 regulatory function .............................. 295
Table A3.6. List of parameters values for post-translational reactions ................... 296
Table A4.1. Strains used in this study ........................................................................ 308
Table A4.2. Primers used in this study ...................................................................... 309
Table A4.3. Model parameters for transcriptional regulation .................................... 310
Table A5.1. Sporulation in the KinC strain (MF4531) in the presence of different IPTG concentrations ................................................................. 333
Table A5.2. Sporulation Efficiency and Spo0A activity in the KinC-Spo0A double induction (MF4317) strain ................................................................. 333
Table A5.3. KinC can activate Spo0A and trigger sporulation in Δspo0F background........................................................................................................... 334
Table A5.4. Strains used in this study ........................................................................ 335
Table A5.5. Plasmids used in this study ..................................................................... 335
Table A5.6. Oligonucleotide primers used in this study ........................................... 336
Table A5.7. Parameter values used for the model of transcriptional regulation of the native spo0A promoter ................................................................. 336
Table A5.8. Parameter values used for the models of transcriptional regulation of Spo0A target genes.................................................................337

Table A5.9. Parameter values used for post-translational interactions in the model of Spo0A activation.................................................................337

Table A6.1. *B. subtilis* strains used in this study...........................................352

Table A6.2. Parameter values used for post-translational interactions in the model of sporulation phosphorelay.........................................................353

Table A6.3. Parameter values used for gene regulatory interactions in the model of sporulation phosphorelay. .........................................................354

Table A6.4. Genome sizes and chromosomal locations of phosphorelay genes spo0F and kinA in 46 spore forming bacteria. ........................................355

Table A7.1. *B. subtilis* strains used in this study...........................................371

Table A7.2. Parameter values used for gene regulatory interactions in the model of sporulation phosphorelay. .........................................................371

Table A7.3. Parameter values used for the population dynamics model of growth dynamics during starvation........................................................372

Table A8.1. List of parameters values used in the model for σB network............383
Chapter 1

Introduction

Cellular differentiation is a catch-all term for the processes by which cells specialize phenotypically and morphologically to either adapt to their environmental conditions or perform specific functions (1). This ability to generate specialized cell types from multipotent precursor cells is of course central to the development of multi-cellular eukaryotes (1-3) but it is also ubiquitous in single cell organisms like bacteria (4-7). Decades of work has made it increasingly clear that the functional differentiation of a cell is primarily the result of specialization of its gene expression repertoire (6-9). This specialization in turn is wrought by Gene Regulatory Networks (GRNs) that sense changes in the cellular environment, detect and process differentiation and stress related signals and execute the appropriate gene expression response (10, 11). As a result, to mechanistically understand cell differentiation programs we have to understand the structure and function of the regulatory networks that control them.

A typical differentiation regulatory network looks like an inscrutable mess of interconnected components (2). A cursory glance at this type of network makes meaningful mechanistic analyses seem an impossible task. Crucially however, in analyzing different GRNs we frequently find similar patterns of network structure or design (12).

Structurally, these regulatory networks are the result of evolution under selective pressures to successfully execute particular differentiation programs.
As a result, it is reasonable to expect that the structure or design of these networks is not arbitrary but rather dictated by the functional requirements placed upon them. This immediately suggests that that if we parse all the regulatory tasks of a particular differential program, we should be able to explain the structural design of the regulatory network that controls it.

Notably comparisons of different differentiation programs show that they can share functional requirements (12-14). As a result, the derivation of this type of structure-function map for even one particular differentiation program can yield design principles – general concepts that relate how functional requirements influence the selection of network designs – that can inform our understanding of a GRN structures in the wider context of other differentiation programs (13, 15, 16). Over the past decades, research guided by this line of thought has uncovered a variety of design principles governing the structure of regulatory networks controlling cellular metabolism and stress-responses (13, 16). In comparison, progress for differentiation related GRNs has been less prolific due to experimental limitations and the size and complexity of these networks (17, 18). The major objective of this thesis is to use the methods of mathematical modeling to exploit recent developments in single cell microcopy and synthetic biology and advance the knowledge of design principles for GRNs that control cellular differentiation.

1.1. Structure of Gene Regulatory Networks: Parts, Assemblies, Modules and Systems

The derivation of design principles for differentiation GRNs can be non-trivial owing to the inordinate complexity of these networks. Fortunately these networks bear a striking similarity to the hierarchical structure of electronic circuits (19, 20). This conceptual analogy has prompted researchers to borrow approaches and methods from systems engineering to analyze, characterize and unravel the complexity of biological GRNs (13, 15, 16, 19, 21, 22). In recent years, the influx of these engineering approaches has led to the development of
new fields like systems biology (23-25) and synthetic biology (19, 26) that aim to reverse-engineer naturally occurring biological networks and forward-engineer artificial ones. These new developments have also profoundly changed the way we think about biological networks.

Similar to electronic circuits, GRNs that regulate differentiation and stress response are highly hierarchical (12, 19, 27). At the bottom of this hierarchy are individual biological parts. Generally, biological part refers to the individual proteins, the genes and mRNAs that encode them, small molecules and non-coding DNA elements like promoters, ribosome binding sites, terminators etc. that together make up the GRN (28). Analogous to the physical layer of transistors, capacitors and resistors in electronics, these components are individually non-functional and need to be specifically assembled to be operationally useful (27, 28).

Part assemblies are the simplest functional units of a gene regulatory network. Assemblies can take the form of transcriptional units that consist of a promoter, the gene(s) expressed from that promoter, and the regulatory proteins (and their cognate DNA binding sites) that affect the expression of that gene (28). Alternatively assemblies can be entirely post-translational and consist of a signal sensing kinase, an associated response regulator protein that is phosphorylated by the kinase and a small signaling molecule that binds to and controls the activity of the kinase (29). Relative to individual parts, a crucial distinguishing feature of part assemblies is that they have defined inputs and outputs. Transcriptional assemblies respond to the input levels of transcriptional regulators (activators and repressors) by changing their output levels of target gene expression. Similarly post-translational assemblies respond to changes in the concentration of signaling molecules by changing the output levels of response regulator activity.

In any GRN the inputs and outputs of different functional part assemblies are found to be specifically interconnected to each other via shared components forming sub-circuits or network modules (10, 11, 15). Different network modules
can be typically distinguished by virtue of the fact that parts within each module are more highly interconnected to each other than parts in other modules (13). Thus each network module can be viewed as an operational device within this GRN (12, 27, 28). These biological devices are associated with and perform specific regulatory functions for the GRN. For example, signaling modules interface with the environment and act as information processors and transcriptional modules act as actuators that dynamically regulate the gene expression patterns of the GRN components.

Seminal work from Uri Alon’s group has revealed that specific structural designs of network modules are highly enriched in bacterial transcriptional networks (30, 31). Crucially their results have led to the wide adoption of the very attractive working hypothesis that network modules that are structurally similar can have specific similar functional properties, despite being composed of different biochemical parts and genes (16).

One of the simplest and most prevalent modules are autoregulatory circuits (32, 33), which involve either direct or indirect feedback loops which can be positive or negative. Studies of these modules in a variety of contexts have shown that positive feedback circuits often lead to ultrasensitive switch-like behavior and even bistable and hysteretic responses (13). In contrast, negative feedback loops have been shown to be capable of producing oscillatory and excitable behavior (34, 35). Further studies have shown the contrasting nature of these feedback loops. Specifically positive feedback loops are known to act as amplifiers of noise generated by stochastic biochemical fluctuations (13) whereas negative feedbacks are frequently act as noise suppressors (36). Negative feedbacks also accelerate system response activation whereas positive feedbacks tend to slow down system activation (15, 16, 33).

Another class of extensively studied and prevalent network modules is the three species feed-forward loop (FFL) in which an upstream network species regulates the target both directly and indirectly via an intermediate (13). Based on the regulatory sign (activation/repression) a wide variety of FFL architectures
are known to be possible each associated with useful functions. These functions include ultrasensitive signal amplification, noise suppression, signal persistence detection and fold-change detection (13, 37). Studies of these and a number of other network modules have led effectively to the development of a library of known functional modules that may serve as elemental building blocks for GRNs (38, 39).

Despite this progress in the identification and characterization of functional modules the decomposition of complex differentiation GRNs into a finite set of interacting modules has not been very successful (17, 18). The limited success of this approach suggests that we are still missing information about global organizational design principles that govern the arrangement of modules in GRNs. Our lack of knowledge of these principles is what currently limits our ability to unravel complex cellular behaviors like cellular differentiation.

To advance past this hurdle we need to take a more holistic view of differentiation GRNs. Such a top-down systems approach has the potential to uncover new types of functional requirements that shape GRNs and can help us better understand their global organizational design principles.

1.2. Functional Requirements of Cellular Differentiation

*Switch-like gene expression control*

Growing a new cell type typically requires the activation of a one battery genes and the down-regulation of a different battery (6, 9). As a result, one of the most basic requirements for a cellular differentiation GRN is the ability to dramatically shift gene expression profiles within a cell. Many early studies of GRNs have focused on this functional requirement for this switch-like behavior (40). Two types of GRN responses are known to achieve this switch-like behavior: ultrasensitivity and bistability (40-42). Both bistable and ultrasensitive systems convert graded signal levels into all-or-none responses based on a
signal threshold (43). However, there is one important difference – ultrasensitive systems always have a unique stable state whereas two distinct stable states are possible for bistable systems.

Networks that have two (or more) stable states are described as bistable (multi-stable). In biological context, this refers to a GRN’s ability to exhibits two (or more) discrete levels of gene expression (a high state and a low state) (16, 44). The structural requirements for a regulatory network to exhibit bistability have been studied extensively (45-47). Briefly, a network needs high non-linearity and positive feedback to be bistable. This non-linearity typically results from specific design features like multimerization, multi-site phosphorylation, cooperative binding to DNA and phosphorylation-dephosphorylation futile cycles. A common feature of bistability is hysteresis (48). Hysteresis imposes memory-like characteristics onto the network making the gene expression response of GRNs dependent on their input signal history. This cellular memory is a useful functional property for differentiation programs since it can act as a buffer and prevent accidental switching between states due to minor perturbations (49).

In the absence of positive feedback, the same design features described above like multimerization, multi-site phosphorylation, cooperative binding etc. can endow a GRN with an ultrasensitive response (47, 48). Unlike bistable systems, ultrasensitive networks can only buffer fluctuations far from the switching threshold, near-threshold fluctuations can cause these systems to switch back and forth between the two distinct gene expression levels (50). However bistability and hysteresis can dramatically slow down the response times of a GRN (48, 49). As a result, ultrasensitivity is a preferable alternative when response timing is critical for a differentiation program. Consequently many differentiation GRNs like the MAPK cascade in Xenopus oocytes, the mating decision network in yeast and the circuit controlling patterning of Drosophila embryos employ ultrasensitive designs (51-53).

Although switch-like behavior is perhaps the best studied functional requirement for cellular differentiation, novel network designs that achieve this
behavior are continually being discovered (54, 55). These results highlight the fact that to explain the diversity of GRN designs that can achieve ultrasensitivity or bistability it may be necessary to better understand functional properties associated with different switch-like response networks.

**Bet-hedging and noise management**

Genome-wide studies of cell-to-cell variability in mRNA or protein levels in *E. coli* and *Saccharomyces cerevisiae* have revealed the tremendous variation that can exist in the molecular make-up of even genetically identical cells (56, 57). This type of cell-to-cell variability originates from the intrinsically stochastic nature of biochemical reactions and is commonly referred to as intrinsic biological noise (58). Intrinsic noise in cells represents both opportunities and challenges for GRNs that control cellular differentiation. On the one hand, variability in network component levels can affect the reliability with which GRNs can control signal-dependent cellular differentiation programs. On the other hand, intrinsic variability offers a way for GRNs to create phenotypic heterogeneity in populations of genetically identical cells.

Under stress, generating variable phenotypes can actually be beneficial for the survival of populations, and stimulating noisy is an elegant way of achieving this (59, 60). The ultrasensitive and bistable GRN designs discussed above have been shown to amplify small fluctuations into bimodal gene expression profiles (49, 50). As a result, these designs can bifurcate cell populations into differentiated and un-differentiated clusters and thereby increase their survival likelihood in fluctuating environmental conditions.

Uncertainty of differentiation outcomes can significantly and adversely affect the survival of cells (59, 61). Consequently the ubiquity of intrinsic noise can be expected to be reflected in the structures of differentiation GRNs. Accordingly a number of studies have suggested that intrinsic noise might be the reason why homeostatic, perturbation-resistant negative feedbacks are widely prevalent in differentiation networks (59, 62). sRNA and miRNA based noise
regulating modules have also been identified as a common design feature (63-65). Although recent work has shown that there exist strict limits on the extent to which intrinsic noise can be regulated (66), how GRNs manage its effects on phenotypic variability remains an ongoing area of research.

**Signal integration and information processing**

The standard paradigm for understanding cell-fate decisions in cellular differentiation programs involves relating the choice to the level of one or two driving environmental or metabolic signals (40). This paradigm has been upended recently by the discovery that most differentiation GRNs are sensitive to a diverse constellation of signals that control their associated differentiation program.

Dramatic examples of multi-signal driven GRNs include the general stress response pathways of bacteria. These pathways include the RpoS regulatory network in *E. coli* (67) and the network in *B. subtilis* (68). Both networks can sense starvation (for C and P) as well as stress inducing conditions such as high or low pH, high or low temperature and high or low osmolarity. In response these networks activate large regulons of genes that result in changes in the stress-resistance, membrane permeability, morphology, metabolism and virulence of their respective bacteria (68, 69).

Other examples include the starvation sensing sporulation phosphorelay in Bacillus (70, 71), the yeast Msn2/4 network that controls the response to glucose starvation, osmotic shock and calcium stress (72-74) and p53 response network in mammalian cells that is activated by DNA damage, hypoxia, osmotic shock and ribonucleotide depletion (75-79).

How these GRNs manage to sense these myriad signals has been carefully worked out through extensive molecular genetic studies. Despite this wealth of available information how GRNs integrate and process these myriad signals is poorly understood even for the best studied networks. Nevertheless it is clear
that mechanistic principles for signal integration and information processing must exist and they likely play major role in shaping the design of these networks.

*Non-stochastic origins of cell-to-cell variability*

Contrasting with a decade of work focused on the role of intrinsic noise in differentiation, recent studies have begun to reveal a remarkable degree of determinism in the control of cellular differentiation (80-87). These studies have shown that the heterogeneity of phenotypic outcomes previously ascribed to biochemical noise is actually the result of cell-to-cell variability in network extrinsic features like cell size, local cell density and micro-environmental differences in differentiation signals and nutrient availability. A striking example of this new found determinism includes reports that the lysis-lysogeny decision of phage in *E. coli* cells – long viewed to be stochastic and unpredictable – is primarily dictated by the size of the infected bacterial cell (86). Similarly, it was recently shown that the probability of SV40 infection in HeLa cells is strongly predetermined by cell nucleus size and local cell density (88).

These results highlight how GRNs differ from electronic circuits in one crucial way. These networks and their parts are animated by shared cellular resources such as RNA polymerases, ribosomes and powered by energy currencies such as ATP, GTP etc (89, 90). As a result, GRNs operate in a biological milieu and are sensitive to fluctuations in their cellular environment.

How the cellular environment impinges upon the function of differentiation programs and GRNs is still being understood. Some recent studies have made important first steps in this direction by characterizing the impact of cell growth speed on the general patterns of gene expression in bacteria (91, 92). Other studies have shown how cell size and DNA content variability affects transcription rates and mRNA content in eukaryotic cells (93). Simultaneously various groups have begun to reexamine well-studied differentiation programs with the aim of cataloging the types of cell-to-cell variability that appear to impact
differentiation outcomes (86-88, 94). These efforts will undoubtedly lead to a revision of the design principles that govern the organization of GRNs.

Dynamics and cell-cycle control

Despite their obvious role in the regulation of cellular lifecycle, the control of cell biology related functions is frequently the least well understood aspect of differentiation related GRNs. In several well-studied cases of cellular differentiation, the process needs to be carefully coordinated with either cell division or DNA replication or both (12, 76, 95-97). In many of these cases the exact outputs of the GRN that specifically control cell cycle activity have been determined. Similarly it is known how the cell-cycle regulatory network connects to the differentiation GRN. Nevertheless the regulatory logic that governs the interfacing between these two critical types of regulatory networks is rarely clear. The problem arises because the cell-cycle by definition is a highly dynamic process (98-100). Accordingly its regulatory control must also be time-dependent. As a result, the standard approaches of understanding GRN function that focus on static diagrams of network structure and steady state analyses of its response fail to capture the underlying regulatory logic and make the network structure appear inscrutably complex.

Recently, however, single-cell experiments have begun to reveal a very different picture of these GRNs. In this view, many GRNs actively and spontaneously generate complex dynamic responses that include oscillations, spikes and pulses in the activity of key regulators (101-104). These dynamic responses possess the necessary temporal resolution to coordinate differentiation with critical cellular functions like DNA replication (105). That being said, design principles that can crystallize our understanding of how temporal coordination between cellular differentiation and the cell-cycle can be achieved have as of yet remained elusive.
1.3. Melding experimental and systems modeling approaches to understand cellular differentiation

Despite the progress of recent decades the complexity regulatory networks that control cellular differentiation still represents a major obstacle (17, 18). The biggest challenge is that standard network perturbation techniques—knockdown, overexpression and mutation have proven insufficient. These techniques while extremely effective at identifying the proteins involved in a differentiation program are significantly less effective at extracting mechanistic design principles (106, 107).

In response to this challenge, researchers have increasingly moved towards the application of quantitative methods drawn from engineering and physics for the analysis of differentiation networks (106, 108). The application of these methods has enabled the design of novel experiments that allow both the development of quantitatively accurate detailed models of regulatory networks as well as the identification of lower order, simplified representations of these networks.

These new experiments involve eschewing the standard paradigms of switching extracellular stimuli ON and OFF and investigating regulatory networks by successively inactivating each network component individually. Instead this new paradigm uses the methods of synthetic biology to tune the concentrations of network components and study how this impacts the transmission of signals and information down through the network (101, 109, 110). This mode of analysis is uniquely suited to uncovering detailed information about regulatory networks such as the critical nodes for feedback control or ultrasensitivity.

These methods also offer the opportunity to probe differentiation regulatory networks with periodic or arbitrary time-varying input signals (108, 111, 112). The use of time-varying signals to identify critical feedback modes and time-scale separation effects is a standard technique in the study of electronic circuits (113). These methods have till recently been unavailable to the study of biological
networks. However with the proliferation of synthetic gene regulatory tools and fluorescence based activity measurement methods, it is finally possible to bring these engineering techniques to bear on cellular differentiation problems.

Finally these new experimental developments enable the quantitative characterization of combinatorial gene expression regulatory elements like eukaryotic enhancers. In eukaryotes, or more specifically multi-cellular vertebrates, a large fraction of the genome is devoted to the precise spatial and temporal patterning of gene expression during differentiation and development (114). The discovery of a functional role for this ‘junk-DNA’ has spurred computational, empirical, and comparative studies that are producing an ever increasing collection of cis-regulatory elements commonly called enhancers (115, 116). These enhancers are complex regulatory modules with multiple protein–protein, protein–DNA and protein–promoter interactions (114). The reliability of models of transcriptional regulation in differentiation depends heavily on accurate characterization of these combinatorial regulatory enhancers (116-119). The use of synthetic biology methods has dramatically expanded our ability to create libraries of artificial enhancers created by mutating specific combinations of binding sites within an enhancer. Combined with a statistical thermodynamic based mathematical modeling approach, measurements of transcription from these mutant enhancers can be used to exhaustively characterize gene regulation and build truly predictive models.

The application of all of these experimental developments goes hand-in-hand with the development of mathematical models of the differentiation regulatory network under study. These models serve not only to store, catalogue, and condense the rapidly accumulating mass of experimental data but also to explicitly combine it with essential non-quantitative information that form a part of the model as assumptions and simulation frameworks (21, 22, 38). An appropriately constructed model is perhaps the most valuable tool in studying differentiation network because it is simultaneously a hypothesis generating engine and a testing-bed for experimental design. Crucially, mathematical
models enable the decomposition of the apparent complexity of regulatory networks (38, 106). By synthesizing experimental data, these models enable the identification of the core functional modules and signal transmission pathways in the regulatory network that are most critical to the differentiation program. This allows the study of one differentiation program in one particular model system to yield design principles that may be independent of specific biochemical details and therefore broadly applicable (13, 16). In this thesis we harness this approach and study three model cellular differentiation/stress-response networks to glean design principles that have are widely relevant to other programs and cell types.

**Organization of the Thesis**

This thesis is divided into three parts:
I. Regulation of Embryonic Hematopoiesis (Chapters 2-4).
II. Sporulation in Bacillus subtilis (Chapters 5-8).
III. General Stress Response in Bacillus subtilis (Chapter 9).

In Part I, we focus on the emergence of Hematopoietic Stem Cells (HSCs) from less differentiated pluripotent stem cells during embryogenesis. Hematopoietic stem cells are stem cells with self-renewal capacity that are the source of all blood cell lineages in vertebrate organisms. *In vitro* derivation of HSCs from Embryonic Stem Cells (ESCs) or induced Pluripotent Stem Cells (iPSCs) has huge potential in regenerative medicine (120). However the limited production of HSCs in current experimental protocols prevents in vitro HSCs from becoming a viable therapeutic option (121-123). Understanding the molecular mechanisms that drive HSC formation in the developing embryo and limit it in *in vitro* cultures will be crucial in tapping the potential of this therapeutic strategy.

The ontogeny of HSCs has been actively studied for decades using mouse, zebrafish and embryonic stem cells as model systems (124-126). These studies have demonstrated that emergence of HSCs in the embryo involves the interplay of many different transcription factors in the underlying Gene Regulatory
Network (GRN). In Chapter 2, we focus on a recently identified triad of transcription factors (Gata2, Fli1, Scl/Tal1) that form part of the GRN controlling HSC emergence (118, 127). This triad module also consists of the three regulatory enhancers through which Gata2, Fli1 and Scl positively regulate each other's transcription forming multiple direct and indirect positive feedback loops. Two signaling pathways, Bmp4 and Notch1, also control the response of Scl-Gata2-Fli1 triad (118). It has been suggested that the dense connectivity and positive feedback loops within stem cell GRN modules play important roles in stabilizing the stem cell phenotype (118). We investigate this hypothesis by constructing a mathematical model of the Scl-Gata2-Fli1 triad module and characterizing its dynamical properties. In addition we propose a novel no-contact mechanism whereby distant transcriptional enhancers may control gene expression by controlling chromatin state. Finally we develop a general thermodynamic framework to characterize combinatorial gene regulation using the measurements of the reporter transcription in a enhancer mutant library. Using these methods we identify a functional role for the design of the Scl-Gata2-Fli1 triad during HSC emergence. Our results also suggest that similar network architectures could play a more general role in the development of other major organ systems.

In Chapter 3, we investigate the no-contact mechanism of enhancer function and show that this mechanism is highly versatile. We find that this mechanism is capable of performing all the same logic gate regulatory functions as an enhancer-promoter contact based regulatory mechanism. We also show that the differences in the biophysical mechanism between the no-contact and contact based enhancer mechanisms lead to significant differences in their sensitivities to mutations and dynamic properties like stochastic wait-time distributions of transcriptionally active and inactive states.

In Chapter 4, we expand the model of the GRN regulating HSC emergence by explicitly including the essential hematopoietic gene Runx1, as well as Smad1 and Smad6 which are components of the Bmp4 signaling
pathway. Using this expanded model, we elucidate the role of Runx1 in the network and explain the early emergence of blood progenitors observed in Runx1 haploid embryos as well as other *in vitro* and *in vivo* experimental observations.

Part 2 of the thesis focuses on sporulation in *Bacillus subtilis*. Upon starvation soil bacterium *B. subtilis* initiates a differentiation program that results in the formation of stress-resistant spores (4). This sporulation program has been intensely studied for over 50 years as a model for cellular differentiation. Molecular genetic studies have been very successful in identifying the many components of the sporulation network controlling the execution of this differentiation program. These studies have shown that at the heart of the sporulation network is a phosphorelay in which phosphate is transferred from multiple histidine protein kinases (KinA-E) to a master transcription regulator Spo0A through two intermediate phosphotransferases Spo0F (0F) and Spo0B (0B) (70, 128). Despite the important advances made by these molecular genetic studies, several critical features of the sporulation decision making program have remained unexplained. In Chapters 5-8 we attempt to resolve four of these issues.

The master regulator Spo0A (0A) in its phosphorylated form Spo0A~P (0A~P) controls the expression of over a hundred sporulation specific genes and a threshold level of 0A~P is assumed to control the decision to sporulate in starvation conditions (129, 130). However it remains unclear how this 0A~P threshold is set by the sporulation network. In Chapter 5, we use a combination of a synthetic phosphorelay (which allows us to artificially control 0A~P levels) and mathematical modeling to answer this question.

In Chapter 6, we investigate why a gradual increase in 0A~P is essential for the successful execution of the sporulation program (131). To this end, we use an artificially engineered two-component system to control 0A~P accumulation and mathematical modeling to identify 0A~P target genes that may be perturbed by rapid 0A activation thereby resulting in sporulation failure.
Departing from the approach used in Chapters 5 and 6, in Chapter 7 we focus on the wildtype phosphorelay. Recent studies have shown that 0A~P levels pulse in a cell-cycle coupled fashion during starvation (102, 132) but the mechanism behind these pulses is not clear. We use a mathematical model of the phosphorelay to mechanistically explain the origin of these pulses, their cell-cycle coupling and their role in ensuring successful sporulation. Further we employ a combination of synthetic biology and time-lapse microscopy methods to verify our modeling predictions.

In Chapter 8, we address one of the most basic questions about the sporulation response: what is the environmental/metabolic signal that triggers sporulation initiation? Despite several decades of efforts, no specific nutritional trigger sporulation has been identified. Taking a new approach, we focus on the role of cell growth in controlling the response of the phosphorelay. We use mathematical modeling to demonstrate that changes in cell growth can affect accumulation of phosphorelay proteins and thereby 0A~P levels. We verify this model experimentally to establish growth rate as the primary starvation signal that drives cells to sporulate during starvation.

In Part 3 of the thesis we study the general stress response network of *Bacillus subtilis*. Chapter 9 focuses on the stress response sigma factor $\sigma^B$ and the regulatory network that controls its activation. $\sigma^B$ is activated in response to a wide variety of stresses including ethanol, starvation and energy stress agent like MPA and CCCP. This active sigma factor in turn activates the expression of over a hundred genes that protect the cell from these stresses. Recent studies have shown that under stress, the $\sigma^B$ network shows a pulsatile rather than a persistent response. We use a mathematical model of the $\sigma^B$ network to identify the design features that enable this pulsatile response. Further we use this model to demonstrate how the design of the $\sigma^B$ network enables it to successfully handle competition for RNA polymerase availability.
Regulation of embryonic hematopoiesis
Bistability and low-pass filtering in the network module determining blood stem cell fate

2.1. Introduction

Hematopoiesis has long served as a powerful model to study the specification and subsequent differentiation of stem cells (133). Sophisticated cell purification protocols coupled with powerful functional assays have allowed a very detailed reconstruction of the differentiation pathways leading from early mesoderm via hemangioblasts and hematopoietic stem cells (HSCs) to the multiple mature hematopoietic lineages. Transcriptional regulators (TRs) have long been recognized as key hematopoietic regulators but the wider networks within which they operate remain ill defined (134). Detailed molecular characterization of regulatory elements (enhancers/promoters) active during the early stages of HSC development has identified specific connections between major regulators (135-138) and has led to the definition of combinatorial regulatory codes specific for HSC enhancers (117, 139, 140). Moreover, these studies identified a substantial degree of cross-talk and positive feedback in the
connectivity of major HSC TRs (141). In particular, a triad of HSC TRs (Gata2, Fli1, Scl/Tal1) forms a regulatory module that appears to lie at the core of the HSC GRN (117). This module consists of the three transcription factor proteins as well as three regulatory elements through which they are connected via cross-regulatory and autoregulatory interactions (118, 135) (Fig. 2.1A). The details of regulatory interactions in this triad are shown in Fig. 2.1B; only significant binding sites in the enhancers are shown for simplicity. Gata2-3 and Fli1+12 enhancers contain multiple Gata2 (GATA), Fli1 (ETS) and Scl (E-BOX) binding motifs. The Scl+19 enhancer contains ETS and GATA binding motifs. Scl, Gata2 and Fli1 are all essential for normal hematopoiesis in mice (135) suggesting that the triad is an important sub-circuit or kernel of the GRN that governs hematopoiesis.

Figure 2.1 Regulation of gene expression in the Scl-Gata2-Fli1 triad.
A. Scl, Gata2 and Fli1 form a triad module of TRs in the GRN of hematopoietic stem cells. The triad architecture consists of multiple positive feedback loops. Signals activating or deactivating the network are shown in magenta. Notch activates the transcription of Gata2 and Bmp4 activates the transcription of Gata2 and Fli1 by acting
at the promoters. Gata1 binds to the Gata2 enhancer and downregulates Gata2 expression.

**B.** The triad proteins regulate each other’s transcription by acting at the Scl+19, Gata2-3 and Fli1+12 enhancers. These enhancers contain multiple binding sites that allow combinatorial control of gene expression. Only sites significantly affecting expression are shown.

**C.** Enhancer libraries similar to the one shown for Scl were constructed for all three proteins and sub-cloned with a suitable reporter in and in triad expressing cells to characterize the combinatorial control of gene expression. Typical results show the enhancement of gene expression from TR binding sites individually and in combination relative to enhancer-less expression of the reporter.

The triad architecture (Fig. 2.1A) is very dense in regulatory connections and possesses multiple direct and indirect positive feedback loops. Such network topologies are rare in prokaryotes (30) but have been identified in other stem cell systems such as the Nanog-Oct4-Sox2 triad in the embryonic stem cell GRN (142, 143). These observations suggest that the triad design may be associated with stem cell behavior. This idea prompted further investigation of combinatorial control by the triad TRs (118). Generation of an enhancer library with wild type and mutant enhancers allowed the construction of different combinations of binding motifs in each enhancer. Wild type and mutant enhancers were sub-cloned into a SV minimal promoter and lacZ reporter vector and tested using stable transfection of hematopoietic progenitor cell lines (118). This analysis produced results such as those schematically illustrated in Fig. 2.1C.

It has been suggested that the dense connectivity and positive feedback loops within stem cell GRN modules play important roles in stabilizing the stem cell phenotype (117). However, the dynamical nature as to how this self-enforcing circuit may be initiated or indeed exited remains unclear. In this paper we construct a mathematical model of the Scl-Gata2-Fli1 triad module and characterize its dynamical properties using continuous ODE modeling approaches. We first propose a thermodynamic method of estimating free energies of different configurations of the enhancer regions from the
measurements of the transcriptional reporter libraries. This method together with a proposed biochemical mechanism of distant transcriptional enhancement significantly reduces dimensionality of the network parameter space. Measurements of protein lifetimes provide experimentally informed timescales to model transient behavior of the network. We analyze the network response to physiologically relevant signals such as Notch, Bmp4 and Gata1 and show that the network behaves as an irreversible bistable switch in response to these signals. Our model also predicts the results of various mutations in the enhancer sequences and shows that the triad module can ignore transient differentiation signals shorter than threshold duration. The combination of a bistable switch with short signal filtering not only provides new mechanistic insights as to how the Scl-Gata2-Fli1 triad may function to control HSC specification and differentiation but also suggests a possibly more general role for this network architecture in the development of other major organ systems.

2.2. Results

2.2.1. Thermodynamic model for enhancement of gene expression

Full quantitative characterization of the combinatorial nature of transcriptional regulation requires measurements of binding affinities between the DNA and TRs as well as interaction strengths among TRs. Moreover, the contribution of each individual TR and each possible combination to the transcriptional rate must be assessed. This information is extremely tedious to measure due to the combinatorial multiplicity of TR configurations and does not exist for the majority of experimental systems. Experimental data for synthetic libraries of transcriptional reporters that contain the gene regulatory elements is more readily available. We develop thermodynamic methods to characterize the combinatorial transcriptional regulation by distal enhancers based on this type of data and apply it to model the Scl-Gata2-Fli1 triad - a core module of the GRN of hematopoietic stem cells. Recently this system has been experimentally
characterized (118). In this study distal enhancer regions regulating the transcriptional rate of network proteins were identified and the relative contributions of each of the regulatory motifs were thereafter assessed individually and in combination by the use of a suitable transcriptional reporter (e.g., luciferase, lacZ). The typical results from these experiments are illustrated in Fig. 2.1; see Table A1.1 for the full data used. We use this data to obtain the functional form describing the transcriptional rate of the reporter-enhancer constructs and estimate the biochemical parameters characterizing this function. Below we illustrate our approach for the Scl+19 enhancer; the full model is derived in the methods section.

We assume that the distant enhancers increase the transcriptional rate via modulation of chromatin remodeling rather than through direct interaction with transcriptional machinery. This assumption is motivated by the observations that activation of the Scl+19 enhancer is only revealed upon integration of the enhancer-promoter construct into chromatin and that the activity of the enhancer is independent of its position (upstream or downstream) relative to the reporter gene (118, 144). Moreover, when integrated as single copy reporters into the genome of embryonic stem cells and assayed following 5 days of in vitro differentiation, the difference between wild type and mutant enhancer constructs lies in the number of cells that express the transgene rather than the level at which it is expressed (cf. Fig. A1.1 and Appendix 1). Taken together, these observations suggest that chromatin dynamics play a significant role in the action of TRs at the enhancers. In the absence of enhancer binding, the gene can be in either open or a relatively stable closed chromatin state. In the closed chromatin state the binding regions for the TRs and the transcriptional machinery are wrapped in nucleosomes and are inaccessible; thus no gene expression is possible from this state. The closed chromatin state can spontaneously unwrap to an open state where the binding sites become accessible to allow polymerase to bind to the promoter and initiate transcription. Since most promoters bind RNA polymerase weakly, the probability of RNA polymerase binding and subsequently transcription rate \( I \) is proportional to the probability of the chromatin being in the
open state \( I = I_o p_o \); see Methods Eqs (15)-(17)). This probability depends on the equilibrium between open and closed chromatin states. Binding of the TRs at the enhancer stabilizes the open conformation thus shifting the equilibrium towards the open state (cf. Fig. A1.2). This way the probability of open conformation increases with increase in TR concentration or increase in binding affinity. The rate of gene expression is still given by \( I_o p_o \) but \( p_o \) is now defined by a more complicated thermodynamic expression accounting for all the possible configurations of TR binding. Mutations in the enhancer site eliminate the configuration of TR binding thereby affecting \( p_o \) but not \( I_o \). Below we illustrate this formalism for the Scl+19 enhancer.

The Scl+19 enhancer contains binding sites for Gata2 and a Fli1 dimer and therefore can exist in closed and four different open states (enhancer empty, Gata2 bound, Fli1 dimer bound, both Gata2 and Fli1 bound). The cumulative probability of all open state configurations is then given by \( p_o = 1 - p_{\text{closed}} \), where \( p_{\text{closed}} \) is the probability of the closed state given by

\[
p_{\text{closed}} = e^{-\beta G^C_s} / Z_s \quad (1)
\]

where subscript \( s \) denotes the Scl+19 enhancer: \( G^C_s \) is the effective closed state energy, and \( Z_s \) is the partition function given by the sum of exponentiated free energies \( G^\alpha_s \) of each state \( \alpha \): \( Z_s = \sum e^{-\beta G^\alpha_s} \). \( \beta = 1 / kT \) is an inverse temperature and hereafter all free energies are in its units. For TR-bound states, free energies are concentration dependent due to the loss of entropic degrees of freedom, e.g. for the Gata2-bound state \( G^\alpha_s = G^\alpha_{\text{Gata2}} - \log([\text{GAT}]) \), where \([\text{GAT}]\) denotes concentration of Gata2. (Similarly \([\text{SCL}]\) and \([\text{FLI}]\) denote concentrations of Scl and Fli1 respectively). Since the free energies are only defined up-to a constant we can choose the free energy of the open state to be zero and thus obtain the following expression for the partition function:
\[ Z_s = 1 + e^{-G^c_s} + [GAT]e^{-G^{Gata2}_s} + [FLI]^2 e^{-G^{FLI}_s} + [FLI]^2 [GAT]e^{-G^{FLi1Gata2}_s} = e^{-G^c_s} + Z^E_s \] (2)

where \( G^{FLi1}_s \) and \( G^{FLi1Gata2}_s \) represent the free energies of Fli1 dimer and Gata2-Fli1 multimer binding and \( Z^E_s \) is the partition function for all open chromatin states. We use the subscript \( s \) in all these terms to specify that they are associated with the \( \text{Scl}+19 \) enhancer and the superscript to specify the binding configuration (cf. Table A1.2 for notation).

Direct measurements of the binding free energies in this expression may be tedious but these can be straightforwardly computed from the ratios of the transcription rates from synthetic reporter libraries with full or mutated enhancer sites. Ratios of the reporter expression levels of cell lines with wild-type (wt) and mutated (mut) enhancers can be used as constraints on the values of the binding free energies.

\[
\frac{I^{wt}}{I^{mut}} = \frac{p^{wt}_o}{p^{mut}_o} = \frac{1 - e^{-G^c_s}}{Z^{wt}_s} / \frac{1 - e^{-G^c_s}}{Z^{mut}_s} \] (3)

Equations similar to (3) can be constructed for all reporter-enhancer libraries and used to recursively compute the binding free energies (cf. Eqs (22)-(27) in Methods and Eqs (1)-(11) in Appendix 1).

**2.2.2. Mathematical model for the Scl-Gata2-Fli1 Network triad module**

Scl, Gata2 and Fli1 form an interconnected triad of positive interactions and play an important role in hematopoietic differentiation (118, 135). To understand the role of the unique architecture of the triad module we construct a dynamical model of the system.

Assuming first-order degradation kinetics, deterministic rate equations for the change in TR concentrations take the form
where the functions $V_{S}^{+}$, $V_{G}^{+}$ and $V_{F}^{+}$ describes the rates production whereas $k_{d}^{S}$, $k_{d}^{G}$ and $k_{d}^{F}$ denote degradation rate constants for Scl, Gata2 and Fli1 respectively. Rate constants for protein degradation are estimated from known half-lives of the proteins. Since proteins are long-lived relative to mRNA, we can assume that production rates are directly proportional to the respective transcription rates $I' = I_{o}p_{o}'$ (cf. Eq (28)).

In addition to distant enhancers, Notch and Bmp4 are known to serve as activators of the promoters of Gata2 and Fli1, Gata2 respectively (145, 146). These activators increase the rate of transcription by increasing the recruitment of RNA polymerase to the respective promoter. In particular, Notch and Bmp4 increase Gata2 expression by 3.5 fold (146) and 4 fold (147) respectively. In this case, to compute $V_{G}^{+}$ one needs thermodynamic expressions of the probabilities of multiple open conformations corresponding to binding of Notch or Bmp4. These probabilities depend upon Notch and Bmp4 concentrations ($[N]$ and $[B]$ respectively) and their binding energies $G^{N}$ and $G^{B}$ via the full partition function $Z_{g}$ (subscript g stands for Gata2-3 enhancer):

$$Z_{g} = K_{g} + (1 + [N]e^{-G^{N}} + [B]e^{-G^{B}})Z_{E}^{g},$$

$$Z_{E}^{g} = 1 + [GAT]e^{-G_{Gata2}^{<}} + [FLI]^{2} e^{-G_{Fli1}^{<}} + [FLI]^{2} [GAT] e^{-G_{Gata2Gata2}^{<}} + [SCL][GAT][FLI]^{2} e^{-G_{SCLGATAFLI}^{<}}$$

(5)

Here $K_{g} = e^{-G_{g}^{C}}$ is the equilibrium constant for chromatin transitions between open and closed states for Gata2 enhancer (similarly $K_{s} = e^{-G_{s}^{C}}$ and...
\[ K_i = e^{-G_i^C} \] for Scl+19 and Fli1+12 enhancers respectively). These equilibrium constants are dimensionless quantities characterizing the maximum possible fold enhancement of gene expression by the respective enhancer. The partition functions are used to compute Gata2 synthesis rate \( V_{\text{G}} \) (cf. Eq (20)). The same procedure is used to describe the rate of expression of Fli1, although in this case only Bmp4 acts at the promoter (cf. Eq. (21)).

Conversion to dimensionless form can greatly simplify the model allowing easy interpretation of simulation results. We normalize the species concentrations of Scl, Gata2 and Fli1 as \([scl] = [SCL]/[SCL], [gat] = [GAT]/[GAT], [fli] = [FLI]/[FLI]\). \([SCL], [GAT], [FLI]\) represent the mean observed concentrations of Scl, Gata2 and Fli1 in wildtype HSCs where the triad is actively expressed. In addition, \([n]\) and \([b]\) are Notch and Bmp4 concentrations normalized with respect to their promoter dissociation constants. With these normalizations, wildtype HSCs in the absence of signals would have \([scl] = [gat] = [fli] = 1\) and \([n] = [b] = 0\). We choose this state as a reference state for the estimation of free-energies (cf. Methods Section for details). The dimensionless form of equation (4) is then given by

\[
\frac{1}{k_a^s} \frac{d[scl]}{dt} = \frac{p^s_i([gat],[fli])}{p^s_o(1,1)} - [scl]
\]

\[
\frac{1}{k_a^G} \frac{d[gat]}{dt} = \frac{p^G_i([scl],[gat],[fli],[n],[b])}{p^G_o(1,1,1;0,0)} - [gat]
\]

\[
\frac{1}{k_a^F} \frac{d[fli]}{dt} = \frac{p^F_i([scl],[gat],[fli],[b])}{p^F_o(1,1,1;0)} - [fli]
\]

Where \(p^s_o\) are dimensionless synthesis rates (cf. Eq 25). Note that in the final form of our model equations the wild-type state of HSCs \([scl] = [gat] = [fli] = 1\); is always a steady state in the absence of signal
...\text{free} \quad G_{Gata2Fli1}sG \quad \text{free} \quad G_{Gata2Fli1}ssG \quad \text{free} \quad G_{Gata2Fli1}ssG \quad \text{free} \quad G_{Gata2Fli1}ssG 

The matching is only possible for the values of equilibrium constant above a threshold: \( K_s > 819.51 \). This lower bound is simple a consequence of the fact that in the proposed thermodynamic framework the maximal possible enhancement is given by \( K_s + 1 \) and the experimentally
measurable enhancement is 820.51. Similarly the free energies for the Gata2-3 and FlI1+12 enhancers are estimated based on the experimental results and $K_s$ and $K_f$, respectively (cf. Methods section and Appendix 1 for details). The values of these constants are also limited from below by the respective maximal measured enhancer factors.

Figure 2.2. Steady state signal-response analysis of the triad module to Notch, Bmp4 and Gata1 signals demonstrates irreversible bistability.

A. The action of Notch and Bmp4 at the promoters of switches the triad module from a low expression (OFF) state to a high expression (ON) state. Only Gata2 concentrations are shown for brevity. Solid lines represent stable and dotted lines represent unstable steady states. (Notch and Bmp4 concentrations are normalized by their respective binding affinities). Once the triad is in the ON state, the positive feedback loops in the modules architecture ensure that it remains in that state without signals (inset: the same plot in the linear scale). The switchability of the triad steady state response is sensitive to the values of $K_f$ and $K_s$. 
B, C. Effect of different values for chromatin equilibrium constants on steady state response to Bmp4 and Notch. For $K_g = 233.5$ in B, only Bmp4 can switch the triad from OFF to ON. For $K_g = 235$ (C) neither Notch nor Bmp4 can switch the triad to ON state.

D. Bistable response of the triad module to Gata1 repressor signal. Gata1 competes with Gata2 for binding sites on the Gata2-3 enhancer and can switch the triad from ON state to OFF by decreasing the recruitment of RNA polymerase to the Gata2 promoter by a factor $f([\text{Gata1}])$. As a result the system irreversibly switches from ON to OFF. (note that this figure is shown in linear scale, the inset shows the deactivation in log-log scale for comparison with A). To evaluate the steady state dose response of each signal individually the concentrations of other of other signals were kept fixed at zero during simulation.

In addition qualitative information about system behavior, namely its switchability as a response to physiologically relevant Notch and Bmp4 signals, places an upper bound on chromatin equilibrium constants values. For a different set of $K$ values the computed free energies are such that Notch and/or Bmp4 cannot cause the switch between low and high steady states (Fig. 2.2BC). As a result the system remains switchable in the very narrow range of two equilibrium constants $\left(819.51 \leq K_s \leq 819.69, \ 233.38 \leq K_g \leq 233.47\right)$ where the full enhancer brings the transcriptional rate to a nearly saturated value. The resulting narrow ranges do not indicate lack of model robustness but rather are a consequence of strict constraints placed on free energy values by the exact matching to the experimental reporter data (cf. equations (1)-(11) in Appendix 1). In fact without these constraints the range of $K_s$ and $K_g$ for switchable bistable response extends over several orders of magnitude (cf. Fig. A1.3 and below). If we tolerate some deviation from the experimentally measured transcriptional data we can relax these constraints and significantly enhance the range of parameter values for which the system is bistable and switchable. For example, if we allow up to 20% deviation from transcriptional reporter measurements then the values of chromatin equilibrium constants can vary by 20% and still result in switchable
response (data not shown). It is quite reasonable to tolerate such levels of deviation from the experimental results because the experimental results usually have a margin of error. Therefore we find that the qualitative predictions of the model (switchable bistable response) are robust however the quantitative predictions (transcriptional data) are only as accurate as the experimental data one which the model is based.

We expect the triad to be switchable in response to both Notch and Bmp4. Therefore we choose the chromatin equilibrium constants from within the narrow ranges shown above and calculate the TR-enhancer binding free energies using these chosen values. For this chosen set of parameter values the model shows an irreversible bistable response to Notch and Bmp4 (Fig. 2.2A). Bmp4 concentrations were set to zero for evaluating the Notch dose response and vice versa. The presence of one signal reduces the threshold concentration of the other signal at which the triad switches from OFF to ON (data not shown). The calculated free energies are shown in Table A1.3 and used through the remaining simulations. Once the free energies of TR binding are fixed at Table A1.1 values, the system becomes robust to variability of chromatin equilibrium constants (Fig. A1.3). Such changes may biologically correspond to histone modification or other physical perturbations. In response to changes over a large range the triad shows switchable and irreversible bistable responses to Notch and Bmp4 (Fig. A1.3). Therefore the switchable nature of triad bistability is robust to several fold parameter changes.

Gata1 can displace Gata2 from its binding sites in the Gata2-3 enhancer. Through competition for binding sites and subsequent chromatin remodeling Gata1 can switch the triad from high expression back to the low expression state. We represent the chromatin remodeling effect of Gata1 by including a factor $0<f([\text{Gata1}])<1$ in our expression for the rate of Gata2 gene transcription $I_g = I_0 p_o f([\text{Gata1}])$. Because the exact biochemical mechanism of the Gata1 action is not established we choose a decreasing function of Gata1
and make no other assumptions about the functional form of $f([Gata1])$. We therefore, plot Gata1 dose-response curves with $f([Gata1])$ as the x-axis where its values decrease left to right (Fig. 2.2D). This phenomenological description of the effect of Gata1 captures the effect it has on RNA polymerase recruitment to the promoter by initiating chromatin remodeling. Inclusion of Gata1 in our model (Fig. 2.2D) allows the system to switch from ON to OFF states. The switching is irreversible – the system will remain OFF even after Gata1 signal is gone ($f([Gata1])=1$). Notch and Bmp4 concentrations were fixed at zero for evaluating the Gata1 response because the concurrence of Notch/Bmp4 and Gata1 signals is physiologically unlikely.

Interestingly, Gata1-deactivation is far more susceptible to noise than the activation by Notch/Bmp4. This can be concluded from the dotted line representing the unstable steady state that separates the stable ON and OFF states (compare Figs. 2.2A and D). This line characterizes the magnitude of concentration fluctuations required for spontaneous transitions. For sub-threshold signals, this line is much closer to the stable steady state in Gata1 dose-response curves (Fig. 2D) as compared to Notch or Bmp4 curves (Fig. 2.2A). A more rigorous investigation of the magnitude of stochastic effects and their relation to separatrix of deterministic model requires a full stochastic model of the network and will be conducted elsewhere.

2.2.4. Enhancer mutations change the steady-state response of the triad

We expect the steady state response of the Scl-Gata2-Fli1 module depends on the triad architecture and design of enhancers. The model presented above allows us to verify this claim by introducing changes in the triad design corresponding to mutations of enhancer sequence and gene knockouts and examining the effects on the steady state response. To this end, we systematically deleted TR-binding sites from each enhancer in silico and
analyzed the steady state response of the system. We also analyze the steady state response of \textit{Scl}, \textit{Gata2} and \textit{Fli1} deletion mutants.

Mutations in the triad enhancer sequences can produce many modules with simpler architecture as shown in Fig. 2.3. Notably, since some TR-enhancer configurations do not make a significant contribution to the enhancer activity, removal of a single enhancer binding site might effectively eliminate multiple TR-enhancer interactions. For example, the effect of Scl on the Gata2 and Fli1 enhancers is only significant when both Gata2 and Fli1 are bound to the enhancer. Therefore the probability of Scl bound enhancer configurations for these enhancers is negligible for any motif where the Gata2 or Fli1 sites on these enhancers are deleted.

Keeping this in mind we analyze 10 different triad module designs that can be obtained by selective single and double mutations of enhancer binding sites. The model described above is suitably altered to predict the steady state response of these alternate designs. All relevant parameter values are taken from the full triad model. Of the 10 “mutant” designs, all 6 modules where the \textit{Scl+19} or \textit{Gata2-3} enhancers are mutated show only a single steady state with the expression of Scl, Gata2 and Fli1 comparable to the low expression state of the full triad (cf. Fig. 2.3A). On the other hand, high levels of expression can still be observed in 4 modules with mutations in the \textit{Fli1+12} enhancer (see Figs. 2.3BC). However, in contrast to wild-type (Fig. 2A), this high level of expression cannot be maintained in the absence of Notch and Bmp4. Even when the E-BOX binding site for Scl is eliminated from the \textit{Fli1+12} enhancer the system remains bistable for a range of signal. For the designs in which the GATA site in the \textit{Fli1+12} enhancer is eliminated (Fig. 2.3C) Fli1 expression is uncoupled from Gata2 and Scl and is monostable while the responses of Scl and Gata2 are still bistable. This is expected because Fli1 autoregulation is not strong enough to produce bistability.
Figure 2.3. Selective deletion of enhancer binding sites can change the steady state response characteristics.
A. Deletion of any of the enhancer binding sites from the Scl+19 or Gata2-3 enhancers eliminates the high expression state of Scl, Gata2 and Fli1 seen in the wildtype HSCs. Black crosses mark the deleted sites, red crosses mark the interactions that are no longer significant as a result of the deletion.

B. Mutations in the Scl or Fli1 binding site in the Fli1+12 enhancer allow triad activation but lead to reversible bistability—the ON state switches back to OFF in the absence of Notch and Bmp4.

C. Deletion of the primary Gata2 binding site from the Fli1+12 enhancer makes the Scl interaction with the enhancer insignificant. This effectively makes Fli1 independent of external regulators Scl and Gata2. Fli1 expression is low for these mutants and monostable. Notch has no effect on Fli1 concentration. Gata2 and Scl show reversible bistability in response to Notch and Bmp4 in these mutants.

Complementarily, we can also assess the effects from alterations of TRs rather than their binding sites. Simulations show that Scl\(^{-/-}\), Gata2\(^{-/-}\) and Fli1\(^{-/-}\) knockout mutants cannot support the high expression state of the triad. These mutants produce a phenotype similar to the enhancer mutations in Fig. 2.3A. Comprehensive analysis of knockout mice has shown that hematopoiesis is severely impaired in all three deletion mutants (148-151). Our model suggests that the knockout of any of the triad proteins prevents the switch to ON state which is likely to affect the specification of HSCs during early embryonic development and therefore compromise the development of all mature blood cell types as seen experimentally. On the other hand, the irreversible bistability of triad response is preserved if we delete one chromosomal copy of any one of the three triad genes; however the heterozygotic mutants are expected to be more prone to differentiation (cf. Fig. A1.4 and Appendix 1). This could explain why these mutants have reduced repopulation capacity (152, 153).

2.2.5. Dynamical response of the triad module architecture

The dynamics of the response of the bistable triad module to a pulse of Notch is illustrated in Fig. 2.4A. The step increase in Notch concentration almost
immediately increases Gata2 concentration slightly. However Fli1 concentration remains stagnant because Scl level rises very slowly. The slow speed of Scl response is governed by its slow degradation rate (half-life ~8 hrs). Once enough Scl has accumulated, the probability of Scl being present on the Gata2 and Fli1 enhancers becomes significant. This results in a rapid increase of expression rates and the triad switches to the high expression state. The rate limiting step for switching ON the triad expression levels is therefore the slow accumulation of Scl.

To further investigate the dynamics of triad switching in response to transient stimuli we have computed the minimal pulse duration that can cause irreversible switching as a function of signal amplitude (Fig. 2.4BC; black lines). The results indicate that the system can be switched ON by signal pulses longer than a certain threshold level (~42 hrs for a Notch pulse and ~21 hrs for a Bmp4 pulse). This threshold is a few fold larger than Scl-lifetime, the longest timescale for the system. Our simulations therefore indicate that the triad module is capable of filtering transient signals that are shorter than the threshold simulation. We refer to this property as low-pass filtering – a term accepted for similar phenomena in engineering literature (154). This filtering appears to be related to the slow turnover of Scl and the feedback loops connecting Scl with Gata2 and Fli1.
Figure 2.4. Comparison of dynamical responses of the triad and the reduced module to Notch, Bmp4 and Gata1 signals.

A. A time course of the switching from low expression to high expression state in response to a pulse of Notch. The inset shows that there is an increase in Gata2
concentration immediately after the introduction of Notch. Scl starts accumulating slowly in response to this increase in Gata2 concentration but Fli1 concentration is stagnant because enough Scl is not present to appreciably increase Fli1 expression. Once Scl has reached the required concentration ($t_{acc}$ after start of Notch pulse) Gata2 and Fli1 concentrations increase rapidly to the ON state level. Thus switching of the triad to the high expression ON state is rate-limited by the slow accumulation of Scl.

B. The minimum Notch pulse-duration required for OFF->ON switching as a function of pulse amplitude. Black line is the full triad the red curve is the reduced module with constitutive Scl (cf. text for details).

C. Same as (B) but for Bmp signal.

D. Steady-state response of the reduced module (red) with the Scl concentration fixed at the value that ensures that the switching threshold is identical to that of the wild-type triad (black). Note that the unsteady state (separatrix-dotted curves) for the reduced module is much closer to the ON state.

E. Controlled comparison for deactivation by Gata1 steady-state response with Scl concentration fixed to ensure that the deactivation thresholds for both modules are identical.

F. Transient filtering of Gata1 signals is very similar for the two designs since Scl does not limit the rate of response to Gata1.

To understand how slow Scl dynamics contributes to the filtering of transient Notch and Bmp4 signals we compare the dynamics of the triad module to that of a simpler network module where the Scl+19 enhancer has been deleted. We call this module the reduced module. In this reduced module Scl is assumed to be under an external regulator that controls Scl concentration. With this reduction, Scl concentration is constant and the dynamics of Gata2 and Fli1 response are not limited by the slow accumulation of Scl. For a controlled comparison of the dynamics (155) we assume that all relevant parameters have the same values as they do in the full triad model. This leaves the Scl concentration as the only free parameter. The reduced module shows irreversible bistable response to Notch, Bmp4 and Gata1 for a range of Scl values. We constrain the Scl concentration such that the threshold for OFF to ON transitions is the same for the reduced module and the full triad (Fig. 2.4D). Notably, the
separatrix between the two stable states (dotted line, Fig. 2.4D) is much closer to the ON state for the reduced module. This suggests that the reduced module is more susceptible to fluctuations in TR levels as compared to the full triad.

We now use the reduced module as described above for a controlled comparison of the dynamics of the OFF to ON and ON to OFF switching. Both bistable switches act as filters for transient signals above the threshold (Fig. 2.4BC). We compared this dynamic response of the triad and reduced modules to Notch and Bmp4 pulses. The models for the two modules have the same Notch/Bmp4 thresholds and close to the threshold the minimum pulse duration for both modules is high. However at higher concentrations of Notch and Bmp4, the minimum pulse duration is much higher for the triad module than for the reduced module (16 hrs and 9.5 hrs for Notch and Bmp4 pulses respectively). These results show how the slow dynamics of Scl allow the full triad module to act as a better low pass filter function for activation as compared to the reduced module.

For a controlled comparison of the response of the two modules to Gata1 we fix the Scl concentration of the reduced module such that the threshold level of Gata1 is identical (Fig. 2.4E). This fixed concentration of Scl is 4 fold higher for deactivation than for activation. Gata1 acts at the Gata2-3 enhancer to shut off transcription through chromatin remodeling. The slow dynamics of Scl do not affect the Gata2 concentration during this deactivation. As a result the deactivation dynamics and the minimum pulse duration for ON to OFF switching at high Gata1 concentrations (~8 hrs) of the reduced module and the full triad are identical. The triad and reduced module are equivalent low pass filters for deactivation signals such as Gata1.
2.3. Discussion

A new method for determining free energies of TR-DNA interactions

Combinatorial gene regulation is ubiquitous in eukaryotes with complex DNA regulatory regions acting as integration points for multiple signals and pathways involved in gene regulation. The characterization of these regulatory regions through mathematical models is an important step towards understanding the functionality of gene regulatory networks. In order to fully characterize each regulatory element, one needs to determine dynamical functions that describe the rate of transcription as a function of TR concentrations. The most biochemically and biophysically realistic method of characterizing transcriptional regulation is rooted in statistical thermodynamics where each state of the regulatory region is assigned a free-energy so that the probability of each state can be computed from Boltzmann distribution (156). These methods have been previously applied to bacterial systems (157) but rarely used for eukaryotic gene regulatory networks as a lack of reliable parameter measurements prevents researchers from undertaking detailed modeling approaches. Here we have developed a method for the quantitative characterization of combinatorial gene regulation by multiple TRs in eukaryotic distant enhancers. Our proposed method extends the thermodynamic approach of (158) in order to relate it to experimental transcriptional reporter assays. We develop a recursive method to estimate relevant free energies from the measurements of combinatorial libraries of transcriptional reporters. There are multiple benefits of computing free-energies of TR-DNA configurations. First, these parameters allow straightforward construction of mathematical models for quantitative analysis of system behavior with no or just a few free parameters. Second, free energies can be used for model reduction by specifically excluding thermodynamically unfavorable states and subsequent model reduction. Third, the parameters provide important qualitative insights into gene regulatory mechanisms such as cooperativity of TRs. We further reduce the number of...
parameters required to characterize the distant transcriptional enhancers by proposing a detailed mechanism based on the modulation of chromatin remodeling activity.

Chromatin structure is known to play an important role in eukaryotic gene regulation. The organization of DNA into nucleosomes can prevent the transcriptional machinery and regulatory factors from accessing regulatory regions. The detailed mechanism of action of distant enhancer sites has not been established. It has been suggested however that its action may involve modulation of chromatin remodeling dynamics (159). For instance, regulatory elements of the Scl-Gata2-Fli1 triad were shown to be critically dependent on integration into chromatin (135). Here we propose a ratchet mechanism of enhancer action (cf. Fig. A1.2). We propose that DNA can be in a dynamic equilibrium between open (promoter site accessible) and closed (promoter site inaccessible) conformations. Such a dynamic equilibrium between wrapped and unwrapped nucleosomal DNA has also been discussed elsewhere (160). In the absence of enhancer TRs, the equilibrium is heavily shifted towards a closed state resulting in very low transcription probability. We hypothesize that binding of TRs to the enhancer site stabilizes an open conformation and thereby shifts the equilibrium towards it. This mechanism therefore allows the TRs to ratchet the spontaneous unwrapping of nucleosomal DNA and trap it in a state accessible to the transcriptional machinery. We apply this thermodynamic framework to a regulatory module hypothesized to play a pivotal role in hematopoiesis. Under this assumption the binding of Fli1, Gata2 and Scl to their enhancer sites activates gene transcription by increasing the probability of transcription rather than the rate of transcription. This hypothesis is consistent with previously reported results of studies focused on enhancer function in mammalian cells (159, 161, 162) and with our flow cytometry experiments with cells containing the Scl+19 enhancer-reporter constructs (cf. Fig. A1.2).

The proposed mechanism assumes that the unwrapping of DNA from nucleosomes is independent of all triad factors and thus effectively spontaneous.
However chromatin modification and chromatin remodeling factors can affect these nucleosome dynamics. In particular, factors such as the Gata2 repressor Gata1 may regulate the expression by modulating free energies of DNA unwrapping through chromatin modification. By shifting the equilibrium further towards the closed state, Gata1 can suppress transcription to such an extent that TR concentrations are too low to ratchet the very short-lived open state.

**Steady state characteristics of the Scl-Gata2-Fli1 triad**

The recently characterized Scl-Fli1-Gata2 triad module includes a large number of transcriptional interactions resulting in multiple positive feedback loops. The complex enhancer structure makes it rather difficult to phenomenologically deduce dynamical expressions for Scl, Gata2 and Fli1 transcription. However, with our newly developed approach based on transcriptional reporter data, construction of a mathematical model of the triad becomes a straightforward task. The resulting model of the triad exhibited bistability in response to the action of Notch and Bmp4. We have chosen the free energy values for DNA unwrapping to ensure that the action of these two activators at the promoters switches the triad from low expression (OFF) state to high expression state (ON). The model predicts this switching to be irreversible – the triad will remain ON even after the signals are gone (Fig 2.2A). The development of HSCs in the hemogenic endothelium is known to be a Notch regulated event (163). Notch is known to be expressed in endothelial cells and act as a regulator of Gata2 expression during the onset of hematopoiesis (146). Bmp4 expression has also been observed in the dorsal aorta region where HSCs first develop in the embryo (117, 164). Notch and Bmp4 are known to be mediators of HSC specification during embryonic development (163). Our model shows how the action of Notch and Bmp4 is crucial for the OFF to ON switch of the Scl-Gata2-Fli1 triad. Since HSC specification requires Scl, our model predicts that in the absence of Notch and Bmp4, newly generated HSCs are trapped in a low expression state and hematopoietic development is compromised. The network also irreversibly switches from the ON to OFF state when reaching a
threshold value of repression of Gata2 transcription by Gata1. The network will then remain in the OFF state in the absence of other signals. Interestingly, in ref. (142) the authors use a mathematical model to predict that a similar triad module in embryonic stem cells is also bistable. However their module is expected to be bistable only in the presence of activating or deactivating signals unlike Scl-Gata2-Fli1 triad that shows irreversible bistability.

Our analysis indicates essential roles of all the enhancer sites included in the model in maintaining irreversible bistability in steady state dose-response curves of the triad. Elimination of any binding sites in Scl or Gata2 enhancers leads to complete elimination of bistability with only the OFF state remaining. Mutations in the Fli1 enhancer may lead to a reversible bistability phenotype in which the triad is activated only in the presence of Notch and/or Bmp4 signals above a certain threshold. We emphasize, however, that these predictions do not indicate that simpler triad networks with less autoregulation are incapable of achieving irreversible bistable switching behavior. Our goal was to predict the behavior of the triad to the mutations of the regulatory regions. If one allows compensatory changes in other model parameters one can restore the irreversible switching behavior and even set the switching threshold to be equal to that of wild-type triad. However, as indicated below the reduced modules may still display physiologically important differences in other aspects of dynamic behavior.

**Transient responsiveness of the Scl-Gata2-Fli1 triad**

In order to characterize the transient responses of the triad module one needs the values of kinetic parameters – lifetimes of triad proteins. Scl and Fli1 are known to be relatively stable proteins with measured half-lives of 8 hours and 2 hours respectively (165, 166). Gata2 is comparatively unstable with a half-life less than 30 minutes (166). This combination of short-lived and long-lived transcription regulators allows the triad to respond quickly to changes in mRNA transcription rates and at the same time, act as memory modules for history-
dependent switches into and out of the HSC regulatory state. Analysis of the dynamical response of the triad to Notch/Bmp4 indicates that slow accumulation of Scl acts as a rate-limiting step for OFF-ON switching. As a consequence the triad must be exposed to Notch/Bmp4 signals for significant time periods for switching to occur. Physiologically this means that the triad motif works as a low-pass filter that responds only to transient stimuli longer than threshold duration and ignores brief, transient signals shorter than the threshold duration. All bistable switches show this type of threshold filtering of transient signals but to a different degree (167). In our case, the response rate for the triad is limited by slow Scl dynamics and therefore multiple features of the triad network contribute to this property. For example, Scl is the slowest in degradation among the TRs and Notch/Bmp4 signals affect its accumulation only indirectly (Fig. 2.1A). In addition, we hypothesized that the positive feedback loops involving Scl play a significant role in determining the threshold for low-pass filtering. We have confirmed this hypothesis by comparing the response of the triad to a hypothetical reduced module wherein the Scl+19 enhancer is deleted and Scl acts as an external TR for the Gata2-Fli1 feedback loop (155). We therefore conclude that the full triad is a better low pass filter because of the rate-limiting nature of Scl accumulation and Scl-mediated positive feedbacks significantly affect the signal filtering properties of the triad.

Studies in heterogeneous cell populations derived from differentiating ES cells or mouse fetal liver had suggested low level binding of Scl itself to the Scl+19 enhancer (117). However, more recent analysis in a clonal population of blood stem/progenitor cells did not detect any binding of Scl to this element (138). Positive autoregulatory feedback through the Scl+19 enhancer is therefore unlikely to play a significant role in stem cells, especially as the Scl+19 element does not contain a bona fide binding site for Scl which would necessitate indirect binding. Nevertheless, we have considered the addition of a positive auto-feedback loop on Scl but simulations demonstrated that it does not generate a qualitatively different scenario with the only major consequence being a further
slow-down of the switching rate due to the retardation of response by positive feedback (data not shown).

Gata1 acts at the Gata2-3 enhancer and is reported to actively promote chromatin modification (168). The decrease in Gata2 concentrations is not limited by Scl dynamics because Gata1 directly affects Gata2 transcription by reducing RNA polymerase recruitment. We therefore expected that filtering characteristics of the full and reduced triad motif would be the same. We performed a controlled comparison choosing a concentration of Scl in the reduced module (with the Scl+19 deleted) that ensures the same switching threshold. The results indeed show essentially identical low-pass filtering properties of the two modules because Scl dynamics are not rate limiting in this case.

The Scl-Gata2-Fli1 triad as a central regulator of stem cell fate

Experiments have shown that the knockout of any one of the genes Scl, Gata2 or Fli1 affects the development of HSCs and leads to severely impaired hematopoiesis. Thus the expression of these TRs is critical for hematopoiesis. More recent studies have shown that these three genes regulate each other by acting at distant enhancers as activators. Results from our model provide insight into the function of this module of TRs and suggest that the triad is a central regulator that controls the specification of HSCs during early hematopoiesis and the generation of progenitors committed to differentiation from these cells.

The bistable switch properties of the triad are hallmarks of a decision module. The triad switches irreversibly from the low to high expression state in response to external cues such as Notch and Bmp4 that are important for establishing definitive HSCs in the hemogenic endothelium. The bistable response predicted by the model is robust to fluctuations in parameter values. Experimental results also support this prediction (169). The model shows that the knockout mutants are unable to reach the activated high expression state due to the all or none nature of this bistable response. Additionally the slow turnover of
Scl retards the triad response to Notch and Bmp4 and thus makes it a highly effective low pass filter for noise in these signals.

The response to deactivation by Gata1 is not affected by Scl dynamics. As a result the $ON$ to $OFF$ switch for the triad is much faster than the $OFF$ to $ON$ switch. Deactivation by Gata1 is also more sensitive to stochastic fluctuations in triad protein concentrations. The cells can be switched to the $OFF$ state to produce progenitor cells committed to differentiation by fluctuations in triad TR concentrations. Thus asymmetric partitioning of these proteins during cell division can allow sub-threshold Gata1 concentrations to silence Gata2 expression in one of the daughter cells by chromatin modification. The probability of this stochastic exit from the pluripotent HSC state of the cell is governed by the Gata1 concentration in the cell. This observation is consistent with experimental analysis of a multipotent hematopoietic progenitor cell line which demonstrated that these cells exist in two distinct subpopulations when cultured under self-renewal conditions with the more differentiation prone subpopulation expressing higher levels of Gata1 (170). Of note, the triad switches between states in an all or none fashion where overexpression of exogenous Gata2 for example could prevent deactivation of the triad by Gata1. In line with these predictions, it has been demonstrated that overexpression of Gata2 in differentiating ES cells increases the production of hematopoietic progenitors and slows down their differentiation (171).

Our model of the triad module shows that it responds differently to activation and deactivation signals. This allows the $OFF$ to $ON$ and $ON$ to $OFF$ switches to fulfill different functional requirements. The activation response is slow, irreversible and robust to fluctuations in external signals to allow the development of HSCs in a noisy intercellular signaling environment. Simulation results for the dynamics of deactivation suggest that it may be faster than the $OFF$ to $ON$ switch and may exploit stochastic intracellular fluctuations during the cell cycle to maintain the HSC population and guarantee a continuous supply of lineage committed progenitors at the same time.
From a model based on the quantitative experimental characterization of the triad enhancers we have predicted several qualitative features of the steady state and transient response of the triad as well as its sensitivity to mutations and over-expression. We favored a deterministic model for our analysis of the triad function because of the reliability and robustness of the predictions that we have been able to extract from this approach. Even so, a stochastic model can potentially offer additional information about noise properties of the system and we intend to use results presented here to guide the construction of a full stochastic model in the future. Taken together the results presented here are consistent with prior experimental data and provide new mechanistic insights into potentially critical features of the regulatory networks that govern the specification and subsequent differentiation of hematopoietic stem cells. Moreover, our strategy of exploiting experimental data to infer biophysical properties should be widely applicable to aid regulatory network reconstruction in a wide range of cellular and developmental systems.

2.4. Methods

*Modeling regulation at the enhancer level*

We extend the Shea-Ackers (172) description of gene regulation to construct the deterministic models discussed above. The following assumptions are the foundation of this modeling approach,

The TR-DNA binding and unbinding processes are fast compared to transcription and translation and can be assumed to be at equilibrium. We note that the equilibrium assumption may only be applicable for the population-average deterministic model we construct here and may fail to accurately describe single-cell data.

The rate of gene transcription is linearly related to the probability of RNA polymerase \( R_p \) being bound to the promoter.
The assumption of equilibrium allows us to calculate the probability of finding TRs bound to DNA using the Boltzmann weighting factors for all configurations (occupied and unoccupied) of the DNA regulatory element (156). The sum of the Boltzmann factors for all configurations is the partition function

\[ Z = \sum_{\alpha} e^{-\beta G^\alpha} \]  

Here \( G^\alpha \) is the free energy of the state \( \alpha \) (we measure free energies \( G^\alpha \) in units of \( kT \) and use \( \beta = 1 \)). The partition function is used to calculate the probability of each of configuration. We distinguish three different types of regulatory element configurations based upon our model of nucleosome dynamics.

Closed chromatin configuration for the enhancers: The DNA of the regulatory enhancer is tightly wrapped around histones. No DNA binding proteins (including RNA polymerase) can access binding sites when the DNA is in this configuration. No gene transcription occurs while the gene is in closed chromatin state.

Open chromatin configurations: Spontaneous unwrapping of DNA from the histones produces a configuration where none of the TRs are bound to the enhancer but binding of RNA polymerase to the promoter is allowed. Gene transcription can happen in this state.

Occupied enhancer configurations: DNA is unwrapped from the histones and enhancers are occupied by TRs. This set of configurations includes all possible configurations of TRs at the enhancer. RNA polymerase can bind to the promoter in this state leading to gene transcription.

We use these definitions to formulate the probabilities \( p_{\text{open}}, p_{\alpha}^\text{open} \) and \( p_{\text{closed}} \) of open chromatin with no TR binding, different enhancer bound states \( \alpha \) and closed chromatin respectively:
Here the energy for all states is measured relative to the open chromatin state (which is set to zero).

The Gottgens group cloned the Scl+19, Gata2-3 and Fli1+12 enhancers upstream of a SV promoter controlling a lacZ reporter gene and integrated this construct into the genome of wild-type HSCs that show high expression of Scl, Gata2 and Fli1 (118). In the presence of all three TRs, the enhancer can be occupied in many different TR configurations and reporter expression is significantly higher than constructs with no enhancer. Mutant enhancers where certain TR binding sites have been deleted were also used with reporter gene constructs to measure the gene expression enhancement. The results from these experiments show that only the deletion of certain critical enhancer binding sites affects gene expression enhancement. These critical sites are shown in Fig. 2.1 and the experimental results from (118) are included in Table A1.1. We use these results to simplify the model of combinatorial gene regulation in the triad.

The expression of Scl is under the control of two TRs Gata2 and Fli1 with different binding sites in the Scl+19 enhancer. The Scl+19 enhancer can therefore be in either closed state, open state, bound by Gata2, bound by Fli1 dimer or bound by Gata2 and Fli1 dimer simultaneously. Given the various configurations of the enhancer, the derivation of the partition function is straightforward (cf. Eq. (2)).

We define $Z^E_i$ as the sum of the Boltzmann weights of all open state enhancer configurations for ease of representation of the probability of open chromatin states in equation (9). Note, that the binding energy $G^{Fli1,Gata2}_i$ includes the TR-TR interaction of Gata2 and 2 Fli1 TRs while bound to DNA.
The *Gata2*-3 enhancer includes binding sites for *Gata2*, *Scl* and the 2 *Fli1* TRs. Many TR binding sites can be deleted without affecting the reporter gene expression enhancement (118). The binding sites for *Scl*, *Gata2* and *Fli1* shown in Fig. 2.1 are critical for gene expression enhancement. Gene expression is decreased but still significantly enhanced if only the *Gata2* or *Fli1* sites are present. Deletion of all sites except *Scl* binding site makes the expression enhancement negligible. However deletion of only the *Scl* site significantly decreases the expression enhancement from the full enhancer. These results suggest that although *Scl* binds weakly to the incomplete enhancer by itself, the *Scl*-*Gata2*-*Fli1* complex has great affinity for the *Gata2*-3 enhancer. Of all possible configurations of *Gata2*-3 enhancer occupation only the *Gata2* bound, *Fli1* bound, *Gata2*-*Fli1* bound and *Scl*-*Gata2*-*Fli1* bound configurations are therefore included in the partition function $Z_g$ for *Gata2*-3.

$$Z_g = e^{-G_g} + Z^F_g,$$

where

$$Z^F_g = 1 + e^{-G_S} + [GAT]e^{-G_{Gata2}} + [FLI]^2 e^{-G_{Fli1}} + [FLI]^2 [GAT]e^{-G_{Fli1/Gata2}} + [SCL][GAT][FLI]^2 e^{-G_{Scl/Gata2/Fli1}}$$

Fig. 2.1 also shows the critical *Fli1+12* enhancer binding sites. This enhancer includes two *Gata2* binding sites (primary site at 5′ end). The *Scl* binding site and the secondary *Gata2* site (3′ end) cannot enhance gene expression by themselves. The primary *Gata2* site and the *Fli1* dimer sites have some effect on gene expression and together they raise gene expression ~20 fold. Single mutation of either the *Scl* or secondary *Gata2* binding sites has a negligible effect on gene expression. Deletion of both sites together reduces the gene expression enhancement from ~60 fold to ~20 fold. Thus the *Gata2* bound, *Fli1* bound, *Gata2*-*Fli1* bound and *Gata2*-Scl-*Fli1*-Gata2 bound configurations have a significant effect on the gene expression. Incorporating these experimental results simplifies the partition functions $Z_f$ for *Fli1*. 
50

\[ Z_f = e^{-G_f} + Z_f^E, \text{ where} \]

\[ Z_f^E = 1 + e^{-G_f^C} + [GAT]e^{-G_f^{Gata2}} + [FLI]^2 e^{-G_f^{Fli1}} \]

\[ + [FLI]^2 [GAT]e^{-G_f^{Fli1/Gata2}} + [SCL][GAT]^2 [FLI]^2 e^{-G_f^{Gata2/Fli1}} \]  \hspace{1cm} (11)

**Modeling regulation at the promoter level**

So far we have enumerated all configurations of the enhancers. Notch ([N]), Bmp4 ([B]) and RNA polymerase ([R]) each can bind at different promoters in the triad when chromatin is on the open state with binding affinities that are represented here as free energies \( G^N, G^B \) and \( G^P \) respectively. These free energies can vary for different promoters and also depend upon energy of interactions between different proteins bound to DNA. We note that the triad enhancers bind TRs to regulate gene expression in a chromatin integration dependent manner (135). Moreover the position of the enhancer does not affect its ability to regulate transcription. These results suggest that the enhancer bound TRs do not physically interact with promoter bound factors such as Notch, Bmp4 and RNA polymerase to affect transcription. Therefore we assume that the free energy of interaction between enhancer and promoter bound proteins is zero. We assume that the binding of Notch/Bmp4 and RNA polymerase at the promoter is cooperative. Under this assumption the binding of RNA polymerase at the promoters is enhanced by the free energies of its interaction with Notch \( (G^{NP}) \) and Bmp4 \( (G^{BP}) \). In our partition functions, we now account for configurations where either the enhancer or the promoter or both or neither are occupied by the various factors. We assume that \([R_p]e^{-G^P} \ll 1 \) because typical promoters bind RNA polymerase weakly and use this assumption to simplify the equations below.

\[ Z_i = e^{-G_i} + Z_i^E(1 + [R_p]e^{-G^P}) \approx e^{-G_i} + Z_i^E \]  \hspace{1cm} (12)
\[ Z_g = e^{-G^C_g} + Z_g^E (1 + [N]e^{-G^N_g} + [B]e^{-G^B_g} + [R_p]e^{-G^P_g} (1 + [N]e^{-G^N_{g-P'}} + [B]e^{-G^B_{g-P'}})) \]
\[ \approx e^{-G^C_g} + Z_g^E (1 + [N]e^{-G^N_g} + [B]e^{-G^B_g}) \]
(13)

\[ Z_f = e^{-G^C_f} + Z_f^E (1 + [B]e^{-G^B_f} + [R_p]e^{-G^P_f} (1 + [B]e^{-G^B_{f-P'}})) \]
\[ \approx e^{-G^C_f} + Z_f^E (1 + [B]e^{-G^B_f}) \]
(14)

Interestingly, even though we assumed in our derivation that there is no physical interaction between enhancer bound and promoter bound TRs we find that the partition functions of the Gata2-3 and Fli1+12 enhancers are not separable \((Z \neq Z^E Z^{P'})\) into distinct factors \(Z^E\) and \(Z^{P'}\) representing the partition functions for the enhancer states and promoter states respectively. Therefore the binding of TRs at the enhancers and the promoter is not independent. This emergence of cooperativity from competition of TRs with nucleosomes has been observed experimentally (173) and incorporated into mathematical models (174).

We define \(K_s = e^{-G^C_s}\), \(K_g = e^{-G^C_g}\) and \(K_f = e^{-G^C_f}\) to be the equilibrium constants of chromatin rewrapping for the Scl, Gata2 and Fli1 respectively. Using equation (12), the probability of RNA polymerase being bound to the Scl promoter can be written as

\[ p^s(R_p) = \frac{Z_g^E [R_p] e^{-G^P_g}}{K_s + Z_g^E} \]  
(15)

Similarly we can write the expressions for the probability of Gata2 and Fli1 promoters being occupied by polymerases.

\[ p^s(R_p) = \frac{Z_g^E (1 + [N]e^{-G^N_{g-P'}} + [B]e^{-G^B_{g-P'}})[R_p]e^{-G^P_g}}{K_g (1 + [N]e^{-G^N_g} + [B]e^{-G^B_g})Z_g^E} \]  
(16)

\[ p^f(R_p) = \frac{Z_f^E (1 + [B]e^{-G^B_{f-P'}})[R_p]e^{-G^P_f}}{K_f (1 + [B]e^{-G^B_f})Z_f^E} \]  
(17)
We note that the effect of Notch and Bmp4 on the probability of transcription from the Gata2-3 enhancer is saturable because Notch and Bmp4 concentrations ([N] and [B] respectively) appear in both the numerator and denominator of the expression for \( p^g(R_p) \) (cf. Eq. (16)). Similarly the effect of Bmp4 on the probability of transcription from the Fli1+12 enhancer (\( p^f(R_p) \)) is also saturable (cf. Eq. (17)). The rate of gene expression for gene \( i \), \( I^i = k p^i(R_p) \) is assumed to be proportional to the probability of promoter occupation by RNA polymerase. The proportionality constant \( k \) is the rate of isomerization of RNA polymerase to the open conformation. We rearrange the rate of gene expression as

\[
I^i = I_o p^i, \text{where } I_o = k \left[ R_p \right] e^{-G^F}
\]

(18)

\( I_o \) represents the maximal rate of expression from the promoter in the open state. \( p^i_o \) is a dimensionless rate of transcription that represents the cumulative regulatory effect of all enhancer and promoter bound TRs. Using equations (15)-(17) we can now write the expressions for \( p^i_o, p^g_o \) and \( p^f_o \).

\[
p^i_o = \frac{Z^E_i}{K_g + Z^E_i}
\]

(19)

\[
p^g_o = \frac{Z^E_s (1 + \left[ N \right] e^{-G^N} e^{-G^{NSp}} + \left[ B \right] e^{-G^B} e^{-G^{BP}})}{K_g + (1 + \left[ N \right] e^{-G^N} + \left[ B \right] e^{-G^B}) Z^E_i}
\]

(20)

\[
p^f_o = \frac{Z^E_i (1 + \left[ B \right] e^{-G^B} e^{-G^{BP}})}{K_f + (1 + \left[ B \right] e^{-G^B}) Z^E_i}
\]

(21)

Recursive estimation of free energies from experimental results

Deletion of binding sites from the enhancer \( i \) modifies \( Z^E_i \), the partition coefficient for all bound configurations of that enhancer. Experimental results from the Gottgens group describe the fold-change in gene expression

\[
E^s = E^o + E^{Eg} + E^{EB} + E^{EBg} + E^{EBp} + E^{EBg} + E^{EBp}
\]

(19)

\[
E^f = E^o + E^{Eg} + E^{EB} + E^{EBg} + E^{EBp} + E^{EBg} + E^{EBp}
\]

(20)

\[
E^o = \frac{Z^E_i}{K_g + Z^E_i}
\]

(21)

Recursive estimation of free energies from experimental results

Deletion of binding sites from the enhancer \( i \) modifies \( Z^E_i \), the partition coefficient for all bound configurations of that enhancer. Experimental results from the Gottgens group describe the fold-change in gene expression
enhancement due to the selective mutation of certain enhancer binding sites (118). Using their results for deletion of critical binding sites we can estimate the free energies of each TR-DNA interaction for the three enhancers. We use Scl+19 as an illustrative example. Fig. 2.1 shows the Scl+19 enhancer and the fold expression enhancement for the reporter construct in the presence of the wildtype (wt) enhancer and three mutant enhancers: Mutant enhancer 1 (mut1)- Fli1 binding site deleted, Mutant enhancer 2 (mut2)-Gata2 binding site deleted, Mutant enhancer 3 (mut3) - all binding sites deleted. The transcription rates $I_{wt}^s, I_{mut1}^s, I_{mut2}^s, I_{mut3}^s$ are normalized with the expression rate $I_{mut3}^s$ of the reporter when all enhancer binding sites have been deleted. We assume that the lacZ reporter transcription rates are proportional to the fluorescence intensities measured in these experiments because all experiments were performed in the presence of excess fluorescent substrate and wild-type and mutant constructs were assayed at the same time using the same reagents. Moreover the experimental conditions were controlled to ensure that the proportionality constants that relate various transcription rates $I$ to the fluorescent intensities are the same for different experiments.

Note that the experimental results were obtained in HSCs which show high expression levels of Scl, Gata2 and Fli1 (118). Notch and Bmp4 signals are expected to be absent in these cells (175). We accordingly exclude all Notch and Bmp4 states from our partition functions. We can see from equations (19)-(21) that $p_o^s, p_o^d$ and $p_i^d$ are the probabilities of the Scl+19, Gata2-3 and Fli1+12 enhancers being in open state in the absence of Notch and Bmp4.

The introduction of the mutant enhancer reporter construct is not expected to affect the growth rate or availability of RNA polymerases in a significant manner. Thus $I_o^i$ is unaffected by the deletion of binding sites. However the deletion of Fli1 binding sites eliminates the Fli1 bound state in the enhancer partition function $Z_o^s$ in equation (15). Therefore $p_o^s$ is affected by deletion of binding sites. Since $I^s = I_o^s p_o^s$, using equations (22)-(24) we can relate the fold...
enhancement in gene expression to the free energies of TR-DNA interaction.

\[ I_{\text{mut1}} = \frac{[GAT] e^{-G_{\text{Gata}2}} + 1}{1/(K_s + 1)} \]  \hspace{1cm} (22)

\[ I_{\text{mut3}} = \frac{[GAT] e^{-G_{\text{Gata}2}} + 1}{1/(K_s + 1)} \]

\[ I_{\text{mut2}} = \frac{[FLI]^2 e^{-G_{\text{Fli}1}} + 1}{1/(K_s + 1)} \]  \hspace{1cm} (23)

Equations (22) and (23) can be solved analytically for \( G_{s_{Gata2}} \) and \( G_{s_{Fli1}} \) as functions of \( K_s \) and the concentrations \([GAT]\) and \([FLI]\).

\[ G_{s_{Gata2}} = \log([GAT]) - \log \left( \frac{(1 - I_{\text{mut1}}/I_{\text{mut3}})(K_s + 1)}{1 - I_{\text{mut1}}/I_{\text{mut3}}} \right) \]  \hspace{1cm} (25)

\[ G_{s_{Fli1}} = 2\log([FLI]) - \log \left( \frac{(1 - I_{\text{mut2}}/I_{\text{mut3}})(K_s + 1)}{1 - I_{\text{mut2}}/I_{\text{mut3}}} \right) \]  \hspace{1cm} (26)

The solution for \( G_{s_{Fli1Gata2}} \) is dependent on \( G_{s_{Gata2}} \) and \( G_{s_{Fli1}} \). Using (25) and (26) we can solve for \( G_{s_{Fli1Gata2}} \) and reduce it to a function of only \( K_s \), \([GAT]\) and \([FLI]\).

\[ G_{s_{Fli1Gata2}} = \log([GAT][FLI]^2) - \log \left( \frac{(1 - I_{\text{mut1}}/I_{\text{mut3}})(K_s + 1)}{1 - I_{\text{mut1}}/I_{\text{mut3}}} \right) \]

\[ - \log \left( \frac{(1 - I_{\text{mut2}}/I_{\text{mut3}})(K_s + 1)}{1 - I_{\text{mut2}}/I_{\text{mut3}}} \right) \]  \hspace{1cm} (27)

We apply this recursive procedure to uniquely determine in a similar fashion all free energies of \( Gata2-3 \) and \( Fli1+12 \) enhancers. The full equations for all free energies are presented in Appendix 1.
**Dynamical equations**

Since mRNA is labile relative to stable cellular proteins, we assume that the mRNA concentration for the triad proteins is at steady state. We can thus directly relate the rate of transcription \( I_i = I_o p_o^i \) to the rate of production of the proteins

\[
V_i^+ = k_i I_o p_o^i = I_i p_o^i
\]  
\( (28) \)

(here \( k_i \) represents the number of protein molecules produced per mRNA lifetime). The ODEs for change in protein concentration can be written as a balance between the rate of production \( V_i^+ \) and the degradation/dilution rates that are linear in protein concentration (cf. Eq (4)).

The major hurdle in the analysis of this ODE model is the determination of all TR-enhancer interaction free energies from the equations described above (25)-(27) and in the supplement (Appendix 1). The free energies can be determined from these relations if the concentrations of Scl, Gata2 and Fli1 in the wildtype cells and the constants \( K_s, K_g, K_f \) are known. However the TR concentrations are difficult to measure *in vivo*. We make our equations dimensionless to avoid the measurement of actual Scl, Gata2 and Fli1 concentrations. We normalize these TR concentrations by their wildtype concentrations. In wildtype HSCs the Scl, Gata2 and Fli1 concentrations are at steady state. Let these steady state wildtype concentrations be \([SCL],[GAT]\) and \([FLI]\). Normalizing Scl, Gata2 and Fli1 concentrations with \([SCL],[GAT]\) and \([FLI]\) we can rewrite equation (4) as a system of ODEs in dimensionless variables \([scl],[gat]\) and \([fli]\) (cf. Eq. (6)).

Rates \( p_o^i \) for all three enhancers as given by equations (19)-(21) can be recalculated in terms of the dimensionless variables by adjusting the free energies of each state with the appropriate concentrations. For example,
Dimensionless rates \( p_o^s(1,1), p_o^s(1,1,1;0,0) \) and \( p_o^f(1,1,1;0) \) the wild-type, steady state dimensionless rates of transcription can be evaluated from the expressions in (29) by using adjusted free energies and \([scl]=[gat]=[fli]=1\). Then \([SCL]=I_1p_o^s(1,1)/k^s_\delta\), \([GAT]=I_1p_o^s(1,1;1,0)/k^s_\delta\) and \([FLI]=I_1p_o^f(1,1;1;0)/k^f_\delta\).

The parameter space of free energies can now easily be explored by tuning \(K_s,K_g,K_f\). Since the free energies can be determined by fixing \(K_s,K_g,K_f\), we can also analyze the system response to Notch and Bmp4 by substituting the full expressions of \(p_o^s, p_o^f\) and \(p_o^f\) in equation (6).

\(e^{-G^{NP}}\) and \(e^{-G^{Bp}}\) represent the strength of the interaction between RNA polymerase and Notch and Bmp4 respectively. Notch and Bmp4 increase Gata2 expression in wildtype HSCs by 3.5 (146) and 4 fold (147) respectively. At
saturating concentrations of Notch (high \([n]\)) \( \frac{Z^E_g (1+[n]e^{-Gsp})}{K^g_r + (1+[n])Z^E_g} \approx e^{-Gsp} \). This implies \( e^{-Gsp} = 3.5 p^g_s (1,1,1;0,0) \). And similarly, \( e^{-Gsp} = 4 p^g_o (1,1,1;0,0) \). Thus \( K_s, K_g, K_f \) are the only unknown parameters in our model.

The model offers both a quantitative means of analysis of combinatorial regulation of gene expression by TRs and a succinct mathematical description of the biophysics of the regulation. The model can easily be extended to regulation involving repressors and many other situations.

The reduced model where Scl is not under regulation by Gata2 or Fli1 represents a simplification of this system where the concentration of Scl is kept constant. The reduced system then comprises only the equations for Gata2 and Fli1. The time normalization is carried out relative to the Scl half-life (~8 hrs) (165). Gata2 and Fli1 have half-lives of ~10 minutes and 2 hours respectively (166). Accordingly \( k^G_d = 0.00144 \text{min}^{-1}, k^G_d = 0.07 \text{min}^{-1} \) and \( k^F_d = 0.0057 \text{min}^{-1} \). Our method for estimation of binding affinities reduces the number of unknown parameters in the system to three chromatin rewrapping equilibrium constants. These constants have been reported to be in the range 10 - 10000 (176). We find that for irreversible bistable behavior with switchability our parameter estimation scheme restricts two of these equilibrium constants to a narrow range.

\[
\text{819.51} \leq K_s \leq 819.69 \\
233.38 \leq K_g \leq 233.47 \\
59 \leq K_f 
\]

We chose the following values for the equilibrium constants: \( K_s = 819.54 \), \( K_g = 233.44 \) and \( K_f = 61.0 \) from within the ranges. The free energy values are thereafter calculated as described above to complete the parameter set for the triad model (cf. Table A1.3). The same parameter values are retained for the reduced model, however the Scl concentration in this case is fixed such that the
threshold concentration (the concentration at the bifurcation point) of Notch/Bmp4 is identical for both the full triad and reduced model.

Simulations

The system of equations for the triad described in the previous section was analyzed using a number of numerical methods and tools. The steady state characterization of both the triad and reduced modules was carried out using XPPAUT and the associated bifurcation analysis package AUTO (177). Parameter sensitivity analysis for the chromatin equilibrium constants was also done with AUTO. The analysis of the dynamics of the ODE model was carried out using the ODE45 solver of MATLAB 2008a(R) (The MathWorks, Natick, Massachusetts). To compute the minimum pulse duration for Notch/Bmp4 signals, the integration was initiated at the low steady state and a step input of Notch/Bmp4 was introduced. The pulse duration to switch the system was minimized using the `fmincon` function (Optimization toolbox) in MATLAB. In all simulations only the dimensionless models were used.

Notes

This chapter is based on work that was published in the following article:

Chapter 3

Thermodynamic models of combinatorial gene regulation by distant enhancers

3.1. Introduction

Differential regulation of gene expression is key to cellular diversity in complex organisms. Its dynamical properties are controlled by underlying gene regulatory networks (GRNs) consisting of transcription factor genes and their cis-regulatory elements that, together with basic transcriptional machinery, control the expression levels of each gene (2). Complexity of genetic regulation in higher organisms is related to the complexity of the underlying networks rather than the number of genes (178). In particular, this complexity often manifests itself in combinatorial regulation of gene expression with multiple inputs converging on regulatory control elements. Binding sites for transcriptional regulators are found either in the immediate vicinity of a transcription initiation site or in the enhancer sequences situated several kilobases upstream or downstream (179).
The molecular mechanisms of gene regulation via distant enhancers are not very well understood. The proposed mechanisms of distal regulation of gene expression can be broadly characterized into two classes – contact mechanisms and non-contact mechanisms. Contact mechanisms involve DNA looping or packing that brings the enhancer-bound proteins close to the promoter (Fig. 3.1A) (180). Non-contact mechanisms do not rely on direct physical contact of the enhancer-bound proteins and transcriptional machinery (160, 174). Proposed mechanisms of noncontact enhancer action include super-helical tension in negatively supercoiled DNA, nuclear localization and nucleosome remodeling (179). The nucleosome-remodeling hypothesis is particularly attractive as it explains why chromatin integration is often essential to observe any enhancer action (117, 181). It also explains why enhancers in some single-cell measurements affect the probability of transcription rather than the rate of transcription (159, 181) and how in many cases enhancers regulate transcription in a manner that is independent of their orientation and distance relative to the transcription initiation site (182).

Dynamical modeling of GRNs is often essential to understand their functionality as it provides information about network steady states and its responsiveness to physiologically important inputs and perturbations. Construction of such models requires functional expressions that relate concentrations of transcription factors to the rate of transcription of regulated genes. Common approaches to constructing such input functions include the use of Boolean functions (such as logical gates), Hill functions and thermodynamic models. Boolean models and ordinary differential equation models that use Hill functions provide useful qualitative information about the behavior of regulatory networks. However, these approaches are based on phenomenological information about the networks rather than a specific biophysical mechanism of gene regulation (183-185). In contrast, thermodynamic treatment of transcriptional regulation provides a rigorous method to translate hypotheses about the mechanism of transcriptional regulation into quantitative models (156, 172, 186). This approach has been extensively used to model bacterial gene
regulation but has not been widely adopted for combinatorial regulation in higher organisms.

We recently developed a thermodynamic model of distant enhancer activation via chromatin disruption and applied it to the dynamic modeling of the core network module in hematopoiesis (181). The model assumes that the structure of chromatin in the gene neighborhood is in either an unstable open state that allows binding of the transcriptional machinery and gene transcription or a relatively stable closed state that does not allow transcription (see Figure 3.1B). In the closed chromatin state the binding regions for the transcriptional machinery are wrapped in nucleosomes and are inaccessible and no gene expression is possible from this state. The closed chromatin state can spontaneously unwrap to an open state where the binding sites become accessible and allow the transcriptional machinery to bind to the promoter and initiate transcription. This model of chromatin structure dynamics is based upon experimental results that show (1) that the structure of chromatin, in particular nucleosomes, can impede transcriptional initiation (187, 188), (2) chromatin exists in a dynamic equilibrium of open and closed states (176) and (3) transcription factors (TFs) can disrupt chromatin structure by displacing nucleosomes to control gene transcription (189, 190). The central idea of our model is that by binding at the enhancer and modulating chromatin structure, TFs can control the rate of gene expression without any physical interactions with the transcriptional machinery.

In this paper, we further develop a general thermodynamic framework to construct input functions of combinatorial gene regulation. We generalize this mechanism to include the possibility of negative regulation via stabilization of the closed chromatin conformation. With that generalization, we show that this chromatin mechanism is capable of generating the same logical input functions as the direct contact mechanism of transcriptional regulation (156, 186). We further compare the sensitivities of the resulting input functions with respect to changes of the parameters and indicate important distinctions between contact
and chromatin mechanisms. In addition, we describe an approach that uses gene expression reporter measurements to estimate thermodynamic parameters and thereby characterize the complete response function for any enhancer design and apply it to characterize the regulation of Gata2, an essential hematopoietic stem cell gene, by a distant enhancer. Finally, we compare the dynamic properties of the two mechanisms with respect to wait-time distributions of transcriptionally active and inactive states.

Our results indicate that the chromatin mechanism of gene regulation can perform the same logic gate type input functions for transcriptional regulation as the contact mechanism. However, the differences in the biophysical mechanism (direct contact vs. chromatin) lead to differences in the design of regulatory elements, in sensitivities to mutations and in dynamical properties.

Figure 3.1. Two mechanisms of combinatorial gene regulation by distant enhancers.
A. Contact mechanism: DNA looping brings the distant enhancer bound TFs A and B close to the promoter bound transcriptional machinery to allow protein-protein interactions. TFs A and B can activate transcription by stabilizing promoter-bound transcriptional machinery or repress transcription by destabilizing or sterically hindering the binding of the transcriptional machinery to the promoter.

B. Chromatin mechanism: RNA polymerase binding sites are inaccessible in the closed chromatin state where DNA is tightly wrapped into nucleosomes. The binding sites become accessible when DNA unwraps from nucleosomes and forms the open chromatin state. TFs A and B (activators) bind to enhancer binding sites in the open chromatin state to shift the equilibrium towards the open state and increase the probability of gene transcription. TF C (repressor) binds to the enhancer in the closed chromatin state and shifts the equilibrium away from the open state to decrease the probability of transcription.

3.2. Results

3.2.1. Thermodynamic formalism to model combinatorial gene regulation via chromatin mechanism

Quantitative characterization of gene regulation requires a mathematical expression relating the rate of gene transcription to the concentrations of TFs that regulate its expression. Because the initiation of transcription is usually the rate limiting step in gene expression (191), at thermodynamic equilibrium the rate of gene expression \( I \) is given by the product of the binding probability \( P_B \) of transcriptional machinery to the promoter and the rate of RNA polymerase isomerization \( I_0 \):

\[
I = I_0 P_B \quad (30)
\]

We assume that in both contact and chromatin mechanisms of enhancer action, TFs at the enhancer modulate the transcriptional rate via probability \( P_B \). Binding of TFs and the transcriptional machinery at DNA binding sites in the
regulatory region generates multiple protein-bound DNA configurations or microstates. At thermodynamic equilibrium, the probability $p(\alpha)$ of each microstate $\alpha$ is given by a Boltzmann distribution:

$$p(\alpha) = \frac{e^{-G_\alpha}}{Z}$$  \hspace{1cm} (31)

Here $Z = \sum_\alpha e^{-G_\alpha}$ is a partition function that represents the sum of the Boltzmann weights of all possible configurations and $G_\alpha$ is the dimensionless free energy of each configuration in the units of $kT$. In order to compute $p_B$, we then simply sum-up probabilities of all configurations where the transcriptional machinery is bound to the promoter (we denote this set by $\alpha_T$):

$$p_B = \sum_{\alpha \in \alpha_T} p(\alpha)$$  \hspace{1cm} (32)

Using equations (31) and (32) we obtain the following expression for the cumulative probability for transcriptional machinery bound to the promoter:

$$p_B = \frac{Z_{ON}}{Z_{ON} + Z_{OFF}}$$  \hspace{1cm} (33)

where we split the partition function $Z$ into two parts $Z_{ON}$ and $Z_{OFF}$ corresponding to transcriptional-machinery bound and not bound states respectively

$$Z_{ON} = \sum_{\alpha \in \alpha_T} e^{-G_\alpha}; \quad Z_{OFF} = \sum_{\alpha \in \alpha_T} e^{-G_\alpha}$$  \hspace{1cm} (34)

The free energies depend on the binding affinities, cooperative interaction energies and concentrations of all bound proteins in that configuration. Therefore TFs can activate gene transcription by increasing $Z_{ON}$ or repress the transcription rate by increasing $Z_{OFF}$. Concentrations of TFs and RNA
polymerase/transcriptional machinery enter Eqs. (33) and (34) via entropic contributions to free energies of each bound configuration

\[ G_\alpha = G_\alpha^0 - \sum_i \log([C_i]) \]  

(35)

Here \([C_i]\) stands for the concentration of the \(i\)-th protein and summation is over all the bound protein monomers in the configuration and \(G_\alpha^0\) is the standard free energy of the configuration at unit TF concentration.

Thus far the formalism is very general and can be applied to both mechanisms of enhancer action depicted in Figure 1. The chromatin mechanism (Figure 3.1B) includes microstates corresponding to both open and closed chromatin configurations. Transcription activators that attach to DNA binding sites in the open chromatin state shift the equilibrium towards the open state and increase the probability of gene transcription \((p_B)\). Similarly repressors bind to and stabilize the DNA in the closed chromatin state thereby decreasing \(p_B\). This can occur for instance if the binding site consists of the sequence motifs that are only brought in close physical contact by DNA packaging (see Figure 3.1B) (192, 193). As a result, under the chromatin mechanism TFs modulate the rate of transcription even without direct physical interaction with the transcriptional machinery.

We first consider open and closed chromatin states in the absence of TF or transcriptional machinery binding and define their respective free energies as \(G_0\) and \(G_i\). Thereafter we set \(G_0 = 0\) and measure all the other free energies from this reference state. We define an equilibrium constant of “spontaneous” DNA opening (equilibrium constant for transitions between open and closed chromatin in the absence of TF bound) as

\[ e^{-G_i} = K \]  

(36)
In most cases, the probability of spontaneous opening is low resulting in a large equilibrium constant, \( K \gg I \) (176). Binding of the transcriptional machinery only occurs in an open state and therefore only these states contribute to \( Z_{ON} \).

On the other hand, \( Z_{OFF} \) includes contributions from open chromatin enhancer configurations \( (\alpha \in \alpha_{\text{Open}}) \) and closed chromatin enhancer configurations \( (\alpha \in \alpha_{\text{Closed}}) \) and is represented as follows:

\[
Z_{OFF} = Z_{\text{Closed}} + Z_{\text{Open}}
\]

where

\[
Z_{\text{Open}} = 1 + \sum_{\alpha \in \alpha_{\text{Open}}} e^{-G_{\alpha}} \quad \text{and} \quad Z_{\text{Closed}} = K + \sum_{\alpha \in \alpha_{\text{Closed}}} e^{-G_{\alpha}}
\]

We assume that enhancer-bound TFs do not interact directly with the promoter-bound transcriptional machinery. Therefore, for each open chromatin microstate of the enhancer, the binding free energy for transcriptional machinery is the same value denoted \( G_T \). As a result the partition function \( Z_{ON} \) is factorized as:

\[
Z_{ON} = e^{-G_T} Z_{\text{Open}}
\]

These general equations can be used to model any gene regulatory element that functions via the chromatin mechanism. We will use these equations to model the distant regulatory elements shown in Figure 3.2 that implement input functions corresponding to various logic gates.

### 3.2.2. Implementation of cis-regulatory logic gate functions with chromatin mechanism

Buchler et al. showed that AND, OR, NAND and XOR logic type \( \text{cis} \)-regulatory functions can be implemented with the contact model (186). In this section we show that enhancer bound TFs can produce similar logic gate input functions without TF-transcriptional machinery interactions based on the
thermodynamic formalism given by equations (33), (34) and (37)--(39). The parameter values for each gate were numerically determined to minimize the mean-square difference from the corresponding gate function of the contact mechanism (Ref. (186) and Figure A2.1).

**Figure 3.2. Designs of distant enhancers that exhibit a logic gate response.**

A. AND gate response. The binding sites for the TFs in the distant enhancer are weak and cooperative interaction between TFs is strong.

B. OR gate response. The TF binding sites in the enhancer are strong and there is no cooperativity.

C. NAND gate response. The TFs bind weakly to the enhancer sites in the closed chromatin configuration.

D. XOR gate response. Only one TF can bind to the enhancer in the open chromatin state because the binding sites overlap. Binding of the TFs to the closed chromatin state is weak but highly cooperative. Dashed lines indicate direct physical interaction. Weaker binding sites are hatched.
Our implementation of the AND gate type input function is schematically shown in Figure 3.2A. Both TFs are activators and we assume that they only bind to DNA in the open chromatin state and thereby increase $p_B$. The probability of transcription for the enhancer element is calculated by using the following expressions for $Z_{ON}$ and $Z_{OFF}$ in equation (33).

$$Z_{ON} = e^{-G_{f}} (1 + [A]e^{-G_{A}} + [B]e^{-G_{B}} + [A][B]e^{-G_{A} - G_{B} - G_{AB}})$$
$$Z_{OFF} = K + I + [A]e^{-G_{A}} + [B]e^{-G_{B}} + [A][B]e^{-G_{A} - G_{B} - G_{AB}}$$

Clearly the probability of transcription for an AND gate is maximum at saturating TF concentrations, i.e. $[A],[B] \rightarrow \infty$ (cf. equations (33) and (40)). We calculate the transcription rate normalized to this maximum level of expression and the results are shown in Figure 3.3A. Note that because promoter-bound transcriptional machinery and enhancer-bound TFs do not interact, the response function value at saturating levels of $A$ is always the same as the value at saturating levels of $B$. This is not generally true for the contact model as the response function value at saturating concentrations depends on the free energy of the TF-transcriptional machinery interaction. However, for an appropriate choice of free energies this saturation effect will not be observed and the resulting input functions are very similar.

The implementation of the OR logic input function is shown in Figure 3.2B. The design of the OR logic gate is similar to the AND gate in that both TFs increase the rate of gene expression by binding to the enhancer. The expressions for $Z_{ON}$ and $Z_{OFF}$ for the OR logic enhancer are the same as the ones specified in equation (11) for the AND gate.

A crucial difference between the AND logic and OR logic gate designs is the strength of TF-enhancer binding. In the AND gate TFs bind to the enhancer weakly and the TF-TF interactions stabilize the TF-DNA complex. On the other hand in the OR gate each TF binds very strongly to it binding site in the enhancer. We substitute the expressions from (11) into (33) to calculate the
transcription rate for the OR gate normalized relative to the maximum rate at $[A],[B] \rightarrow \infty$. Figure 3.3B(b) shows the transcriptional input function of the OR-gate. The OR gate type logic of the input function shown here and the contact model input function are very similar (compare Figures 3.3B and Figure A2.1B).

**Figure 3.3.** Gene expression response functions for the logic gates of the chromatin mechanism.

A. For the AND gate response, the normalized rate of transcription is calculated relative to the transcription rate at high transcription factor concentrations $[A]=[B]=10^3 \ (G_A = G_B = 5.26, \ G_{AB} = -0.70, \ K = 39.4, \ e^{-G_T} = 1.16)$.

B. For the OR gate type response, the normalized rate of transcription is calculated using equation **Error! Reference source not found.**, relative to the transcription rate at $[A]=[B]=10^3 \ (G_A = G_B = -3.38, \ G_{AB} = 1.4, \ K = 24.14, \ e^{-G_T} = 2.15)$.

C. Normalized rate of gene transcription for the NAND response function relative to the transcription rate at $[A]=[B]=0 \ (G_A = G_B = -1.61, \ G_{AB} = 0, \ K = 19, \ e^{-G_T} = 2000)$. 


D. Normalized rate of transcription from equation (42) relative to the rate of transcription at $[A] = 10^7, [B] = 0$ shows XOR logic ($G_A = G_B = 4.52, G_A^2 = G_B^2 = -0.10, G_{AB}^2 = -6.22, K = 7.51, e^{-G_T} = 1.15$).

TFs $A$ and $B$ act as repressors of transcription in the NAND gate input function. Therefore, we assume that both TFs bind to the enhancer in the closed chromatin state (Figure 3.2B) and decrease the probability of transcription. This NAND gate is implemented with binding sites in the enhancer (silencer), unlike the implementation for the contact model NAND gate where the TF binding sites must overlap with the binding site of the transcriptional machinery (186). $p_B$ depends on the following $Z_{ON}$ and $Z_{OFF}$,

$$Z_{ON} = e^{-G_I};$$
$$Z_{OFF} = K + I + [A]e^{-G_A} + [B]e^{-G_B} + [A][B]e^{-G_{AB}}$$  \hspace{1cm} (41)

The resulting analytical expression of the chromatin mechanism NAND is identical to that of the contact mechanism because of the lack of TF-transcriptional machinery interaction energies in either mechanism. Therefore, we can analytically find parameter values for the chromatin mechanism such that its normalized transcription rate shown in Figure 3.3C is identical to the normalized transcription rate of the contact mechanism (see equations (12) and (13) in Supplementary Information and Figure A2.1C).

The XOR gate input function is obtained with the chromatin mechanism as shown in Figure 3.2D). In this design, TFs $A$ and $B$ both have two binding sites in the enhancer. One pair of binding sites is only accessible to the TFs in the open chromatin state (binding affinities $G_A^l, G_B^l$) and the two sites overlap such that only one TF can be bound at a time. The other pair of binding sites is weak and only accessible to TFs for binding in the closed chromatin state (binding affinities $G_A^2, G_B^2$). The two TFs bind to the closed chromatin sites cooperatively...
(high free energy of interaction $G_{AB}^2$). $Z_{ON}$ and $Z_{OFF}$ in this case are given by the following equations:

$$Z_{ON} = e^{-G_i} \left( I + [A]e^{-G_A} + [B]e^{-G_B} \right)$$

$$Z_{OFF} = K + [A]e^{-G_i} + [B]e^{-G_B} + [A][B]e^{-G_A - G_B - G_{AB}} + I + [A]e^{-G_A} + [B]e^{-G_B}$$

We use the above equations with numerically estimated parameters (see Methods) to calculate the normalized transcription rate relative to the maximum transcription rate for the case $[A] \to \infty, [B] = 0$ (or equivalently $[A] = 0, [B] \to \infty$). Despite our attempt to match the response functions of the contact mechanism XOR gates, the shape of the resulting response functions is slightly different (compare Figures 3.3D and Figure A2.1D). However, we argue that the chromatin mechanism’s design mimics an XOR gate better than the contact mechanism’s design. In fact the response function of the XOR gate of the chromatin mechanism is similar to the response of a more complicated contact mechanism XOR gate that involves two promoters (cf. Ref (186)).

3.2.3. Sensitivities of logic input functions to free energy values

The logic gate response functions discussed above depend on the values of free energies $G_i$ of TF binding and interactions. Even though the chromatin mechanism is capable of matching the response function gates, it still may possess different sensitivities to parameter variation. To quantify these differences we calculate the logarithmic sensitivity as follows (155):

$$S_{G_i} = \frac{\partial \log(p_b)}{\partial \log G_i}$$

Here $p_b$ is the probability of transcription as given by equations (33) and (40), and the index $i = A, B, AB$ indicates a specific free energy. These free energies are easily affected by mutations in the DNA-TF binding sites. Therefore,
the sensitivities are important indicators of the evolutionary robustness and adaptability of the transcriptional response.

The AND gate response is most sensitive to the free energies $G_A$, $G_B$ and $G_{AB}$ at high concentrations of TFs $A$ and $B$, respectively. The sensitivity of the AND gate response to these free energies is similar for the chromatin mechanism and the direct contact mechanism (see Figure A2.2A-D).

The sensitivities of the OR gate response to free energies $G_A$, $G_B$ and $G_{AB}$ were calculated using equation Error! Reference source not found.. The chromatin mechanism OR gate response is sensitive to $G_A$ and $G_B$ in a larger range of TF concentrations than the direct contact mechanism (see Figure 3.4AB). However, the direct contact mechanism shows more sensitivity to $G_{AB}$ near saturating concentrations of TFs $A$ and $B$ (see Figure 3.4CD).

The sensitivities of the NAND gate response to variations in the TF binding and interaction energies are identical for the two mechanisms. This is expected because the models for the two systems are exactly the same as shown above (see equations (12) and (13) in the Supplementary Information and Figure A2.2E-H).

The sensitivities of the XOR gates to various free energies differ significantly between the two mechanisms (see Figure A2.2I-L). However, these differences in sensitivity are mainly due to the dissimilarity of the response functions themselves.

In summary, we found that the sensitivities to free energies for AND and NAND logic gate responses of the chromatin mechanism do not differ significantly from the corresponding sensitivities of the contact mechanism. However, the chromatin mechanism OR gate response is more sensitive to free energies of TF-enhancer binding. This suggests that the OR gate response of the chromatin mechanism is more sensitive to mutations in the TF binding sites.
Figure 3.4. Sensitivity of OR gate response to variations of free energies values.

A,B. Sensitivity of the transcription probability to the free energy of TF binding, $G_A$, for the contact and chromatin mechanism, respectively. The chromatin mechanism has a larger region of high sensitivity than the response of the contact mechanism.

C,D. Sensitivity of the OR gate response to the interaction energy between two transcription factors $G_{AB}$ for the contact mechanism and chromatin mechanism, respectively. For both mechanisms, the response is sensitive to $G_{AB}$ only at high TF concentrations. In this region, the response for the contact mechanism is more sensitive.

3.2.4. Parameter estimation for chromatin model from experiments

Statistical thermodynamic models can be used to predict the transcriptional response combinatorial cis-regulatory enhancers have over a range of TF concentrations and quantitatively characterize different designs of gene regulation as shown above. But these models usually have a large number of independent parameters – the free energies of all the configurations. Direct measurement of these parameters can be very cumbersome and without the
parameter values it is difficult to relate results from these models to experimental information about gene expression. This problem greatly limits the utility of thermodynamic models. In this section we outline an approach that reduces the dimensions of the unknown parameter space for the chromatin mechanism using experimental measurements of gene expression from enhancer-reporter constructs. As a result, a handful of reporter measurements allow us to quantitatively reconstruct the full transcriptional response function.

To illustrate our approach for parameter estimation we develop a thermodynamic model of the regulation of Gata2, a gene that regulates the specification and differentiation of hematopoietic stem cells (HSCs) (118, 147, 152, 153, 194). Enforced over-expression and knockout experiments have shown that that the control of Gata2 expression has major implications for hematopoietic stem cell function (118, 153, 168, 195). Gata2 gene expression is an ideal example for the illustration of our parameter estimation approach because its regulation is dependent on the presence of multiple TFs as well as the chromatin organization of distant upstream regulatory regions (118, 168, 195). Moreover, experimental gene expression measurements for the Gata2 enhancer-reporter constructs have recently become available (118).

As shown in Figure 3.5, Gata2 binds to an enhancer 3 kb upstream (Gata2-3) along with another TF Fli1 to up-regulate its own transcription (118). Both TFs enhance Gata2 gene expression (118) therefore we assume that they bind to the Gata2-3 enhancer only in the open chromatin state. The effect of the Gata2-3 enhancer on gene expression was recently measured experimentally and reported in (118). The authors cloned the Gata2-3 enhancer upstream of a SV40 promoter controlling a LacZ reporter gene (118, 181). Thereafter, this construct was integrated into the genome of hematopoietic progenitor cells that show high concentrations of Gata2 and Fli1. The cells were then disrupted and analyzed for β-galactosidase activity. Assuming that the reporter protein is stable, the level of β-galactosidase activity in cells with the enhancer-reporter construct is directly proportional to the rate of reporter transcription in these cells.
The measured rate of transcription of the reporter $I^G$ is proportional to the probability $p_B$ that the promoter is bound by the transcriptional machinery (see equation (33)). Since RNA polymerase binds typical core promoters very weakly (156, 196-198) we find from equations (37)-(39) that

$$Z_{ON} \ll Z_{OFF}$$

(44)

Accordingly we keep only $Z_{OFF}$ in the denominator of equation (33) for $p_B$ to obtain:

$$p_B = \frac{e^{-G_f} (1 + e^{-G_G} + e^{-G_f} + e^{-G_{FG}})}{K_e + 1 + e^{-G_G} + e^{-G_f} + e^{-G_{FG}}}$$

(45)

Here $G_G, G_F$ and $G_{FG}$ represent the free energies of the Gata2 bound, Fli1 dimer bound and by Gata2-Fli1 dimer bound enhancer configurations respectively. $G_{FG}$ includes the binding affinities $G_G, G_F$ as well as the free energy of the Gata2-Fli1 protein-protein interaction. These free energies follow the definition in equation (35) and include entropic contributions from wild-type concentrations of Gata2 and Fli1 and the concentrations of Gata2 and Fli1 are normalized with these wild-type concentrations. $K_G$ represents the equilibrium constant for transitions between open and closed chromatin. The general idea behind the approach is that if the binding site of a TF is mutated or deleted, the binding of that TF to the mutated enhancer becomes energetically unfavorable and the corresponding terms are excluded from both the numerator and denominator in the expression for $p_B$. This allows us to compute one of the remaining free energies from the ratio of transcription rates of reporters with wild-type and mutated enhancers.
Figure 3.5. Application of the parameter estimation method to the Gata2-3 enhancer.

A. Schematic representation of Gata2-3 and mutant enhancer-reporter constructs. The wild-type enhancer contains both Gata2 and Fli1 binding sites, Enhancer 1 (E1) contains only a Gata2 binding site and Enhancer 2 (E2) contains only Fli1 binding sites. The numbers show the fold expression enhancement relative to the expression from Enhancer 3 (E3) which does not have any TF binding sites (data taken from (118, 181)). These measurements are used in equations (49)-(51) to calculate the parameters of the Gata2 response function.

B. Gata2 enhancer response function shows AND type logic for a chromatin equilibrium constant of K=300. We have normalized the Gata2 and Fli1 concentrations with the respective wild-type concentrations and the white lines demarcate this physiologically relevant range of TF concentrations. Note that the transcription rates are normalized relative to the minimum transcription rate at [Gata2]=[Fli1]=0. The fold change under over-expression of Gata2 and Fli1 is ~K.
C. Sensitivity of the Gata2 response to the value of the chromatin equilibrium constant K. The Gata2 response function is not sensitive to the value of K within the range of wild-type concentrations of TFs (demarcated by white lines). However the response is sensitive to the value of the chromatin equilibrium constant when TFs are overexpressed.

Figure 3.5A shows the fold expression enhancement for the reporter construct in the presence of the wild-type (wt) Gata2-3 enhancer and three reduced versions of this enhancer: Enhancer 1 (E1)-Fli1 binding sites deleted, Enhancer 2 (E2)-Gata2 binding site deleted, Enhancer 3 (E3)- all binding sites deleted. All the experimental data have been abstracted from refs (118, 181). The fold change in gene expression for different enhancer-reporter constructs was calculated by normalizing the level of $\beta$-galactosidase activity of cells with enhancer-reporter constructs with the $\beta$-galactosidase activity levels of cells with enhancerless-reporter constructs. Therefore fold enhancements of gene expression reflect the ratio of $p_b$ in the presence and absence of enhancers. Note in (45) that the factors $e^{-G_G}$ will cancel as ratios of transcription rates are computed.

We use the equations above to relate the free energies of different configurations to the fold enhancement of gene expression. The enhancer $E1$ can only bind Gata2. Accordingly the ratio of transcription rates $I_{E1}^G / I_{E3}^G$ depends only on the free energy $G_G$ of TF Gata2 and the equilibrium constant $K_G$:

$$\frac{I_{E1}^G}{I_{E3}^G} = \frac{p_b^{E1}}{p_b^{E3}} = \frac{(e^{-G_G} + 1) / (K_G + e^{-G_G} + 1)}{1 / (K_G + 1)}$$

Similarly, only Fli1 can bind to the enhancer $E2$ and the rate of gene transcription from this enhancer relative to the expression rate from $E3$ only depends upon $G_F$ and $K_G$. 
The ratio of transcription rates from the wild-type Gata2-3 enhancer and the enhancer $E3$ is easily constructed using equation (45) and this ratio depends on the free energies $G_G, G_F$ and $G_{FG}$.

$$\frac{I_{E2}^G}{I_{E3}^G} = \frac{p_B^{E2}}{p_B^{E3}} = \frac{(e^{-G_G} + 1)(K_G + e^{-G_F} + 1)}{1/(K_G + 1)}$$ (47)

Equations (46) and (47) are solved analytically for $G_G$ and $G_F$ in terms of the equilibrium constant $K_G$.

$$G_G = -\log \left( \frac{(1 - I_{EI}^G/I_{E3}^G)(K_G + 1)}{(K_G + 1 - I_{EI}^G/I_{E3}^G)} \right)$$ (49)

$$G_F = -\log \left( \frac{(1 - I_{E2}^G/I_{E3}^G)(K_G + 1)}{(K_G + 1 - I_{E2}^G/I_{E3}^G)} \right)$$ (50)

These solutions are used in equation (48) to solve for $G_{FG}$ as a function of only $K_G$.

$$G_{FG} = -\log \left( \frac{(1 - I_{wt}^G/I_{E3}^G)(K_G + 1)}{(K_G + 1 - I_{wt}^G/I_{E3}^G)} \right)$$

Thus, using experimental data from (118, 181) for fold enhancement of gene expression in equations (49)-(51), we reduce the dimensions of the parameter space to one. If we know $K_G$ we can uniquely determine the free energies $G_G, G_F$ and $G_{FG}$. The unknown parameter $K_G$ can only be experimentally determined through overexpression of one of the TFs but this data is currently unavailable. We assume an appropriate value for $K_G$ to calculate the free energies and the response function of the Gata2-3 enhancer. The response function is shown in Figure 5b and indicates that the Gata2-3 enhancer functions
as an asymmetric AND gate. Cooperative binding of Fli1 and Gata2 predicted from our estimations ensures that a high expression level is achieved only in the vicinity of maximal concentrations of both TFs.

As the exact value of $K_G$ is unknown, we explore the sensitivity of the Gata2 response to the value of the chromatin equilibrium constant. The logarithmic sensitivity was calculated with equation (43) and is shown in Figure 3.5C. We found that the AND logic property of the response function is not sensitive to the value of the equilibrium constant $K_G$ (see Figure 3.5C and Figure A2.3). We can choose any value from the range $K_G > \frac{I_{net}}{I_{E3}} = 120$ (here we choose $K_G = 300$). Note that the maximum fold change in gene expression is approximately $K_G$, thus showing that the chromatin equilibrium constant can be measured by overexpression of TFs. Note that the choice of this equilibrium constant can affect the dynamic properties of the transcriptional response. In the construction of dynamical ODE type models, the value of the chromatin equilibrium constant may also be constrained by qualitative phenotypic requirements (cf. Ref. (181)).

3.2.5. Comparison of stochastic kinetics of gene regulation by direct contact and chromatin mechanisms

So far we have focused on the steady state transcriptional response for combinatorial gene regulation via the chromatin mechanism and found that the chromatin mechanism can mimic the transcriptional response of the contact mechanism when the effect of the TFs is symmetric. In such cases it might be difficult to distinguish the two mechanisms based on steady state measurements of gene expression levels. However, the two mechanisms can still be distinguished based upon the differences in their dynamics. Recent advances in single molecule experimental techniques offer a wealth of data about the dynamics of transcriptional regulation in single cells (199-201). In this section, we will use a simple example to show how single molecule experimental data about
the dynamics of gene regulation can be used to infer the mechanism of gene regulation. For simplicity, we use a “toy model” with a single transcription activator to demonstrate two differences in the microscopic kinetics of these two models that can be experimentally observed.

Consider a gene that is regulated by a single TF \( A \) that binds to a distant enhancer. TF \( A \) up-regulates gene expression via direct physical contact with the transcriptional machinery or by shifting the equilibrium of local chromatin structure to an open conformation in which the promoter is accessible to the transcriptional machinery. The probability of gene transcription in thermodynamic equilibrium for both mechanisms can be calculated using the framework discussed in the section 2.1. Note that throughout this section the superscripts \( \text{con} \) and \( \text{chr} \) denote the direct contact mechanism and the chromatin mechanism, respectively. For the contact mechanism, the probability of transcription \( p^\text{con} \) is calculated using equation (33):

\[
p^\text{con} = \frac{e^{-G^\text{con}} \left(1 + \omega[A]/K^\text{con}_A\right)}{1 + [A]/K^\text{con}_A + e^{-G^\text{con}} \left(1 + \omega[A]/K^\text{con}_A\right)}
\]

(52)

Here \( K^\text{con}_A \) is the dissociation equilibrium constant of TF-enhancer binding and \( \omega \) represents the strength of TF-transcriptional machinery interaction. The probability of transcription \( p^\text{chr} \) for the chromatin mechanism is also calculated using equation (33):

\[
p^\text{chr} = \frac{e^{-G^\text{chr}} \left(1 + [A]/K^\text{chr}_A\right)}{K + 1 + [A]/K^\text{chr}_A + e^{-G^\text{chr}} \left(1 + [A]/K^\text{chr}_A\right)}
\]

(53)

Similar to the contact model, \( K^\text{chr}_A \) represents the enhancer binding energy of TF \( A \) and \( G^\text{chr}_R \) represents the free energy of transcription machinery binding. Note that there is no interaction energy between the TF and transcriptional machinery in the chromatin mechanism. It can easily be shown that the
probabilities of transcription $P_B^{\text{con}}$ and $P_B^{\text{chr}}$ are equal for all TF concentrations if the parameters are chosen according to

$$K = \omega - 1$$
$$K_A^{\text{chr}} = K_A^{\text{con}} / \omega$$
$$G_T^{\text{chr}} = G_T^{\text{con}} - \log(\omega)$$

When these three conditions are satisfied, the steady-state rate of gene expression is the same for both mechanisms. However, there are still differences in the kinetics of the binding and dissociation of the transcriptional machinery in these two mechanisms.

Figures 3.6AB show the kinetic schemes for the direct contact and chromatin mechanisms, respectively. The model for the direct contact mechanism involves 4 configurations of the regulatory region – empty (O), TF bound (O_A), transcriptional machinery bound (O_R) and both TF and transcriptional machinery bound (O_AR). The model for the chromatin mechanism involves 5 configurations: closed chromatin (C), open chromatin-empty (O), TF bound (O_A), transcriptional machinery bound (O_R) and TF and transcriptional machinery bound (O_AR). We assume that only the rate constants of TF and transcriptional machinery dissociation from DNA are affected by their respective affinities for the binding sites. Using this assumption and equation (54), the rate constants of TF/transcriptional machinery binding and dissociation reactions are set to the values shown in Figure 3.6AB. Additionally, in the chromatin model $k_o$ and $k_c$, the rate constants of spontaneous closed to open chromatin (C to O) and open to closed chromatin (O to C) transitions, respectively, are related to the chromatin equilibrium constant defined in equation (54) as: $K = k_c / k_o$. 
Figure 3.6. Waiting time distributions for transcriptional machinery bound and unbound states.

A, B. Kinetic schemes of the direct contact and chromatin mechanisms for transcriptional regulation by a single activator. O, OA, OR and OAR denote empty enhancer, activator bound, transcriptional machinery bound and both activator and transcriptional machinery bound configurations of the enhancer-gene locus in both mechanisms. In addition, these four configurations of the chromatin mechanism represent the open chromatin configurations whereas the closed chromatin configuration is denoted by C. The transcription machinery bound configurations OR and OAR together represent the ON
state for both mechanisms. All other configurations are part of the OFF state. The rate constants of the transitions between different states are shown above the respective arrows (see text for details).

C. PDF for waiting times in the ON state for the contact mechanism (dashed line) shows two different timescales whereas the PDF for the chromatin mechanism (solid line) shows only one timescale. The time axis is normalized by the timescale of dissociation of the transcriptional machinery in the chromatin mechanism \((k_d^R / \omega)\)^{-1}.

D. PDF for the waiting time in the OFF state for the chromatin mechanism (solid line) shows three timescales whereas the PDF for the contact mechanism (dashed line) shows only one. The time axis is normalized by the timescale of binding of the transcriptional machinery \((k_R)^{-1}\).

E. Mean waiting times in the ON state \(\langle \tau_{ON}^{con} \rangle, \langle \tau_{ON}^{chr} \rangle\) (solid line) is not a function of TF concentration while \(\langle \tau_{ON}^{con} \rangle\) (dashed line) increases with an increase in TF concentration.

F. Mean waiting times in the OFF state \(\langle \tau_{OFF}^{con} \rangle, \langle \tau_{OFF}^{chr} \rangle\) decrease with an increase in TF concentration while \(\langle \tau_{OFF}^{con} \rangle\) is independent of TF concentration.

We use the methods discussed in (202) to calculate the probability density functions (PDFs) of waiting times in the transcriptional machinery bound (ON) states and unbound states (OFF) for the two mechanisms (see Methods). \(P_{ON}(\tau)\) represents the PDF that the first exit from the ON state lies in the interval \((t_0 + \tau, t_0 + \tau + \Delta \tau)\), given that the system entered the ON state at time \(t_0\).

Similarly \(P_{OFF}(\tau)\) represents the PDF of the first exit times from the OFF state. The fraction of time spent in the ON state is directly related to the rate of transcription. The ON and OFF times are related to transcriptionally active and transcriptionally inactive states and therefore may be obtained from the time-series data of single-cell gene expression. In addition, several groups have already shown that the time spent in transcriptional machinery bound and
unbound states can be tracked *in vivo* by adding fluorescent protein-binding hairpin loops to the mRNA tail-end (203-205) or by localization enhancement that can detect protein binding and dissociation from a specific location (200, 201). The waiting time distributions $P_{\text{ON}}(\tau)$ and $P_{\text{OFF}}(\tau)$ can be determined from these types of experiments and qualitative features of these distributions can be used to determine whether the gene regulation mechanism involves direct interactions between TFs and transcriptional machinery.

Figures 6(c) shows the PDF for time spent in the *ON* state for the half-saturated TF concentration $[A] = K_A^{\text{con}}$. The waiting time PDF $P_{\text{ON}}^{\text{chr}}(\tau)$ for the chromatin model depends on only one rate constant of dissociation of transcriptional machinery $k_d^R / \omega$:

$$P_{\text{ON}}^{\text{chr}}(\tau) = \frac{k_d^R}{\omega} e^{-k_d^R \tau / \omega}$$

(55)

This happens because the rate of transcriptional machinery dissociation is the same for the $O_R$ and $O_{AR}$ states without direct interactions with the activator. In contrast, for the contact mechanism the rate of exit from the *ON* state depends on whether the system is in sub-state $O_R$ or $O_{AR}$ because the rate of exit from the two states is different. Accordingly, the PDF $P_{\text{ON}}^{\text{con}}(\tau)$ is a sum of two exponential terms:

$$P_{\text{ON}}^{\text{con}}(\tau) = w_1 r_1 e^{-\tau} + (1 - w_1) r_2 e^{-\tau \omega}$$

(56)

$r_1$ and $r_2$ represent two different time-scales for the waiting time in the *ON* state and $w_j \in [0,1]$ is a weighing factor that represents the probability of observing the $r_j$ timescale. The weighting factors and $P_{\text{ON}}^{\text{con}}(\tau)$ can be modulated with the TF concentration. In the absence of TF $A$ ($[A] \to 0$) the probability of being in the sub-state $O_{AR}$ is zero. As a result, $w_j \to 1$ and $r_j \to k_d^R$ resulting in an exponential decay of $P_{\text{ON}}^{\text{con}}(\tau)$ with the characteristic rate $k_d^R$. At saturating levels of
$A ([A] \to \infty)$, the sub-state $O_{AR}$ dominates, $w_i \to 0$ and $r_2 \to k_d^R / \omega$, and this results in an exponential decay of $P_{ON}^{\text{con}} (\tau)$ with the characteristic rate $k_d^R / \omega$. The contact model assumes a strong interaction of transcriptional machinery and the activator resulting in separation of these time-scales ($\omega > 1$). For intermediate concentrations of TF $A$ two timescales are visible (see Figure 3.6C). The range of concentrations for which this time-scale separation can be observed depends on the strength of the TF-transcriptional machinery interaction $\omega$ and $k_d^R / k_d^A$.

The mean waiting times in the ON state are:

$$
\langle t_{ON}^{\text{con}} \rangle = \frac{w_i}{r_i} + \frac{1 - w_i}{r_2}
$$

$$
\langle t_{ON}^{\text{chr}} \rangle = \frac{\omega}{k_d^R}
$$

Notably $\langle t_{ON}^{\text{chr}} \rangle$ is independent of the TF concentration while $\langle t_{ON}^{\text{con}} \rangle$ increases with TF concentration till $\langle t_{ON}^{\text{con}} \rangle = \langle t_{ON}^{\text{chr}} \rangle$ at saturating concentrations of $A$ (see Figure 6(e)). Moreover, because there are no TF-transcriptional machinery interactions in the chromatin model, $\langle t_{ON}^{\text{chr}} \rangle$ has only one timescale and is independent of TF concentration. This result will hold even when multiple TFs bind to the enhancer to regulate gene expression (not shown).

The situation is different for the distribution of transcriptionally inactive states (Figure 6d). In this case the waiting time distribution in the OFF state for the contact model is exponential for the contact model

$$
P_{OFF}^{\text{con}} (\tau) = k_R e^{-k_d^R \tau}
$$

This is a consequence of the assumption that the binding rate of the polymerase does not change with the presence of an activator (only dissociation rate does). Therefore, the decay of PDF is determined by the rate constant of the binding transcriptional machinery $k_R$. In contrast, up to three distinct timescales
can be present in the waiting time distribution in the \textit{OFF} state for the chromatin mechanism (Figure 3.6D, solid line). The three different timescales of the chromatin mechanism are reflected in the PDF $P_{\text{OFF}}^{\text{chr}}(\tau)$, which consists of three exponentials:

$$P_{\text{OFF}}^{\text{chr}}(\tau) = c_1 r_1 e^{-r_1 \tau} + c_2 r_2 e^{-r_2 \tau} + (1 - c_1 - c_2) r_3 e^{-r_3 \tau}$$ (59)

Here $r_1$, $r_2$ and $r_3$ represent three different time-scales for the waiting time in the \textit{OFF} state and $c_1, c_2 \in (0, 1)$ are weighing factors that represent the probabilities of the $r_1$ and $r_2$ timescales, respectively. $P_{\text{OFF}}^{\text{chr}}(\tau)$ is modulated by changing the TF concentration. In the absence of TF $\mathcal{A}$ there are only two timescales. The fast timescale corresponds to the direct exit from the open state (O) and the slow timescale involves switching between the open and closed (C) chromatin states before exiting from the open state. As the TF concentration is increased, three timescales become visible due to the presence of an O$_A$ state. At saturating concentrations of $\mathcal{A}$ the equilibrium of the open and closed chromatin states shifts almost completely towards the open state. As a result the chromatin mechanism resembles the contact mechanism and there is only a single timescale ($k_r$) in the waiting time distribution.

$$P_{\text{OFF}}^{\text{con}}(\tau) = k_r e^{-k_r \tau}$$ (60)

The mean waiting times in the \textit{OFF} state are:

$$\left< \tau_{\text{OFF}}^{\text{chr}} \right> = \frac{c_1}{r_1} + \frac{c_2}{r_2} + \frac{(1 - c_1 - c_2)}{r_3}$$

$$\left< \tau_{\text{OFF}}^{\text{con}} \right> = \frac{1}{k_r}$$ (61)

We find that $\left< \tau_{\text{OFF}}^{\text{con}} \right>$ is independent of TF concentrations while $\left< \tau_{\text{OFF}}^{\text{chr}} \right>$ decreases with TF concentration till $\left< \tau_{\text{OFF}}^{\text{con}} \right> = \left< \tau_{\text{OFF}}^{\text{chr}} \right>$ at saturating levels of TF $\mathcal{A}$ (see Figure 6(f)). Note that while changing the TF concentration affects only
\( \langle \tau_{\text{chr}} \rangle \) in the chromatin mechanism and only \( \langle \tau_{\text{con}} \rangle \) in the contact mechanism, the fractional time spent in the ON state is the same for both mechanisms \( \langle \tau_{\text{chr}} \rangle / (\langle \tau_{\text{chr}} \rangle + \langle \tau_{\text{chr}} \rangle) = \langle \tau_{\text{con}} \rangle / (\langle \tau_{\text{con}} \rangle + \langle \tau_{\text{con}} \rangle) \). This fractional time in the ON state is proportional to the probability of transcription \( p_h \). Because we assumed that the probability of transcription is the same for both mechanisms, this result shows that our analysis is self-consistent.

Our results show that qualitative differences in the ON and OFF state waiting time distributions can be used to identify the biophysical mechanism of gene regulation. While a relatively simple model was chosen to illustrate the effect, many of the results can be generalized for combinatorial regulation by multiple TFs. A more detailed investigation of the weight-time distributions will be reported elsewhere.

### 3.3. Discussion

Our results show that the chromatin mechanism and the direct contact mechanism are capable of creating functionally similar logic transcriptional gates. Notably, the NAND gate is a universal gate that can be used to create any logic gate. Moreover, transcriptional logic gates are continuous functions that can be adapted to more complicated combinatorial operations by tuning the TF binding affinities through binding site mutations as shown by the sensitivity analysis. This adaptability suggests that virtually any response function can be constructed with a combination of different logic gates and appropriate manipulation of TF binding affinities.

Although chromatin and direct contact mechanisms can show functionally equivalent transcriptional responses, the designs of regulatory elements for any transcriptional input function are very different between the two mechanisms. These differences in design may have important implications.
The chromatin mechanism is more flexible in the design of enhancers. Specific interactions between TFs and the transcriptional machinery are unnecessary to produce the same response as the contact mechanism. The only requirement is that the chromatin structure at the enhancer and the gene unpack together. This allows a lot of flexibility in enhancer location because chromatin domains as large as several kilobases in length can open as a whole (206). It is important to note that according to the chromatin mechanism, enhancers can act in a non-specific manner to activate the transcription of genes in the neighborhood of the target genes. In fact, this type of non-specific transcriptional activation is consistent with a number of reports regarding the effects of distant enhancers and locus control regions in eukaryotes (207-209). Each TF binds to the enhancer and disturbs the equilibrium between open and closed chromatin states and changes the probability of binding of other TFs in a manner similar to the Monod-Wyman-Changeux model for allosteric enzymes (210). As a result, physical TF-TF interactions are unnecessary for cooperativity between TFs. For example, the NAND gate response functions of the two mechanisms are identical (cf. equations (12) and (13) in Appendix 2) but the response equation for the contact mechanism has an explicit TF-TF interaction term whereas the equation for the chromatin mechanism does not. The effective cooperativity that emerges from the equilibrium between open and closed chromatin in this case is equivalent to an effective free energy of interaction (see Appendix 2). The emergence of cooperativity without direct physical interaction between TFs means that any two DNA binding proteins can be used as TFs under the chromatin mechanism.

In contrast, the direct contact mechanism restricts the location of binding sites for transcriptional regulation. First, transcriptional repression requires binding sites in the promoter vicinity. For example, in the NAND gate of the contact model (186) binding sites for repressors $A$ and $B$ must be in the promoter region so that they can occlude the RNA polymerase binding site. Second, for all contact gates the free energies of DNA looping and TF-transcriptional machinery interaction affect the possible enhancer location. Third,
each response function requires specific domains on TFs that are responsible for
the appropriate TF-transcriptional machinery interactions.

Both the contact mechanism and chromatin mechanism can be utilized for
combinatorial gene regulation in higher organisms and it might be necessary to
investigate the particulars of the mechanism for each gene. Our models suggest
several experimental designs to distinguish between the alternatives. Although
the two mechanisms are functionally equivalent within the operating range of TF
concentrations, we could distinguish the two through forced over-expression of
any one of the TFs. Saturating concentrations of any TF will show the same level
of gene expression for the chromatin mechanism. On the other hand, expression
rates at saturating concentrations of different TFs might be different for the direct
contact mechanism. Another method involves shifting the position of the
enhancer relative to the promoter. Regulation by contact mechanism is sensitive
to such translocations because the free energy of the enhancer bound TFs and
promoter bound transcriptional machinery depends on the distance between
them. Regulation by the chromatin mechanism will likely be unaffected by
translocation of the enhancer because local accessibility of DNA at the enhancer
can be propagated over long distances (several kB) to establish an open
chromatin state (206).

Interestingly, the sensitivities of the AND and NAND logic gate designs to
free energies of TF-enhancer binding are not very different for the two
mechanisms. However, the chromatin mechanism OR gate is more sensitive to
TF-enhancer binding free energies. This increased sensitivity of the OR gate
response suggests that this design is more sensitive to binding site mutations
than the equivalent design of the contact mechanism. The thermodynamic
approach that we have developed allows us to characterize gene expression
input functions based on a handful of transcriptional reporter measurements.
From this perspective, gene regulation via the chromatin mechanism is easier to
quantify because it does not involve binding energies between TFs and the
transcriptional machinery and therefore involves fewer parameters. In this case,
the method that we have proposed can use experimental results directly in parameter estimation.

While the chromatin mechanism and the direct contact mechanism can produce functionally equivalent time-averaged transcriptional responses, there are intrinsic differences between the two mechanisms that nevertheless lead to differences in the stochastic kinetics of gene expression. We have shown that the chromatin mechanism can be distinguished from the contact mechanism based on single-molecule gene expression data. The chromatin mechanism can easily be identified from such data from the single characteristic timescale in the PDF of time spent in the transcriptionally active state. On the hand, multiple time-scales are present in the PDF of time spent in transcriptionally inactive state. We have also found that the mean waiting time in the ON state is independent of TF concentration for the chromatin mechanism. These distinguishing dynamical properties highlight the irreducible differences between the two mechanisms. While these results were obtained using a somewhat oversimplified model of transcriptional activation, we expect our observations to hold even for more complex kinetic schemes. This will be a subject of a separate investigation.

We also note that transcriptional regulators using the chromatin mechanism have potentially promising applications in synthetic biology and genetic engineering. At present, synthetic biology circuits use simple promoter architectures with a single regulator to control gene transcription. This clearly limits the transcriptional response of the gene and functional properties of the circuits. This limitation exists because combinatorial gene regulation that follows the contact mechanism requires specialized TFs with appropriately interacting domains. At the same time, cis-regulatory modules of living systems, especially eukaryotes, are typically dauntingly complex. The increase in complexity of gene regulation is associated with the evolutionary emergence of complex multicellular organisms (178, 211). Moreover, the increase in proteins that control chromatin structure and nucleosome remodeling correlates well with the increase in complexity of cis-control elements in metazoans (178). This adoption of the
chromatin mechanism of gene regulation in higher organisms reflects the advantages of the flexibility in design of complex combinatorial regulation. Synthetic designs of combinatorial regulation based on the chromatin mechanism can harness this flexibility to avoid the limitations of the contact mechanism regulation. The designs of logic gates with combinatorial regulation via the chromatin mechanism that we have discussed in this paper are only an indication of how this mechanism can help simplify the design of synthetic circuits for any transcriptional response function.

Notes

This chapter is based on work that was published in the following article:

4.1. Introduction

Hematopoietic stem cells (HSCs) are a rare population of cells with self-renewal potential to divide and contribute cells to all blood lineages throughout the life of an organism. The ontogeny of HSCs has been carefully studied in terms of anatomical locations and stages of cellular progression (133, 212). Studies using mouse, zebrafish and embryonic stem cells have demonstrated that blood progenitor cells (with limited self-renewal ability) are formed early during embryogenesis, initially in the yolk sac and then in the embryo (212). This is followed by the emergence of definitive HSCs (with long-term self-renewal potential) initially in the aorta-gonads-mesonephros region of both mice and humans (212). Furthermore, it has been shown that a specialized part of the blood vessel network termed the ‘hemogenic endothelium’ undergoes an endothelial-to-hematopoietic transition (EHT) to form blood stem/progenitor cells.
(126, 213-216). Understanding the molecular mechanisms that drive HSC and blood formation in the developing embryo will be crucial in designing novel regenerative medicine protocols.

Tight spatial and temporal control of gene expression is vital for the proper development of an organism (9). Gene expression programs are coordinately regulated by the combinatorial binding of tissue-specific transcription factors (TFs) and external cues that are communicated to cells via signaling pathways. Several TFs regulating key stages of blood cell development have been identified (217). Scl, Gata2 and Fli1 act early during development to specify the hemogenic endothelium and are necessary for HSC emergence (149, 151, 218-220). On the other hand, Runx1 is required in the hemogenic endothelium for the EHT but not subsequently (221-223). TF activities and signaling pathways are integrated by cis-regulatory modules such as promoters and enhancers which have been characterized for numerous TFs involved in HSC emergence (224). Enhancers for Gata2, Fli1 and Scl are bound by themselves and each other to form a fully connected triad (118), and the HSC enhancer for Runx1 is bound by all three triad proteins (127).

Two signaling pathways, Bmp and Notch, are required for HSC and progenitor development (163, 164, 217, 225). The Notch1 intracellular mediator binds at the Gata2 locus, whereas Bmp-induced signaling mediator, Smad1, binds at the Runx1 promoter and at the Gata2 and Fli1 enhancers (Fig. 1.3.1A; (117, 145, 146)). Smad6, an inhibitory Smad, participates in the Bmp4-signaling pathway by hindering Smad1 activation and targeting it for proteolytic degradation (226). The Smad6 enhancer is bound by the triad proteins, Smad1 and Runx1, and a negative feedback loop from Smad6 regulates Runx1 by promoting its proteosomal degradation (226). Runx1 binding at the Smad6 enhancer is mediated by triad TFs, thus triad activation temporally balances Runx1 activity by up-regulating both Runx1 and its negative regulator Smad6 (226). Altogether these interactions form a gene regulatory network (GRN) that controls hematopoietic stem and progenitor cell emergence in the developing
embryo (Fig. 4.1B). Multiple feed-forward and feedback loops present in the GRN (Fig. 4.1A) lead to complex dynamical properties that allow tight control over the network’s response to external and internal cues. Understanding these complex emergent properties with purely experimental approaches is challenging; mathematical modeling of networks can serve as an important complementary approach. Models can combine qualitative and quantitative information about network architecture and parameters, and thereby serve as an integrative platform for understanding the results of various genetic perturbations and for making novel predictions.

In this study, we build a mathematical model of the GRN shown in Fig. 4.1B based on previously published details of cis-regulatory modules, TF-binding and protein-protein interactions. The model integrates Runx1 regulation as well as Bmp4 and Notch1 signaling with the Scl-Gata2-Fli1 triad module. Using this model we elucidate the role of Runx1 in the network. Dynamical properties of the network predicted by the model are in good agreement with in vitro and in vivo experimental observations. Moreover, in silico perturbations of Runx1, Notch1 and Bmp4 in the simulations closely match the observations in knockout and over-expression phenotypes. Importantly, our model provides mechanistic insight into the early emergence of blood progenitors observed in Runx1 haploid embryos. Taken together these results suggest that the GRN analyzed here can act as a master-level switch in the signal pathway controlling definitive hematopoiesis.
Figure 4.1. GRN responsible for regulating HSC specification.

A. The GRN responsible for regulating HSC cell specification contains TFs Scl, Gata2, and Fli1 that are connected via multiple positive feedback loops (dashed box). This triad is regulated directly via Notch1 and indirectly via Bmp4 through a peripheral circuit containing Smad1, Smad6, and Runx1. Bmp4 affects the triad indirectly by regulating the Smad1 phosphorylation rate. Smad6 negatively regulates pSmad1 and Runx1 (blunted arrows) by targeting them for proteasomal degradation. Arrows represent positive transcriptional regulation.

B. Detailed representation of the regulatory connections in the GRN that explicitly shows the various promoters and binding sites (using the notation from (8)). The top half of the diagram shows the triad module, whereas the bottom half shows the signaling module.

4.2. Results

4.2.1. Notch1 is necessary for irreversible activation of the triad

Definitive hematopoiesis is the production of blood progenitor cells with the potential to form mature erythroid and myeloid cells, and occurs in multiple sites of the developing embryo including the yolk sac, placenta, AGM and head regions (125, 227-229). The Scl-Gata2-Fli1 triad (Fig. 4.1A, dashed box) is at the core of the GRN analyzed here; its activation with Notch1 and Bmp4 signals is known to play an important role in definitive hematopoiesis (118, 164, 230,
Previously, we used a mathematical model to show that Notch1 and Bmp4 cause an irreversible switch to high levels of triad gene expression and thereby explained their role in the activation of these master regulatory genes of definitive hematopoiesis (232). Here we extend this model to incorporate recently uncovered interactions between components of the Bmp4 signaling pathway and Runx1, another key regulator of definitive hematopoiesis (117, 226).

In this extended model we explicitly include the components involved in Bmp4 signaling - Smad1, Smad6 and Runx1. We briefly outline the major interactions and assumptions of the model (see Methods and SI for details). Bmp4 promotes the phosphorylation of Smad1, following which pSmad1 translocates to the nucleus and upregulates the transcription of the triad genes as well as of Runx1 and Smad6 (233-235). Runx1 forms a complex with pSmad1 in the nucleus (236). We assume that the formation of this complex enhances the effect of pSmad1 on triad gene expression although it is not essential for triad upregulation. As a result, in our model, Runx1 participates in triad regulation but is not essential for triad gene expression. Smad6 post-translationally modulates Bmp4 signaling by forming complexes with Runx1 and pSmad1, and thereby targeting them for proteolytic degradation (226, 237). In addition the triad feeds back to the signaling module by transcriptionally upregulating Runx1, Smad6 and Smad1 (see Methods and Fig. A3.1; (118, 127, 226, 234, 238)). It should be noted that our model focuses specifically on the emergence of HSCs from the hemogenic endothelium and as such cannot be used to infer the effects of either Bmp4 and Notch1 signals or triad gene expression levels on the eventual fate (i.e. differentiation and/or proliferation potential) of these cells.
Figure 4.2. Notch1 is necessary for triad activation.

Steady-state concentrations of the triad proteins are plotted as a function of Smad1 phosphorylation rate, (kp), a proxy for Bmp4 signal (panels A and B). Red, blue and green curves represent the steady-state concentrations of Scl, Gata2 and Fli1 respectively. Solid and dashed lines represent the stable and unstable states respectively. Gray areas represent the region where the triad is bistable and can exist in either LOW or HIGH concentrations.

A. In the presence of Notch1, the triad LOW state only exists when Bmp4 signal is below a certain threshold (kp*). Signals above the threshold (kp > kp*) irreversibly switch the triad to HIGH state (shown by arrow). In this and other figure panels HIGH and intermediate Scl concentrations represent HSPC (Hematopoietic Stem and Progenitor Cell) and HE (Hemogenic Endothelium) respectively.

B. Black curve shows the dependence of kp* on [Notch1]. Below this curve the triad is bistable, whereas above it the triad is monostable and can only exist in the HIGH state. Note that at very low [Notch1] the triad is always bistable. The blue dot shows the operating point [Notch1] = 1 used in panel A. The red dot denotes the amount of Notch1 that can activate the network in the absence of a Bmp4 signal.

C. In the absence of Notch1, the triad can exist in either HIGH or LOW states, both of which exist and are stable for all values of the Bmp4 signal.

To understand the role of the Smad1-Smad6-Runx1 signaling module we first examine the steady-state response of the network to Notch1 and Bmp4 signals. To this end we compute how the steady-state concentrations of the triad proteins depend on the signal levels. In our model the Smad1 phosphorylation rate (kp) is used as a proxy for Bmp4 signaling (239), whereas Notch1 signaling
is modeled explicitly as the binding of Notch1 to the Gata2 promoter site (146). We find that the triad module is irreversibly bistable (Fig. 4.2) which is in line with our previous results (232). As a result, for a range of Notch1 and Bmp4 signals the triad can exist in one of two expression regimes: low expression (LOW) or high expression (HIGH) of all three triad genes. In both regimes, steady state expression levels of triad genes increases with increase in Notch1 and Bmp4 signal levels (Fig. 4.2A). In the presence of Notch1, the LOW state only exists when the Bmp4 signal is below a certain threshold (Fig. 4.2A, grey area). In this region the system is bistable, whereas outside (Fig. 4.2A, white area) the system is monostable and can exist only in the HIGH state. Therefore, increasing the Bmp4 signal irreversibly switches the triad from LOW to HIGH (indicated by arrow in Fig. 4.2A). The multiple positive feedback loops in the triad maintain it in the HIGH state even after the signals (Bmp4 and/or Notch1) are removed. This again is in accord with our previous results (232). It should be noted that due to the multiple cooperative interactions between the triad genes, at steady state all three triad gene are predicted to simultaneously be either in the LOW or HIGH expression states.

Although it not known what level of triad gene expression is required to facilitate the transition of a pre-hemogenic endothelial cell to a hemogenic endothelial (HE) cell we assume here that the irreversible switch from triad LOW to triad HIGH is not required for this transition and the increase in LOW state triad expression in response to Bmp4 is sufficient for transition to HE (Fig. 4.2A). Moreover we believe that the irreversible switch that occurs past a threshold level of Bmp4 signal (hereafter referred to as the activation threshold, \( k_p^* \)) represents the transition of HE cells to Hematopoietic Stem/Progenitor cells (HSPCs).

The model predicts that the Bmp4 signal threshold for triad activation (\( k_p^* \)) depends on the level of Notch1 (Fig. 4.2B, black curve). Increasing Notch1 lowers this activation threshold. For the parameter values used in the model, finite amounts of Notch1 result in triad activation even in the absence of Bmp4
signal (Fig 4.2B, red dot). Thus, the model predicts that significant overexpression of Notch1 can lead to triad activation even in the absence of Bmp4 signal. On the other hand, decreasing Notch1 below a certain level causes the activation threshold to approach infinity, thereby making the activation impossible in the absence of Notch1. Fig. 4.2C shows that in the absence of Notch1 the triad is always bistable i.e., both HIGH and LOW states exist, regardless of the Bmp4 signal level. As a result, while Bmp4 can cause an increase in triad gene expression it is impossible to switch the cells that start in the triad LOW state to HIGH state by increasing Bmp4 signal in Notch1 mutants. This prediction is consistent with the observation that blood progenitor emergence does not occur in the Notch1-null embryo proper (163, 240).

4.2.2. The model explains accelerated emergence of HSCs in Runx1 heterozygotes

Next we analyze the role of Runx1 in triad activation. In the Runx1-null mutant (Runx1<sup>-/-</sup>) the triad is bistable (Fig. 4.3A), and as with the Notch1-null mutant (Fig. 4.2C) both LOW and HIGH states exist regardless of the Bmp4 signal level. As a result, while Bmp4 can increase triad gene expression and lead to the appearance of the hemogenic endothelium (HE), it cannot cause the irreversible transition to HIGH state (HSPC) and cells remain stuck in the HE stage. This effect is related to Runx1’s ability to enhance the transcription of triad genes by forming a nuclear complex with their regulator pSmad1 (236).
Figure 4.3. The role of Runx1 in triad activation.

A. In the Runx1 null mutant (Runx1−/−), the triad is bistable, but Bmp4 signaling is incapable of switching the triad from LOW to HIGH (HSPC). However triad expression can still increase leading to the appearance of HE. Red, blue, and green curves represent Scl, Gata2 and Fli1 concentrations respectively.

B. Activation threshold (kp*) changes non-monotonically with increase in the Runx1 production rate (black-green-blue curve). Gray area represents the region where triad is bistable. Red and green dots denote the activation thresholds for WT (Runx1+/+) and Runx1 haploid mutant (Runx1+/−) respectively.

C. Steady-state response of triad to Bmp4 signal in WT (Runx1+/+—red curve) and haploid mutants (Runx1+/−—green curve). Only [Scl] is shown for clarity. The activation threshold for haploid Runx1 mutant is ~1/2 of that for WT. Dashed curves show unstable states.

D. Dynamics of triad activation in response to a step-increase in Bmp4 signal from kp=0 to kp=50hr−1. The dynamics depend on Runx1 gene dosage. The lower activation threshold allows haploid mutant (green curve) to speed-up triad activation by ~2 days compared to WT (red curve). Grey region represents $\Delta T_{90}^{[Scl]}$, the difference in time to reach 90% of [Scl] level in HIGH state. Note that the Runx1 null-mutant only switches to HE in the presence of the same signal (yellow curve).
We further explore the effect of Runx1 on the triad in Fig. 4.3B by examining how activation threshold \((k_p^*)\) changes with the maximum Runx1 production rate \((\nu^r)\). As expected from Fig. 4.3A, Bmp4-mediated triad activation is not possible if the Runx1 production rate is below a minimum level (see Fig. A3.2). Increasing production rate above this minimal level has a non-monotonic effect (i.e. both increasing and decreasing) on the activation threshold. In this case the effect is first decreasing, then increasing and finally again decreasing (black-green-blue curve, Fig. 4.3B). This is a consequence of the contrasting roles that Runx1 plays in modulating triad activation – positive via complex formation with pSmad1 and negative via upregulation of Smad6. Initially, increasing Runx1 production rate decreases the activation threshold (black branch, Fig. 4.3B) due to the cooperative effect of Runx1 on pSmad1’s upregulation of the triad. However, an opposite trend is observed later where increasing Runx1 production rate increases the activation threshold (green branch, Fig. 4.3B) due to Runx1’s negative effect on pSmad1 via up-regulation of Smad6-mediated degradation. Further increasing Runx1 production makes its level high enough to sequester all of Smad6, thereby saturating the negative effect on pSmad1. As a result, increasing Runx1 production rate decreases the Bmp4 signal required for triad activation (blue branch, Fig. 4.3B).

For model simulations, we assume that wild-type Runx1 (WT, Runx1\(^{+/+}\)) production rate lies in the intermediate portion of the curve in Fig. 4.3B (red circle). Hence, a reduction in Runx1 production rate lowers the activation threshold. This model assumption is crucial to explain the phenotype observed in Runx1 heterozygous deletion mutants (Runx1\(^{+/-}\)). In Runx1\(^{+/-}\) mutants there is a 50% decrease in Runx1 production rate (green circle, Fig. 4.3B), due to which the activation of the triad can occur at a lower Bmp4 signal threshold (compare \(k_p^*\) for red and green circles in Fig. 4.3B). Comparing the steady-state triad responses reveal that the activation threshold for Runx1\(^{+/-}\) mutant is \(\approx 1/2\) of that for the WT (Fig. 4.3C). This decrease greatly affects the dynamics of triad activation.
Runx1 is known to be essential for HSC emergence, however surprisingly HSC emergence is accelerated in the Runx1+/− mutants (241, 242). To understand this difference in the dynamics of HSC emergence between WT and Runx1+/− mutant we simulate the kinetics of triad LOW-to-HIGH activation. Specifically, we study the kinetics of triad activation in response to a step-up in the Bmp4 signal (k_p) to a value exceeding the WT activation threshold (five times the value of k_p* for the red circle in Fig. 4.3B). As shown in Fig. 4.3D, the decrease in the activation threshold for the Runx1+/− mutant (observed in Figs. 4.3B and 4.3C) results in much faster activation of the triad. The speed-up in the activation can be quantified by ΔT_{Scl}^{90} - the difference in the times between WT and haploinsufficient mutant at which Scl reaches 90% of its HIGH steady state value. For the chosen parameter values, ΔT_{Scl}^{90} is about 2 days, which is in quantitative agreement with the experimental observation that HSC emergence occurs earlier in Runx1+/− embryos (241, 242). Though the precise value of ΔT_{Scl}^{90} is affected by changes in kinetic parameters the qualitative effect is robust (Figs. A3.3 and Appendix 3) as it is a consequence of a fundamental property of bistable switches – critical slowdown in dynamics near ghost steady states (243-245).

4.2.3. Runx1 overexpression can compensate for Notch1 deletion

Recent in vitro experiments have shown that artificially-induced Runx1 overexpression can rescue definitive hematopoiesis in Notch1 null mutants (246). To mimic this experiment and to understand the relationship between Runx1 and Notch1 in the context of HSC emergence we introduce an additional Runx1 production term (v_{rr}) in our model (see Methods). Simulations show that, similar to its dependence on Ω_v, the activation threshold k_p* changes non-monotonically as a function of the Runx1 overexpression rate v_{rr} (black curve, Fig. 4.4A). To uncover the consequences of Runx1 overexpression, we fix Ω_v at the level shown in Fig. 4.3B and choose a high v_{rr} (=100 hr⁻¹). We find that this level of Runx1 overexpression allows the triad to be activated by Bmp4 signal alone.
([Notch1]=0 results in finite $k_p^*$ in Fig. 4.4B; compare to Fig. 4.2B). Therefore, our model explains the observation in Ref. (246) that artificially-induced Runx1 overexpression rescues definitive hematopoiesis in Notch1-null mutants by reducing the activation threshold to physiologically relevant levels.

**Figure 4.4.** Runx1 overexpression compensates for Notch1 deletion.

**A.** $k_p^*$ changes non-monotonically as a function of Runx1 overexpression rate $v_{rr}$ (black curve).

**B.** At high $v_{rr}$, Notch1 is not essential for Triad activation as indicated by a finite value of $k_p^*$ (black curve) at [Notch1] = 0.

### 4.2.4. Timing of Runx1 conditional deletion or Smad1 overexpression determines triad activation dynamics

Several studies have used Cre-mediated recombination and induced overexpression to highlight the dynamical roles of the Bmp4-signaling circuit components Smad1 and Runx1 in definitive hematopoiesis (222, 247-249). In particular, these studies have shown that (i) Runx1 is only essential up to a certain stage in HSC emergence and can be conditionally deleted after this stage without compromising HSC maintenance (222, 247) and (ii) transiently induced artificial overexpression of Smad1 expands hematopoietic progenitor populations only in a specific time-window of HSC emergence (249). These studies shed light on previously unknown dynamical properties of the regulatory system controlling
hematopoiesis, however, the mechanistic underpinnings of the transient roles of Runx1 and Smad1 remain unexplained. We use our model to show that these transient roles are the result of the dynamical characteristics of the irreversible bistability in the triad.

First, we address the recent observations that only transient expression of Runx1 is required for HSC emergence (222, 247, 248). In Chen et. al. (222), Runx1 was conditionally deleted at different stages of early embryonic development and it was found that if this deletion occurs sufficiently early, HSCs never emerge (222). However, if Runx1 was conditionally deleted later in the developmental process HSC emergence is unaffected. Similar temporal and cell state requirements for Runx1 were observed using different techniques in other studies, including recent work demonstrating that Runx1 activity is required at E7.5, four days prior to HSC emergence at E11 (247, 248). To explain these experiments we simulate the activation of the triad (only Scl shown in Fig. 4.5A) to a step-increase in Bmp4 signal and then decrease the Runx1 production rate to zero at different times (12h, 24h,.. etc). Our model shows that if Runx1 is knocked out after triad proteins reach sufficient levels for self-activation, there is negligible impact on the triad’s steady state or dynamic behavior (Fig. 4.5A and Fig. A3.4). This is a consequence of the characteristic property of irreversibly bistable switches that they only require transient activation to switch to the HIGH state. Once the triad proteins reach sufficient levels (the region of attraction for the HIGH state), the positive feedback loops in the triad forces the system to quickly reach the HIGH state irrespective of the signal level. Our model shows a distinct time threshold of about 3.5 days, thereby defining how long Runx1 expression is necessary for triad activation and hence HSC emergence (Fig. 4.5B and Fig. A3.4). Therefore, our model is able to quantitatively capture the transient requirement of Runx1 during HSC emergence (222).
Figure 4.5. Timing of Runx1 conditional deletion or Smad1 overexpression determines whether Triad activation is affected or not.

A. Scl dynamics with different timings of Runx1 conditional deletion. Each black circle denotes the time in the trace at which Runx1 production is reduced to zero. Brown curves show Scl dynamics for systems that fail to activate due to Runx1 conditional deletion, whereas red curves represent activated systems. Switch to HIGH state of the triad with Bmp4 signaling is impossible if Runx1 is deleted early; however, cells can still increase Scl expression and reach the HE stage. Runx1 deletion after a certain time has negligible effect on triad activation dynamics or final steady state.

B. Steady-state Scl concentration as a function of the time of Runx1 conditional deletion. Early Runx1 deletion interferes with triad activation and cells cannot fully upregulate Scl; however, Runx1 deletion after day 3.5, has no effect on triad activation.

C. Scl dynamics with different timing of a Smad1 overexpression pulse. Smad1 is overexpressed at a rate of 0.1 hr⁻¹ for a 6hr time-window at specific times during triad activation. Red trajectories show Scl dynamics without Smad1 overexpression. Blue bars indicate the times of Smad1 overexpression for other trajectories. Note that overexpression around day 2 (middle panel) speeds-up triad activation the most, whereas both early and late overexpression have less prominent effects (top and bottom
panels). Gray region represents the difference in time to reach 90% of the final [Scl], $\Delta T_{90}^{[Scl]}$, with and without the pulse.

D. Speed-up of triad activation, $\Delta T_{90}^{[Scl]}$ depends non-monotonically on the time of Smad1 overexpression.

The effects of a pulse of Smad1 overexpression on the dynamics of hematopoietic progenitor cell emergence \textit{in vitro} has been recently described (249). The results indicated that this effect depends non-monotonically on the timing of the Smad1 overexpression pulse. We simulate this experiment \textit{in silico} by including an additional Smad1 production term (rate of production - 0.1 hr$^{-1}$) in our model equations. Consistent with the experimental protocol, this additional production is only active for a 6-hr period starting at a specific time-point in the simulation. Under these modifications, we vary the time of Smad1 overexpression pulse and analyze the triad activation dynamics in response to a step-increase in Bmp4 signal. Fig. 4.5C shows the Scl activation dynamics in the presence (black, green and yellow curves) and absence of pulses (red curves). Blue bars depict the Smad1 pulse timing. The difference in time to reach 90% of the final Scl concentration with and without a pulse, $\Delta T_{90}^{[Scl]}$, is denoted by grey regions. Our model shows that early (~day 0 - top panel) and late pulses (~day 4 - bottom panel) have small effects on triad activation dynamics whereas a pulse of Smad1 around day 2 (middle panel) considerably speeds-up Scl activation (Fig. 4.5C). This non-monotonic dependence of triad activation speed-up ($\Delta T_{90}^{[Scl]}$) on the timing of the Smad1 pulse (Fig. 4.5D) could explain why Smad1 pulses only expand the hematopoietic progenitor population within a limited developmental time window (249). This non-monotonic dependence is a consequence of the characteristic slow-down in switching dynamics of a bistable switch. Bistable systems are known to slow-down in a critical concentration range during a switch between states (243, 250). The coordination of a Smad1 pulse with the time at which the triad reaches the critical slow-down concentration range can significantly speed-up the dynamics. However due to the small pulse
duration, pulses that are timed to occur before the triad reaches this region have little effect on the overall time to reach the steady state. Pulses that come after the triad has passed the slow-down have already been delayed in the critical region and again the contribution to activation time is expected to be small.

4.2.5. Novel predictions about the roles of Runx1 and Smad6 in HSC emergence

Based on the non-monotonic relationship between Runx1 expression rate and triad activation threshold (see Fig. 4.3B) we can make several predictions about the roles of Runx1 and Smad6 in this master-regulatory circuit controlling HSC emergence. Our model predicts that unlike the WT, in Smad6-null mutants the dynamics of HSC emergence will be identical in Runx1+/- and Runx1+/+ backgrounds (see Fig. A3.5A-C). In addition, we predict that stabilization of p-Smad1 activity by inhibiting its proteasomal degradation by Bortezomib (251) would neutralize the proteolytic effect of Smad6 on Runx1 thereby reducing the difference between HSC emergence dynamics in Runx1+/+ and Runx1+/- mutants (see Fig. A3.5DE). Finally, we note that, relative to WT, a mutant with three copies of Runx1 is predicted to increase Smad6 expression and consequently the triad activation threshold. As a result, HSC emergence in triploid Runx1 mutants is expected to be slower than both WT and Runx1+/- mutants (Fig. A3.5F).

4.3. Discussion

We have developed a comprehensive model for the network controlling definitive blood emergence by combining our previous model (232) of the Scl-Gata2-Fli1 triad with the Bmp4 signaling module containing Smad1, Smad6 and Runx1. This framework allows us to systematically assess, analyze and predict the effects of perturbations on dynamical properties of this network. The model shows that the GRN module under consideration is irreversibly bistable and can be activated by the combination of Notch1 and Bmp4 signaling. Once activated,
the triad remains active even when the signals are removed. If high levels of Scl-Gata2-Fli1 gene expression ensure a stable HSC state, the irreversible activation of the triad in the AGM could explain why Bmp signaling later becomes dispensable for the maintenance of HSCs in the fetal live and adult bone marrow (252). We have hypothesized that the irreversible activation of this bistable switch is not required for the formation of hemogenic endothelial cells but occurs in HE cells and both contributes to and is modulated by Runx1 expression which plays a critical role in the emergence of blood stem/progenitors from the HE. It is important that this irreversible bistable switch occurs prior to blood cells emerging from the HE so that maximal expression of these factors, which are required for down-stream applications, is achieved prior to cells losing resident cues from their tissues of origin. Based on this hypothesis, we can compare the predicted dynamical properties with those observed in various in vivo and in vitro genetic perturbations of this network.

Our model predicts that Runx1 is essential for definitive hematopoiesis in part because it mediates Bmp4 signaling and is required, albeit only transiently, for the activation of the triad module. Moreover, the model is able to reconcile this essential requirement of Runx1 for definitive hematopoiesis (247) with the counter-intuitive observation that heterozygous deletion of Runx1 actually accelerates HSC emergence (241, 242). We have predicted that this acceleration is connected to the negative effect that Runx1 has on pSmad1 levels via upregulation of Smad6. Due to this negative effect, the two-fold decrease in the Runx1 production rate that results from a heterozygous deletion decreases the Bmp4 signal threshold and thereby accelerates triad activation. Based on our simulations we also predict that when examined in a Smad6-null mutant background or in the presence of specific proteasome inhibitors, there will be little or no difference in dynamics of HSC emergence between Runx1+/− and Runx1+/−.

Additionally, the model predicts that the transient perturbations to the signaling module will affect the activation dynamics only if they occur sufficiently
early. Current observations of the temporal requirement for Runx1 during hematopoietic development suggest it is required early, either in the E7.5 extra-embryonic mesoderm, and/or later in the hemogenic endothelium (248). We predict that the conditional deletion of Runx1 has no effect on triad activation if it occurs after ~3.5 days from the onset of the hematopoietic program. We also predict that pulsed overexpression of Smad1 leads to the expansion of the hematopoietic progenitor population within a particular time-window and the magnitude of the expansion depends non-monotonically on the timing of the Smad1 pulse. These predictions are in excellent qualitative and quantitative agreement with recently reported and unexplained experimental observations that motivated these simulations (222, 248, 249). Moreover, these predictions are robust to the parameter values used in the model as these dynamical effects are related to the fundamental mathematical properties of bistable switches (see Figs. A3.3, A3.4).

We have also found that these predictions are robust to certain changes in the dynamics of upstream signaling systems such as the Notch pathway. While this manuscript was under review, a new study has shown that Notch signaling is downregulated during HSC emergence (253). By repeating our simulations for a decreasing Notch1 level (see Appendix 3), we found that this change in the dynamics of Notch signaling increases the minimum Bmp4 signal required to switch the triad to a high expression HSC state (Fig. A3.5AB). However decreasing Notch1 levels do not change our results regarding the earlier emergence of HSCs in the haploid Runx1^{+/-} mutants (Fig. A3.5CD). This shows that our predictions are qualitatively robust to dynamical changes in the signaling systems controlling the HSC network.

In our model we have used a statistical thermodynamics framework (156, 232, 254) to model the combinatorial effect of multiple transcription factors on gene regulation. Following recent experimental evidence (117, 118, 127, 145, 226, 232), we assume that transcription factors acting at cis-regulatory elements control gene expression by increasing the probability of a RNA-polymerase
accessible state of chromatin rather than through direct physical interactions with polymerase (see Methods). This no-contact assumption posits a biophysical mechanism for gene regulation which is different from the typical TF-RNA polymerase contact mechanism where DNA looping is assumed (156). However, mathematically the two approaches are inter-convertible (see Methods) and the no-contact mechanism dramatically reduces the number of free model parameters and thereby enhances the utility of this approach when data is limited.

The heterogeneity of assays and functional readouts of definitive blood development limits the direct comparisons between model predictions and experimental data. Nevertheless we have demonstrated that predictions from our model are consistent with experimental observations. Altogether our model shows that the temporal influence of Runx1 and mediators of the Bmp4 signaling pathway on the emergence of HSCs and progenitor cells can be explained by their effects on the activation dynamics of the Scl-Gata2-Fli1 triad. Such an integrative view of the GRN controlling definitive blood cell emergence offers a way to consolidate information from an array of genetic perturbation experiments. However it should be noted that while this model reflects and explains current knowledge of HSC emergence in the AGM it will require updating to accommodate future advances in our knowledge.

Our model also enables us to generate hypotheses about the roles of network components and connections. For instance, our observation that higher Runx1 production rates lead to slower emergence (Fig. 4.3D), may indicate that Runx1 plays a role in the temporal order of stem/progenitor emergence from hemogenic endothelium during embryonic development, i.e. progenitors with restricted lineage potential preceding emergence of definitive HSCs.

Although required for definitive HSC production, neither Runx1 nor Notch1 is essential for primitive erythroid development, whereas components of the triad such as Scl are required for both (149, 247, 255). Our view is that activation of the triad to the HIGH state may not be essential for primitive hematopoiesis. As
such, expression of components of the triad (such as Scl) which are required for this process can occur without activation of the triad to the HIGH state. In this context, our model shows that although Runx1 and Notch1 are essential for switching the triad to the HIGH expression state of the HSCs, Bmp4 can lead to increased expression of triad components in their absence (Fig. 4.3A).

Our results demonstrate that despite the complexity of the HSC network architecture, computational approaches are a useful adjunct to integrating experimental data and help interpret previously unexplained biological observations. Further our results illustrate two important features of this network: the complementary roles of Notch1 and Bmp4 and the vital role of negative regulators such as Smad6. Both of these features could inform how the duration and combination of exogenous stimulation of key signaling pathways is considered when developing protocols for in vitro HSC production.

4.4. Methods

To uncover the role of Runx1, Smad1 and Smad6 in hematopoietic stem cell (HSC) emergence we use an ordinary differential equation (ODE) based deterministic model. This model is based on known transcriptional and post-translational interactions (see Fig. 4.1) and previously published experimental characterizations of the various cis-regulatory interactions in the network controlling HSC emergence in the embryo. These are described in detail in the following sections.

*Thermodynamic modeling of gene regulation*

We combine the processes of transcription and translation involved in the production of a protein from its corresponding gene into a single synthesis reaction. The rate of this reaction depends on the combinatorial binding of transcription factors (TFs) at promoters and enhancers that regulate the expression of each gene. Further, we assume based on available experimental
results (118, 144) that the increase in the transcriptional rate occurs via modulation of the state of chromatin rather than through direct interaction with transcriptional machinery. We call this the No-contact model of transcriptional regulation by cis-regulatory elements (232, 254, 256) as opposed to Contact models where transcription factors act by direct physical interaction with RNA polymerase. Specifically, the gene of interest can exist in either a RNA polymerase-accessible (open) chromatin or RNA polymerase-inaccessible (closed) chromatin state. The binding of TFs at the enhancer or promoter shifts the equilibrium towards the open state, thereby increasing the probability of gene expression. As a result, the probability of transcription can be approximated by the probability of the open chromatin state.

To determine how this probability depends on the concentrations of various TFs we use a statistical thermodynamic approach (232, 254, 256) that allows us to derive each gene’s regulatory function on the basis of known combinatorial interactions. The parameterization of such a regulatory function requires measurements of binding affinities between DNA and TFs as well as interaction strengths among TFs. This entails measuring the contribution of each TF to the transcriptional rate, both individually as well as in every possible combination with other TFs. It is extremely difficult to measure all these contributions due to the combinatorial multiplicity of TF configurations at enhancers and promoters in eukaryotic systems, and as a result such information does not exist for most systems. Instead, experimental data typically includes quantification of the effect of mutation of specific TF binding sites on the expression rates from a luciferase or lacZ reporter (118, 232). Previously, we have shown that such measurements can be used to estimate the strengths of TF-TF and TF-DNA interactions by using appropriate normalization for TF concentrations (232). Here we use the same methods to model the GRN in Fig. 1.3.1 and to estimate the various parameters. Below we provide the various regulatory functions and estimated parameter values, however we refer the reader to refs. (232, 256) for details of this method and provide the regulatory functions and parameters used in our model below.
Relationship between no-contact and contact models of transcriptional regulation

Consider a gene \( Y \) regulated by the combinatorial action of multiple activating TFs \( X_i \) \((i \in 1,2,...,n)\) at a cis-regulatory region. In the no-contact model (see (256) for a detailed derivation), the transcription rate of \( Y \) can be written in terms of the concentrations of \( X_i \)s:

\[
v_Y = \frac{v_Y^o \, z_Y}{(K_Y + z_Y)}; \tag{1}
\]

\[
z_Y = 1 + \sum_{i=1}^{n} [X_i] e^{-G_i} + \sum_{i=1}^{n} \sum_{j=1}^{n} [X_i][X_j] e^{-G_{ij}} + ... \]

where \( v_Y^o \) is the maximum rate of expression of gene \( Y \), \( K_Y \) is the equilibrium constant of chromatin remodeling, \([X_i]\) is the concentration of TF \( i \), and \( G_i \) is the binding free energy of TF \( i \). The fraction \( z_Y/(K_Y+z_Y) \) is the probability of the RNA-polymerase accessible state of chromatin and its exact form is determined by the partition function \( z_Y \) which sums over all the possible TF-bound states of the regulatory region. In this model we assume that TFs only control the probability of the polymerase accessible state of chromatin at the promoter and therefore the rate of transcription is proportional to the probability of finding DNA in one of the accessible conformations. In this mechanism, regulation of transcription by TFs does not require their physical interaction with RNA polymerase. However, as shown below, the regulatory function of this No-contact mechanism (Eq. [1]) can be rewritten in a form similar to regulatory functions for a contact mechanism of gene regulation where TFs do interact with RNA polymerase.

Dividing both numerator and denominator in Eq. [1] by \( K_Y+1 \) we obtain,

\[
v_Y = \frac{v_Y^o \, z_Y}{K_Y} \frac{1 + \sum_{i=1}^{n} [X_i] e^{-G_i} + \sum_{i=1}^{n} \sum_{j=1}^{n} [X_i][X_j] e^{-G_{ij}} + ...}{1 + \sum_{i=1}^{n} [X_i]\frac{e^{-G_i}}{K_Y+1} + \sum_{i=1}^{n} \sum_{j=1}^{n} [X_i][X_j]\frac{e^{-G_{ij}}}{K_Y+1} + ...} \tag{2}
\]
The above equation can be rewritten as:

\[
v_Y = \frac{v_{Y,\text{basal}}}{1 + \sum_{i=1}^{n} f[X_i] e^{-G_i} + \sum_{i=1}^{n} \sum_{j=1}^{n} f[X_i][X_j] e^{-G_{ij}} + \ldots}
\]

where,

\[
v_{Y,\text{basal}} = \frac{v_Y^o}{K_Y + 1}, \quad G_i = G_i + \log(K_Y + 1) \quad \text{and} \quad f = K_Y + 1,
\]

Here the fold-change f can be interpreted as the maximum increase in transcription rate resulting from the binding of each TF or TF complex. Eq. [3] resembles the typical cis-regulatory function derived in the contact models of combinatorial transcription regulation where TFs directly interact with RNA polymerase (156). As seen from Eqs. [2] and [3] the contact and no-contact models are mathematically equivalent even though they represent different biophysical mechanisms of gene regulation. However, in the contact model the fold-changes resulting from binding of different TFs or TF complexes can be different but Eq. [2] and Eq. [3] are only mathematically equivalent only under the assumption of equal fold-changes. It should be noted though that the experimental data that is generally available (measurements of lacZ expression from deletion libraries of cis-regulatory elements) is usually insufficient to simultaneously determine the fold-changes and binding free-energies of TFs and TF complexes (232, 256). Therefore, we believe that using the no-contact model with its implicit assumption of equal fold-changes is a reasonable approximation. This formulation (Eq. [1]) is used throughout the following sections to derive the cis-regulatory functions of various genes in our model.

**Notation**

The notation used throughout these Methods is summarized in Table A3.1. Brackets [..] denote concentrations. Note that all concentrations are in
dimensionless units and are assumed to be normalized by their steady-state values in the absence of Bmp4 and Notch1 signals. In addition, $G^x_\gamma$, denotes a free energy of binding of TF $x$ to an enhancer of gene $y$.

**Triad transcriptional regulation**

The triad genes Scl, Gata2 and Fli1 regulate each other by acting at the Scl+19, Gata2-3 and Fli1+12 enhancers (118, 144). In addition, the Gata2 promoter is known to be regulated by Bmp4 (via pSmad1) and Notch1 and the Fli1 promoter is regulated by Bmp4 (via pSmad1). Using available details of TF binding sites in these enhancers and promoters we have previously shown that the following equations can be used to describe the rates of transcription $v_s$ (Scl), $v_g$ (Gata2) and $v_f$ (Fli1) for the triad:

$$v_s = v^o_s \frac{z_s}{(K_s + z_s)};$$  \hspace{1cm} [4]$$

$$z_s = 1 + [G]e^{-G^s_\gamma} + [F]^2 e^{-2G^s_\gamma} + [G][F]^2 e^{-G^f_\gamma};$$

$$v_g = v^o_g \frac{(1 + f_1([pS1-R] + \alpha[pSI])e^{-G^w_\gamma} + f_2[N1])z_g}{(K_g + (1 + ([pS1-R] + \alpha[pSI])e^{-G^w_\gamma} + [N1])z_g)};$$  \hspace{1cm} [5]$$

$$z_g = 1 + [G]e^{-G^g_\gamma} + [F]^2 e^{-2G^g_\gamma} + [G][F]^2 e^{-G^f_\gamma} + [S][G][F]^2 e^{-G^f_\gamma};$$

$$v_f = v^o_f \frac{(1 + f_1([pS1-R] + \alpha[pSI])e^{-G^w_\gamma})z_f}{(K_f + (1 + ([pS1-R] + [pSI])e^{-G^w_\gamma})z_f)};$$  \hspace{1cm} [6]$$

$$z_f = 1 + [G]e^{-G^f_\gamma} + [F]^2 e^{-2G^f_\gamma} + [G][F]^2 e^{-G^f_\gamma} + [S][G][F]^2 e^{-G^f_\gamma};$$

Here, $v^0_s$, $v^0_g$ and $v^0_f$ are the maximum rates of transcription from the Scl, Gata2 and Fli1 promoters in the absence of Notch1 and Bmp4 signals. $K_s$, $K_g$ and $K_f$ are the equilibrium constants of the chromatin state of the Scl, Gata2 and Fli1 regulatory regions respectively. $f_1 (\approx 4)$ (147) and $f_2 (\approx 3.5)$ (146) denote the fold-change in gene expression rate resulting from the effect of Bmp4 and Notch1 on the appropriate promoters. Based on refs. (145, 236), we have modified equations [4]--[6] from those in (232) to specify that Bmp4 affects Gata2 and Fli1 expression via the binding of pSmad1 and pSmad1:Runx1 complex. We assume
based on the observations of ref. (236) that the pSmad1:Runx1 complex is more effective than pSmad1 alone at increasing the transcription from Gata2 and Fli1 promoters. Accordingly, we assume that pSmad1 alone can only effect a fold-change $\alpha f_1$ where $\alpha < 1$. [N1] represents the concentration of Notch1 acting on the Gata2 promoter normalized by its dissociation constant of binding to the promoter. The parameters used in equations [4]-[6] were derived in (232) based on the results from (118, 135) and are reproduced in Table A3.2.

**Smad1 transcriptional regulation**

Gene expression from the Smad1 promoter is regulated by the Smad1-7 upstream enhancer. This enhancer includes multiple Fli1(Ets) binding sites as well as a Gata2 binding site (Fig. A3.1 and data not shown). We assume that this enhancer modulates Smad1 transcription by controlling the probability of transcription from the Smad1 promoter. Following this assumption and the deletion analysis of the Smad1-7 enhancer (Fig. A3.1) the rate of Smad1 transcription, $v_{s1}$, was modeled using the following equation:

$$v_{s1} = v_{s1}^o \frac{z_{s1}}{K_{s1} + z_{s1}};$$  

$$z_{s1} = (1 + [F]e^{-G_{1}})(1 + [F]e^{-G_{2}})(1 + [G]e^{-G_{1}})$$

Here, $v_{s1}^o$ is the rate of transcription from the Smad1 promoter and $K_{s1}$ is the equilibrium constant of the chromatin state of the Smad1 regulatory region. Using lacZ-reporter expression assay data (not shown) we have determined the various free energies in Eq. [7] in terms of the $K_{s1}$. The parameter values used in model for the Smad1-7 enhancer are given in Table A3.3.

**Smad6 transcriptional regulation**

Smad6 expression is regulated by the Smad6 promoter as well as the upstream Smad6-57 enhancer (Fig. A3.1). Both the promoter and enhancer include binding sites for the triad proteins as well as Runx1 and pSmad1. We assume that this enhancer modulates the probability of transcription whereas the
promoter controls the rate of Smad6 transcription. Following these assumptions and the deletion analysis of the Smad6 enhancer and promoter (117, 226), the rate of Smad6 transcription, $v_{s6}$, was modeled using the following equation:

$$v_{s6} = \frac{z_{s6p}}{(K_{s6p} + z_{s6p})} \cdot \frac{z_{s6e}}{(K_{s6e} + z_{s6e})};$$

$$z_{s6p} = (1 + [F]^2 e^{-2G_{s6p}})(1 + ([R] + [pSI - R]) e^{-G_{s6p}})(1 + [S]^2 G [F]^2 e^{-G_{s6p}}),$$

$$z_{s6e} = 1 + [R] e^{-G_{s6e}} (1 + [G] e^{-G_{s6e}}).$$

Here, $v_{s6}^o$ is the maximum rate of transcription from the Smad6 promoter and $K_{s6p}$ and $K_{s6e}$ are the equilibrium constants of the chromatin state of the Smad6 promoter and enhancer respectively. Using the data from (117, 226) we can determine the various free energies in Eq. [8] in terms of $K_{s6p}$ and $K_{s6e}$. The parameter values used in model for the Smad6 regulation are given in Table A3.4.

### Runx1 transcriptional regulation

Runx1 expression is regulated by the two promoters P1 and P2 as well as the intronic $Runx1+23$ enhancer (Fig. A3.1; (117, 127, 234, 238)). These promoters and enhancer include binding sites for the triad proteins as well as Runx1 and pSmad1. It has been shown that the $Runx1+23$ enhancer specifically directs gene expression from both promoters in hemogenic tissue (234). Moreover $Runx1+23$ can modulate the gene expression from each promoter independent of the other (238). It has also been recently shown that transcription from promoter P1 largely depends on the binding of Runx1 and pSmad1 rather than triad proteins (234). We again assume that the enhancer modulates the probability of transcription whereas the promoters control the rate of Runx1 transcription. Taking all these facts and assumptions into account and using the deletion analyses of the Runx1 enhancer and promoters (117, 127, 234, 238), the rate of Runx1 transcription, $v_{s6}$, was modeled using the following equation:
Here, $v_r^o$ is the maximum rate of transcription from the Runx1 promoter and $K_{rp1}$, $K_{rp2}$ and $K_{re}$ are the equilibrium constants of the chromatin state of the promoters P1 and P2 and the Runx1+23 enhancer respectively. We assume the same $v_r^o$ for both promoters since transcription from both promoters has been shown to be comparable (234). Using the data from (117, 127, 234, 238) we can determine the various free energies in Eq. [9] in terms of $K_{rp1}$, $K_{rp2}$ and $K_{re}$. The parameter values used in model for the Runx1 regulation are given in Table A3.5. $v_r^o$ was multiplied by 0, 0.5 and 1.5 to model mutations such as Runx1-/-, Runx1+/- and three copies of Runx1 respectively.

**Protein degradation/dilution**

We assume that the rate of degradation for all model species follows first-order kinetics. Degradation rates $k_{deg}^i$ for each species $i$ in the model were calculated based on experimental data about protein stability (166). We used a half-life of 24hrs for Scl, 4hrs for Fli1 and 2hrs for the remaining species (Gata2, Smad1, pSmad1, Runx1, Smad6, pSmad1:Runx1, pSmad1:Smad6 and Runx1:Smad6).

**Post-translational Reactions**

The Bmp4 signaling network includes many post-translational interactions involving the regulatory Smads, Smad1 as well as the inhibitory Smad, Smad6 (237, 257, 258). Phosphorylation of Smad1 by Bmp4 bound receptors leads to the formation of pSmad1 which translocates into the nucleus and affects gene expression of Bmp4 signaling target genes (257). Smad6 inhibits Bmp4 signaling by (i) binding to Bmp4 bound receptors and targeting them degradation (257), (ii)
binding pSmad1 and inhibiting its translocation into the nucleus (257) and (iii) binding pSmad1 and facilitating Smurf-mediated ubiquitination and subsequent proteosomal degradation of pSmad1 (237, 258). We assume, for simplicity, that primary inhibitory effect of Smad6 on pSmad1 is the Smurf-mediated degradation of pSmad1. Runx1 can participate in Bmp4 signaling by binding to pSmad1 in the nucleus and affecting its transcription factor activity (236). Moreover Smad6 can bind to Runx1 and target it to the proteasome similar to its effect on pSmad1 (259). We include the following post-translational reactions in our model to reflect a simplified version of the Bmp4 signaling pathway and Runx1’s interaction with Smads.

Bmp4 binds to a cell-surface receptor leading to the phosphorylation of the regulatory Smad Smad1 (260, 261). We model the phosphorylation of Smad1 as a simple first-order reaction and use the rate of phosphorylation, $k_p$, as a proxy for the Bmp4 dose.

\[
\text{Smad1} \xrightarrow{k_p} \text{pSmad1} \tag{10}
\]

Smad1 localization in the nucleus has been shown to be Runx1 dependent (236). In our model, we assume that Runx1 binds, reversibly, to phosphorylated Smad1 (pSmad1) and that this complex can affect transcriptional regulation (see section 4.1.1; (236)).

\[
p\text{Smad1} + \text{Runx1} \xrightleftharpoons[k_{d}]{k_{a}} p\text{Smad1-Runx1} \tag{11}
\]

Smad6 binds to pSmad1 and Runx1 to form complexes that are targeted for Smurf-mediated proteolytic degradation (237, 257, 258). Proteolytic degradation is modeled as a first-order reaction.

\[
p\text{Smad1} + \text{Smad6} \xrightarrow[k_d]{k_b} p\text{Smad1-Smad6} \xrightarrow{k_{\text{prot}}} \emptyset \tag{12}
\]

\[
\text{Runx1} + \text{Smad6} \xrightarrow[k_d]{k_b} \text{Runx1-Smad6} \xrightarrow{k_{\text{prot}}} \emptyset \tag{13}
\]
We also include a partner-switching reaction in which Smad6 replaces Runx1 in the pSmad1-Runx1 complex. The resulting pSmad1-Smad6 is targeted for proteolytic degradation as above.

\[ \text{pSmad1-Runx1 + Smad6} \xrightarrow{k_h} \text{pSmad1-Smad6 + Runx1} \quad [14] \]
\[ \text{pSmad1-Smad6} \xrightarrow{k_{prot}} \emptyset \quad [15] \]

The parameter values for all post-translational reactions are given in Table A3.6. To model the effect of proteosomal inhibition by Bortezomib (BTZ) the proteolytic degradation rate, \( k_{prot} \), was set to zero.

**Model Equations**

All model details are used to derive the following system of ODEs that describe the GRN under consideration:

\[
\begin{align*}
\frac{d[S]}{dt} &= v_s - k_{deg}^{S}[S] \\
\frac{d[G]}{dt} &= v_g - k_{deg}^{G}[G] \\
\frac{d[F]}{dt} &= v_f - k_{deg}^{F}[F] \\
\frac{d[S1]}{dt} &= v_{s1} - k_p[S1] - k_{deg}^{SI}[S1] \\
\frac{d[pS1]}{dt} &= k_p[S1] - [pS1](k_b[S6] + k_{b1}[R]) + k_{d1}[pS1-S6] + k_{d1}[pS1-R] - k_{deg}^{SI}[pS1] \\
\frac{d[S6]}{dt} &= v_{s6} - k_b[S6]([pS1]+[R]+[pS1-R]) + k_d([pS1-S6]+[R-S6]) - k_{deg}^{SI}[S6] \\
\frac{d[R]}{dt} &= v_r + v_r - [R](k_b[S6] + k_{b1}[pS1]) + k_d[R-S6] + k_{d1}[pS1-R] + k_{d1}[pS1-R][S6] - k_{deg}^{R}[R] \\
\frac{d[pS1-S6]}{dt} &= k_b[S6]([pS1]+[pS1-R]) - k_d[pS1-S6] - k_{prot}[pS1-S6] - k_{deg}^{SI}[pS1-S6] \\
\frac{d[R-S6]}{dt} &= k_b[R][S6] - k_d[R-S6] - k_{prot}[R-S6] - k_{deg}^{SI}[R-S6] \\
\frac{d[pS1-R]}{dt} &= k_{s1}[R][pS1] - k_{d1}[pS1-R] - k_b[pS1-R][S6] - k_{deg}^{SI}[pS1-R]
\end{align*}
\]
Here $v_r$ in the differential equation for [R] is included to model Runx1 overexpression (see Fig. 4.4). $v_r$ was set to $100\text{hr}^{-1}$ for Fig. 4.4B.

ODEs for all models (WT and mutants) were solved using the command-line bifurcation package CL-MATCONT (262) to obtain the steady-state signal-response curves (one-parameter bifurcation diagrams) containing range of signals for which two different steady states are possible. The package was also used to obtain the two-parameter bifurcation diagrams which depict the range of parameter values for which two different steady states are possible. For all models, dynamics to step-input in the Bmp4 signal were analyzed using the ode23s solver in MATLAB 2010a(R) (MathWorks, Natick, Massachusetts). Dimensionless models were used in all simulations.

Notes

This chapter is based on work published in the following article:

Regulation of the Bacillus subtilis sporulation decision
Chapter 5

Ultrasensitive dependence of sporulation decision on Spo0A~P

5.1. Introduction

In response to nutrient deprivation, B. subtilis cells undergo asymmetric cell division and then follow a cell-differentiation program resulting in formation of metabolically inert spores (Fig. 5.1A)(1, 2). Progression of the sporulation program is under the control of a large regulatory network (hereafter called the sporulation network). This network involves the sporulation master regulator Spo0A and five alternative sigma factors (σ^H, σ^F, σ^E, σ^K and σ^G), that are activated in precise temporal order (8). Initiation of the sporulation program is controlled by Spo0A (4, 9). The activity and concentration of this master transcription factor is regulated by a phosphorelay through both post-translational and transcriptional interactions (10). Post-translationally, phosphoryl groups are transferred from one of the five auto-phosphorylating kinases (KinA-KinE) to Spo0A via the phosphotransferases Spo0B and Spo0F (Fig. 5.1B). Moreover, phosphorylated Spo0F and Spo0A are subject to negative regulation by phosphatases, RapA and Spo0E, respectively (11). Transcriptionally, expression
of genes for KinA, Spo0F, and Spo0A are regulated by the activated form of Spo0A (phosphorylated form of Spo0A, Spo0A~P hereafter) directly and indirectly via multiple feedback loops (Fig. 5.1B).

As Spo0A~P accumulates, it activates the expression of sporulation genes and promotes polar septation, resulting in the formation of a small forespore compartment and a larger mother cell compartment inside the cell (Fig. 5.1A) (12). Subsequently, compartment-specific sigma factors $\sigma^F$ and $\sigma^E$ are activated in the forespore and the mother cell, respectively (8) (Fig. 5.1C). $\sigma^F$- and $\sigma^E$-bound RNA polymerases (RNAP) direct the compartment-specific gene expression programs including the activation of the late-stage sigma factors $\sigma^G$ and $\sigma^K$. Morphological changes resulting in the production of a mature spore are directed by each of these sigma factors in a temporally and spatially controlled manner (5, 6).

**Figure 5.1. Regulation of sporulation in B. subtilis.**

**A.** Starvation triggers a switch from vegetative growth to sporulation and activates the sporulation master regulator Spo0A~P (0A~P). 0A~P promotes the formation of asymmetric septa and activates sigma factors $\sigma^F$ and $\sigma^E$ in the forespore and the mother cell respectively. Only cells with active $\sigma^F$ and $\sigma^E$ are committed to progress through several additional stages before a mature spore appears.
B. The sporulation phosphorelay which transfers phosphoryl groups from the kinases KinA-E (only KinA is shown) to the master regulator Spo0A via two phosphotransferases Spo0B (0B) and Spo0F (0F). 0A~P controls the expression of multiple genes in the phosphorelay through transcriptional feedback. KinA expression is indirectly regulated by 0A~P (dashed arrow) in the wild-type phosphorelay but in the ASI system the KinA promoter is replaced with an IPTG inducible promoter (solid arrow).

C. The sporulation network is hierarchically organized. 0A~P directly controls the expression of $\sigma^F$ and indirectly controls its activation (via SpoIIE, anchored in the polar septum, grey bar). 0A~P also controls the expression of $\sigma^E$ and its activation via the $\sigma^F$ regulated expression of SpoIIR (black and red arrows show transcriptional and post-translational regulatory interactions respectively).

D. At a threshold level of KinA induction, spore counts increase dramatically (~20-fold increase in spore count between 4 and 10 $\mu$M IPTG) to match wild-type sporulation levels. The blue circles represent experimentally measured spore counts. The solid and dashed lines represent Hill-equation fit and the 95% confidence intervals respectively.

Although the molecular interactions in sporulation network have been mapped out, it is not clear how this network processes environmental and metabolic signals to achieve distinct cell-fate decisions. Moreover, even in an isogenic population of starving cells, only a fraction of them form spores (13). This variability in cell fate is likely to be part of a bet-hedging strategy to manage the risks and benefits of sporulation decision (13, 14) but the implementation mechanism for this strategy are unknown. In particular, it is not clear which sporulation network component can serve as a reliable predictor of cell fate.

In this paper, systematically interrogate the sporulation network using a synthetic Artificial Sporulation Initiation (ASI) system which allows for tunable control of Spo0A~P activity by artificially tuning KinA level from an IPTG inducible promoter (see Fig. 5.1B) (29, 30). By combining quantitative experiments using ASI system with detailed mathematical modeling, we uncover the relationship among the levels of KinA activity, Spo0A~P, and the activities of $\sigma^F$ and $\sigma^E$. Comparing the fractions of cells with each regulator activated with the fraction of
sporulating cells at different levels of KinA activity, we determine the point at which the sporulation decision is finalized. Finally the relevance of the uncovered facts is verified in the wild-type cells under starvation conditions. Thus our results provide system-level picture on how sporulation network makes appropriate cell-fate decision.

5.2. Results

5.2.1. Sporulation in ASI is induced in ultrasensitive manner

Previously, we have shown that ASI system can efficiently induce sporulation in rich medium when KinA is induced to a level that matches KinA concentrations in wild-type cells in sporulation media (129). Notably, in these experiments that the number of cells forming spores increases sharply in a small range of IPTG concentration (129). As a result, increase in IPTG concentration between 4 and 10µM IPTG leads to about 20-fold increase in the number of spore-forming cells. Such responses are often referred to as ultrasensitive; mathematically they can be characterized by large value of slope in log-log coordinates (x is an input, y is an output): \( \frac{d\log(y)}{d\log(x)} = \frac{(dy)/y}{(dx)/x} \gg 1 \). Or equivalently they are characterized by a large effective Hill coefficient, \( n \), when fitted using the Hill-equation: \( y = b + f x^n / (K^n + x^n) \).

By repeating these experiments we verified that spore counts increase ultrasensitively in response to the increasing IPTG concentrations (Fig. 5.1D). If fitted by the Hill-equation, the curves can be characterized by the Hill-exponent \( n \sim 12 \) with a 95% confidence interval of \( (7.5, 16.5) \). The threshold amount of IPTG triggering sporulation is even more robust with a mean level of \( K=7.3 \mu M \) and 95% confidence interval of \( (7.0 - 7.6 \mu M) \). Ultrasensitive increase in spore counts cannot be attributed to the response of IPTG-inducible promoter since the average KinA concentration (in cell culture) increases less than two-fold between 4 and 10 µM IPTG (Fig. A4.1A; effective Hill-coefficient of KinA induction \( n \sim 2.0 \)). Notably single-cell distributions of KinA concentration measured
at 4 and 10 µM IPTG are unimodal but noisy and partially overlapping (Fig. A4.1B).

**Figure 5.2. The phosphorelay response is not ultrasensitive, due to response time requirements.**

**A.** Modeling results show that the steady state Spo0A activity computed as rate of *PspoIIG* transcription can be either graded (green curve) or bistable and ultrasensitive (purple curve).

**B.** Stochastic simulations show that at 10µM IPTG *PspoIIG* expression increases significantly by three hours after induction for the graded (green curves) phosphorelay whereas bistable (purple curves) phosphorelay shows little change in expression from *PspoIIG*. Thin curves are individual stochastic simulation trajectories and thick curves indicate the average of 400 such trajectories.

**C.** Measurements of the increase in Spo0A activity at T3 as a function IPTG with a *PspoIIG-lacZ* reporter (black squares) matches the predicted response of the graded phosphorelay (green curve) but not bistable phosphorelay (purple curve). All values are normalized by the value at 0 µM IPTG.

**D.** In agreement with the graded phosphorelay model, single cell measurements of GFP expression from the spoIIIG promoter at T1 show no bimodality at either 4 or 10 µM IPTG and can be fit with gamma distributions (solid curves).
Ultrasensitivity in the population response can be explained if we assume the level of some downstream gene essential for sporulation increases ultrasensitively as a function of KinA in single cells (Fig. A4.1). As a result, noisy unimodal distribution of KinA will be converted into bimodal distribution of the downstream gene and the fraction of cells in each peak will ultrasensitively depend on the mean level of KinA (Fig. A4.1C,D). To confirm this, we used a toy theoretical model that introduced a hypothetical downstream gene which ultrasensitively depends on KinA (a Hill function with $n = 20$, Fig. A4.1D). We tune the parameters of KinA induction to match the mean level and distribution of KinA observed in the experiments and then calculated expected population distribution of the hypothetical downstream gene. Resulting fractions of simulated cells in which these downstream gene activated are in good agreement with observed fraction of sporulating cells (Fig. A4.1E). Based on this, we hypothesize that the abrupt increase in the number of sporulating cells above the threshold level of KinA originates from ultrasensitivity of the sporulation network in individual cells. In the following sections, we test this hypothesis and uncover the mechanism for generating such an ultrasensitive transfer function.

5.2.2. Activation of Spo0A is not ultrasensitive

To establish how induction of KinA expression affects Spo0A activation, we first built a model for the phosphorelay module. The model includes the transcriptional regulation of module components and their post-translational modifications by phosphotransfer reactions. The model inputs concentration of IPTG which is converted to parameters of KinA transcription to ensure agreement of the model predictions with observed mean and standard deviation of KinA concentration at all IPTG levels. To maintain consistency with experimental results, we report Spo0A~P levels in terms of transcriptional activity of the high-threshold Spo0A~P target spolIG promoter ($P_{spolIG}$) (44). Throughout this section we define Spo0A activity as $P_{spolIG}$ transcription level by experimentally measuring β-galactosidase activity derived from $P_{spolIG-lacZ}$
or GFP intensity derived from *PspollG-GFP* or computing transcription rate in the model.

Our model showed that the increase in Spo0A activity as a function of KinA can be either graded (green curve in Fig. 5.2A) or ultrasensitive (purple curve in Fig. 5.2A) depending on the parameter values selected (in Fig. 5.2A). The sensitivity of the phosphorelay response to IPTG (IPTG-Spo0A~P transfer function) critically depends on the strength of the transcriptional positive feedbacks from Spo0A~P to Spo0A and Spo0F (see Fig. A4.2A). Parameter sets with strong positive feedback, for example those with a high Hill coefficient for the effect of Spo0A~P on the activation of *Pspo0A* and *Pspo0F* (transcription from the promoters of *spo0A* and *spo0F*, respectively) display ultrasensitivity, whereas weaker feedback leads to graded IPTG-Spo0A~P transfer functions (Fig. A4.2A). Notably, ultrasensitive increases in Spo0A~P as a function of IPTG occur for phosphorelay module parameter values for which the module is a bistable switch (purple curve in Fig. 5.2A; also see Fig. A4.2A). Thus, if the decision to sporulate was directly determined by the level of Spo0A~P, a bistable phosphorelay response could explain why the spore count rises so sharply from 4µM to 10µM IPTG.

However our simulations indicated that the bistable phosphorelay would show negligible increase in Spo0A activity even up to 6 hours (purple curve - Fig. 5.2B; also see Fig. A4.2B). Since a large number of phase-bright endospores are formed at IPTG=10µM by T6 (Tn denotes a time n hours after IPTG addition) (129), these results (Fig. 5.2B) suggest that in the bistable parameter regime the phosphorelay response is unrealistically slow. We note here that even the graded phosphorelay does not achieve steady state within 3 hours of induction (green curves in Fig. 5.2B). Therefore, we must use the computed dynamical transfer function (rather than steady-state) transfer function of the phosphorelay to compare the modeling results with experimental data. We have found in our experiments with the ASI system that most cells reach the point of engulfment by three hours after induction with IPTG. This indicates that most cells have reached
a decision about cell-fate by T3. Therefore it is a suitable time-point to evaluate the output of our models. Focusing on transcriptional activity of *PspollG* at T3, we found that the bistable phosphorelay model showed little increase in it even at high IPTG concentrations (Fig. 5.2C, purple curve) whereas the graded phosphorelay model predicted a 30-fold increase over the range of IPTG simulated (Fig. 5.2C, green curve). However, the graded model of the phosphorelay showed only a 5-fold change in *PspollG* expression between 4 and 10µM (Fig. 5.2C, green curve-average of stochastic simulations; Hill-coefficient for fit to experimental data: n~2.4), which is insufficient to explain the ultrasensitive response of spore formation (Fig. 5.1D).

Taken together, the modeling results suggest that only graded phosphorelay model is sufficiently fast to explain the sporulation dynamics. To confirm this prediction, we experimentally analyzed the input-output dynamics of the phosphorelay module. We measured Spo0A activity using a *PspollG-lacZ* (output) reporter at different IPTG concentrations (input). As shown in Fig. 5.2C (black squares), the experimentally measured Spo0A activity closely matched the predictions of the graded phosphorelay model, suggesting that the phosphorelay module response is not bistable and thus not the determinant of the ultrasensitive increase in spore formation. To further test the graded phosphorelay model, we quantified Spo0A activity in single cells with a *PspollG-gfp* reporter to determine whether the distributions show bimodality indicative of a bistable or ultrasensitive phosphorelay response. While the GFP expression level from *PspollG* at T1 was highly heterogeneous, we found that the distributions are not bimodal (Fig. 5.2D). Moreover the peaks corresponding to low (4µM) and high (10µM) IPTG levels are not well separated.

Thus, our results agree with several recent reports that have shown that cell-fate heterogeneity during sporulation is not associated with bimodality of Spo0A~P activity (263, 264). Altogether, these results lead us to conclude that the phosphorelay response or Spo0A~P concentration does not increase steeply around the KinA threshold and that the ultrasensitivity in the sporulation response
must be associated with network components that reside in the downstream of the phosphorelay.

5.2.3. Activation of $\sigma^F$ in the forespore is ultrasensitive but occurs below the sporulation threshold

To test whether the $\sigma^F$ module is the determinant of the ultrasensitive cell-fate decision, we first modeled this module separately to compute its input-output properties and then integrated it with the phosphorelay module to establish its response to KinA induction.

The $\sigma^F$ activation module has been studied extensively both experimentally and theoretically. These studies have shown that $\sigma^F$ (SpoIIAC) expression from the $spolIA$ operon is turned on at a relatively low concentration of Spo0A-P prior to septation (265). The anti-sigma factor SpoIIAB and anti-anti-sigma factor SpoIIAA are expressed along with $\sigma^F$ from the $spolIA$ operon (266). The SpoIIAB kinase inactivates SpoIIAA and forms a complex with $\sigma^F$, preventing it from interacting with RNAP (see Fig. A4.3A and Methods for details) (267). Compartment-specific activation of $\sigma^F$ in the forespore is achieved when the polar-septum forms and the subsequent preferential localization of septum-bound phosphatase SpoIIE on the forespore side leads to an effective increase in its concentration in the forespore compartment. As a result, SpoIIE dephosphorylates SpoIIAA leading to the activation of $\sigma^F$ in a forespore-compartment specific manner (268). Mathematical models of the $\sigma^F$ activation network have shown that the activity of $\sigma^F$ increases ultrasensitively with an increase in SpoIIE concentration (269, 270).
Figure 5.3. $\sigma^F$ activation overestimates the fraction of cells that sporulate.

A. Stochastic simulations of a mathematical model integrating phosphorelay and $\sigma^F$ activation modules shows ultrasensitive increases of $\sigma^F$ in single cells as a function of IPTG (mean response-solid blue line, std. deviation-shaded area, Hill-equation fit-dashed black line). However, only a 2-fold increase in the mean active $\sigma^F$ level is observed between 4µM and 10µM IPTG (grey arrow) as $\sigma^F$ activation has a low half-maximal threshold (~4µM IPTG).

B-C Bimodal distributions of active $\sigma^F$ level in model (B, upper panel) and single-cell experiments with PspoIIQ-mCherry reporter (C, upper panel). Bottom panels of (B) and (C) represent cumulative distributions corresponding to the data in top panels, i.e. total fraction of cells with active $\sigma^F$ level below the given value. Threshold values separating two peaks (grey bars) are chosen to predict the fraction of cells that activate $\sigma^F$ for Panel D.

D. Model predictions for fraction of sporulating cells based on the threshold of active $\sigma^F$ level (black curve) are computed using the distributions for various IPTG levels and the threshold value shown in Panel B. Experimental data (red triangle and green square for 4µM and 10µM respectively) is obtained using the threshold and distributions in Panel C. Fractions of $\sigma^F$ active cells computed from experimental and simulation data are in excellent agreement with one-another but both exceed observed spore fractions (purple dots; calculated from spore counts shown on Fig. 2.1D, see Methods).

E. Examples of the microscopy data of strain MF3765 used to construct distributions in Panel C. A significant fraction of cells show $\sigma^F$ activity (measured by PspoIIQ-mCherry false-colored magenta forespore in the image) even at low IPTG concentrations (left
This fraction increases at high IPTG (right image, 10 µM) but the increase is not ultrasensitive. Spo0A activity was measured by *PspollG-gfp* (green).

To determine how phosphorelay activity impacts σ^F activation, we extended our earlier models to include regulation of *spolla* and *spollE* gene expression by Spo0A~P. We found that the threshold of SpoIIE concentration, around which σ^F activity increases, is dependent on the expression levels of the *spolla* operon (Fig. A4.3B). Moreover, since Spo0A~P controls both *spolla* and *spollE* transcription, the regulation of the σ^F module can be defined as an AND-type coherent feed-forward (see Fig. 5.1C) (271). AND-type coherent feed-forward loops combine the sensitivity of each branch (272) which can potentially lead to ultrasensitive responses (see equation [2] in Methods). In our case, we note that Spo0A~P regulates both *spolla* and *spollE* cooperatively [Hill exponents: n_{IIA}, n_{IIIE} >2 (273)]. This cooperativity is amplified by the feed-forward loop because σ^F activation itself is highly ultrasensitive as a function of SpoIIE concentration (Fig. A4.3B). Based on these facts, we expect that even the modest increase in the level of Spo0A~P seen in the previous section could result in an ultrasensitive increase in σ^F activity.

The results of our simulations corroborated this rationale but surprisingly, the model predicted that σ^F activation increases ultrasensitively at low IPTG concentrations (Fig. 5.3A). Hill-equation fit for σ^F activity as a function of IPTG concentration (dashed black curve in Fig. 5.3A) indicates a threshold of 4µM IPTG (95% confidence interval: 3.9-4.2 µM) and a Hill-coefficient n~4.6. At the population level, σ^F activity has a bimodal distribution and about half of cells showing significant σ^F activity at 4 µM IPTG when less than 1% of cells actually end up as spores (Fig. 5.3D, purple dots). Moreover our model predicted that the fraction of cells activating σ^F (we choose 2µM σ^F as a threshold) increases less than two-fold from 50% to 77% between 4 and 10µM IPTG (Fig. 5.3D, black curve). This is in stark contrast to the experimentally observed increase in the
fraction of spores (Fig. 5.3D, blue dots) which increases 20-fold between 4 and 10µM IPTG.

These results indicate that the threshold of KinA for $\sigma^F$ activation falls below the observed value for the sporulation response because both *spoIIA* and *spoIIE* transcription occur at relatively low Spo0A~P threshold (130). As a result, both operons are expressed at high levels even at low IPTG concentrations (4µM IPTG) for KinA induction. Inevitably then, under such conditions, cells accumulate sufficient SpoIIE to activate $\sigma^F$ after septation. Our prediction that a large fraction of cells activate $\sigma^F$ even at low IPTG is robust to the choice of threshold concentration for which we consider $\sigma^F$ to be “activated”. This robustness is due to the bimodality of active $\sigma^F$ concentration in the population - see Fig. 5.3B lower panel; the fractions of activated cells would not change much as long as the threshold level is between the peaks.

Our model also demonstrated that despite the ultrasensitivity of the $\sigma^F$ activation module, $\sigma^F$ is activated within ~10 minutes of septation (Fig. A4.3D). This result contrasts the dynamic behavior of the phosphorelay network where there is a trade-off between the ultrasensitivity and response time of Spo0A~P activation (Fig. 5.2A-C and Fig. A4.2AB). This tradeoff is avoided in the $\sigma^F$ activation module because ultrasensitivity is achieved through coherent feed-forward rather than positive feedback architecture (Fig. 5.1C). Moreover, increase in *spoIIA* and *spoIIE* transcription does not slow down the response of the $\sigma^F$ module because the gene products of both accumulate before septation under Spo0A~P control and thus the rate of $\sigma^F$ activation is only determined by relatively fast posttranslational reactions in the forespore. In fact, the rapid activation of $\sigma^F$ is known to be essential for its role in inter-compartmental signaling after septation (274, 275).

We confirmed our model predictions for the $\sigma^F$ module experimentally by tracking Spo0A and $\sigma^F$ activities simultaneously using *PspollG-gfp* and *PspollQ-mCherry* as reporters (MF3765; here and below strain numbers are show as in Table A4.1), respectively (Fig. 5.3E). We found that the $\sigma^F$ activity displays a
bimodal distribution and only a fraction of cells activate significant levels of $\sigma^F$ (see Fig. 5.3CD and Fig. A4.3E). In agreement with model predictions, our results indicated a relatively large fraction of $\sigma^F$-active cells (55 ± 2.5%; average ± standard deviation of five independent measurements; see Methods for details) even at 4 µM IPTG (sub-threshold level of KinA). Accordingly, the increase in fraction of cells that activate $\sigma^E$ between 4-10 µM ($\sigma^F$-active cells increase from 55% of cells counted at 4 to 74.25 ± 2.5% of cells counted at 10) matched our predictions of the response of the $\sigma^F$ module to KinA induction. As shown on Fig. 5.3D, the red triangle (4 µM) and the green square (10 µM) corresponded to the fraction of cells with $\sigma^F$ reporter intensity above the threshold selected in Fig. 5.3C, and matched the model predictions (black curve) but not the measured spore fractions (purple dots; calculated from spore counts on Fig. 5.1D and viable cell counts - see Appendix 4). Based on these results we can rule out $\sigma^F$ activation as a determinant of cell-fate ultrasensitivity around the KinA threshold.

5.2.4. Ultrasensitive $\sigma^E$ activation in the mother cell determines cell-fate

We extended our approach of modular analysis of the sporulation network further downstream to determine if activation of $\sigma^E$ is the decision point for the ultrasensitive cell-fate response. Analogously to previous sections, we first model the $\sigma^E$ module in isolation and thereafter integrate it with the models of the $\sigma^F$ activation module and the phosphorelay.

The $\sigma^E$ module is regulated at several different levels as follows: (i) Spo0A~P stimulates transcription of the spollG operon encoding both pro-$\sigma^E$ (inactive form of $\sigma^E$) and its processing enzyme SpoIIGA preferentially in the mother cell (276), (ii) in the forespore, the gene for the SpoIIGA activation signal, SpoIIR, is transcribed by $\sigma^F$-RNAP (thus indirectly controlled by Spo0A~P) (277) (see Fig. A4.4A). As a result, the active form of $\sigma^E$ is produced at a level sufficient to transcribe genes in its regulon only after the completion of polar septation and $\sigma^F$ activation (278). Therefore, similar to the $\sigma^F$ module, the $\sigma^E$
module is controlled both directly and indirectly by Spo0A~P via an AND-type coherent feed-forward loop that can amplify the degree of response sensitivity (265).

**Figure 5.4.** $\sigma^E$ activation is the ultrasensitive switch that controls cell fate.

**A.** Stochastic simulations show that $\sigma^E$ activation (mean response-solid black line, std. deviation-shaded area) increases ultrasensitively in single cells as a function of IPTG and that this threshold coincides with the KinA threshold for sporulation. Mean active $\sigma^E$ level increases ~30-fold between 4 µM and 10µM IPTG.

**B-C** Bimodal distributions of active $\sigma^E$ in the model (B, upper panel) and single-cell experiments (C, upper panel). In C PspolID-gfp was used as a reporter to track active $\sigma^E$ level in single-cell experiments; fluorescence is shown in arbitrary units. Bottom panels of (B) and (C) represent cumulative distributions corresponding to the data in top panels. Threshold values separating two peaks (vertical grey bars) are chosen to predict the fraction of cells that activate $\sigma^E$ for Panel D.

**D.** Model predictions for fraction of sporulating cells based on the threshold of active $\sigma^E$ level (black curve) are computed using the distributions for various IPTG levels and the threshold value shown in B. Experimental data (red triangle and green square for 4µM and 10µM respectively) are obtained using the threshold and distributions in panel C. Both experimental and computationally computed fractions of $\sigma^E$ active cells are in excellent agreement with observed experimental spore fraction (purple dots; calculated from sporulation efficiency same data as in Fig. 5.2C).

**E.** Examples of microscopy data of strain MF1957 used to construct distributions in panel C. Only a small fraction of cells show $\sigma^E$ activity (green mother cell in the image) at
low IPTG concentrations (left image, 4µM). This fraction increases ultrasensitively at high IPTG (right image, 10 µM).

The level of activated σ^E depends on the rate of proteolytic cleavage of its pro-peptide and hence on the concentrations of both SpoIIGA and SpoIIR as well as the amount of pro-σ^E available for activation. Our model supported the above notion that a combined increase in both SpoIIGA and SpoIIR concentrations dramatically increases the rate of pro-σ^E cleavage (Fig. A4.4B). We therefore expected that the modest increase in Spo0A~P activity around the KinA threshold could result in a switch-like ultrasensitive increase in σ^E activity (see Fig. A4.4B).

A combined model of the phosphorelay, σ^F and σ^E modules showed that σ^E activation does increase ultrasensitively as a function of KinA levels induced by different concentrations of IPTG between 4 and 10µM (Fig. 5.4A). Hill-equation fit to σ^E activity as a function of IPTG shows a half-maximal threshold of 7.4 µM IPTG (with 95% confidence interval: 7.28-7.63) and a Hill-coefficient n~5. Around this threshold (from 4 to 10µM IPTG), we observed only a modest increase in PspolIG transcriptional activity and σ^F activity (see sections above: mean PspolIG transcriptional activity increases ~5-fold while mean σ^F activity increases ~2-fold). However, the combined increase in spolIG and spolIR expression was synergistically amplified by the feed-forward loop, resulting in an approximately 30-fold increase of active σ^E. As a result of this ultrasensitive increase in σ^E activity at the single cell level, the distribution of σ^E in the population became bimodal (Fig. 5.4B). This allows us to define a threshold of active σ^E concentration (1 µM active σ^E, Fig. 5.4B dashed line) and robustly predict the fraction of cells that activate σ^E and thereby commit to sporulation. The fraction of cells committed to sporulation, as predicted by a threshold concentration of 1 µM active σ^E by T3, increased ultrasensitively between 4 and 10µM IPTG, in good agreement with the change in sporulation faction [(Fig. 5.4D;
compare the model prediction (black curve) with the fraction of sporulated cells (purple dots, same as in Fig. 5.3D)].

To test our modeling predictions of a bimodal distribution of $\sigma^E$ activity, we employed single-cell measurements of $\sigma^E$ activity at T3 using *PspollD-gfp* (MF1957) as a reporter using fluorescence microscopy (Fig. 5.4CDE). These experiments showed that the fraction of cells with active $\sigma^E$ changes drastically between 4 and 10$\mu$M IPTG (Fig. 5.4D). $\sigma^E$-active cells increase from $1.075 \pm 0.26\%$ (average $\pm$ standard deviation of 4 measurements; see Methods for details) of cells counted at 4$\mu$M to $75.5 \pm 3\%$ of cells counted at 10$\mu$M. This dramatic increase agrees well with our model as well as the measurements of spore fraction in experiments (Fig. 5.4D). As shown of Fig. 5.4D; the red triangle (4$\mu$M) and the green square (10$\mu$M) show the fraction of cells with $\sigma^E$ activity above the threshold from Fig. 5.4C, and match the model predictions (black curve) and the measured fraction of sporulated cells (purple dots). These results show that activation of $\sigma^E$ can be used as an accurate predictor of cell fate during IPTG-induced sporulation.

The difference in ultrasensitive activation thresholds for $\sigma^F$ and $\sigma^E$ (compare Fig. 5.3D and 5.4D) suggests that in the 2-5$\mu$M IPTG range $\sigma^E$ activation and consequently sporulation is limited primarily by the inadequate expression of *spollG*. To test this hypothesis, we first modified our model so that *spollG* was expressed constitutively at a level comparable to that seen at 10$\mu$M IPTG (see Appendix 4 for details). Repeating simulations described above, we found that in this case the fraction of cells that activate $\sigma^E$ at 4$\mu$M IPTG increases substantially to $\sim 40\%$ (Fig. A4.5A, solid curve). These results suggest that high level expression of *spollG* reduces the IPTG threshold for $\sigma^E$ activation and the fraction of cells that activate $\sigma^E$ is determined by the number of cells that activate for $\sigma^F$ (compare solid and dotted curves, Fig. A4.5A). Next, to experimentally confirm these predictions, we constructed a modified ASI strain (MF4883) where the *spollG* operon (including both *spollGA* and *spollGB*, gene for pro-$\sigma^E$) is expressed constitutively from the *Pspac* promoter (59). In this
strain, the LacI operator binding site in the *Pspac* promoter was deleted to allow *spollIG* to be expressed at high-levels during growth independent of the IPTG concentration used to induce *kinA* expression (59). Using the same experimental protocol as above, $\sigma^E$ activity at T3 for this strain was quantified using a *PspollID-gfp* reporter. We found that, similar to Fig. 5.5B-C, $\sigma^E$-activity is bimodally distributed in the population at both 4 and 10 µM IPTG (see Fig. A4.5B,C) even though *spollIG* is expressed constitutively. However, in contrast to Fig. 5.5C, the fraction of $\sigma^E$-active cells in this strain increases only approximately twofold from 37% to 78% between 4-10 µM IPTG (mean of two independent measurements at each IPTG concentration; see Fig. A4.5A,C). These measurements are in excellent agreement with the predictions from our model (Fig. A4.5A). Thus, even at low level of KinA, the constitutive expression of *spollIG* increases the fraction of $\sigma^E$-active and consequently sporulating cells. Taken together, these results reinforce the model that the increase in sporulation at the KinA threshold is mainly due to the ultrasensitive increase in $\sigma^E$ activation via feedforward loops involving both transcriptional and post-translational processes (Fig. 5.1A).

### 5.2.5. Activation of $\sigma^F$ does not ensure successful sporulation in wild-type *B. subtilis* cells

Thus far, from our data based on a combination of mathematical modeling with ASI experiments, we uncover two important properties of sporulation network response: (i) Activities of $\sigma^F$ and $\sigma^E$ increase ultrasensitively with increase in Spo0A activation and (ii) at some sub-threshold levels of Spo0A activity, cells that activate $\sigma^F$ may fail to activate $\sigma^E$. We therefore ask if these predictions can be tested using wild-type cells in sporulation-inducing conditions. To this end, we employed fluorescence microscopy to measure gene expression in individual wild-type cells. We note that, in this setup, we have no direct control of the Spo0A activity and therefore cannot probe the input-output response of the network directly at the population level. Nevertheless, since Spo0A activity is broadly heterogeneous during starvation (102, 263, 264, 279), correlations between activities of Spo0A and sigma factor can be examined in single cells.
over a wide range of values. Therefore, to determine if activation of $\sigma^F$ and $\sigma^E$ is ultrasensitive to increases in Spo0A activity, we simultaneously measured activities of Spo0A and sigma factors during starvation. We used a triple reporter strain (MF4859) with a $\sigma^F$ activity reporter ($P_{spoIIQ}$-YFP) and a $\sigma^E$ activity reporter ($P_{spoIID}$-YFP), in addition to the Spo0A activity reporter $P_{spoIIA}$-CFP. To ensure that the measured activities is not hindered by differences in local micro-environments or by heterogeneity in sporulation initiation times, we chose to study the effects in liquid sporulation media and thereafter measure activities of the reporters in several time-points following the sporulation. Our results indicated that activity of Spo0A first increased up to T2.5 (for wild-type cells T$n$ refers to $n$ hours after cells were place in the sporulation media) and then started to decrease, as observed in general (129, 280). This decrease might be associated with the transcriptional repression that occurs at later times, when there is enough active $\sigma^E$ for the downstream sporulation events.

We therefore focused on the analysis of the reporter activities at T2.5; the time at which many cells have already formed an asymmetric septum. In agreement with our predictions from the ASI system, starving wild-type cells showed distinctly bimodal patterns of $\sigma^F$ and $\sigma^E$ activity (Fig. A4.6A,B). Such distributions allowed us to choose a threshold for $\sigma^F$ and $\sigma^E$ activity separating two peaks in bimodal distributions (see horizontal dashed grey line in Fig. A4.6AB) and thereby to classify cells as either $\sigma^F/\sigma^E$ active or inactive. Strikingly, the number of $\sigma^F/\sigma^E$ active cells depended on ultrasensitivity of Spo0A activity (Fig. 5.5AB and Fig. A4.6AB). In other words, we could define clear threshold value (vertical dashed grey line in Fig. A4.6AB) for Spo0A activity required for the activation of the downstream sigma factor. To quantify the ultrasensitivity of $\sigma^F$ and $\sigma^E$ activities in response to Spo0A activity cells were binned according to their activity levels of Spo0A and the fraction of $\sigma^F/\sigma^E$ active cells in each bin are calculated. This analysis yields the fraction of cells with active $\sigma^F/\sigma^E$ as a function of Spo0A activity. (Fig. 5.5AB) . Using Hill-equation fits, we found that the fraction of $\sigma^F$ active cells increases ultrasensitively (Hill-exponent $n$~4) around a Spo0A activity threshold ~147 (in arbitrary units of $P_{spoIIA}$-CFP fluorescence; 95%
confidence interval: 120-174). Using the same fitting procedure, the fraction of $\sigma^E$ active cells was found to increase with even greater sensitivity (Hill-exponent $n \approx 10.67$) around a Spo0A activity threshold ~198 (95% confidence interval: 178-218.5). Notably, the threshold of Spo0A activity and effective Hill coefficient for $\sigma^E$ activation exceed their corresponding values for $\sigma^F$ activation. Thus, based on these results, we can conclude that $\sigma^F$ and $\sigma^E$ activity increase ultrasensitively as a function Spo0A activity, with $\sigma^E$ activation occurring more ultrasensitively and at a higher threshold.

Figure 5.5. Ultrasensitivity and cell-fate decision in wild-type cells.

A-B Triple reporter strain (MF4859) was used to simultaneously measure Spo0A and sigma factors activities in wild-type cells grown in starvation media. Cells were binned based on CFP fluorescence for Spo0A activity ($P_{spoIIA-cfp}$) and the fraction of cells in each bin that are $\sigma^F$ active and $\sigma^E$ active (based on appropriate $P_{spoIIQ-mCherry}$ and $P_{spoIID-yfp}$ fluorescence thresholds) are shown in panels (A) and (B), respectively. The solid lines represent Hill-equation fits. These fits indicate that fractions of cells displaying both $\sigma^F$ and $\sigma^E$ activities increase ultrasensitively at around different Spo0A activity thresholds.

C-D Time-lapse microscopy was used to track $\sigma^F$ activities and cell-fates in wild-type cells (MF1027) under starvation conditions. (C) Time-lapse trajectories of $\sigma^F$ activity in
cells that activate $\sigma^F$ but fail to engulf the forespore and resume growth. (D) Time-lapse trajectories of $\sigma^F$ activity in cells that activate $\sigma^F$, engulf the forespore and form a phase-bright spore.

**E-F** Representative examples of cells that activate $\sigma^F$ and fail to engulf (E) or successfully engulf and form a spore (F). $\sigma^F$ activity ($P_{spoIIQ}$-gfp) is false-colored magenta. Trajectories of $\sigma^F$ activity for the specific cells shown in E and F are indicated by thick lines in panels C and D respectively. All fluorescence intensities are reported in arbitrary units (a.u.).

Going further, we tried to determine if starving wild-type cells can indeed fail to sporulate despite activation of $\sigma^F$. This question required our ability to track cell fate in a long-time sequence of observations and we therefore employed time-lapse microscopy of cells growing in starvation conditions on a solid agarose medium. We note that, while we attempted to use a strain expressing two fluorescence reporters (YFP and CFP) for $\sigma^F$ and $\sigma^E$ activities, one of the intensities was unstable on the agarose medium (but stable in liquid culture), for unknown reasons. Thus, we decided to use the single color GFP reporter system for each time-lapse microscopy experiment. With this setup, we first tracked $\sigma^F$ activity simultaneously with sporulation cell-fate. Notably, we found that during starvation, a fraction of wild-type cells (MF1027) form an asymmetric septum and activate $\sigma^F$ but never engulf the forespore (see Fig. 5.5C,E and Fig. A4.6C). In this fraction, $\sigma^F$ activity subsequently subsides and cells either become dormant or resume vegetative growth (see Fig. 5.5C,E). Since engulfment of the forespore depends entirely on activation of $\sigma^E$ (281), these results indicate that some fractions of the wild-type population fail to activate $\sigma^E$ despite activating $\sigma^F$ and, as a result, are unable to complete sporulation. It should be noted that failure to activate $\sigma^E$ in these cells is not manifested by insufficient $\sigma^F$ activity. Peak $\sigma^F$ activity in these cells is comparable to that in cells that do engulf the forespore and complete sporulation (compare Fig. 5.5C,E vs D,F). Therefore failure to activate $\sigma^E$ must be attributed to other factors. Our results with ASI system (Figs. 5.3 and 5.4) and the results of Fig. 5.5B suggest that insufficient
Spo0A activity may be responsible for failure to activate $\sigma^E$. Thus, these results verify our prediction that $\sigma^F$ activation is not always an accurate predictor of cell-fate since its activation can be reversed in some circumstances.

By repeating the time-lapse measurements with the strain containing $\sigma^E$ activity reporter (MF248), we found that expression patterns of the reporter gene show quite distinct from those for $\sigma^F$ activity reporter (Fig. A4.6DE). We find that the majority, if not all, of the cells that activate $\sigma^E$ successfully engulf the forespore and proceed to sporulation. While some cells that activate $\sigma^E$ died before to the completion of sporulation, little or no $\sigma^E$ active cells were observed to resume vegetative growth. These results are consistent with previous reports that indicate that $\sigma^E$ activation is the sporulation commitment point (274, 282, 283). Notably, in these reports it was also suggested that all cells wild-type cells that form an asymmetric septum eventually activate $\sigma^F$ and $\sigma^E$ and reach the commitment point unless the environment is changed (e.g. by resuspension in nutrient rich media). In contrast, our results indicate that, even under unchanging starvation conditions, a fraction of cells activate $\sigma^F$, but never reach the point of $\sigma^E$ activation.

Taken together, our results with wild-type cells in sporulation conditions confirm the predictions made based on the ASI system and mathematical modeling. We suggest that different thresholds for $\sigma^F$ and $\sigma^E$ activation may allow cells to reverse their developmental progression, even after activating $\sigma^F$, and thereby protect them from unnecessary time and energy.

5.3. Discussion

Examination of our results in light of the architecture of the sporulation network (Fig. 5.1AB) suggests that the design of the ultrasensitive switch that determines cell fate in sporulation is quite unique. In many other organisms, cell-fate decision has been shown to be controlled by the ultrasensitive activation of the master regulator through the positive feedback, resulting in a bistable
response. Thus, two stable states of the responses correspond to two distinct cell fates (52, 232, 244, 284-287). Surprisingly despite the presence of positive feedback in the phosphorelay architecture, we have found that the phosphorelay response to KinA induction is graded and non-ultrasensitive. This is a unique and crucial design feature, with which the sporulation network avoids the slowdown in the dynamical response that is characteristic of bistable systems (43, 232, 250). Instead, the sporulation network generates an ultrasensitive response by using a cascade of coherent feed-forward loops combining transcriptional and post-translational interactions. Specifically, Spo0A~P, in a concentration dependent manner, directly controls the expression of both $\sigma^F$ and $\sigma^E$ and indirectly controls their activation. The mother-cell specific activation of $\sigma^E$ is therefore under the control of a cascade of two inter-compartmental coherent AND-type feed-forward loops (see Fig. 5.1C). Around the KinA threshold, all the branches of each of the cascades are activated, leading to a highly ultrasensitive increase in the level of active $\sigma^E$. Perturbations of this architecture decrease the the threshold and sensitivity of the sporulation response (Fig. A4.5).

Our results also show that although most cells fail to activate sufficient levels of $\sigma^E$ at sub-threshold KinA levels, many cells express sufficient levels of Spo0A~P to form an asymmetric septum and even activate $\sigma^F$. From a design perspective, this response at sub-threshold levels of KinA appears somewhat wasteful and unnecessary. However, in contrast to the well-controlled conditions of the ASI system, under unpredictable starvation conditions, wild-type cells need to make cell-fate decisions based on noisy information that is encoded into a widely fluctuating Spo0A~P level. To make a timely and accurate decision under such conditions, the cell has to collect information about the environment for a certain period of time to average out the noisy fluctuations of signals and noise of its own biochemical processes. By deferring the decision to a stage downstream of Spo0A~P, the cells gains a time-averaging interval for filtering out noisy fluctuations while retaining responsiveness to environmental signals at the Spo0A~P and phosphorelay levels. It is not surprising then that the sporulation network decision is in fact deferred to the stage of $\sigma^E$ activation. By deferring the
decision to this point of sporulation commitment, cells maximize the time-averaging interval that they can use to filter out noisy signals.

Despite our initial focus on the sporulation decision in the ASI system, we had several reasons to expect that these results apply even to wild-type cells. In particular, we have previously shown that ASI cells under nutrient-rich conditions expressing KinA levels similar to those in the starving wild-type cells lead to efficient sporulation, similar to the starving wild-type cells (129, 279). Furthermore, our results (Fig. 5.2D) regarding the unimodality of the expression of Spo0A-P target genes agrees with previous reports that studied the phosphorelay response of wild-type cells in starvation conditions (263, 264, 279). Taking together all these results, we expect that the ASI system can effectively mimic the natural sporulation network despite differences in KinA regulation and lack of other environmental inputs.

In accordance with these expectations, our experiments with wild-type strain in sporulation conditions are consistent with the major predictions made using the ASI. $\sigma^F$ and $\sigma^E$ activation in single wild-type B. subtilis cells under starvation conditions does indeed increase ultrasensitively at a threshold level of Spo0A activity. Although our measurements indicate that $\sigma^E$ is activated at a higher threshold, it is difficult to accurately determine these thresholds in the wild-type setting due to the large heterogeneity of Spo0A activity. This illustrates the utility of the ASI system since it offers a mode of control over Spo0A activation. Nevertheless, using time-lapse microscopy, we were able to verify the occurrence of cells that activate $\sigma^F$ and yet fail to successfully complete sporulation. Since these cells never engulf the forespore, we concluded that the failure to complete sporulation results from a lack of $\sigma^E$ activity. On the other hand, most of cells that activate $\sigma^E$ were found to complete engulfment and proceed to sporulation. Therefore, even in the wild-type setting, $\sigma^F$ activation alone overestimates sporulation response of the population, albeit far less strikingly than in the ASI at low KinA. Given these ultrasensitive cell-fate determining relationships between activities of Spo0A and $\sigma^F$ and $\sigma^E$, we believe
more detailed comparisons between the ASI and wild-type dynamics in future studies may allow us to uncover how specific environmental signals influence sporulation.

5.4. Materials and Methods

5.4.1. Modeling Methods

To uncover how the increase in KinA expression affects sporulation, we built a detailed mathematical model of the sporulation network based on its known topology. For this, the network was divided into three modules that play essential roles in the early phases of sporulation in a hierarchical order: 1) the phosphorelay that controls Spo0A activation, 2) the $\sigma^F$ activation module in the forespore compartment and 3) the $\sigma^E$ activation module in the mother cell compartment (Fig. 5.1C). We briefly discuss the model here and refer the reader to the Appendix 4 for details.

To study the function of each module, we modeled their input-output functions separately. The input and the output signals for the phosphorelay module can be defined as the level of KinA and the resulting Spo0A$\sim$P level, respectively (Fig. 5.1B). In turn, Spo0A$\sim$P, as the input, governs the forespore-specific activation of $\sigma^F$ by controlling the transcription of the spolI A operon (encoding $\sigma^F$ and its two regulators) and the spolIE gene (encoding the serine phosphatase required for the activation of $\sigma^F$; see Fig. 5.1C)(265). Immediately after forespore formation, Spo0A$\sim$P in the mother cell and active $\sigma^F$ in the forespore act as inputs for the $\sigma^E$ module and thus control expression of the $\sigma^E$ regulon as the output (265, 276). Details of all post-translational interactions, transcriptional regulation as well as relevant parameter values were extracted from the literature (70, 130, 263, 288-290), (265, 270, 278, 291) (269).

We studied both the steady state and the dynamical properties of each module. First we used the ode45 simulator of MATLAB (©MathWorks) to study the deterministic response of each network module. Global parameter sampling
for the phosphorelay was used to determine if steady state Spo0A~P concentrations can increase ultrasensitively between 4 and 10 µM IPTG (see Appendix 4 section on Logarithmic Gains for definition of ultrasensitivity). Subsequently, stochastic simulations for different IPTG doses were performed using the SSA algorithm of the Stochkit package (292). An ensemble of 400 trajectories of the phosphorelay response to each IPTG concentration was used to estimate average response times (Fig. A4.2) and distributions of spolIIA, spolIE and spolIG gene expression levels.

Variability in the expression of these genes combined with the ultrasensitive activation of $\sigma^F$ and $\sigma^E$ produces bimodal distributions of sigma factor activities (see Fig. 5.3B, 5.4B, A4.3 and A4.4). A threshold level of $\sigma^F$ can be used to divide the population into $\sigma^F$ active and $\sigma^F$ inactive fractions. We choose a threshold of 2µM of $\sigma^F$, which lies between the two modes of the $\sigma^F$ distribution to ensure that the predicted fractions of $\sigma^F$-active cells are least sensitive to variations in this threshold. To predict cell fate based on $\sigma^F$, all cells above this threshold are counted as sporulating cells (black line in Fig. 5.3D). Similarly we define 1µM of $\sigma^E$ (between the modes of $\sigma^E$ distributions) as a threshold to robustly predict fractions of active cells. To predict cell fate based on these distributions all cells above the threshold of 1µM of $\sigma^E$ were counted as sporulating cells (black line in Fig. 5.4D). The fraction of sporulating cells predicted based on $\sigma^F$ and $\sigma^E$ activities was subsequently compared to the fraction of sporulating cells calculated from measurements of sporulation efficiencies as described in the next section.

5.4.2. Estimation of fraction of sporulating cells from spore counts

Heat-shock resistant spores counts and total viable cell counts (spores + vegetative cells) at each IPTG dose were estimated as described in experimental methods. Spore counts were measured at least thrice at each IPTG concentration indicated in Fig. 5.1D. To calculate fraction of sporulating cells (as shown Figs. 5.3D and 5.4D), we first compute sporulation efficiency, i.e. the ratio
of spores to viable cells that is observed at the end of the experiment. To calculate the efficiency each spore count was normalized by the total viable cell counts measured at the respective IPTG concentration. We note that total viable cell count varies only ~10% between 0 and 20 μM IPTG whereas spore counts change more than 20-fold. As a result, there is relatively little effect of total cell count variability on sporulation efficiency and the KinA-threshold effect is reproducible. We therefore normalize the three independent measurements of spore counts at each IPTG concentration by the average total viable cell count to determine three measurements of sporulation fraction at each IPTG dose. These calculated sporulation efficiencies can be subsequently used to reliably determine the fraction of sporulating cells.

The fraction of sporulating cells (f) is under-represented in the final viable cell counts since non-sporulating cells can further divide and increase their numbers whereas spores do not divide. To account for this in our estimate of the fraction f, we establish a relation between the sporulation (p) to the actual fraction of cells that chose to sporulate (f). Assuming that the population size was N before nutrients became too scarce for sporulation, the number of spores in the final population is Nf and the number of non-spores is 2^nN(1-f) since the non-spores divide n times. Accordingly the sporulation efficiency (p) is related to the fraction of sporulating cells (f) as:

\[
p = \frac{Nf}{Nf + 2^n N(1-f)} = \frac{f}{f + 2^n (1-f)} \Rightarrow f = \frac{2^n p}{1 + (2^n -1)p}
\]

Where n is the average number of times that a cell that does not sporulate will divide. Based on experimental observations we find that n~2. We use this relationship to calculate the fraction of sporulating cells and compare f to the fraction predicted by our simulations (Fig. 5.3D and 5.4D).
5.4.3. Strain construction

The parental strain for all experiments was \textit{B. subtilis} PY79. Strains used and generated in this study are listed in Table A4.1. Detailed information of Artificial Sporulation Initiation system (ASI, \textit{Phy-spank-kinA}) can be found in the published papers (131, 279). CFP, YFP, mCherry reporter genes were amplified by PCR using the primers listed in Table A4.2. DNAs of pDR201 harboring the gene for mCherry, \textit{B. subtilis} strain BTD217 (\textit{amyE::PspoIIQ-RBSopt-cfp-spoIIQ}) and \textit{B. subtilis} strain BKM1563 (\textit{ycgO::PspoIVF-spoIVFB-yfp}) (all gifts from David Rudner, Harvard Medical School) were used as templates for PCR. Promoter DNA fragments for \textit{spoIIQ} and \textit{spoIID} were amplified from \textit{B. subtilis} PY79 strain using the primers listed in Table A4.2. After restriction enzyme digestion (as listed in Table A4.2), each of the DNA fragments was cloned into either pDG1730 (\textit{amyE} integration vector) (293) or pDR1664 (\textit{thrC} integration vector) (293). Plasmids generated were inserted by double crossover recombination into either the \textit{amyE} or \textit{thrC} locus of the chromosome of \textit{B. subtilis} PY79 strain. \textit{PspoIIA-cfp} (a gift from Jan-Willem Veening, University of Groningen, The Netherlands) was constructed as described (294). GFP reporter strains were published previously (276). Details of the constructions are available upon request.

5.4.4. Media and sporulation conditions

Sporulation of the ASI strain was induced in Luria–Bertani (LB) medium by adding IPTG at the indicated concentrations as described previously (279). Sporulation in the wild-type strain background was induced by the procedure of Sterlini and Mandelstam (295).

5.4.5. Sporulation and β-galactosidase assays

Assays for sporulation and β-galactosidase activity were performed as described previously (279).
5.4.6. Fluorescence and time-lapse microscopy

Fluorescence microscopy was performed as described previously (296). In brief, cells from 0.2 ml of culture were centrifuged and suspended in 50 μl of 1 x PBS. The concentrated cell suspension (2 μl) was placed on a 1% (w/v) agarose pads containing 1 x PBS within a 25 μl Gene Frame (AB Gene, UK) covered by a clean microscope slide cover slip and was viewed using an Olympus BL51 fluorescence microscope with Slidebook software (Intelligent Imaging Innovations). For Figs. 5.3C and 5.4C, fluorescence measurements of σF and σE activities at 4 and 10 μM IPTG were made for multiple colonies (5 colonies for σF and 4 colonies for σE; each colony had ~100 cells). These fluorescence measurements were used along with appropriate thresholds to count the number of cells that are sigma-active in each colony. The mean and standard deviation of the fractions are reported as percentages. See Fig. A4.3F and A4.4E for histograms of sigma factor activities in individual colonies.

Time-lapse microscopy was performed by the procedure of de Jong et al. (297) using the fluorescence microscope with a temperature controlled chamber from Solent Scientific (UK). In brief, sporulating cells incubated for 1 h at 37°C in the medium of Sterlini and Mandelstam was spotted on a 1% (w/v) agarose pads containing the identical medium within a 25 μl Gene Frame (AB Gene, UK) and covered by a clean microscope slide cover slip. The prepared slide glass was placed in the pre-warmed (37°C) environmental chamber of the microscope and monitor single cells over time. Intensities of fluorescent proteins (arbitrary units/pixel) in cells were analyzed electronically by the software Slidebook (Intelligent Imaging Innovations, Inc.) and custom MATLAB® code.

Notes

This chapter is based on work that was published in the following article:

Chapter 6

The importance of proper Spo0A dynamics in sporulation

6.1. Introduction

Upon nutrient starvation, a majority of *Bacillus subtilis* cells differentiate to produce spores. Extensive studies have elucidated a detailed genetic network of genes involved in an early stage of sporulation (70, 298-300). At the heart of this network is a phosphorelay (Fig. 6.1A), in which phosphate is transferred from multiple histidine protein kinases (KinA-E) to a master transcription regulator Spo0A through two intermediate phosphotransferases (Spo0F and Spo0B) (70, 301-303). Transcriptionally, the phosphorelay genes encoding KinA, Spo0F, and Spo0A are controlled by the phosphorylated form of Spo0A (Spo0A–P) directly and indirectly via multiple feedback loops (Fig. 6.1B). This complex network appears to be the integration point for putative extracellular and intracellular signals that trigger entry into the sporulation program in response to starvation (298, 299).
The initial phase of starvation is marked by a gradual increase in the level of Spo0A~P, which triggers the expression of many genes essential for sporulation (130, 131, 276). These observations are the motivation for the long-standing hypothesis that sporulation cell-fate can be predicted based on whether or not a cell has a threshold level of Spo0A~P (298, 299). We have recently demonstrated that such a Spo0A~P threshold is justified considering the ultrasensitive (switch-like) response of a cascade of feed-forward loops downstream of Spo0A~P (304). However, this Spo0A~P threshold model of cell-fate determination does not explain the experimental observation that rapid accumulation of active Spo0A actually has an adverse effect on sporulation (130, 131). These studies raise two important questions about the mechanistic connection between Spo0A~P dynamics and sporulation cell-fate: (1) Why does accelerated Spo0A~P accumulation impair sporulation efficiency?, and (2) How does the design of the sporulation phosphorelay ensure proper Spo0A~P accumulation dynamics? In this study we took a synthetic biology approach to answer these questions.

To control the temporal dynamics of Spo0A~P accumulation, in this study we employed a system in which the kinase KinC is expressed from an IPTG-inducible promoter and Spo0A is expressed from a xylose-inducible promoter. Since KinC can transfer phosphate to Spo0A directly (305, 306), both the level of Spo0A and its phosphorylation rate can be easily modulated externally in this system. Using this system, we first determined the relationship between sporulation efficiency and the level of expression of KinC and Spo0A. Surprisingly, we found that sporulation efficiency depends non-monotonically on the levels of these proteins despite the fact that Spo0A activity increases monotonically with increases in Spo0A and KinC.

Analyzing these results within the framework of a mathematical model, we found that only a mechanism coupling the dynamics of Spo0A activation to the cell-fate can explain the impairment of sporulation at high levels of KinC/Spo0A.
Further, using a combination of modeling and experiments, we were able to predict and confirm the specific biochemical mechanism underlying the adverse effect of accelerated Spo0A~P accumulation on sporulation in our artificial system. Finally, we modified our inducible KinC-Spo0A and demonstrated that the positive transcriptional feedback in the wild-type phosphorelay is sufficient to ensure proper temporal accumulation of Spo0A~P and efficient sporulation.

Figure 6.1. The sporulation network phosphorelay.

Solid arrows and T-shaped bars indicate positive and negative gene regulatory interactions, respectively. Dashed arrows indicate phosphotransfer interactions.
A. Post-translational interactions in the wild-type phosphorelay regulatory network. Sensor histidine kinases (KinA and KinC) autophosphorylate and provide phosphate to activate the master regulator Spo0A (0A). KinA transfers phosphate to 0A via Spo0F (0F) and Spo0B (0B), whereas KinC can transfer phosphate to 0A directly.
B. Transcriptional regulatory interactions. Phosphorylated Spo0A (0A~P) becomes a positive regulator for sporulation genes, including those for 0A itself and 0F. 0A~P also indirectly activates the expression of the gene for $\sigma^H$, which is essential for sporulation, by repressing AbrB, the repressor of $\sigma^H$. 
6.2. Results

6.2.1. KinC-Spo0A artificial two-component system triggers entry into sporulation

To investigate the importance of the dynamics of Spo0A~P accumulation, we perturbed these dynamics artificially. To control the dynamics of accumulation of Spo0A~P we needed to directly control both the level of Spo0A and its phosphorylation rate. To this end, we constructed a strain in which genes for the kinase KinC and Spo0A are independently placed under the control of IPTG-inducible (P_{hy-spank} for KinC) or the xylose inducible (P_{xylA} for Spo0A) promoters (KinC-Spo0A strain hereafter) (Fig. 6.2A). We chose KinC (instead of other kinases, such as KinA, etc.) for two reasons. First, it has been previously shown that induced expression of KinC can be used to achieve high sporulation efficiencies irrespective of culture conditions (131); see also Supplementary Table A5.1). Second, since KinC can directly phosphorylate Spo0A, thus bypassing the need for Spo0F and Spo0B (305, 306), it offers an easy and direct method of modulating the rate of Spo0A phosphorylation. We also introduced a spo0F deletion into this KinC-Spo0A-inducible strain to ensure direct phosphotransfer from KinC to Spo0A. By restricting phosphate flow in this manner and conducting experiments in rich media we were able to minimize the effects of starvation related signals and modulators acting on the phosphorelay.

Using this system we evaluated the effect of different concentrations of IPTG and Xylose (that control the synthesis of KinC and Spo0A, respectively) on sporulation efficiency in nutrient-rich conditions. To this end we defined and measured the fraction of total colony-forming units (total CFU=viable cells and spores) that were heat-resistant (heat-resistant CFU=spores; see Experimental procedures for more details). We found that in the double-induction strain sporulation efficiency depends non-monotonically on the levels of both IPTG and xylose (Fig. 6.2B): sporulation efficiency was low in the absence of inducers and increased to reach a maximum of about 30% at the optimal combination of 10μM
IPTG and 0.04% xylose (see Fig. 6.2B). This sporulation efficiency is comparable to that observed in the wild-type cells in sporulation medium (50-70%, Table A5.1). However, our results show that sporulation was impaired by further increasing either the concentration of IPTG or xylose (Fig. 6.2B and Table A5.2). Going from 10μM IPTG to 500μM while keeping xylose at the optimal level of 0.04% decreases sporulation efficiency to ~9%. Similarly, going from 0.04% xylose to 1% xylose while keeping KinC induction at the optimal level (at 10μM IPTG) decreases efficiency to ~5%. Moreover, we found that simultaneously overexpressing both KinC and Spo0A by adding excess amounts of two inducers resulted in even greater impairment of sporulation (sporulation efficiency ~0.1% at 500μM IPTG and 1% xylose – Fig. 6.2B and Table A5.2). Thus, we established an artificial two-component sporulation system (harboring independently inducible copies KinC and Spo0A) and found that only an optimal combination of Spo0A and KinC expression can be used to trigger an efficient entry into sporulation.
Figure 6.2. Simultaneous overproduction of KinC and Spo0A reduces sporulation efficiency.

A. The KinC-Spo0A artificial-two component system (strain MF4317).

B. Bars indicate the results of sporulation efficiency measurements at various combinations of IPTG and xylose inducer concentrations (see Methods). The black bar indicates the optimal combination of IPTG and Xylose for sporulation in this strain.

C-D PspoIIG-lacZ reporter was used to measure the effects of KinC and 0A induction on 0A~P levels in the KinC-Spo0A strain. (C) Induction of KinC expression from a Phy-spank promoter using IPTG leads to a monotonic increase in 0A~P levels (dots and errorbars show the mean and standard deviation of 3 independent measurements of β-galactosidase activity in Miller units). (D) Induction of Spo0A expression from a PxylA promoter using xylose leads to a similar monotonic increase in 0A~P levels (dots). Solid lines in (C) and (D) show Hill-equation fits to the measured responses. Insets in (C) and (D) show the same data on a log-scale for inducer concentrations.
To understand the non-monotonic dependence of sporulation efficiency on IPTG and xylose levels, we quantified the cellular levels of KinC and Spo0A in the KinC-Spo0A strain. We constructed strains expressing a functional KinC-GFP and the wild-type Spo0A under these two inducible promoters, respectively. Using this strain, we quantified the amount of KinC and Spo0A proteins at 2h after the addition of inducers with immunoblot analysis using anti-GFP and anti-Spo0A antibodies, respectively. For comparison, the exponentially growing and sporulating cells expressing KinC-GFP from native promoter were cultured in LB and SM media, respectively, and the cell extracts were prepared and quantified with immunoblotting. As shown in Fig. A5.1A, at 10µM IPTG concentration KinC-GFP levels were approximately 5-fold higher than those in the wild-type strain cultured either in rich (LB) or sporulation (SM) conditions. The Spo0A level at 0.04% xylose was approximately 2-fold higher than that in the sporulating wild-type strain (Fig. A5.1B). At higher concentrations of the inducers (500µM IPTG and 1% xylose), the synthesis of KinC and Spo0A proteins was increased further (Fig. A5.1AB). Thus, the conditions that trigger efficient sporulation correspond to overexpression of both proteins and therefore may not be relevant for wild-type sporulation induction. Nevertheless, the double induction system can enable insights into how level and dynamics of Spo0A~P affect sporulation.

Considering the above results, the low sporulation efficiency seen in our experiments at low IPTG and xylose concentrations (<10µM IPTG and <0.04% xylose) was expected and may be explained by insufficient expression of KinC and/or Spo0A for sporulation. However, the decrease in sporulation efficiency at high inducer concentrations (>10µM IPTG and >0.04% xylose) was somewhat surprising. One possible explanation for this negative effect is that overexpression of KinC and Spo0A with excess IPTG and xylose (>10µM IPTG and >0.04% xylose) may somehow result in a decrease of Spo0A activity, thereby impairing sporulation. To test this, we used the P_{spolIG} lacZ reporter and determined Spo0A activity in this strain at different inducer concentrations. As shown in Fig. 6.2CD, in the KinC-Spo0A system, β-galactosidase activity
increases monotonically with IPTG and xylose concentrations. Furthermore, we found that the Spo0A activity is higher under conditions of simultaneous overexpression of both KinC and Spo0A (500μM and 1% xylose), as compared to the optimal conditions for sporulation (10μM IPTG and 0.04% xylose; see Table A5.2). Therefore, the impairment of sporulation by the induced overexpression of KinC and Spo0A is not due to a decrease in Spo0A activity.

6.2.2. The role of 0A~P accumulation kinetics in efficient sporulation

To understand how simultaneous overproduction of both KinC and Spo0A impairs sporulation, despite an increase in Spo0A activity, we constructed a mathematical model of the KinC-Spo0A artificial two-component system. In our model the parameters for the expression of KinC and Spo0A from the inducible promoters P_{hy-spank} and P_{xylA} respectively, were constrained using immunoblotting measurements of KinC and Spo0A protein levels and the single-cell measurements of transcriptional reporters P_{hy-spank-yfp} and P_{xyl-cfp} (see Supplementary Methods and Supplementary Fig. A5.2). The post-translational interactions in the model include KinC autophosphorylation, direct phosphorylation of Spo0A by KinC and dephosphorylation of Spo0A by the Spo0E phosphatase.

The results of deterministic simulations of our model showed that increasing the concentration of IPTG from 10 to 500μM to increase KinC levels accelerates the accumulation of Spo0A~P (Fig. 6.3A-C). The modeling results also showed that increasing Spo0A levels by increasing xylose concentration from 0.04% to 1% leads to a similar acceleration in Spo0A~P dynamics (Fig. A5.3).
Figure 6.3. Dynamic Threshold model explains the dependence of sporulation on IPTG and xylose.

**A-C** Dynamics of accumulation of Spo0A–P in the KinC-Spo0A strain at 0.04% xylose and varying IPTG concentrations. Spo0A–P accumulation is sped up by increase in IPTG (KinC expression). Green regions define the Spo0A–P threshold models. (A) Single Threshold: [Spo0A–P]>0.91µM at T5. (B) Spo0A–P Range: 0.91<[Spo0A–P]<1.1 at T5. (C) Dynamic Threshold: 0.91<[Spo0A–P]<1.1 at T5 and 0.46<[Spo0A–P]<0.55 at T2.

**D-F** Range of IPTG concentrations at for which cells can sporulate as function of the xylose concentration is indicated by green regions. These regions are calculated using Single Threshold (D), Spo0A–P Range (E) and Dynamic Threshold (F) mechanisms respectively. The orange and yellow regions in (F) correspond to the range of IPTG satisfying the constraints 0.91<[Spo0A–P]<1.1 at T5 and 0.46<[Spo0A–P]<0.55 at T2 respectively.

We combined these results with phenomenological thresholds for Spo0A–P levels to determine how the decrease in sporulation during simultaneous overproduction of KinC and Spo0A can be explained. Three types of Spo0A–P threshold mechanisms were used:
1. **Single Threshold**: For sporulating cells, Spo0A~P concentration at steady state must be above some threshold level \(a_1\) (our simulations indicate that Spo0A~P reaches steady state at 5 hours after induction, hereafter T5), i.e. \(a_1<[\text{Spo0A}\sim\text{P}]\) at T5.

2. **Spo0A~P Range**: For sporulating cells, Spo0A~P concentration at steady state must be greater than some level \(a_1\) but below another threshold \(a_2\), i.e. \(a_1<[\text{Spo0A}\sim\text{P}]<a_2\) at T5.

3. **Dynamic Thresholds**: For sporulating cells, the steady-state (T5) concentration of Spo0A~P must be greater than \(a_1\) but less than \(a_2\) and greater than \(b_1\) but less than \(b_2\) at the point of asymmetric septation (T2, 2 hours after induction in our conditions), i.e. \(a_1<[\text{Spo0A}\sim\text{P}]<a_2\) at T5 and \(b_1<[\text{Spo0A}\sim\text{P}]<b_2\) at T2.

These thresholding mechanisms, although phenomenological, represent different cell-fate decision mechanisms and can be used to discriminate between different hypotheses regarding the mechanistic basis of decisions during sporulation. The Single Threshold mechanism (as above) is the traditional model used to predict sporulation outcomes. It assumes that cells only need to achieve a minimum Spo0A~P level to sporulate. The dynamics of Spo0A~P accumulation do not affect the sporulation outcome in this case. The dynamics are also irrelevant to the outcome of the Spo0A~P Range mechanism (as above). However, under this thresholding mechanism, excess Spo0A~P at steady state can prevent sporulation. Finally, in the Dynamic Threshold mechanism, sporulation decision depends on Spo0A~P levels at two time-points. As a result, cell-fate is determined not only by the Spo0A~P level achieved but also by the rate of accumulation of Spo0A~P. We applied these thresholding mechanisms to the results of our deterministic simulations to determine which combination of IPTG and xylose concentrations would allow sporulation. As shown in Fig. 6.3D-F, these three thresholding mechanisms make qualitatively distinct predictions about the range of inducer concentrations that allow sporulation (green regions indicate inducer concentrations where cells
sporulate according to criteria listed above). The Single Threshold mechanism (#1) predicts that there exists a minimum IPTG and xylose concentration for successful sporulation and any IPTG and xylose concentration above these minimum levels leads to sporulation (see green region in Fig. 6.3D). The Spo0A~P Range mechanism (#2) predicts that there exists a range of IPTG concentrations for successful sporulation and that this range depends on the level of xylose (green region in Fig. 6.3E). The Dynamic Thresholds mechanism (#3) combines two such xylose dependent IPTG ranges (orange and yellow regions indicate inducer concentrations where $a_1<[\text{Spo0A}\sim P]<a_2$ at T5 and $b_1<[\text{Spo0A}\sim P]<b_2$ at T2 respectively; Fig. 6.3F). As a result, the dynamic thresholding mechanism (#3) predicts that there is an optimal combination of IPTG-xylose concentrations for triggering sporulation efficiently. We note that the regions of inducer concentrations where sporulation can occur (green regions in Figs. 6.3D-F) are distinctly shaped for the three models. These shapes are robust to the exact values of the thresholds used or of the time-points these are applied.

Comparing these results with Fig. 6.2B we concluded that only the Dynamic Thresholds mechanism (#3) can explain the experimentally-demonstrated relationship between sporulation efficiency and the inducer concentrations (compare Fig. 6.3F and Fig. 6.2B). Therefore, we concluded that entry into sporulation depends on both the Spo0A~P level and the dynamics of Spo0A~P accumulation in a fashion similar to that suggested by the Dynamic Thresholds mechanism (#3). In addition, these results also suggest that the impairment of sporulation by overproduction of KinC and Spo0A is the result of premature activation of Spo0A, as trajectories corresponding to these conditions overshoot the T2 threshold from above (Fig. 6.3C). We conclude that premature induction of high levels of Spo0A at the earlier time-point may impair sporulation, and began to explore the possible molecular mechanism(s) underlying this effect.
6.2.3. Accelerated accumulation of 0A\textasciitilde P perturbs the temporal order of 0A regulon expression

The Dynamic Thresholds mechanism used in the previous section explains the effect of overproduction of KinC and Spo0A by assuming that accelerated Spo0A\textasciitilde P accumulation impairs sporulation. To justify this assumption and validate The Dynamic Thresholds mechanism mechanistically, we examined the effects of accelerated Spo0A\textasciitilde P accumulation on the induction and repression of genes in the Spo0A regulon.

Based on the known network architecture of the Spo0A regulon, we predicted two consequences of the accelerated accumulation of Spo0A\textasciitilde P: 1) no effect on the indirectly controlled, low Spo0A\textasciitilde P threshold genes (Fig. 6.4A, left panel); 2) premature activation/repression of the directly controlled high Spo0A\textasciitilde P threshold genes (Fig. 6.4A, right panel). The first consequence mainly refers to genes that are indirectly activated by Spo0A\textasciitilde P through the repression of AbrB and the resulting derepression of $\sigma^H$ (130). The kinetics of activation of these genes is determined mainly by dilution and degradation rate of AbrB and thus Spo0A\textasciitilde P is not rate-limiting. Therefore, the indirectly controlled low-threshold genes in the Spo0A regulon are independent of Spo0A\textasciitilde P accumulation rate. By contrast, in the latter case, we expected that the accelerated accumulation of Spo0A\textasciitilde P could activate/repress the Spo0A-directly controlled high-threshold genes in a premature fashion, resulting in the abnormal control of low threshold (indirect) and high threshold (direct) events in the sporulation gene expression program.

To test the above predictions, we extended our mathematical model of the KinC-Spo0A system from the previous section to include Spo0A regulated genes. In this extended model, two types of Spo0A\textasciitilde P regulated genes were included: 1) directly activated/repressed high threshold genes: spollGB (the second gene in a spollG operon encoding pro-$\sigma^E$, the precursor of $\sigma^E$, hereafter SpollGB) and
divIVA (encoding DivIVA involved in cell division site selection), and 2) an indirectly activated low threshold gene citG (encoding $\sigma^H$-controlled fumarate hydratase gene) (Fig. 6.4A) (130, 303). It should be noted that even though we use these specific Spo0A~P target genes, our results reflect the effects of perturbations on the temporal expression of low- and high-threshold genes in general.

Figure 6.4. Rapid Spo0A~P accumulation reverses the temporal order of low and high threshold genes.

A. Two types of gene expression control exerted by Spo0A~P: direct activation/repression of high threshold genes (spoIIG/divIVA) and indirect activation of genes like $\sigma^H$ by relieving repression by AbrB.

B. Accumulation dynamics of directly- and indirectly-activated Spo0A target genes (SpoIIGB-black curves, $\sigma^H$-gray curves) in the KinC-Spo0A strain as predicted by our model. Two different conditions are shown: 10µM IPTG, 0.04% Xylose (solid) and 500µM IPTG, 1% Xylose (dashed). Note that higher inducer concentrations shift the onset of [SpoIIGB] accumulation to an earlier time-point but do not significantly affect [$\sigma^H$] accumulation.
C-D Accumulation of $\sigma^H$ (C) and transcription of $spolIG$ (D) in the KinC-Spo0A strain were measured every 30mins after induction using $P_{citG}\cdot lacZ$ and $P_{spoIIG}\cdot lacZ$ reporters, respectively. In both (C) and (D), points on dashed and solid curves indicate measurements for 10µM IPTG, 0.04% xylose and 500µM IPTG, and 1% xylose, respectively. β-galactosidase activity (Miller units) was normalized to the level at T2 (C) and T0 (D) respectively after subtracting appropriate background.

E. Predicted dynamics of a directly-repressed Spo0A target DivIVA in the KinC-Spo0A strain under conditions of no inducer (solid curve), 10µM IPTG, 0.04% xylose (dotted curve) and 500µM IPTG, and 1% xylose (dashed curve).

F. Level of DivIVA was assayed with immunoblotting in the KinC-Spo0A strain expressing a DivIVA-CFP fusion protein (MF4812) at 0µM IPTG, 0% xylose (circles), 10µM IPTG, 0.04% xylose (diamonds) and 500µM IPTG, 1% xylose (squares) at indicated time-points (see Fig. A5.4). DivIVA-CFP measurements were divided by the immunoblot measurements of constitutively expressed $\sigma^A$ and then normalized by the T0 values in each case. Error bars show the standard deviation of 3 independent measurements at each time-point.

Results of deterministic simulations of this model showed that simultaneous overproduction of KinC and Spo0A by increasing inducer concentrations to 500µM IPTG and 1% xylose (solid curves Fig. 6.4B) from 10µM IPTG and 0.04% xylose (dashed curves Fig. 6.4B) accelerates the accumulation of Spo0A~P, resulting in the early expression of the Spo0A-directly activated $spolIGB$ gene (black curves in Fig. 6.4B). Furthermore, our simulations show that the increases in both KinC and Spo0A have no significant effect on the response profile of the $\sigma^H$-regulated $citG$ gene that is indirectly regulated by the low threshold level of Spo0A~P (Fig. 6.4B - gray curves).

To verify these modeling results, we experimentally measured gene expression profiles of the low- and high-threshold Spo0A-regulated genes in the KinC-Spo0A strain. For the measurement of the expression of the low-threshold Spo0A indirectly-regulated and the high-threshold Spo0A directly-controlled genes, β-galactosidase activities from a $\sigma^H$ controlled $citG$ P2 promoter (308) and
a Spo0A controlled spollG operon promoter (130) fused to the lacZ gene (P_{citG-lacZ} and P_{spollG-lacZ}) were separately assayed. The results indicated no significant effect on the response profile of the σ^H-regulated gene expression (Fig. 6.4C) despite the accelerated accumulation of Spo0A~P, which is evident from the premature activation of the high threshold P_{spollG-lacZ} reporter (Fig. 6.4D). We also found that our model also predicted that accelerated Spo0A~P accumulation resulting from overproduction of KinC and Spo0A causes the earlier onset of reduction in the expression of high threshold Spo0A-directly repressed genes such as divIVA (Fig. 6.4E). To verify this experimentally we measured the protein level of endogenous DivIVA using immunoblotting for a functional DivIVA-CFP fusion protein (309, 310) in the KinC-Spo0A strain (MF4812). We found that, in contrast to the optimum conditions of 10µM IPTG and 0.04% xylose (dotted curve in Fig. 6.4F), DivIVA expression at T3 is significantly reduced under conditions of overproduction of KinC and Spo0A (500µM IPTG and 1% xylose; see dashed curve in Fig. 6.4F and Fig. A5.4).

We conclude that these experimental results agree with the mathematical modeling, thus confirming our prediction that overproduction of KinC and Spo0A reverses the regular temporal order of expression of the Spo0A regulon: spollG expression increases and DivIVA is repressed significantly before σ^H activation.

These results can be used to provide a mechanistic basis for the Dynamic Thresholds mechanism. As indicated in the previous section, according to the Dynamic Thresholds mechanism, sporulating cells must have a_1<[Spo0A~P]<a_2 at T5 and b_1<[Spo0A~P]<b_2 at T2. The minimum Spo0A~P thresholds a_1 and b_1 can be explained by the requirement for the high-threshold Spo0A~P-activated genes like spollGB. The Dynamic Threshold mechanism explains that there is a sufficient level of Spo0A~P for the activation of the essential downstream master-regulatory factors such as σ^E (spollGB product). Accordingly, the minimum Spo0A~P threshold may be replaced by the requirements c_1<[SpollGB] at T5 and d_1<[SpollGB] at T2. Furthermore, if we assume that there is a minimum
threshold concentration for the high-threshold Spo0A~P repressed genes like \textit{divIVA}, the maximum Spo0A~P thresholds \(a_2\) and \(b_2\) must be available during the early phase of sporulation. Thus, the upper bounds on Spo0A~P threshold can be replaced by lower bounds on DivIVA as: \(c_2<\text{[DivIVA]}\) at T5 and \(d_2<\text{[DivIVA]}\) at T2. Moreover since \(\sigma^H\) accumulation rate is robust to changes in Spo0A~P accumulation kinetics, \(\sigma^H\) can be used as a time-keeping device in this strain. As a result, evaluating the requirements of \(d_1<\text{[SpoIIG]}\) and \(d_2<\text{[DivIVA]}\) at a specific concentration of \(\sigma^H\) is equivalent to evaluating these requirements at a fixed time-point. Taken together, the three types of Spo0A~P thresholding mechanisms can then be reformulated as: Single Threshold: \([\text{SpoIIGB}]>c_1\) at T5 (steady state); Spo0A~P Range: \([\text{SpoIIGB}]>c_1\) and \([\text{DivIVA}]>c_2\) at T5; and Dynamic Thresholds: \([\text{SpoIIGB}]>c_1\) and \([\text{DivIVA}]>c_2\) at T5 and \([\text{SpoIIGB}]>d_1\) and \([\text{DivIVA}]>d_2\) when \([\sigma^H]=s_0\).

To validate these mechanisms, we performed stochastic simulations of the KinC-Spo0A model using the Gillespie SSA algorithm. In the stochastic simulations of our model, cellular responses at the single-cell level are characterized by noisy trajectories. To separate these trajectories into sporulating and non-sporulating phenotypes, we apply the three threshold models described above and calculate the fraction of cells that sporulate at each inducer condition. The values of the thresholds \(c_1, c_2, d_1\) and \(d_2\) were chosen by optimizing the fit of the Dynamic Thresholds mechanism to the experimentally measured sporulation efficiencies (see Fig. 6.5 and Appendix 5).

As expected from the results in Fig. 6.3, our stochastic simulations show that only the Dynamic Thresholds mechanism predicts that the fraction of sporulating cells is maximal at 10μM IPTG and 0.04% xylose (compare Fig. 6.5A-C and Fig. 6.2B). Moreover, the sporulating fraction predicted by the Dynamic thresholds mechanism matches the experimentally observed non-monotonic dependence of sporulation efficiency on IPTG and xylose concentrations (Fig. 6.5D-E). The combination of mathematical modeling and experimental results
suggests that a minimal threshold concentration of each of the proteins at two different time-points is required for a successful entry into sporulation.

Figure 6.5. Dynamic Threshold model explains the existence of optimal IPTG and xylose concentrations for sporulation.

A-C Contour diagrams of fraction of sporulating cells as predicted by using Single Threshold (A), Spo0A~P Range (B), and Dynamic Thresholds (C) mechanisms on stochastic simulation results. Single Threshold: [IIGB]>13.5µM at T5. Spo0A~P Range: [IIGB]>13.5µM and [DivIVA]>0.84µM at T5. Dynamic Threshold: [IIGB]>13.5µM and [DivIVA]>0.84µM at T5 and [IIGB]>2µM and [DivIVA]>0.97µM when [σH]=0.8. Note that only the Dynamic thresholds mechanism predicts that the sporulating fraction is optimum at 10µM IPTG and 0.04% xylose.

D-E Comparison of model predicted sporulating fraction and experimentally measured sporulation efficiencies. Bars indicate the results of sporulation efficiency measurements at various combinations of IPTG and xylose inducer concentrations (same data as Fig. 2.2.2B). Xylose=0.04% in (D) and IPTG=10µM in (E). The lines show the fraction of sporulating cells predicted by using the Single Threshold (red), Spo0A~P Range (blue), and Dynamic Threshold (black) on the trajectories from stochastic simulations of the KinC-Spo0A model.
6.2.4. Accelerated 0A-P accumulation leads to premature repression of DivIVA which impairs chromosome segregation

The thresholds for DivIVA can be explained by the essential requirement for this protein during sporulation. In the early phase of sporulation, accumulation of $\sigma^H$ leads to asymmetric septation by promoting increased expression of FtsZ (311). DivIVA plays an important role during sporulation since it is essential for proper chromosome segregation; divIVA mutants have impaired chromosome segregation, which leads to formation of chromosome-free compartments and poor sporulation efficiency (312, 313). Based on these facts, we reasoned that early repression of DivIVA in conditions of KinC and Spo0A overexpression may mimic the phenotype of divIVA mutants, resulting in the low sporulation efficiency due to lack of the proper chromosome segregation during sporulation.

To evaluate the effect of Spo0A and KinC overexpression on DivIVA levels and on chromosome segregation in single cells, we performed fluorescence microscopy experiments with a functional DivIVA-CFP fusion construct (309). We found that DivIVA-CFP levels in single cells were significantly lower at 500µM IPTG and 1% xylose compared to 10µM IPTG and 0.04% xylose (Fig. 6.6A). Moreover, DivIVA-CFP was irregularly localized in the cells under the high inducer conditions (Fig. A5.5). Further, using DAPI (4-,6-diamidino-2-phenylindole dihydrochloride) to stain for DNA, we found that DNA was improperly segregated more frequently at 500µM IPTG and 1% xylose compared to 10µM IPTG and 0.04% xylose (Fig. 6.6B-G and Fig. A5.5 and A5.6).

Our results suggest that low sporulation efficiency under conditions of overproduction of KinC and Spo0A in the artificial two-component system can be explained at least in part by the improper chromosome segregation resulting from premature repression of DivIVA.
Figure 6.6. Overproduction of KinC and Spo0A results in early repression of DivIVA and improper chromosome segregation.

A. Histograms for DivIVA levels at T2 after the addition of 0.04% xylose and 10μM IPTG (black) and 1% xylose and 500μM IPTG (red). DivIVA level in single cells was measured by expressing a functional DivIVA-CFP fusion in the KinC-Spo0A strain. DivIVA-CFP fluorescence levels were normalized to the maximum level seen in both conditions.

B-G The KinC-Spo0A strain harboring both a gfp reporter gene under the control of the spoIIG promoter and stained with DAPI for DNA was cultured in the presence of (B-D) 0.04% xylose and 10μM IPTG; and (E-G) 1% xylose and 500μM IPTG and examined with fluorescence microscopy. Typical cells under each condition are shown. GFP and DAPI (DNA) are pseudo-colored with green and magenta, respectively. Yellow carats indicate cells with improper chromosome segregation. PC, phase contrast image. Scale bar: 2μm.
6.2.5. Autoregulation of Spo0A ensures proper temporal coordination of high and low threshold gene expression

Thus far, our results show that the impairment of sporulation by rapid accumulation of Spo0A~P can be explained by the premature repression of its targets such as DivIVA before $\sigma^H$ is fully activated. We also noted that in the wild-type phosphorelay the $\sigma^H$-mediated positive feedback to $spo0A$ may limit the accumulation of Spo0A~P until $\sigma^H$ is fully activated. Taking all this into account we hypothesized that putting $spo0A$ under native promoter may be sufficient to ensure proper Spo0A~P dynamics and prevent premature activation/repression of the Spo0A regulon.

To test this hypothesis, we first constructed a mathematical model for a strain where KinC is expressed from the IPTG inducible P_{hy-spank} promoter, $spo0A$ is regulated by its native promoters, and the phosphotransferase $spo0B$ is deleted to ensure only direct phosphotransfer from KinC to Spo0A ($\Delta spo0B$, Fig. 6.7A). The $spo0A$ expression is regulated by multiple Spo0A binding sites that regulate to both a $\sigma^A$-dependent and a $\sigma^H$-dependent promoters (288). Thus native $spo0A$ regulation results in direct feedback from Spo0A to itself (both positive and negative loops) and indirect positive feedback via $\sigma^H$. Deterministic simulations of this model showed that increase in KinC expression can lead to earlier onset of the high Spo0A~P threshold program ($spolIGB$ activation and divIVA repression; Fig. A5.7), similar to the KinC-Spo0A strain. However, the autoregulation of $spo0A$ expression ensures that this onset does not precede the activation of $\sigma^H$ (Fig. A5.7). Consequently, sufficient levels of high threshold repressed genes like DivIVA are always available at the appropriate time-point in the sporulation program (Fig. A5.7).

The results of stochastic simulations (Fig. 6.7B) for this strain showed that the fraction of sporulating cells predicted based on the Dynamic Thresholds mechanism (black dashed line) increases monotonically with IPTG concentration.
These predictions are very similar to the fraction of sporulating cells computed based on the Single Threshold mechanism (gray solid line). Therefore, the fraction of sporulating cells in this strain is mainly limited by the requirements for activation of high-threshold genes (such as spoIIGB). The increase in sporulation efficiency at higher IPTG concentrations can then be explained as more cells achieving sufficiently high Spo0A~P level (Fig. 6.7B and Fig. A5.7). Moreover, $\sigma^H$ accumulation always precedes the repression of the high threshold genes (such as DivIVA), and as a result, sufficient levels of these proteins are available for normal cellular functions ([DivIVA] $\geq$ 0.97µM when [$\sigma^H$]=0.8µM; see Fig. 6.7B and Fig. A5.7).

Figure 6.7. $\sigma^H$-mediated feedback prevents decrease in sporulation efficiency at high KinC levels.

A. KinC induction strain with spo0B deletion and spo0A expressed from native promoter (MF4419).
B. Predicted fraction of sporulating cells as a function of IPTG in this strain. Fraction of cells that sporulate in this strain was calculated using the results of stochastic simulations at different IPTG concentrations and the Single threshold (gray curve) and Dynamic Thresholds (black dashed curve) models. The [SpoIIGB] and [DivIVA] thresholds used are the same as Fig. 6.5. Predictions based on the Dynamic Thresholds model are largely equivalent to predictions based on a single [SpoIIGB] threshold (Single Threshold model) for this strain.

C. Results of sporulation efficiency measurements at various concentrations of IPTG in the KinC induction strain harboring spo0B deletion mutation (MF4419). Note that sporulation efficiency increases with IPTG.

Next we computationally examined the roles of the individual autoregulatory feedbacks in Spo0A~P accumulation by removing either the negative direct feedbacks or both positive and negative direct feedbacks from Spo0A to itself. Our results showed that the indirect $\sigma^H$-mediated positive feedback to spo0A alone is capable of ensuring the gradual accumulation of sufficient levels of Spo0A~P required for sporulation (Fig. A5.7). These results agree with the observations of Chastanet and Losick (2011) who showed that the elimination of all Spo0A binding sites from the spo0A promoters does not hamper sporulation.

To verify these predictions, we experimentally constructed this strain (MF4419, Fig. 6.7A) and examined how sporulation depends on the level of KinC. We tested for sporulation efficiency in LB media in the presence of varying concentrations of IPTG. As expected from our modeling results, we found that with spo0A under its native promoter, sporulation efficiency at 500μM IPTG was significantly higher ($9 \times 10^{-1}$) than that in the KinC-Spo0A strain ($10^{-2}$) at high 500μM IPTG and 1% xylose (compare Fig. 6.7C and Table A5.2).

Taken together these results suggest that $\sigma^H$-mediated positive feedback to spo0A transcription in the sporulation phosphorelay is sufficient to ensure
proper timing of activation/repression of high-threshold genes with the sporulation program.

6.3. Discussion

We have constructed a genetically engineered strain that decouples and rewires the sporulation phosphorelay network thereby establishing an artificial two-component system of sporulation comprising KinC and Spo0A. In contrast to the previously reported systems (129, 131, 314), our artificial two-component system offers unique control over the kinetics of Spo0A~P accumulation and thus is a powerful tool to investigate why Spo0A~P accumulation needs to be gradual for efficient sporulation.

We found that at a certain combination of Spo0A and KinC protein expression levels (10µM IPTG and 0.04% xylose, Fig. 6.2B), Spo0A~P gradually accumulated to high enough levels so that cells sporulate efficiently. However, increases in KinC or Spo0A expression beyond this condition can significantly speed-up the rate of Spo0A~P accumulation, shift the onset of high threshold gene expression to earlier time-points, and as a result, decrease sporulation efficiency (500µM IPTG and/or 1% xylose, Fig. 6.2B).

These experimental results can be explained using mathematical modeling: the onset of increase/decrease in expression of high threshold activated/repressed genes (such as spoIIIG and divIVA respectively) is shifted to earlier time points by overproduction of KinC or Spo0A (Fig. 6.4). This shift of the timing of developmental gene expression events is especially significant considering that the dynamics of indirect targets of Spo0A~P, such as σ^H and its controlled genes, are not affected by KinC or Spo0A overproduction (Fig. 6.4BC). In fact, experimentally, at the optimal combination of Spo0A and KinC induction, high threshold gene expression is initiated after the onset of σ^H activation whereas σ^H is activated after the onset of the expression of high threshold genes
if Spo0A and/or KinC are overproduced (Fig. 6.4BCD). As a result, the accelerated accumulation of Spo0A~P reverses the temporal order of the sporulation gene expression program. We hypothesize that this change in temporal order of the sporulation gene expression program has this profound effect on sporulation efficiency because repression of essential genes occurs abnormally early.

Several genes repressed by Spo0A~P are involved in cell division, DNA replication, cell-shape determination, and protein synthesis (130, 303, 309). Thus, overproduction of KinC and Spo0A in the artificial two-component system could cause abnormalities in cell division, DNA replication, and protein synthesis, and as a result, adversely affect sporulation. Among those genes repressed by Spo0A~P, divIVA may be the most obvious candidate to explain importance of proper timing. Although it is repressed during sporulation, DivIVA plays an important role in chromosome segregation in the early phase of sporulation (313, 315). In conditions of Spo0A and KinC overexpression, divIVA is repressed prematurely (Fig. 6.4EF), resulting in improper chromosome segregation and increased incidence of chromosome-free compartments (Fig. 6.6 and Fig. A5.5 and A5.6). Similar phenotypes have been observed in sporulation-defective divIVA mutants (313). Thus, premature repression of DivIVA could at least partially explain our observation of decreased sporulation efficiency in conditions of high KinC and Spo0A expression. As divIVA is repressed at high levels of Spo0A~P (130), gradual accumulation of Spo0A~P in the wild-type strain under starvation conditions ensures that it is repressed only after it has served its function during the early phase of sporulation - a period which also includes the activation of low-threshold genes such as $\sigma^H$ by Spo0A~P (130).

We also found that simultaneous overproduction of KinC and Spo0A results in swollen cell morphologies and changes the staining pattern of cell membranes with lipophilic fluorescent dye, FM4-64 (Fig. A5.5). The swollen morphologies may be explained by the early repression of a gene mreB. This
gene is repressed (indirectly) by Spo0A~P during sporulation (130) and is responsible for cell shape determination as a prokaryotic homologue of actin (316, 317). The abnormal staining pattern of the membrane can be attributed to inappropriate fatty acids synthesis, since Spo0A~P has been shown to control fatty acids synthesis and maintain membrane lipid homeostasis during early stages of sporulation (318). Thus accelerated and inappropriate activation of Spo0A appears to have pleiotropic effects on various cellular functions. However, the mechanism by which the resulting aberrant morphologies could interfere with sporulation remains unclear.

The artificial induction KinC-Spo0A two-component system demonstrates the importance of gradual accumulation of Spo0A~P. To explain how the structure of the phosphorelay network plays a role in ensuring gradual Spo0A~P accumulation we put these results in a broader context. In principle, there could be two mechanisms to slow-down Spo0A~P accumulation during the onset of sporulation: (1) Limit the phosphate flux that activates Spo0A, (2) Limit the amount of Spo0A available for activation. It is not clear whether one of the two or both mechanisms are required.

Chastanet et al. (2011) and more recently Levine et al.(2012), found that overexpressing σ^{H}, Spo0F and Spo0A had little effect on the dynamics of Spo0A~P accumulation whereas overexpressing of KinA and KinC speeds up sporulation. Based on these results, it was concluded that Spo0A~P accumulation during starvation is limited by the phosphate flux (i.e. mechanism 1). However, accumulation of Spo0A activity can be made even faster by artificially inducing the expression of the constitutively active mutant form of Spo0A: Spo0A-sad67. But with induced expression of Spo0A-sad67 cells sporulate poorly (131) whereas kinase overexpression strains sporulate efficiently (102, 129, 131, 264, 304, 314). This indicates that Spo0A activation in kinase overexpression strains may still be slower as compared to Spo0A-sad67.
induction strain. Thus mechanism (2) may prevent Spo0A~P from accumulating too quickly when kinase is overproduced.

Our experiments with the inducible strains reported here concur with the importance of mechanism (2) in modulating Spo0A activation. Results for KinC-Spo0A induction strains indicate that kinase overexpression strains are limited by the amount of Spo0A available for activation. The simultaneous overproduction of both Spo0A and KinC can overcome this limitation and further speed up Spo0A activation and thereby reduce sporulation efficiency. However, the high sporulation efficiency of KinC induction strain with native spo0A promoter (Fig. 6.7) suggests that the Spo0A feedback loops ensure gradual Spo0A~P accumulation. Notably, feedback architecture is a common feature of many phosphorelays (319, 320) and may thus represent a common strategy for proper dynamical control in variety of functional settings.

How does Spo0A autoregulation ensure gradual Spo0A~P accumulation? Positive feedback is known to slow-down response dynamics (155) and our modeling results suggest that the σH-mediated positive feedback from Spo0A to itself is sufficient to ensure that high and low threshold Spo0A~P targets are expressed in the proper order for sporulation (Fig. 6.7 and Fig. A5.7). Additional direct positive and negative feedback loops originated from Spo0A~P binding to its promoters may provide further control over of Spo0A~P accumulation dynamics. Alternatively, these loops may affect other aspects of the phosphorelay response such as sensitivity to kinase expression or even act as amplifiers of noise (314).

6.4. Experimental Procedures

6.4.1. Strains, plasmids, and oligonucleotides

All strains for experiments were derived from the *B. subtilis* prototrophic strain PY79 (321). Details of strains used in this study are listed in
Supplementary Table 2.2.S4 in the supplementary material. The strain harboring the \textit{divIVA}\text\_\textit{linker-cfp} was a gift from Prahathees Eswaramoorthy and Kumaran S. Ramamurthi. All plasmid constructions (Table A5.5) were performed in \textit{Escherichia coli} DH5\(\alpha\) using standard methods. The oligonucleotide primers used for plasmid construction are listed in Table A5.6.

6.4.2. Media and culture conditions

To induce the synthesis of the protein of interest in \textit{B. subtilis} cells, an isopropyl-\(\beta\)-D-thiogalactopyranoside (IPTG)-inducible \textit{hyper-spank} promoter (\textit{P\_\textit{hy-spank}}) (131) and/or a xylose inducible promoter (\textit{P\_\textit{xylA}}) (322, 323) were used (kind gifts from D. Rudner, Harvard Medical School). The inducer (IPTG and/or xylose) was added at the indicated concentration to the culture in LB medium during the exponential growth phase (optical density at 600 nm, 0.5). Sporulation of the wild-type strain was induced by the procedure of the resuspension method (295).

6.4.3. Sporulation efficiency and \(\beta\)-galactosidase assays

Sporulation efficiency was determined in overnight (16-18 h) culture, as CFU per ml (spores) after heat treatment by incubation at 80°C for 10 min, compared with CFU per ml (viable count) of the pre-heat treatment sample. Assays of \(\beta\)-galactosidase activity were performed as described previously (296).

6.4.4. Immunoblot analysis

For KinC-GFP, cells of the KinC strain (a gene for KinC is placed under the control of an IPTG-inducible \textit{P\_\textit{hy-spank}} promoter) harboring \textit{kinC-gfp} (MF2734) were cultured to mid-exponential phase (OD\(_{600nm}\)=0.5) in LB, and IPTG was added at indicated concentrations. Cells were harvested at 2h after induction. Cells of the wild-type strain harboring \textit{kinC-gfp} (MF2659) were cultured the same as above, except for the IPTG addition, as nutrient rich conditions. Cells of MF2659 were cultured in hydrolyzed casein (CH) growth medium and resuspended in
Sterlini & Mandelstam (SM) medium as sporulation conditions (295). For Spo0A, the KinC-Spo0A strain (MF4318, a gene for KinC is placed under the control of an IPTG-inducible $P_{hy-spank}$ promoter and a gene for Spo0A is placed under the control of a xylose-inducible $P_{xylA}$ promoter) was cultured as above, and xylose was added at indicated concentrations. MF2659 was used for the wild-type Spo0A protein. Cell extracts for immunoblot analysis were prepared from 2 h after IPTG addition in LB or 2 h culture in SM medium by sonication. Immunoblot analysis was performed as described previously (291). Polyclonal anti-GFP (324), anti-Spo0A (325), and anti-σ$^A$ (325) antibodies were used to detect corresponding proteins or GFP-tagged proteins. The intensities of each band after immunostaining were quantified with a FluorChem digital imaging system (Alpha Innotech). σ$^A$ was served as an internal standard control for normalization (325). The protein levels were normalized to both the levels of σ$^A$ and then the levels of each of the corresponding proteins in the wild-type strain (MF2659).

6.4.5. Fluorescence microscopy

Fluorescence microscopy was performed as described previously (296). The membrane and DNA were stained with FM4-64 and DAPI (4', 6-diamidino-2-phenylindole), respectively. Image acquisition and analysis were performed with Slidebook (Intelligent Imaging Innovations). Single cell fluorescence intensities (arbitrary pixel units) were quantified using MicrobeTracker Suite (326) and custom MATLAB code.

6.4.6. Mathematical modeling

Our model focused on the KinC-Spo0A two component system and the dynamics of Spo0A~P and its targets in the KinC-Spo0A double induction strain (MF4318) and the Δspo0B, $P_{hy-spank}$-kinC induction strain (MF4419). SpoIIQB and DivIVA were included in the model as examples of high threshold activated and high threshold repressed genes, respectively. σ$^H$ was included as an example of an indirectly activated low threshold gene. We assumed that in both strains KinC directly phosphorylates Spo0A and that KinC was the sole kinase capable of
activating Spo0A, as the activities of KinA and KinB under rich-media conditions are negligible. For the double induction strain, the transcription of Spo0A and KinC depended on the xylose and IPTG concentrations, respectively. Appropriate transcription rates for these promoters as a function of inducer concentration were determined from the results of the immunoblotting experiments described above and previously reported measurements of Spo0A and KinC concentrations. For the native spo0A promoter in the MF4419 strain, we modeled the rate of transcription $v_{spo0A}$ as a Hill-function dependent on both Spo0A~P and $\sigma^H$ concentrations:

$$v_{spo0A} = k_v \frac{1}{1 + [0A \sim P]^2 / K_1^2} + k_s \frac{[\sigma^H]}{K_H + [\sigma^H]} \frac{1 + f_A [0A \sim P]^2 / K_3^2}{1 + [0A \sim P]^2 / K_2^2}$$

Here $k_v$ and $k_s$ represent the transcription rates for spo0A from the vegetative and sporulation-specific promoters $P_v$ and $P_s$, respectively. $K_H$ represents the binding affinity for $\sigma^H$. $K_1$, $K_2$, and $K_3$ represent the binding affinities for the three Spo0A~P binding sites in Spo0A promoter. $f_A$ represents the fold-change in gene expression effected by binding of Spo0A~P to site 3 in the Spo0A promoter. Transcription rates for all target genes are similarly modeled with Hill-functions and binding affinities based on the results in (130). The details of post-translational phosphorylation/dephosphorylation reactions as well as all parameter values used are provided in the supplementary information. The ode45 solver of MATLAB was used for deterministic simulations of the model, and the Gillespie algorithm in COPASI was used for all stochastic simulations.
Notes

This chapter is based on work that was published in the following article:


*=Equal contributions
Chapter 7

Chromosomal locations of phosphorelay genes controls the Bacillus subtilis response to starvation

7.1. Introduction

Sporulation is the last resort of starving B. subtilis cells. In response to starvation, B. subtilis cells cease vegetative growth and asymmetrically divide to initiate a multistage program that produces stress-resistant and metabolically inert spores (Fig. 7.1A) (4). At the molecular level, the sporulation program is controlled by the master regulator Spo0A (0A) which is active in its phosphorylated (0A~P) form (128). Expression of many downstream sporulation program genes is controlled by 0A~P, and it has been shown that a threshold level of 0A~P commits cells to sporulation (129, 131, 304). The activation of 0A itself is regulated by a complex network known as the sporulation phosphorelay (70). Since cells need two complete chromosomes to produce a viable spore (327), the dynamics of 0A activation must be temporally coordinated with the
completion of DNA replication. Previous studies have suggested that Sda, an inhibitor of the phosphorelay kinase that activates 0A may be responsible for coordinating 0A activation with DNA replication (132). However, it has been recently shown deletion of sda does not completely abolish cell-cycle coupling of 0A activation (102). As a result, the central question of how the phosphorelay coordinates 0A activation with the cell-cycle remains unaddressed.

Here we show that the chromosomal arrangement of two phosphorelay genes plays a critical role in coupling 0A activation to DNA replication. The location of these genes on opposite sides of the chromosome – one located close to the origin of replication, the other close to the terminus, leads to a transient imbalance in their gene dosage during chromosome replication. Combined with a delayed negative feedback loop in the phosphorelay this transient imbalance results in the pulsatile activation of 0A that is responsible for coordinating the commitment to sporulation with the cell-cycle.
Figure 7.1. Coordination of the sporulation response with the cell-cycle in *B. subtilis*.

**A.** Cell-cycle of starving *B. subtilis* cells. In starvation conditions cells first complete DNA replication and then make a cell-fate decision. They either divide medially and continue with slower vegetative growth or divide asymmetrically and commit to sporulation producing a stress-resistant spore. The cell-fate decision is based on the level of phosphorylated master regulator Spo0A (0A~P). 0A~P exceeding a threshold commit cells to sporulation whereas lower levels allows cells to continue growth. The decision must be made after the completion of DNA replication phase (yellow bar) since two complete chromosomes are needed to produce a viable spore.

**B-C** Single cell time-lapse microscopy using a P_{0A\_cfp} reporter for 0A~P. Expression level of P_{0A\_cfp} (B) increases in a pulsatile fashion over multiple cell-cycles in starvation media. Its promoter activity (defined as production rate, an indicator of 0A~P level) also shows pulses of increasing amplitude (C) over multiple generations during starvation. In (B-C) vertical dashed lines indicate cell divisions and yellow shaded regions indicate periods of DNA replication (detected by the presence of DnaN-YFP foci). Note that DNA replication is sometimes initiated just before cell division. For each cell cycle, we can determine the time from birth to end of DNA replication (T_r) and time from birth to peak P_{0A\_cfp} Promoter activity (T_p) respectively.

**D.** Measurements of time from birth to end of DNA replication (T_r) and time from birth to peak P_{0A\_cfp} Promoter activity (T_p) show that T_p > T_r for the vast majority of the cell-cycles implying that 0A activity peaks occurs after DNA replication is complete.

### 7.2. Results

#### 7.2.1. 0A activity pulses follow the completion of DNA replication

To understand the dynamics of sporulation network response, we employed time-lapse microscopy and simultaneously tracked 0A activity and DNA replication in single *B. subtilis* cells (see Methods). We used fluorescent reporters to measure gene expression from 0A~P-regulated promoters for *spo0A* and *spo0F* (P_{0A} and P_{0F}). In addition, we fluorescently tagged a replisome component, DnaN, so that periods of DNA replication could be detected by the
presence of fluorescent DnaN foci (328). In agreement with previous studies (102, 329), we found that in these conditions cells do not sporulate immediately upon exposure to starvation, but rather complete multiple divisions before finally producing spores. Measurements of the expression level of \( P_{0A-cfp} \) (Fig. 7.1B) and the cell growth rate (inferred from cell elongation rate) enabled us to compute 0A activity defined as production rate of the reporter protein (Fig. 7.1C). The results revealed that during the multi-cycle progression to spore formation a single pulse of 0A activity is produced every cell-cycle in starvation conditions (Fig. 7.1C). Similar pulsing was observed in the production of fluorescent reporters of other 0A-P-regulated promoters such as \( P_{0F} \) (Fig. A6.1). We measured the pulse timing during the cell-cycle (Fig. 7.1C) and found that 0A activity pulses always follow the completion of DNA replication (Fig. 7.1D).

This type of 0A activity pulsing has also been reported by other recent studies (102, 132, 283). Veening and coworkers reported that this pulsatile response is the result of pulsing of Sda, an inhibitor of the kinase that activates 0A (132). However a more recent study (102), has shown that Sda might not be the only mechanism behind 0A activity pulsing since deletion of \( sda \) does not completely abolish pulsing. As a result, the mechanism underlying this pulsatile response remains unclear.

### 7.2.2. A hidden negative feedback loop in the phosphorelay

Our first goal was to uncover the mechanism underlying the pulsing of 0A-P. To this end, we built a detailed mathematical model of the sporulation phosphorelay network (Fig. 7.2A) that controls 0A production and phosphorylation (see Appendix 6). This network consists of multiple histidine kinases KinA-E, two phosphotransferases Spo0F (0F) and Spo0B (0B), and the master regulator 0A (330). Among the five kinases, the cytoplasmic protein KinA is the major sporulation kinase (in our conditions) and therefore we have only included this kinase in our model (129, 331). Upon nutrient limitation, KinA auto-phosphorylates and indirectly transfers the phosphate group to 0A via the
intermediate proteins 0F and 0B (Fig. 7.2A) (70). The expression levels of \textit{kinA}, 0F and 0A are regulated by 0A∼P via direct and indirect transcriptional feedback loops. These post-translational and transcriptional interactions were described using appropriate mass-action and Hill-function type rate laws to build the model of the phosphorelay network (see Appendix 6).

Analysis of simulations of this model showed that an increase in 0F levels leads to higher levels of 0A∼P and increased expression of 0A targets (Fig. 7.2B). However, this result contradicts previous \textit{in vivo} studies, which indicate that 0F overexpression inhibits sporulation (263, 332, 333). To resolve this discrepancy, we made a change to the conventionally assumed phosphorelay architecture (Fig. 7.2A without red arrow) and incorporated substrate inhibition of KinA by 0F into the phosphorelay model (red arrow, Fig. 7.2A). This substrate inhibition effect was based on an \textit{in vitro} study demonstrating inhibition of phosphotransfer by excess 0F (290). The resulting model predicted a non-monotonic dependence of 0A∼P on 0F as high 0F concentrations blocked 0A activation (Fig. 7.2B). We verified this inhibition in an engineered strain, \textit{i0F\textsubscript{amyE}}, in which the native copy of 0F was replaced by a copy of 0F expressed from an IPTG-inducible promoter (\textit{P\textsubscript{hyperspark}}, or \textit{P\textsubscript{hsp}}) at the \textit{amyE} locus. The results indicate inhibition of 0A activity by excess of 0F induction (Fig. 7.2C).

Since 0F expression is activated by 0A∼P, the substrate inhibition of KinA by 0F results in the negative feedback loop in the phosphorelay. This feature is especially significant since negative feedback loops are known to produce adaptation-like pulsatile responses (334). In addition, the inhibition of KinA by 0F made the flux through the phosphorelay very sensitive to the ratio of KinA and 0F concentrations (Fig. A6.2). As a result, any perturbation of the relative KinA/0F ratio can force the negative feedback loop to produce a pulsed 0A∼P response.
Figure 7.2. Substrate inhibition of KinA by 0F produces a negative feedback in the phosphorelay.

A. Network diagram of the sporulation phosphorelay network that controls the activity of the master regulator Spo0A (0A). The phosphorelay includes both post-translational and transcriptional regulatory interactions. Post-translationally, the kinases KinA-E (only KinA is shown) transfer phosphoryl groups to the master regulator 0A via the two phosphotransferases Spo0B (0B) and Spo0F (0F). Transcriptionally, 0A~P controls the expression of kinA, 0F and 0A both directly and indirectly via AbrB and σH (not shown) forming multiple transcriptional feedback loops. Our model also includes a substrate inhibition interaction (red blunted arrow) whereby excess 0F can bind to unphosphorylated KinA and block its auto-phosphorylation. This substrate inhibition creates a negative feedback loop wherein 0A~P activates 0F expression and 0F inhibits 0A activation by inhibiting KinA.

B. Mathematical model predicts steady state levels of 0A promoter activity as a function of 0F concentration. The results show that, for a phosphorelay with substrate-inhibition of KinA by 0F (blue curve), 0A promoter activity is a non-monotonic functions of 0F concentrations and decrease ultrasensitively for [0F]>5µM. In contrast, for a phosphorelay without substrate-inhibition (orange curve), 0A promoter activity monotonically increase to saturated value.
**C.** Predicted non-monotonic dependence of 0A activity on 0F levels is confirmed by engineering inducible 0F strain, \(i0F^{amyE} (\Delta 0F; amyE::P_{hsp}-0F)\) and measuring maximum 0A promoter activity in the at different levels of 0F induction. Gray empty circles show maximum \(P_{0A}\) promoter activity levels achieved by individual cell lineages over 25 hours in starvation conditions at each IPTG concentration. Blue filled circles and errorbars indicate the mean and standard deviations of these measurements at each IPTG concentration. Maximum \(P_{0A}\) promoter activity decreases at high 0F expression levels (IPTG>10µM) in agreement with the substrate-inhibition effect of 0F overexpression.

### 7.2.3. Chromosomal arrangement of phosphorelay genes provides a pulse triggering perturbation

The sensitivity of the phosphorelay response to the KinA/0F ratio in our model suggested that chromosomal arrangement of \(kinA\) and 0F may affect 0A~P dynamics. On the *B. subtilis* chromosome 0F is located close to the origin (326°-oriC proximal) and \(kinA\) near the terminus (126°-ter proximal) of DNA replication (Fig. 7.3A). As a result, replication of 0F precedes that of \(kinA\) and each DNA replication cycle produces a transient decrease in the \(kinA:0F\) gene dosage ratio (Fig. 7.3A). In light of this, we proposed that this arrangement of 0F and \(kinA\) genes on the chromosome might couple the phosphorelay output to DNA replication.

Including a DNA replication window in our simulations, we found that this imbalance in \(kinA\) and 0F expression inhibits the phosphorelay flux and results in a decrease in 0A~P during replication (Fig. 7.3A). Once DNA replication is completed and the \(kinA:0F\) ratio returns to one, the delayed negative feedback loop comprised of transcriptional feedback from 0A~P to 0F and the postulated substrate inhibition of KinA by 0F produces an overshoot of 0A~P above its steady state (Fig. 7.3A and Fig. A6.3). These overshoots are manifested as pulses of 0A activity occurring once per cell cycle. Thus, our model explains both the pulsing mechanism and the observed correlation between DNA replication and timing of 0A pulses (compare Fig. 7.1C and Fig. 7.3A). Moreover
comparison of the chromosomal locations of *kinA* and *0F* in 46 different species of spore forming bacteria that have both these genes showed little variation in their positions relative to the chromosomal origin (Fig. A6.4), which suggests that the proposed relative gene dosage pulsing mechanism is evolutionarily conserved.

To uncover the essential design features necessary for the pulsatile 0A~P dynamics, we tested the model response to specific network perturbations. First, we tested the effect of translocating *0F* so that both *kinA* and *0F* are close to the chromosome terminus. Simulations of this mutant *Trans-0F* ^amyE (Fig. 7.3B) showed that the translocation of *0F* close to the terminus eliminated the transient *kinA:* *0F* decrease during DNA replication. Consequently, simulations of this modified strain showed no 0A~P pulsing and instead the system remained at steady state.

Next, we investigated the role of the negative feedback loop between 0A~P and *0F* by testing the effect of assuming that *0F* is expressed from an IPTG-inducible *P* _hsp_ promoter, rather than the native 0A~P regulated *P* _0F_ promoter. We tested two such inducible engineered strains: *i0F* ^amyE ( *P* _hsp-0F_ located close to the origin of replication - Fig. 7.3C) and *i0F* ^gltA ( *P* _hsp-0F_ located close to the terminus - Fig. 7.3D). In these inducible strains, the effective *kinA:* 0F gene dosage was assumed to depend on the level of 0F expression. Model simulations showed that in the *i0F* ^amyE strain (Fig. 7.3C), there is a transient decrease in effective *kinA*: 0F. This decrease inhibits the phosphorelay phosphate flux and causes a decrease in 0A~P. However 0A~P does not overshoot before returning to the steady state and there is no pulse (Fig. 7.3C), unlike the results for WT (Fig. 7.3A). Simulations further showed that transient decrease of *kinA*: 0F is eliminated by 0F translocation in the *i0F* ^gltA strain (Fig. 7.3D) and that 0A~P stays at steady state in this case. Simulations also predicted that the 0A~P response in these inducible strains depends on the level of 0F, specifically 0A~P levels decrease with increasing 0F expression in both *i0F* ^amyE and *i0F* ^gltA (Fig. 7.3CD).
Therefore our model results suggest that two design features of the phosphorelay are crucial for the pulsatile response of 0A~P during starvation: (i) negative feedback between 0A~P and 0F and (ii) transient gene dosage imbalance between \textit{kinA} and \textit{OF} resulting from the chromosomal arrangement of these genes. Consequently, our model predicts that disrupting these key features would abolish 0A~P pulsing and thereby affect sporulation (Fig. 7.3B-D and Fig. A6.3).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7.3.png}
\caption{Mathematical model identifies the mechanism of 0A~P pulsing and its necessary conditions.}
\end{figure}

\textbf{Top panels in (A-D)} show chromosomal arrangements of \textit{OF} and \textit{kinA} in (A) Wildtype (WT) \textit{B. subtilis} and synthetic mutant strains: (B) \textit{Trans-OF}^{\text{gltA}}, (C) \textit{i0F}^{\text{amyE}}, (D) \textit{i0F}^{\text{gltA}}. \textit{OF} is located close to the origin of replication in WT and \textit{i0F}^{\text{amyE}} strains and close to the terminus in the \textit{Trans-OF}^{\text{gltA}} and \textit{i0F}^{\text{gltA}}. \textit{kinA} is located close to the terminus in all strains. Note that \textit{OF} is expressed from the IPTG-inducible \textit{P}_{\text{hsp}} promoter, rather than the native 0A~P regulated \textit{P}_{\text{OF}} promoter in the inducible \textit{i0F}^{\text{amyE}} and \textit{i0F}^{\text{gltA}} strains.

\textbf{Middle panels in (A-D)} show changes in \textit{kinA:OF} gene dosage ratio in the WT and mutant strains. In WT (A) and \textit{i0F}^{\text{amyE}} (C), \textit{OF} is replicated before \textit{kinA} due to their arrangement on the chromosome. As a result, there is a transient decrease in the \textit{kinA:OF} gene dosage ratio during DNA replication (yellow bar). The translocation of \textit{OF} close to the terminus in a synthetic \textit{Trans-OF}^{\text{gltA}} (B) and \textit{i0F}^{\text{gltA}} (D) strains eliminates the transient \textit{kinA:OF} decrease. In the inducible \textit{i0F}^{\text{amyE}} and \textit{i0F}^{\text{gltA}} strains, the effective
*kinA:0F* gene dosage depends on whether the level of 0F expression from the IPTG-inducible *P*_hsp promoter is low (solid line) or high (dashed line).

**Bottom panels in (A-D)** show model predictions for the response of 0A~P levels to the changes in *kinA:0F* ratio in WT and mutant strains. Model simulations show that the transient decrease in *kinA:0F* during DNA replication (yellow bar) in WT (A) inhibits the phosphorelay phosphate flux, thereby causing a decrease in 0A~P. Once DNA replication is complete, the phosphorelay produces an overshoot of 0A~P before returning to the steady state resulting in a 0A~P pulse. Model results also predict that that elimination of the transient decrease of *kinA:0F* in the Trans-0F<sub>gltA</sub> strain (B) abolishes 0A~P pulsing and instead the system stays at steady state. In the i0F<sub>amyE</sub> strain (C), the transient decrease in effective *kinA:0F* inhibits the phosphorelay phosphate flux and causes a decrease in 0A~P but 0A~P does not overshoot before returning to the steady state and there is no pulse. This 0A~P response depends on the level of 0F expression. Simulations also show that elimination of the transient decrease of *kinA:0F* in the i0F<sub>gltA</sub> strain (D) abolishes 0A~P pulsing and instead the system stays at steady state. The steady state 0A~P level depends on whether 0F expression from the IPTG-inducible *P*_hsp promoter is low (solid line) or high (dashed line). These results show that 0A~P pulsing is triggered by DNA replication and that both the *kinA:0F* chromosomal arrangement and 0A~P-0F negative feedback are essential for 0A~P pulsing.

### 7.2.4. Experimental tests confirm the role of gene-dose imbalance and negative feedback in 0A~P pulsing

We tested our modeling predictions by engineering two sets of *B. subtilis* strains. The first set of strains was engineered to examine how the chromosomal arrangement of *kinA* and 0F and the resulting transient gene dosage imbalance affects 0A~P pulsing (Fig. 7.4A-D). In the first strain, Trans-0F<sub>gltA</sub> (Fig. 7.4B), we eliminated the transient imbalance in the *kinA:0F* ratio by translocation of the 0F gene to the *gltA* locus close to the terminus. As a control, we engineered a second strain (*Trans-0F<sub>amyE</sub>; Fig. 7.4C) by moving the 0F gene to *amyE* locus near the origin so that the *kinA:0F* imbalance is retained despite the translocation. Because the entire 0F gene with the upstream region containing
regulatory sequences was translocated in both strains, only the relative gene dosage during replication was perturbed whereas the negative feedback regulation remained intact. As predicted by the model (Fig. 7.3B), pulsing was abolished in the Trans-0F<sub>gltA</sub> strain but not in the Trans-0F<sub>amyE</sub>, which exhibited pulsing similar to the WT cells (Fig. 7.4E-G). The lack of pulsing in the Trans-0F<sub>gltA</sub> resulted solely from the change in chromosomal position of 0F and not from reduced 0F expression. Specifically, we measured the activity of P<sub>0F</sub> at amyE and gltA loci and found that it displays the same pulsatile behavior with similar expression levels (Fig. A6.1). Together, these results confirm the prediction that kinA:0F gene dosage imbalance is critical for 0A pulse generation.

To further establish the role of transient imbalance in 0F to kinA expression rates, we constructed a rescue strain, iTrans-0F (Fig. 7.4D). In addition to 0F translocation to the terminus (Trans-0F<sub>gltA</sub>), we integrated an additional IPTG-inducible copy of 0F close to the chromosome origin to recover the transient imbalance in 0F to kinA expression rates during replication. In the absence of IPTG, this strain acted like the non-responsive Trans-0F<sub>gltA</sub> strain and showed no 0A~P pulsing. However, at 5µM IPTG the sporulation in the iTrans-0F strain was restored similar to the WT strain (Fig. 7.4H). Thus, we concluded that the transient imbalance in kinA and 0F expression resulting from their chromosomal locations acts as an essential trigger for 0A~P pulses.
Figure 7.4. 0A~P pulsing depend on *kinA*:0F chromosomal arrangement and transcriptional feedback from 0A~P to 0F negative feedback.

**A-D** Chromosomal arrangements of 0F and *kinA* in Wildtype *B. subtilis* (A), *Trans-0F*gltA (B), *Trans-0F*amyE (C), *iTrans-0F* (D) strains. Blue and orange bars mark strains that have transient *kinA*:0F imbalance and negative feedback respectively.

**E-H** Measurements of *P0A-yfp* activity in single cells during starvation. 0A~P pulses seen in Wildtype *B. subtilis* cells (E) are abolished by the translocation of 0F to the *gltA* locus in the *Trans-0F*gltA strain (F) but not by the translocation of 0F to the *amyE* locus in the *Trans-0F*amyE strain (G). Addition of an IPTG inducible copy of 0F near the origin in the *iTrans-0F* strain (H) recovers the transient *kinA*:0F imbalance lost in *Trans-0F*gltA strain.
and rescues the 0A–P pulses seen in wildtype B. subtilis cells. Note that \( P_{0A} \cdot yfp \) promoter activity level increases in pulsing strains (A, C, D) but not in the Trans-\( 0F^{gltA} \) strain (B). Vertical dashed lines indicate cell divisions.

I-L Chromosomal arrangements of 0F and \( \text{kinA} \) in \( i0F^{amyE} \) (I, K) and \( i0F^{gltA} \) strains (J, L). Note that both strains lack the 0A–P-0F negative feedback.

M-P Measurements of \( P_{0A} \cdot yfp \) activity in single cells during starvation. 0A activity pulsing is greatly decreased in the \( i0F^{amyE} \) (M, O) and \( i0F^{gltA} \) strains (N, P) which lack the negative feedback. 0A activity in the \( i0F^{amyE} \) (M) strain fluctuates due to transient changes in \( \text{kinA}:P_{hsp}:0F \) but does not pulse (Fig. A6.6). If 0F expression is low (at 5µM IPTG; M, N), both \( i0F \) strains accumulate high 0A activity levels. High level expression of 0F (at 20µM IPTG; O, P) blocks 0A activation in both \( i0F \) strains similar to Trans-\( 0F^{gltA} \). Vertical dashed lines indicate cell divisions.

To test the role of negative feedback between 0A–P and 0F in pulse generation, we created a second set of strains, \( i0F^{amyE} \) and \( i0F^{gltA} \) (Fig. 7.4I-L), in which 0F is expressed from the IPTG-inducible \( P_{hsp} \) promoter, rather than the native 0A–P regulated promoter. According to our simulations (Fig. 7.3CD), even though 0A–P pulsing in these strains would be disrupted, cells could still accumulate high levels of 0A–P if 0F expression was below the inhibitory range determined in Fig. 7.2C. Indeed, we found that in \( i0F^{amyE} \) and \( i0F^{gltA} \) strains, the 0A promoter activity increased gradually over time to levels comparable to that in WT at 5µM IPTG (low 0F expression; Fig. 7.4MN). In contrast, at 20µM IPTG (high 0F expression; Fig. 7.4OP) there was no significant increase in 0A promoter. We also found that, consistent with our model predictions (Fig. 7.3C), 0A promoter activity in \( i0F^{amyE} \) fluctuates (Fig. 7.4M) due to transient changes in \( \text{kinA}:0F \) expression ratios in this strain. However, these fluctuations did not resemble the adaptation type pulsatile responses of the WT and Trans-\( 0F^{amyE} \) strains. In fact, unlike the WT and Trans-\( 0F^{amyE} \) strains both \( i0F^{amyE} \) and \( i0F^{gltA} \) strains showed no statistically significant difference between the peak 0A activity during a cell-cycle and the 0A activity at the end of the cell-cycle (Fig. A6.5). This led us to conclude that 0A activity does not pulse in these strains, thereby
confirming our prediction that the negative feedback in the phosphorelay is essential for producing pulses of $0\text{A}^{-}\text{P}$ in response to starvation.

7.2.5. Lack of $0\text{A}^{-}\text{P}$ pulsing leads to sporulation defects

Notably, the lack of $0\text{A}^{-}\text{P}$ pulsing in the inducible $i0\text{F}^\text{amyE}$ and $i0\text{F}^\text{gltA}$ strains did not prevent them from producing spores when $0\text{F}$ expression was low (at 5$\mu\text{M}$ IPTG, see Fig. 7.5A). Thus, $0\text{A}^{-}\text{P}$ pulsing was not essential for sporulation, but we hypothesized that it was necessary for ensuring that the threshold level of $0\text{A}^{-}\text{P}$ activity required for asymmetric septation or $\sigma^F$ activation would only be reached in the cells with two complete chromosomes. Accordingly, we predicted that strains with non-pulsatile accumulation of $0\text{A}^{-}\text{P}$ would exhibit an increased frequency of defective sporulation phenotypes resulting from untimely $0\text{A}$ activation.

To determine whether pulsatile $0\text{A}$ activation plays a role in preventing faulty sporulation, we counted the frequency of defects in the pulsing WT and non-pulsing $i0\text{F}^\text{gltA}$ strain. We specifically focused on two types of sporulation defects (Fig. 7.5B): (i) Asymmetric septation without activation of $\sigma^F$ in the forespore and (ii) Activation of $\sigma^F$ in the mother cell before asymmetric septation. We found that asymmetric septation without activation of $\sigma^F$ causes cells to bud off a small daughter cell that lacks DNA and dies soon after division (Fig. 7.5B). On the other hand, activation of $\sigma^F$ in the mother cell causes cell death (Fig. 7.5B). Thus, both types of defects affect the ability of cells to efficiently produce spores. Counting the number of such abnormalities, we found that the frequency of defects per spore produced over 30hrs in starvation conditions was about three-fold higher in the $i0\text{F}^\text{gltA}$ strain (14.7$\% \pm 1.7\%$; 3 independent measurements $>250$ spores each) relative to the WT strain (5.0$\% \pm 1.5\%$; 3 independent measurements, $>250$ spores each) (Fig. 7.5C). Therefore, we find that $0\text{A}^{-}\text{P}$ pulsing plays a key role in preventing defective sporulation.

Next we examined whether the higher frequency of defects/spore ratio in the non-pulsing $i0\text{F}^\text{gltA}$ strain results from lack of proper coordination of
sporulation with the cell-cycle. To test this idea, we used time-lapse microscopy data for the $iOF^{gltA}$ strain to compute the time of cell-fate decisions both in cell-cycles that successfully produce spores and those that end in defective sporulation. The time of cell-fate decision was defined as the time from the start of the cell-cycle to the time of $P_{\text{spolIR}}\cdot\text{cfp}$ (a $\sigma^F$ reporter) activation in the cases of normal sporulation and the mother cell $\sigma^F$ activation defect. For the defect of asymmetric septation without $\sigma^F$ activation, the time of cell-fate decision was defined as the time from the start of the cell-cycle to the time of asymmetric septation. As shown in Fig. 7.5B and 7.5D, cell-cycles that end in sporulation defects reach cell-fate decisions early in the cell-cycle (2-3hrs after the start of the cell-cycle). We note that unlike rich medium conditions, cell-cycle durations in starvation conditions are typically 5-6 hours long. As DNA replication is incomplete early in the cell-cycle, these early cell-fate decisions appear to arise from the attempt to execute the sporulation program without two complete chromosomes. In contrast, in cell-cycles that successfully produce a spore, the timing of cell-fate decisions is typically late in the cell-cycle (>4hrs after the start of the cell-cycle; Fig. 7.5B, D), after the completion of DNA replication. Thus we concluded that activation of 0A and commitment to sporulation too early in the cell-cycle, before the completion of DNA replication, is responsible for both sporulation defects. Moreover, since these defects occur at a higher frequency in the non-pulsing $iOF^{gltA}$ strain, these results show that the 0A~P pulsing plays a key role in preventing sporulation defects because of its ability to ensure proper coordination of the sporulation program with DNA replication.
Figure 7.5. Loss of coordination of sporulation program with DNA replication in non-pulsing strains could lead to sporulation defects.

A. Fraction of cells that have formed spores at 25hrs into starvation in different strains. Bars and errorbars show the mean and standard deviation of spore fraction respectively for each strain. Means and standard deviations were calculated using 3 independent measurements for each condition.

B. Phase-contrast and fluorescence microscopy (PspollIR-cfp) images from a time-lapse experiment showing the difference in timing of SpoIIIR activation/Asymmetric septation in sporulation and sporulation defects. T0 represents the time of birth for each indicated cell (yellow outline). T2, T3 etc. indicate time after birth in hours. Time-point of SpoIIIR activation/Asymmetric septation in each case is marked by blue box. Asymmetric septation and σ^F activation happen late in the cell-cycle (T5) during normal sporulation as compared to the sporulation defect cases. Early activation of σ^F in the whole cell at T2 results in cell death. Early asymmetric septation at T2 produce a small daughter cell (orange outline) which dies without activating σ^F in the forespore.

C. Quantification of number of defects per spore produced over 30hrs in starvation conditions by the pulsing WT strain (green bars) and the non-pulsing iOP^oIIR^gltA strain (yellow bars). Errorbars indicate the standard deviation of 3 independent measurements. The
defects/spore ratio is significantly higher for non-pulsing \( i0F^{\text{gltA}} \) strain for both types of sporulation defects.

**D.** Time difference between birth and SpoIIR (a \( \sigma^F \) reporter) activation/Asymmetric septation in cell-cycles that produce spores and those that end in lysis due to sporulation defects. SpoIIR activation/Asymmetric septation happens significantly earlier in cell-cycles that end sporulation defects.

### 7.3. Discussion

Taken together, our results reveal a novel mechanism for coupling cell-fate decisions to DNA replication. Using an ultrasensitive, delayed negative feedback loop to detect the transient imbalance of gene dosage resulting from directional chromosome replication allows *B. subtilis* to use DNA replication itself as the trigger for 0A activation, thereby ensuring that these two do not temporally conflict with each other. Moreover, the pulsatile activation of 0A during every cell cycle offers *B. subtilis* cells an opportunity to evaluate their starvation level and decide between sporulation and continued vegetative growth on a cell-cycle by cell-cycle basis.

One of the key design features that underlies the pulsatile 0A~P dynamics appears to be a negative feedback loop, which are known to be one of the few network motifs capable of generating adaptation-like pulsatile responses (334). The crucial component that creates this negative feedback is the substrate inhibition of KinA by 0F through the formation of a dead-end complex that blocks KinA autophosphorylation. This substrate inhibition effect has been demonstrated previously (290), but has received little attention in mathematical modeling studies. This effect is however essential for explaining the inhibitory effect of 0F overexpression on sporulation (263, 332, 333). Our results demonstrate that this negative feedback based on the substrate inhibition of KinA by 0F plays a critical role in coupling 0A~P pulsing to DNA replication. Alternative explanations for 0A~P pulsing suggested by earlier studies invoke either the 0A~P-AbrB-Spo0E negative feedback loop (335) or the inhibition of KinA by Sda (132). However, a
pulsing mechanism based on the 0A~P-AbrB-Spo0E negative feedback loop is unlikely since it cannot explain our observations of the cell-cycle coupling of pulses (Fig. 7.1D). In addition, a recent study has shown the Spo0E deletion does not affect pulsing (102). On the other hand, the cell-cycle dependent oscillations of Sda provides a viable explanation for the DNA replication coupled 0A~P pulses. However, our results showing the lack of pulsing in $i0F^{grA}$ strain where 0A~P-0F negative feedback is perturbed (Fig. 7.3D), suggests that substrate inhibition mechanism based feedback that we propose here plays the key role in controlling 0A~P dynamics.

Notably, dead-end complex based substrate inhibition mechanisms have been previously postulated to act as a source of ultrasensitivity in the response of bacterial two-component systems and sigma factor regulation (270, 336, 337). Our results reveal that these mechanisms can also result in ultrasensitivity in the ratio of the two genes involved in dead-end complex substrate inhibition – specifically the KinA:0F ratio. In the case of sporulation, this ratio sensitive response forms the basis of the coupling of 0A~P pulsing to DNA replication by a gene dosage mechanism. More generally, however, this ratio-sensitive response provides a unique mechanism for the integration of different environmental signals - a feature that may be relevant to a wide variety systems that employ phosphorelay networks (338).

The sensitivity of phosphorelay response to the KinA:0F ratio and as a result to their gene dosages also suggests a unique mechanism for growth rate based control of sporulation. In nutrient rich conditions, for cells with a generation time of less than the time required replication, new rounds of replication begin before the previous round terminates, resulting in multifork replication (339). As a result of multifork replication, cells have multiple copies of the origin-proximal chromosomal region and a single copy of the terminus-proximal region. Consequently the relative gene dosage of origin-proximal and terminus-proximal genes increases at high growth rate (339). Considering that 0F and $kinA$ are origin-proximal and terminus-proximal respectively, this implies that the low
KinA:0F ratio at high growth rates could prevent 0A activation sporulation in nutrient rich conditions. A similar origin-terminus relative gene dosage mechanism has been shown to regulate histidine metabolism in response to growth rate in *Salmonella typhimurium* (340). Our results suggest that in *Bacillus subtilis*, the kinA:0F relative gene dosage sensitive phosphorelay may act similarly and function as a growth rate dependent decision mechanism although this requires further investigation.

Our results also show that to enable the pulsatile 0A−P response strategy, the design of the sporulation network exploits a universal feature of bacterial physiology: the transient imbalance of gene dosage between origin-proximal and terminus-proximal genes during chromosome replication. It is well-known that chromosomal location of genes affects their expression (341, 342) and that clustering of genes facilitates their co-regulation (343). Recent studies have also revealed that origin proximal location of genes can be used to detect DNA replication stress and trigger bacterial competence (344). Our results add to this growing repertoire of functional roles for chromosomal arrangement of regulatory genes and provide a simple cell-cycle coupling mechanism that could very well be employed in a wide range of other microbial species and stress response mechanisms.

### 7.4. Experimental Procedures

#### 7.4.1. Strain construction

Table A6.1 lists *Bacillus subtilis* strains used in this study. All strains are isogenic to *B. subtilis* PY79. Gene spo0F with its native P_{spo0F} promoter (165 bp upstream) and 43 bp downstream sequence were PCR amplified from *B. subtilis* PY79 with addition of a terminator upstream (CCAGAAAGTCAAAAGCCTCCGACCG) and ligated to either ECE173 (Pm<sup>R</sup>) between BHI and XbaI restriction sites, or pLD30 (Sp<sup>R</sup>) between BHI and ERV sites. These constructs were used, respectively, to create Trans-0F<sup>30F</sup> and Trans-
To create \(i0F^{amyE}\) strain, the \(spo0F\) gene was PCR amplified from \(B.\ subtilis\) PY79 with addition of optimal RBS and linker (AAGGAGGAAAAGTCACATT) and including 43 bp fragment downstream. It was ligated to a derivative of PLD30, JDE131 plasmid (Sp\(^R\)) next to the Phyperspank promoter (between HindIII and NheI restriction sites). To create \(i0F^{gltA}\) strain, the region containing Phyperspank promoter, \(spo0F\) and \(lacI\) was subcloned from resulting JDE131 and integrated into ECE173 (Pm\(^R\)) between BHI and ERI restriction sites. For \(spo0F\) deletion construct, the 5’ and 3’ fragments were PCR amplified using the following primers:

\[
\text{GAGGCGCCCCTGTGCCTTCTGTCATCTCCTCAG} \quad \text{and} \quad \text{TCGAATTTCGAAAAATACGAATGCGTATGTGATCATCAACGA}
\]

for 5’ arm, \(\text{GATCTAGAGACATCGACGAAATCAGAGACGC} \text{CGTGCAAAAAATATCTGCCCC}
\]
\(\text{TGAAGTCTAAC} \quad \text{and} \quad \text{TCGTGACCCCTTCGAAACACCAAGGATC} \text{ACTGGGAG}
\]

for 3’ arm, and integrated as recombination arms into either per449 (Erm\(^R\)) or per449 (Kan\(^R\)) vectors between KasI and ERI, or XbaI and SalI restriction sites, respectively. The reporter strains were described previously (283, 345).

The vectors used in this study are as follows: ECE174, integrating into the \(sacA\) locus (constructed by R. Middleton and obtained from the Bacillus Genetic Stock Center); pLD30 designed to integrate into the \(amyE\) locus (kind gift from Jonathan Dworkin, Columbia University); ECE173, designed to integrate into the \(gltA\) locus (346) (constructed by R. Middleton and obtained from the Bacillus Genetic Stock Center); per449, a generic integration vector constructed for integration into the gene of interest (kind gift from Wade Winkler, UT Southwestern); and the bi-functional cloning plasmid pHIP13 carrying the replication origin of the cryptic \(B.\ Subtilis\) plasmid pTA1060 (5 copies per genome) (347). One-step \(B.\ subtilis\) transformation protocol was followed.

### 7.4.2. Culture preparation

For imaging, \(B.\ subtilis\) culture was started from an overnight LB agar plate containing appropriate antibiotics (final concentrations: 5 μg/ml
chloramphenicol, 5 μg/ml neomycin, 5μg/ml erythromycin, 5μg/ml phleomycin and 100 μg/ml spectinomycin). Strains containing multiple resistance genes were grown on a combination of no more than three antibiotics at a time. Cells were resuspended in casein hydrolysate (CH) medium (295) and grown at 37°C with shaking. After reaching OD 1.8-2.0, cells were washed once and resuspended in 0.5 volume of Resuspension Medium (RM) (295). The resuspended cells were grown at 37°C for 1 hour, then diluted 15-fold and applied onto a 1.5% low-melting agarose pad made with RM-MOPS medium with desired IPTG or glucose concentration, if necessary. The pads were covered, left to air-dry for 1 hour at 37°C and placed into a coverslip-bottom Willco dish for imaging.

7.4.3. Time-Lapse Microscopy

Differentiation of B. subtilis microcolonies was monitored with fluorescence time-lapse microscopy at 37°C with an Olympus IX-81 inverted microscope with a motorized stage (ASI) and an incubation chamber. Image acquisition was set to every 20 min with a Hamamatsu ORCA-ER camera. Custom Visual Basic software in combination with the Image Pro Plus (Media Cybernetics) was used to automate image acquisition and microscope control.

7.4.4. Image Analysis

A combination of custom written MATLAB programs, Schnitzcells software (http://cell.caltech.edu/schnitzcells), MicrobeTracker tool (326) and freely available ImageJ plugins (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/, 1997-2014) was used to analyze microscopy data as described below.
7.4.5. Data Analysis

Quantification of cell growth rates

The mean cell-growth rate for individual cell-cycles (Figs. 7.1, 7.4 and Fig. A6.1) was quantified using the measurements of cell-length in the time-lapse data. We first calculated the instantaneous cell-growth rate at every frame as:

$$
\mu(t) = \frac{1}{L(t)} \frac{dL(t)}{dt} = \frac{d\log(L(t))}{dt} = \frac{\log(L(t + \Delta t)) - \log(L(t))}{\Delta t}
$$

where $L(t)$ is the cell-length at time $t$ and $\Delta t$ is the time difference between successive frames (20 minutes). For cell-cycles that result in vegetative division, mean growth rate during a cell-cycle was defined as the average of $\mu(t)$ over the cell-cycle duration. For cell-cycles that end in sporulation, the mean growth rate was the average of $\mu(t)$ over the cell-cycle duration until the asymmetric division. Depending on the strain, the asymmetric division was defined either as $P_{spoIIR}$ activation (for the strains in which a $P_{spoIIR}$ reporter was present) or as the time frame two-hours before the appearance of the phase-bright forespore.

Calculation of Promoter Activity

The measurements of promoter activities for $P_{0A-cfp/yfp}$ promoters (Figs. 7.1C and 7.4) and the $P_{0F-yfp}$ promoter (Fig. A6.1) refer to rate of protein production calculated from fluorescence time-lapse data. Promoter activity for a specific promoter can be estimated from the dynamics of fluorescent reporter expression from that promoter. Given a specific promoter- fluorescent reporter combination, the dynamics of the reporter protein concentration $[F]$ are given by the following equation:

$$
\frac{d[F]}{dt} = \gamma(t) - (\gamma + \mu(t))[F]
$$
Here $P_F(t)$ is the protein production rate or the promoter activity. $\mu(t)$ and $\gamma$ are the protein dilution and degradation rates respectively. Rearranging equation (1), we find the following expression for the promoter activity $P_F(t)$:

$$P_F(t) = \frac{d[F]}{dt} + (\gamma + \mu(t))[F]$$

[2]

The reporter protein concentration, $[F]$, depends on the total amount of fluorescent protein inside the cell, $F$ and the cell volume $V$: $[F] = F/V$. Thus we can rewrite equation (2) as:

$$P_F(t) = \frac{d(F(t)/V(t))}{dt} + (\gamma + \mu(t))\frac{F(t)}{V(t)}$$

[3]

Approximating the cells as cylinders,

$$V(t) = \pi W^2 L(t)$$

[4]

$$F(t) = M(t)WL(t)$$

[5]

where $W$ is cell width assumed to be constant, $L(t)$ and $M(t)$ are cell-length and mean cell fluorescence respectively. This change in variables makes the estimates of promoter activity less sensitive to segmentation errors (102, 348).

Using equations (4) and (5) in equation (3):

$$P_F(t) = \frac{1}{\pi W} \left[ \frac{dM(t)}{dt} + (\gamma + \mu(t))M(t) \right]$$

[6]

Using equation (6) the promoter activity $P_F(t)$ can be determined from fluorescence time-series data by using the dilution rate $k_{\text{dil}}(t)$ and the derivative of $M(t)$. We used the following approximations to calculate the promoter activity (up to proportionality constant):

$$\frac{dM(t)}{dt} = \frac{M(t + \Delta t) - M(t)}{\Delta t}$$

$$\mu(t) = \frac{1}{L(t)} \frac{dL(t)}{dt} = \frac{\log(L(t + \Delta t)) - \log(L(t))}{\Delta t}$$

$$P_F(t) \propto \frac{M(t + \Delta t) - M(t)}{\Delta t} + \gamma M(t) + \frac{\log(L(t + \Delta t)) - \log(L(t))}{\Delta t} M(t)$$

[7]
Here $\Delta t$ is the time difference between successive frames (20 minutes). We used $\gamma = 0.1 \text{ hr}^{-1}$ since in our conditions, fluorescent proteins CFP and YFP are stable.

**Quantification and characterization of promoter activity pulses**

Promoter activity time-series determined from fluorescence microscopy were smoothed using the MATLAB *smooth* function by employing a Savitsky-Golay filter with a 3rd-order polynomial over a sliding window of 5 frames. After smoothing, the time to reach the maximum promoter activity from the start of the cell-cycle and was used to calculate the $T_p$ period (Fig. 7.1D).

To differentiate pulsing and non-pulsing strains (Supplementary Fig. A6.4), the promoter activity in each cell-cycle was quantified at three time points: at the start of the cell-cycle, at the point during the cell-cycle where the promoter activity reaches its maximum value and at the end of the cell-cycle (Fig. A6.4AB). To determine whether increase in 0A promoter activity during each cell cycle occurs in a pulsatile manner we compared $\Delta \text{Max}$, the difference between Peak promoter activity and promoter activity at the start of the cell cycle, and $\Delta \text{End}$, the difference between the promoter activities at the start and end of the cell-cycles. In each strain, we aggregate the data for $\Delta \text{Max}$ and $\Delta \text{End}$ at each cell cycle and compare their distributions with a two sample $t$-test. Statistical significance of the observed differences in $\Delta \text{Max}$ and $\Delta \text{End}$ is used to detect whether pronounced pulsing can be detected for each strain.

**Identification of DNA replication periods using DnaN foci**

To identify DNA replication windows in time-lapse experiments (Fig. 7.1BC) we expressed a fluorescent DnaN-YFP fusion protein from the IPTG inducible $P_{\text{hyper-spank}}$ promoter. During DNA replication, the DnaN forms subcellular foci (132, 328). Using the DnaN-YFP fusion protein, these foci can be detected as diffraction-limited spots. We used the SpotFinderF MATLAB program (326) with a signal-to-background ratio of 40 as the minimum peak height to identify DnaN-YFP foci and active DNA replication periods. DNA replication
periods identified in this way were used to calculate the difference between the start of the cell-cycle and the end of DNA replication or $T_r$ period (Fig. 7.1D) for each cell-cycle in the time-lapse data.

**Estimation of the spore fraction**

To get an estimate of the sporulation efficiency of the various strains in Fig. 7.3 we calculated a spore fraction for each strain after 25 hrs in starvation conditions (see Fig. 7.5A). The spore fraction was calculated by dividing the number of cells that had formed phase-bright spores by the total number of cells. Three independent measurements were made for each strain with more than 200 cells counted in each instance to calculate the mean and standard deviations of spore fractions.

**Quantification of sporulation defects**

We manually counted the number of both types of sporulation defects (Asymmetric septation without $\sigma^F$ activation and Activation of $\sigma^F$ in the mother cell) in the time-lapse images of pulsing WT strain (harboring the same integrated reporters and GltA knockout as $i0F^{	ext{gltA}}$ strain) and of the non-pulsing $i0F^{	ext{gltA}}$ strain (at 5µM IPTG) over 30hrs in starvation conditions. We also counted the total number of spores produced over those 30hrs to calculate the frequency of defects produced for both strains per spore (Fig. 7.5B). Two-sample t-test was used to determine the statistical significance of the observed differences.

To determine the timing of cell-fate decision in sporulation defects and normal sporulation (Fig. 7.5C) we used time-lapse microscopy data from the non-pulsing $i0F^{	ext{gltA}}$ strain. The timing of cell-fate decision was defined as the time from the start of the cell-cycle to the time of $P_{\text{SpoIIR-cfp}}$ (a $\sigma^F$ reporter) activation in the cases of normal sporulation and mother cell $\sigma^F$ activation defect. For the case of asymmetric septation without $\sigma^F$ activation defect, the time of cell-fate decision was defined as the time from the start of the cell-cycle to the time of asymmetric septation.
Notes

This chapter is based on work that was published in the following article:


*=Equal contributions
Slowdown of growth controls cellular differentiation

8.1. Introduction

*B. subtilis* cells survive prolonged starvation by differentiating into stress-resistant and metabolically inert spores (Fig. 8.1A) (4). This differentiation program, known as sporulation, is controlled by the master regulator Spo0A (0A) which is active in its phosphorylated (0A~P) form (128). 0A activity itself is regulated by a complex network known as the sporulation phosphorelay - a more complex version of the bacterial two-component regulatory systems (70). This phosphorelay (Fig. 8.1B), consists of the major sporulation kinase KinA that autophosphorylates and indirectly transfers the phosphoryl group to 0A via the intermediate proteins Spo0F (0F) and Spo0B (0B).

The expression levels of *kinA*, *0F* and *0A* are regulated by 0A~P via direct and indirect transcriptional feedback (289, 349). Crucially, autophosphorylation of KinA can be blocked by high levels of 0F (Fig. 8.1B) through a substrate-inhibition mechanism (263, 290, 350). This substrate-inhibition forms part of a
negative feedback loop in the phosphorelay. Recently, we showed that this negative feedback allows cells to respond to transient gene dosage imbalance during DNA replication with a pulsatile activation of 0A (350). Due to the widely conserved arrangement of 0F (326°-oriC proximal) and kinA (126°-ter proximal) genes on the chromosomes in B. subtilis and other sporulating bacteria, 0F gene is replicated before that of kinA, leading to a transient decrease in the kinA:0F gene dosage ratio. Completion of DNA replication returns the ratio to 1:1 and triggers the phosphorelay to respond with a pulse of 0A~P. Thus in every cell-cycle of starving B. subtilis cells, completion of the DNA replication is followed by a pulse of 0A~P (Fig. 8.1A). The decision to sporulate is based on the amplitude of the 0A~P pulse. Low-amplitude 0A~P pulses allow cells to divide medially and continue growth (Fig. 8.1A – left) whereas when this amplitude exceeds a threshold (Fig. 8.1A – right), cells divide asymmetrically and commit to sporulation (129, 131, 304).

Despite these developments in the understanding of the phosphorelay dynamics, the relationship between starvation and the amplitude of 0A~P pulses remains unclear. It has been suggested that a multi-protein phosphorelay, offering multiple entry points for putative starvation signals, is well-suited for proper 0A activity regulation (330, 351). However, it is unclear how the design of the phosphorelay enables it to combine information about several different essential nutrients (352), evaluate the starvation level and correctly time cell-fate decision to allow complete execution of multi-stage sporulation program. As a result, the central question of how B. subtilis sporulation program senses nutrient levels remains open.
Figure 8.1. Decision making during \textit{B. subtilis} sporulation.

\textbf{A.} Sporulation commitment depends on the amplitude of a cell-cycle coordinated pulse of the sporulation master regulator Spo0A\textasciitilde P (0A\textasciitilde P). Yellow bars indicate the DNA replication phase.

\textbf{B.} The sporulation phosphorelay network that controls 0A\textasciitilde P formation (see text for details).

\textbf{C-E.} Single cell time-lapse microscopy using a $P_{0A\textasciitilde P}$-cfp reporter for 0A\textasciitilde P. (C) Cell length (green) and its cell growth rate, i.e. cell-cycle averaged log-derivative, (gray), over
multiple cell-cycles in starvation media. Expression level of $P_{0A}^{-}cfp$ (D) increases in non-monotonic fashion. Its promoter activity (defined as production rate, an indicator of $0A^{-}P$ level) shows pulses with an increase amplitudes that is coordinated with decrease in growth rate (E). In (C-E) vertical dashed lines indicate cell divisions.

**F.** Measurements of $P_{0A}^{-}yfp$ promoter activity show that $0A^{-}P$ pulse amplitudes and growth rates are anti-correlated. Each dot corresponds to ranked measurements of the $P_{0A}^{-}yfp$ promoter activity pulse amplitude and growth rate of an individual cell-cycle. Red and gray dots indicate cell-cycle that end in sporulation and vegetative division respectively. The resulting Spearman's rank correlation $\rho=-0.8$, $p$-value$<10^{-60}$.

**G.** Two hypothetical mechanisms behind the observed correlation. Starvation may be detected by the sporulation network via growth-rate modulation of phosphorelay concentrations or by modulation of phosphorelay activity by growth-rate-correlated signal/metabolite.

Here we explore the correlation between cell-growth rates and resulting amplitudes of $0A^{-}P$ pulses. With combination of mathematical modeling and quantitative single-cell experiments we uncover the mechanistic basis of such correlation. As a result, we demonstrate a strikingly simple way for the sporulation network to sense and integrate information about nutrient in order to decide between continuing vegetative growth and committing to sporulation.

**8.2. Results**

**8.2.1. $0A^{-}P$ pulse amplitudes are correlated with cell growth rate**

To understand the dynamics of *B. subtilis* starvation response we used time-lapse microscopy to track single cells as they grow and sporulate in nutrient limited media. We found that in these conditions *B. subtilis* cells do not sporulate immediately upon exposure to starvation. Instead cells proceed with multiple rounds of vegetative division before eventually dividing asymmetrically and forming a spore. During this multi-cycle progression towards spore formation cell growth rate (inferred from cell elongation rate) gradually decreases (Fig. 8.1C).
To understand 0A activity dynamics in single cells during this period we used fluorescent reporters to measure gene expression from 0A~P-regulated promoters for 0A and spo0F ($P_{0A}$ and $P_{0F}$; see Methods). As expected from prior work, our measurements revealed that 0A activity (defined as production rate of the fluorescent reporter proteins) pulses during every cell-cycle in starvation conditions (Fig. 8.1E). $P_{0F}$ promoter activity similarly pulses once every cell-cycle in starvation conditions (Fig. A7.1). Our measurements further show that 0A activity pulse amplitudes increase gradually over multiple cell-cycles. Notably this increase in amplitudes coincides with the decrease in cell growth rate (compare Fig. 8.1C and E). Quantification of this relationship using a Spearman’s rank correlation showed a highly significant anti-correlation between 0A activity pulse amplitudes and cell growth rate ($\rho$=-0.8, p-value=4e-64, N=307; Fig. 8.1F). Moreover, as expected the cells that end up as spores (red dots on Fig. 8.1F have the highest 0A~P peaks and correspondingly slowest growth-rates.

What can explain this correlation? Given that the amplitude of Spo0A~P peaks is determined by the phosphorelay, one can propose two (non-mutually exclusive) mechanisms (Fig. 8.1G). In the first one, slow-down of growth directly affects the concentrations of phosphorelay proteins through growth-dependent changes in DNA replication, dilution, cell volume, and transcription/translation rates (342, 353). As a result, slow-down of growth could increase concentrations of phosphorelay proteins and lead to higher 0A activity. This type of indirect sensing is particularly appealing since it bypasses the need for any dedicated metabolite sensing. In the second mechanism, growth slowdown is correlated with accumulation of a certain intracellular signaling molecule (“unknown starvation signal”) which activates one of the phosphorelay components and leads to Spo0A~P accumulation. For example, it has been suggested that the autophosphorylation activity of KinA may be modulated by the binding of ATP and NAD$^+$ to its PAS-A domain (354, 355) although other studies indicate that PAS-domain is dispensable for the autophosphorylation activity (296, 356). The correlation of starvation signal with growth implies that starvation signal regulates
cell growth or cell growth controls accumulation of the starvation signal or that both starvation signal and growth rate are controlled by common upstream signal. In the following subsections we evaluate the importance of these hypothetical mechanisms using a combination of mathematical modeling and single-cell experiments.

8.2.2. Decrease in growth rate leads to accumulation of stable proteins.

As cell volume increases exponentially, growth rate acts as effective first order degradation constant of all cellular proteins (Appendix 7). As a result decrease in rate of growth leads to accumulation of constitutively produced proteins. For example, if a stable protein is produced at a constant rate of 10 molecules per min in a cell with 60 min generation time its steady state concentration would be 600 molecules/cell. If cell growth decreased to generation time of 300 min the concentration would increase 3000 molecules/cell. If on top of this, cell volume decreases for slower growing cells, as is the case for starving *B. subtilis* cells, then that would further increase the concentration of the corresponding proteins. Combining these two effect into a simple mathematical model we would predict the following dependence of concentration $C$ on growth rate $\mu$ (see Appendix 7 for derivation):

$$ C(\mu) = \frac{P}{V(\mu)(k_{\text{deg}} + \mu)} \quad [1] $$

Here $V(\mu)$ is a growth-rate-dependent cell volume, $k_{\text{deg}}$ is the rate of protein degradation or deactivation and $P$ is a rate of protein production which for simplicity is assumed to be growth-independent.

To test this model experimentally, we measured the single-cell level of fluorescent reporter protein at cells growing with different rates and compared the results to the predictions of Eq. [1]. Phosphorelay proteins have been shown to accumulate as cell slow-down their growth in starvation (102, 279, 329) but feedback loops from $0A-P$ do not allow us to separate the direct effects of
growth-rate slowdown from the effects of increased 0A~P. Therefore, we chose to test Eq. [1] by measuring the growth rate dependent change in the level of a stable fluorescent protein, YFP, expressed from an IPTG-inducible promoter ($P_{hsp}$) under different inducer concentration in starving B. subtilis. The results demonstrate that a decrease in growth rate during starvation can increase the level of stable proteins several fold and that relative increase is independent of the rate of production as data collapses upon normalization (Fig. 8.2A). When these results are compared to the model we can see that the trend can be fitted to Eq. [1] leading to the value of $k_{\text{deg}} = 0.12 \text{hr}^{-1}$ (about 6 hr half-life).

8.2.3. Accumulation of phosphorelay proteins with growth slowdown is sufficient to explain observed increase in 0A~P levels

To understand how growth rate affects the amplitude of 0A~P pulses we employed a detailed mathematical model of the sporulation phosphorelay that we have recently developed (350). This model successfully explained both the mechanism and timing of 0A~P pulsing but postulated an increase in KinA autophosphorylation rate to explain the increase in pulse amplitudes during starvation. Here, we instead chose to keep all the biochemical rate constants (including KinA autophosphorylation rate) fixed and instead study the effects of reduction of dilution rate and cell volume on 0A~P pulsing.

Our model simulations showed that these effects lead to increase of 0A~P amplitudes as growth slows down and cell cycles get longer (Fig. 8.2B). The results indicate that growth affects 0A~P pulsing in two major ways. At high growth rates, DNA replication period takes all or most of the cell cycle time, leaves little time for the $\text{kinA}:0F$ gene dosage to be at 1:1 ratio which in our model correspond to the time of 0A~P pulsatile increase. As a result of subsequent round of DNA replication, gene $\text{kinA}:0F$ dosage ratio returns to 1:2 ratio, the pulses are cut short before reaching their maximal amplitude. At lower growth rates, each pulse reaches its peak amplitude before being brought down by the negative feedback in the network. Moreover the decreases in cell volume
and dilution rate at these lower growth rates lead to an increase in phosphorelay protein concentrations and consequently higher 0A~P pulse amplitudes (Fig. 8.2B).

To determine how increase in phosphorelay protein levels affect 0A~P and identify the phosphorelay proteins that play the most significant role in this increases, we performed a sensitivity analysis of our model. To this end we chose to computationally change the production rates of individual phosphorelay proteins and examine how this affects 0A~P pulse amplitudes (Fig. 8.2C). Our model indicated that pulse amplitudes are most sensitive to KinA (~25% change in peak 0A~P with 10% change in KinA). Increase in 0F also has a strong effect, however in contrast to KinA, 0F decreases pulse amplitudes (increase in 0F by 10% decreases 0A~P peak by ~18%). Thus our mathematical model predicts that a slowdown of cell growth over multiple cell cycles leads to a gradual increase of 0A~P peak levels (Fig. 8.2D, solid line) mainly due to an increase in KinA concentration with decrease of cell volume and dilution.

Comparison of the model predictions (Fig. 8.2D, solid line) with experimental data from time-lapse microscopy (Fig. 8.2D, dots) showed that the effects of growth slow-down are sufficient to explain observed increases in peak amplitudes. The ultrasensitive dependence of 0A~P pulse amplitudes on growth rate is consistent with the trends in the data. Moreover examining the distribution of 0A~P peak activity and growth rate in the single cell data (Fig. 8.2D, histograms on the axes) we can see that values for sporulating and non-sporulating cells are well separated. This supports our hypothesis that the slowdown of growth can act as a starvation signal and control the sporulation decision.

Since a 0A activity threshold is known to control sporulation cell-fate (304), we expect from the results in Fig. 2D that this threshold would correspond to a threshold growth rate below which cells will sporulate. To test this, we constructed a cell-fate classifier by performing logistic regression to estimate the
probability of sporulation as a function of peak $P_{0A}$ promoter activity or as a function of growth rate. Our results demonstrated an ultrasensitive dependence of the probability of sporulation on both metrics (Fig. A7.2). This observed ultrasensitive dependence allows us to define a threshold value for $P_{0A}$ promoter activity above which cells sporulate with at least 50% probability. Similarly, we can estimate a threshold growth rate below which cells sporulate. We can use these thresholds to predict whether a given cell cycle will end-up producing a spore based on $P_{0A}$ promoter activity being above threshold (or, growth rate being below its respective the threshold). We found that the resulting predictions work very well, producing an accuracy of 85% for $P_{0A}$ promoter activity and 81% for growth rate. Altogether this analysis confirms that growth rate is indeed an accurate predictor of cell-fate.

Figure 8.2. Decrease in growth rate affects 0A–P levels and controls cell-fate.
A. Fold change in \( P_{\text{hsp}} \)-YFP fluorescence levels as a function of growth rate for different IPTG inducer concentrations. Black line represents a model fit for the effect of growth rate on the concentration of stable proteins.

B. Model time course of 0A~P during starvation showing cell-cycle coordinated pulsing. Upper and lower panels show the growth rate (input) and 0A~P response (output) respectively. Yellow bars and dashed lines represent DNA replication periods and cell division respectively. The effect of growth rate on protein levels was modeled following the results shown in A. The simulation predicts that 0A~P pulse amplitudes increase with decreasing cell growth rate.

C. Sensitivity of the growth rate – 0A~P pulse amplitude relationship to changes in the gene expression of phosphorelay components. Model simulations showed that 0A~P pulse amplitudes were most sensitive to changes in the production rate of kinase KinA. Increase in 0F caused a significant decrease in 0A~P pulse amplitude. 0A~P was found to be relatively insensitive to changes in other phosphorelay components.

D. Measurements of \( P_{0A-yfp} \) promoter activity confirm that 0A activity pulse amplitudes increase as growth rate decreases during starvation. Each dot corresponds to a single cell-cycle. Gray and red dots correspond to cell-cycles that end in vegetative division and spores respectively. Solid lines show the model predictions. Panels to the right and bottom show histograms of \( P_{0A-yfp} \) promoter activity and (growth rate)\(^{-1}\) for cell-cycles that end in vegetative division (gray) and spores (red) for each strain. Vertical green lines show the thresholds that can be used to predict cell fate in each case.

E. Receiver Operating Characteristic (ROC) curve for growth-based cell-fate prediction. Blue line shows the relation between False positive rate and True positive rate for different values of growth rate threshold. The black dot marks the optimal growth rate threshold that minimizes classification error. The high area under the ROC curve (AUC=0.91) indicates that growth rate is a highly robust predictor of sporulation.

To test the robustness of growth rate as a cell-fate predictor for sporulation we examined how the threshold value of this predictor affects its accuracy. We varied the growth rate threshold value (Fig. 8.2E) to calculate the Receiver Operating Characteristic (ROC) curve that provides the relation between false positive rate (fraction of cells below growth rate threshold that remain vegetative) and true positive rate (fraction of cells below growth rate
threshold that sporulate). The overall goodness of a predictor is usually characterized by the area under the ROC curve (AUC; Fig. 8.2E - shaded region). An AUC of 1 represents a perfect predictor whereas AUC=0.5 means the cell-fate prediction is no better than random (357). In our case, the area is 0.91 indicating that growth rate is a highly robust predictor of sporulation. Thus we conclude that a simple growth rate threshold (Fig. 8.2D, green vertical line) is an accurate and robust predictor of sporulation cell-fate.

8.2.4. Growth slowdown dynamics control sporulation deferral

In order to further test the effectiveness of growth rate as a predictor of cell-fates we investigated whether the multi-cell cycle deferral of sporulation can be explained simply by the dynamics of growth slowdown. To test this hypothesis, we tracked lineages of sporulating cells (Fig. 8.3A) to determine the fraction of cells that sporulated in each generation (Fig. 8.3B, red dots). We found that the fraction of cells that chose to sporulate gradually increased from ~10% in the first generation to about ~65% in the fourth generation and remained approximately constant thereafter. To determine whether the growth-rate threshold can explain this sporulation deferral, we quantified the distribution of cell growth rates for each generation of cells (Fig. 8.3C) and calculated the fraction of cells with growth rate below the sporulation threshold determined in Fig. 8.2C. The resulting prediction is in good agreement with experimentally observed fractions (Fig. 8.3B).

Next we perturbed growth dynamics with nutrient addition to test the ability of the growth-threshold model to explain cell-fate decisions in different conditions. Based on a simple population-dynamics model, we expected that the addition of a relatively small dose of nutrients to the medium at the start of the experiment would delay the onset of starvation but lead to a faster decrease in cell growth rates once the increasing cell population had depleted the additional nutrients (see Fig. A7.3, Appendix 7). In agreement with this model, experimental results (Fig.8.3D-F) showed that the addition of 0.0025% glucose indeed made
the cells grow faster initially, followed by the rapid decrease of growth rate. Our model posits that the growth rate threshold is a function of biochemical parameters of the phosphorelay and not affected by nutrient addition. Accordingly, we used the same threshold as before to calculate fractions of cells that should sporulate in each generation. The results are in excellent agreement with the observations (Fig.8.3E) confirming that the dynamics of cell growth slowdown control the deferral of sporulation. Notably, the same threshold value robustly predicts fates of cells sporulating early and late in our experiment regardless of initial nutrient addition. This observation calls into question direct modulation of phosphorelay protein activity by metabolites and instead further reinforce the idea that cell growth rate is the primary signal that determines cell-fate during starvation.

Figure 8.3. Growth slowdown controls sporulation deferral.
A. On normal sporulation media (Resuspension Media-RM) plates, sporulation timing (spores are marked by red dots) is heterogeneous and unsynchronized as demonstrated by a sample lineage.

B. Predictions of fraction of cells sporulating in each generation based on a growth-rate threshold (black line) shows good agreement with observed (red dots) fraction of sporulating cells for each generation. The growth rate threshold used was the same as threshold estimated from Fig. 8.2D.

C. Distributions of growth rates in each cell-cycle during starvation in RM media. Box plots indicate median growth rate (red line), 25-75% quintile (blue box) and the range (black whiskers) of growth rates for each generation. D-F Same panels as (A-C) but with 0.0025% of glucose is added to the plates.

D. Sample lineage shows that addition of a small amount of glucose delays sporulation for several generations but decreases the heterogeneity in sporulation timing.

E. Growth threshold model with the same growth rate threshold as (B) shows excellent agreement between predicted (black line) and observed (red dots) fraction of sporulating cells.

F. Addition of the glucose increases the initial cell growth rate and delays the onset of starvation but the resulting increase in cell number exhausts the nutrients leading to sharper decrease in growth rates.

In light of these results, we reasoned that if cell growth rate controls the cell-fate decision, selective induction of Oa~P pulses under conditions of very low growth rates should lead to immediate and synchronized sporulation without any significant deferral. To test this idea, we used an engineered strain iTrans-OF (Fig. 8.4A) in which we translocated the native copy of OF from its oriC-proximal location to the terminus and integrated an additional IPTG-inducible copy of OF close to the chromosome origin (350). Without IPTG, the inducible oriC-proximal copy of OF is inactive, kinA:OF gene dosage ratio remains 1:1 and DNA replication cannot trigger Oa~P pulses in this strain. As a result, this strain does not produce Oa~P pulses and does not sporulate without IPTG. Our modeling results predict that low-level induction of OF from the oriC-proximal locus introduces transient imbalance of gene dosage during DNA replication and
thereby rescues 0A~P pulsing (Fig. 8.4B). Moreover we found that in iTrans-0F upon 0F induction, 0A~P pulse amplitudes increase gradually over multiple cell-cycles as a function of cell growth rate similar to WT (compare Figs. 8.2B and 8.4B). The model also showed that if induction of 0F from the oriC-proximal locus is delayed till growth has slowed down, 0A~P amplitudes increase immediately without the multi-cell cycle delay (Fig. 8.4C). Thus we expect that cells should sporulate immediately upon delayed induction of 0F.
Figure 8.4. Selective exposure to slow growth eliminates sporulation deferral.

A. Arrangement of \textit{kinA} and \textit{OF} on the chromosome in the wildtype and \textit{iTrans-0F} strains. In the \textit{iTrans-0F} strain, \textit{0A~P} pulsing is only observed when the \textit{oriC}-proximal \textit{P}_{\text{hsp}}-\text{OF} is induced with IPTG.

B-C. Timing of induction of the \textit{oriC}-proximal copy of \textit{OF} affects \textit{0A~P} pulsing in the \textit{iTrans-0F} strain. Mathematical modeling results predict that if the \textit{oriC}-proximal \textit{P}_{\text{hsp}}-\text{OF} is induced at the start of starvation (B) pulse amplitudes increase gradually over multiple cell-cycles similar to WT. If \textit{P}_{\text{hsp}}-\text{OF} is induced later during starvation (C), when growth is
222

slow, 0A~P pulse amplitudes increase immediately without a multi-cell cycle delay. Yellow bars and dashed lines represent DNA replication periods and cell division respectively.

**D-E.** Experimental measurements of $P_{0A}$ activity show that both early and late induction of the oriC-proximal $P_{nsp}0F$ rescue 0A activity pulsing. These measurements also confirm that early induction leads to a gradual increase in 0A activity pulse amplitudes (D) and late induction leads to an immediate switch from no pulsing to high amplitudes (E).

**F-G.** Timing of the induction of 0F induction affects timing of sporulation in the iTrans-0F strain. Example lineage trees of the iTrans-0F strain show that sporulation timing (spores are marked by red dots) is heterogeneous and unsynchronized when the oriC-proximal $P_{nsp}0F$ is induced early at T0 (F) and well-synchronized when 0F is induced late at T36 (G).

**H.** Measurements of growth rate show that slowdown of growth during starvation is noisy but is not affected by whether 0F is induced early (T0 – gray) or late (T36 – orange). Dots and error-bars show means and standard deviations of growth rates for each generation respectively. Note that it takes ~5 generations for the mean growth rate to cross the threshold growth rate below which cells sporulate.

**I.** Histograms of the number of generations for which cells defer sporulation after 0F induction in the iTrans-0F strain. Gray and orange histograms show sporulation deferral for early (T0) and late (T36) induction respectively.

We tested this prediction by varying the timing of IPTG addition for the iTrans-0F strain. As shown in Fig. 8.4D-E, our measurements confirmed that 0A~P pulse amplitudes increase gradually if IPTG is added at the start of starvation whereas they increase rapidly if IPTG is added at later times after the growth rate has fallen below the sporulation threshold. Further, sporulation was deferred and its deferral was heterogeneous when IPTG was added early (Fig. 8.4F, I). In contrast, as predicted, when IPTG was added late, cells sporulated immediately IPTG addition without any deferral in a highly synchronized fashion (Fig. 8.4G, I). Crucially, the timing of IPTG addition affected the 0A activity dynamics but not the average growth dynamics (Fig. 8.4H), which rules out the
possibility that 0A~P controls changes in growth rather than growth controlling 0A~P.

Altogether these results provide further support for the hypothesis that slow-down of growth during starvation is the primary signal for sporulation and that it controls cell fate by modulating the amplitude of 0A~P pulse amplitudes.

8.2.5. KinA activity does not depend on the growth rate

Our results thus far indicate that the growth slowdown mediated increase in the concentration of stable proteins like KinA (Fig. 8.2A) combined with the predicted ultrasensitive dependence of 0A~P on KinA concentration (Fig. 8.2C) is sufficient to explain the observed correlation between growth rate and sporulation decision (Fig. 8.2D). However, these results are insufficient to exclude the possibility that in addition to these effects KinA activity is also modulated by an unknown signal that is correlated with growth rate. To check for this possibility we need to see whether the amount of KinA required for sporulation would be the same or different at different growth rates. However, in wild-type cells these effects are inseparable because of the one-to-one relationship between KinA level and growth rate under the feedback regulation of phosphorelay (289, 349).

To circumvent this problem we can use an engineered strain in which transcription of kinA is independent of 0A-activity. To this end, we can use a strain in which KinA is externally controlled by inducible promoter. Crucially, simulations of our mathematical model of the phosphorelay show that 0A~P-KinA feedback is not essential for pulsing (Fig. A7.4). Thus we can use an inducible system to do this test without affecting the pulsatile nature of 0A~P activation. For each fixed inducer concentration, our model shows that the cellular concentration of KinA increases with decreasing growth rate (similar to the increase in YFP shown in Fig. 8.2A). As a result, for each fixed inducer concentration there will be a KinA threshold and a growth rate threshold for sporulation. Examining how these thresholds depend on the KinA induction level would allow us to test whether KinA activity is also modulated by some unknown
starvation signal. Our modeling results predict that if KinA is activated by a growth-rate-correlated signal, different threshold levels of KinA will be required to reach the same threshold level of 0A~P. Therefore, we expect that KinA threshold would be higher for faster-growing cells (Fig. 8.5A and A7.4F) as compared to slower-growing cells. In contrast, if the growth rate only controls KinA concentration, the same threshold level of KinA is expected across all the different growth rates (Fig. 8.5B and A7.4G).

Following this rationale, we tested the two hypotheses by using a strain in which functional KinA-GFP fusion protein is expressed from IPTG-inducible promoter (\(P_{\text{hsp}}\)). GFP intensity measurements gave us information about single-cell KinA levels whereas tracking cell volume in time-lapse movies allowed us to compute single-cell growth rates. These measurements confirm that KinA levels are indeed sensitive to growth rate and that the resulting growth-rate threshold depends on the inducer level (i.e. expression rate from the \(P_{\text{hsp}}\) promoter; Fig. A7.5A). Moreover the growth dependent increase in KinA-GFP levels is well matched by model (see Eq. [1]) for protein accumulation.

Next using the same procedure employed in Fig. 8.2D, we performed logistic regression on the data for the inducible KinA strain to estimate the probability of sporulation as a function of KinA-GFP level or as a function of growth rate. Our results showed an ultrasensitive dependence of the probability of sporulation on both metrics irrespective of the IPTG level used (Fig. 8.5CD and Fig. A7.5BC). Using these ultrasensitive functions we identified threshold values for KinA-GFP above which and growth thresholds below which cells sporulate with at least 50% probability (Fig. A7.5A). Comparing the two types of thresholds for different IPTG concentrations (Fig. 8.5E) we found that the growth rate threshold for sporulation increases as a function of IPTG whereas the KinA-GFP threshold does not depend on IPTG level. Notably, the several fold increase in the growth rate threshold as a function of increase in \(kina\) expression rate (Fig. 8.5E) indicates that cells are able to sporulate at dramatically different levels of
starvation. However, the constant KinA threshold across these different levels of starvation indicates that KinA activity does not change across these conditions.

Figure 8.5. Decrease in growth rate affects OAp levels and controls cell-fate.

A. Standard hypothesis for the relationship between starvation, OAp and sporulation cell-fate. Starvation affects cell growth and an unknown signal that controls OAp activation by KinA. Growth rate and unknown signal level are inter-related since they both depend on starvation. Under this hypothesis the minimum level of KinA necessary to make cells sporulate is a function of the unknown signal level (middle panel) and by extension the growth rate (right panel). This hypothesis predicts that changes in KinA production rate using an IPTG inducible system should lead to a change in both the KinA threshold and growth rate threshold for sporulation (red dots).

B. Alternative hypothesis: signals controlling KinA activity play a negligible role in determining OAp activation and sporulation cell-fate. Under this hypothesis the KinA threshold for sporulation is independent of the unknown signal level (middle panel) and
by extension the growth rate (right panel). This hypothesis predicts that changes in KinA production rate using an IPTG inducible system should lead to a change in the growth rate threshold for sporulation but not the KinA threshold (red dots).

C,D. Logistic regression curves indicating the probability of producing a spore as predicted using measurements of KinA-GFP levels (C) and growth rate (D).

E. KinA and growth rate thresholds for sporulation in the P_hsp-KinA and WT strains. KinA and growth rate thresholds were calculated using the results of the logistic regression in C and D respectively. Note that the KinA threshold does not depend on IPTG level (One-way ANOVA: p-val=0.48) whereas the growth rate threshold decreases with increasing IPTG (One-way ANOVA: p-val<1e-3) as predicted by the alternative hypothesis in B. Errorbars show standard errors.

Thus our measurements showed that the threshold amount of KinA level necessary to trigger sporulation was not dependent on the level of IPTG. To determine if the same KinA threshold also controls sporulation in the wildtype strain we applied the same logistic regression analysis used above to a strain in which KinA-GFP fusion protein is expressed from its native promoter (Fig. A7.5A). Data for this strain showed an ultrasensitive dependence of probability of sporulation on both KinA-GFP levels and growth rates similar to the inducible strain (Fig. 8.5CD and Fig. A7.5BC). Further we found that threshold-level of KinA for this wild-type strain is the same as that for the IPTG inducible strain (Fig. 8.5E). These results imply that KinA activity is not growth rate dependent in our experimental conditions (Fig. 8.5A) and plays no role in controlling sporulation. Instead, these results show that sporulation is triggered by an increase in KinA concentration due to slowdown in growth rate, but independent of any signals modulating KinA activity (Fig. 8.5B).
8.3. Discussion

Previous studies of *B. subtilis* sporulation have primarily focused on identifying specific environmental and metabolic signal molecules that regulate the phosphorelay response. Here, taking a different approach, we have focused on the role of cell growth in determining the phosphorelay response and found that it actually plays a major part in controlling the response to starvation and sporulation decision.

Our results show that the amplitude of 0A−P pulses increases as cell growth rate decreases. This inverse dependence can be mechanistically understood using two observations about the phosphorelay network controlling 0A−P: (i) the concentration of the cytosolic kinase KinA is the rate-limiting factor that determines 0A−P pulse amplitude (102, 129, 304) and (ii) the concentration of KinA, a stable protein, increases as growth slows down. Here, by using a mathematical model to study the combined effect of these two, we were able to demonstrate how slowdown of growth leads to accumulation of KinA which in turn increases 0A−P amplitudes and consequently sporulation. Moreover since a 0A−P threshold is known to control sporulation cell-fate (130, 304), we found that there is a corresponding threshold growth rate below which cells sporulate. Crucially this growth threshold depends on the KinA production rate and can be increased several fold with an increase in *kinA* expression rate. Nevertheless the same level of KinA triggers sporulation regardless of conditions. Taken together, these results suggest that not only does cell growth play a major role in determining 0A activity and cell-fate but it might be the primary signal by which cells gauge their starvation level.

This role of growth in determining cell-fate offers fresh insight into the variability of sporulation timing during starvation. Several recent reports have shown that *B. subtilis* cells can defer sporulation for multiple generations during starvation and that this deferral period is highly heterogeneous (102, 263, 264). In agreement with these reports, our results show that sister cells frequently
adopt different cell fates during starvation (Fig. 8.3A). It has been suggested that this heterogeneity is the result of the stochasticity of gene expression. However our results show that while gene expression stochasticity may play a role, the heterogeneity of sporulation timing can be largely explained by the variability of cell growth rates during starvation. We have shown that the gradual but noisy decrease in growth rate during starvation is the reason sporulation is delayed. Perturbation of the dynamics of growth rate slowdown can increase or decrease the deferral of sporulation. Moreover the selective exposure of the phosphorelay to slow growth eliminates the deferral entirely. The reliability of growth rate as a predictor of cell-fate also calls into question previous suggestions that bet-hedging and memory of environmental conditions play a role in the sporulation decision (329, 358). Instead, our results suggest that at least in our conditions, B. subtilis cells eschew these complex decision strategies and instead favor making the sporulation decision based mainly on the nutrient availability during the current cell-cycle. Why growth rates are so variable in starvation conditions remains an open question and beyond the scope of this present report.

Even though our results clearly establish the key role played by cell growth rate in determining the starvation response they do not rule out other signals that may also affect the phosphorelay. Indeed, several proteins like Kipl, Sda, PhrA, YheH and Obg have been shown to respond to metabolic, environmental, cell density and cell-cycle related signals and modulate the phosphorelay response (359-364). It has also been proposed that starvation signal(s) acting on three PAS-repeats in the KinA N-terminal sensor domain control the autophosphorylation of the KinA’s C-terminal catalytic domain (354, 355, 365). However we should note that knocking out the genes of the proteins mentioned above does not significantly affect sporulation suggesting that they all act as ‘modulators’ rather than primary regulators of sporulation (102, 359, 364, 366). And recent studies suggest that the major role of the N-terminal PAS domain is to form a stable tetramer as an active form and the kinase is constitutively active regardless of culture conditions (296, 367, 368). In contrast,
as shown previously (129, 304) and recapitulated by our results here, KinA induction can override starvation requirements and force even cells that are not starving to sporulate. Notably, the fact that the KinA concentration necessary to trigger sporulation does not depend on growth rate (Fig. 8.5E) indicates that in our experimental setup KinA activity is mainly unmodulated. Therefore we can conclude that growth-dependent increase in KinA is the primary signal for sporulation whereas other signal and sensors only act as checkpoints and fine-tuners.

In summary, our results reveal a novel decision strategy based on the pulsatile 0A~P response to starvation: the phosphorelay only responds with a pulse upon completion of DNA replication and the amplitude of each pulse encodes the cell’s growth rate which is an indicator of the extent of starvation. This simple strategy allows cells to defer commitment to sporulation for as long as environmental conditions remain conducive to growth while bypassing the need for specific metabolite sensing. Notably, the sensitivity of stable protein levels to cell growth that enables this decision strategy is a universal feature of bacterial physiology and not unique to the sporulation network. As a result, similar growth rate dependent strategies for controlling the starvation response could very well be employed in a wide range of other systems.

8.4. Materials and Methods

8.4.1. Strain construction

Table A7.1 lists Bacillus subtilis strains used in this study. All strains are isogenic to B. subtilis PY79.

iTrans-0F strain construction

For the inducible 0F cassette, the spo0F gene was PCR amplified from B. subtilis PY79 with addition of optimal RBS and linker (AAGGAGGAAAGTCACATT) and including 43 bp fragment downstream. It was
ligated to a derivative of PLD30, JDE131 plasmid (SpR) next to the IPTG-inducible $P_{hyperspank}$ promoter (between HindIII and NheI restriction sites. This cassette was transformed into AmyE locus of a reporter B. subtilis strain constructed previously (345) harboring PspoIIR-YFP in SacA locus and Pspo0A-CFP on PHP13 low-copy plasmid. Then, we knocked out native spo0F in this strain. For the 0F deletion construct, the 5’ and 3’ fragments of 0F-adjacent genomic sequence were PCR amplified using the following primers: 

```
GAGGCGCCCTGTCGCTTTTCGTCACTTCCTCAG
TCGAATTCCGAAAAATACGAATGCCGTATTGATCATCAACGA
```

for 5’ arm, 

```
GATCTAGACATCGACGAAATCAGAGACGCCGTCAAAAAATATCTGCCC
TGAAGTCTAAAC
```

TCGTCGAC

for 3’ arm, and introduced into per449 (KanR) vector between XbaI and Sall restriction sites. The specific native genomic Spo0F knockdown in the resulting $i0F^{amyE}$ strain was confirmed by PCR and sequencing. Finally, the native promoter spo0F-rescue cassette was constructed and introduced as follows. Gene spo0F with its native $P_{spo0F}$ promoter (165 bp upstream) and 43 bp downstream sequence were PCR amplified from B. subtilis PY79 with addition of a terminator upstream (CCAGAAAGTCAAAAAGCCTCCGACCG) and ligated between BHI and XbaI restriction sites to ECE173 (PmR) integrating into GltA locus. This construct was transformed into the $i0F^{amyE}$ strain resulting in iTrans-0F strain.

**Other reporter strains, cloning vectors and transformation method**

The reporter strains for spo0A and spo0F promoters and the $P_{kinA}$-KinA-GFP and $P_{hsp}$-KinA-GFP strains were described previously (131, 283, 345). The strain expressing YFP from the IPTG-inducible $P_{hsp}$ promoter was constructed and characterized for dose-response earlier (369). To identify DNA replication periods in time-lapse experiments (Fig. A7.1F) we used a strain described previously that expresses a fluorescent DnaN-YFP fusion protein from the IPTG inducible $P_{hsp}$ promoter (132, 328, 350).
To facilitate segmentation in the glucose addition experiments (Fig. 8.3), we transformed a plasmid bearing a constitutively expressed promoter \( P_{pRI} \) driving mCherry expression (constructed as described in (370)) into the dual sporulation reporter strain harboring \( P_{spoIIR} \)-YFP at SacA genomic locus and \( P_{spo0A} \)-CFP expressed from the low-copy PHP13 plasmid, also described earlier ((283), "0A (5x)-IIR"). Promoter sequences were defined as follows: \( P_{pRI} \) - chromosomal sequence 2853257 to 2853567; \( P_{spo0A} \) - chromosomal sequence 2518060 to 2518350; \( P_{spoIIR} \) - chromosomal sequence 3794404 to 3794543 (345).

The vectors used in this study are as follows: ECE174, integrating into the \( sacA \) locus (constructed by R. Middleton and obtained from the Bacillus Genetic Stock Center); pLD30 designed to integrate into the \( amyE \) locus (kind gift from Jonathan Dworkin, Columbia University); ECE173, designed to integrate into the \( gltA \) locus (346) (constructed by R. Middleton and obtained from the Bacillus Genetic Stock Center); per449, a generic integration vector constructed for integration into the gene of interest (kind gift from Wade Winkler, UT Southwestern); and the bi-functional cloning plasmid PHP13 carrying the replication origin of the cryptic \( B. Subtilis \) plasmid pTA1060 (347). One-step \( B. subtilis \) transformation protocol was followed.

8.4.2. Culture preparation and Microscopy

Culture preparation

For imaging, \( B. subtilis \) culture was started from an overnight LB agar plate containing appropriate antibiotics (final concentrations: 5 μg/ml chloramphenicol, 5 μg/ml neomycin, 5μg/ml erythromycin, 5μg/ml phleomycin and 100 μg/ml spectinomycin). Strains containing multiple resistance genes were grown on a combination of no more than three antibiotics at a time. Cells were resuspended in casein hydrolysate (CH) medium (295) and grown at 37\(^{\circ}\)C with shaking. After reaching OD 1.8-2.0, cells were washed once and resuspended in 0.5 volume of Resuspension Medium (RM) (295). The resuspended cells were
grown at $37^0C$ for 1 hour, then diluted 15-fold and applied onto a 1.5% low-melting agarose pad made with RM-MOPS medium with desired IPTG or glucose concentration, if necessary. The pads were covered, left to air-dry for 1 hour at $37^0C$ and placed into a coverslip-bottom Willco dish for imaging. For the late IPTG addition experiment, a small drop (5 μl) of IPTG dissolved in RM was applied onto the RM-MOPS pad between image acquisitions through an opening in the dish lid.

*Time-Lapse Microscopy*

Differentiation of *B. subtilis* microcolonies was monitored with fluorescence time-lapse microscopy at $37^0C$ with an Olympus IX-81 inverted microscope with a motorized stage (ASI) and an incubation chamber. Image acquisition was set to every 20 min with a Hamamatsu ORCA-ER camera. Custom Visual Basic software in combination with the Image Pro Plus (Media Cybernetics) was used to automate image acquisition and microscope control.

*Image Analysis*

A combination of custom written MATLAB programs, MicrobeTracker (326) and freely available ImageJ plugins (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/, 1997-2014) was used to analyze microscopy data as described below.

**8.4.3. Data Analysis**

*Quantification of cell growth rates*

The mean cell-growth rate for individual cell-cycles (Figs. 8.1CF, 8.2AD, 8.3CF, 8.4F, A7.1ADE, A7.3 and A7.4A) was quantified using the measurements of cell-length in the time-lapse data. We first calculated the instantaneous cell-growth rate at every frame as:

$$
\mu(t) = \frac{1}{L(t)} \frac{dL(t)}{dt} = \frac{d\log L(t)}{dt} = \frac{\log(L(t + \Delta t)) - \log(L(t))}{\Delta t}
$$
where \( L(t) \) is the cell-length at time \( t \) and \( \Delta t \) is the time difference between successive frames (20 minutes). For cell-cycles that result in vegetative division, mean growth rate during a cell-cycle was defined as the average of \( \mu(t) \) over the cell-cycle duration. For cell-cycles that end in sporulation, the mean growth rate was the average of \( \mu(t) \) over the cell-cycle duration until the asymmetric division. Depending on the strain, the asymmetric division was defined either as \( P_{\text{spolIR}} \) activation (for the strains in which a \( P_{\text{spolIR}} \) reporter was present) or as the time frame two-hours before the appearance of the phase-bright forespore.

**Calculation of Promoter Activity**

The measurements of promoter activities for \( P_{\text{OA}}\text{-cfp/yfp} \) promoters (Figs. 8.1EF and 8.2CD) and the \( P_{\text{OF}}\text{-yfp} \) promoter (Fig. A7.1C) refer to rate of protein production calculated from fluorescence time-lapse data using the same procedure as previously reported (350).

**Quantification and characterization of promoter activity pulses**

Promoter activity time-series determined from fluorescence microscopy were smoothed using the MATLAB **smooth** function by employing a Savitsky-Golay filter with a 3rd-order polynomial over a sliding window of 5 frames. After smoothing, the maximum promoter activity during the cell-cycle was used as the Peak promoter activity (Figs. 8.2CD).

**Growth rate dependence of induced gene expression from \( P_{\text{hyperspank}} \) promoter**

The growth dependence of fluorescent protein levels in \( P_{\text{hsp}}\text{-yfp} \) and \( P_{\text{hsp}}\text{-kinA-gfp} \) were modeled with Equation 1 (Figs. 8.2A and A7.5A). In this equation, \( V(\mu) = A L(\mu) \) where \( A \) and \( L(\mu) \) are the cross-sectional cell area and cell length respectively. The cross-sectional area was assumed to be fixed and growth independent. The dependence of \( L(\mu) \) on growth was determined by fitting the data for change in cell length at division as a function of growth rate (Fig. A7.1E).
We used the following phenomenological expression for $L(\mu)$: $L(\mu) = 3.466 \exp(-0.689/\mu) + 3.743 \ \mu m$.

To isolate the growth dependence of gene expression in the $P_{hsp\_yfp}$ expressing strain, the fluorescence measurements were normalized by the concentration predicted by the fitted Equation at $\mu = \log(2) \text{hr}^{-1}$.

**Estimation of the growth rate, $0A$ activity and KinA thresholds for sporulation**

The thresholds for predicting sporulation cell-fate (Figs. 8.2D and 8.5CDE) were determined independently for each strain (and in the inducible KinA strain for each IPTG level). In every case, observations for individual cell-cycles included the independent variables: (i) the mean cell-cycle growth rate and (ii) either the peak $P_{0A}$ promoter activity (Fig. 8.2D) or the peak KinA-GFP level during the cell-cycle (Fig. 8.5CD); and the dependent variable: cell-fate (i.e. vegetative division=0 or sporulation=1). These observations were fit with a logistic function using the MATLAB `glmfit` function to predict the probability of sporulation as a function of each of the individual independent variables (Fig. 8.5CD, Fig. A7.2 and Fig. A7.5BC). Each outcome of the dependent variable, cell-fate, was assumed to be generated from a binomial distribution for the logistic regression. 95% confidence intervals of the logistic regression curves (Figs. A7.2 and A7.4BC) were also calculated using the `glmfit` function.

Threshold values of the independent variables were defined as the value at which the fitted logistic regression predicts probability of sporulation to be equal to 0.5. The `glmfit` function was also used to calculate the standard error of the thresholds. For the growth rate threshold, cell-cycles with growth rate greater than the threshold were classified as vegetative and those with growth rate lower than or equal to the threshold were classified as sporulating. For the peak $P_{0A}$ promoter activity or the KinA-GFP thresholds, cell-cycles with values greater than the threshold were classified as sporulating. These threshold values were then used to predict cell-fates and the calculate three performance measures: (i) False
Positive Rate (FPR, fraction of vegetative cells that were incorrectly predicted to sporulate), (ii) True Positive Rate (TPR, fraction of sporulating cells that were correctly predicted to sporulate) and (iii) Total error rate (fraction of total cells whose cell-fate was incorrectly predicted).

To determine the effectiveness of each threshold as a cell-fate prediction method we computed the receiver operating characteristic (ROC) curve for each case (Fig. 8.2E, Fig. A7.5D). This ROC curve is a commonly used way to characterize the performance of binary classifier in signal detection theory (371). We compute it by varying the threshold and calculating the False Positive Rate and True Positive Rates as functions of threshold value (371). Next we computed the AUC (area under the ROC curve) to estimate the performance of growth rate, peak $P_{PA}$ promoter activity and KinA-GFP level threshold based cell-fate predictions for the various strains. The computed AUCs and optimal thresholds for each case are summarized in Fig. A7.5D. The results indicate that all three variables can be used to robustly predict sporulation cell-fate for the WT and inducible KinA strains.

*Quantification of growth rate and sporulation fraction dynamics*

To calculate the growth and sporulation fraction dynamics based on generations (Fig. 8.3), first we binned the observations for individual cell-cycles of WT cells by generation number (the cell-cycle at the start of the time-lapse experiment was labeled generation 0). The fraction of cell-cycles in each generation bin that ended in sporulation were used to calculate the experimental sporulation fraction dynamics (red dots in Figs. 8.3B,E). Next, the WT growth rate threshold calculated from the data in Fig. 8.2D was used to classify the cell-cycles as vegetative (growth rate>threshold) or sporulating (growth rate<=threshold). The ratio of the number of cell-cycles classified as sporulating to the total number of cell-cycles in each generation bin was used as the predicted sporulating fraction (black curves in Figs. 8.3B,E).
To calculate the growth dynamics based on time (Fig. A7.3), first we divided the whole time-span of the time-lapse experiment into 2 hour time bins. Next, for the each cell-cycle in the experiment, its mean cell-cycle growth rate was calculated in all 2 hour time bins that were spanned by that cell-cycle. The mean and standard deviations of cell-cycle growth rates in each time bin were calculated to determine the colony growth rate dynamics. These results were then used with a population dynamics model to explain how nutrient addition affects the growth slowdown dynamics (Fig. A7.3).
Notes

This chapter is based on work that the following article that is currently under review:


*=Equal contributions
Regulation of the Bacillus subtilis general stress response
Chapter 9

Design and functional properties of the general stress response network in Bacillus subtilis

9.1. Introduction

Ability of cells to simultaneously respond to diverse stresses with an increase in large array of stress resistance genes is key to their survival and therefore present across all kingdoms of life (67, 68, 372). The common feature of these general stress response pathways is a single master regulatory factor that senses and coordinates a diverse stress regulon. For example, in mammals, p53 is activated by both DNA damage (373) and hypoxia (374), and coordinates expression of genes controlling DNA repair (375), cell cycle progression (75) and
angiogenesis (376). Similarly, in yeast Msn2/4 controls the response to glucose starvation (73), osmotic shock (74) and calcium stress (72). The two well-studied examples of general stress-response in bacteria include the alternative sigma factors RpoS in *Escherichia coli* (67, 69) and σ^B^ in *Bacillus subtilis* (68).

The σ^B^-mediated response is triggered by diverse energy and environmental stress signals and activates expression broad array of genes needed for cell survival in these conditions (68). σ^B^ activity is tightly regulated by a partner-switching network (Fig. 9.1A) comprising σ^B^, its antagonist, anti-sigma factor RsbW, and anti-anti-sigma factor RsbV. In the absence of stress, RsbW-dimer (RsbW_2) binds to σ^B^ and prevents its association with RNA polymerase thereby turning OFF the σ^B^ regulon. Under these conditions RsbW_2 inactivates most of RsbV by using its catalytic to phosphorylate RsbV. Phosphorylated form (RsbV~P) has a low affinity for RsbW_2 and can’t effectively interact with it (377). However in the presence of stress signals, RsbV~P is dephosphorylated by the stress-sensing phosphatases RsbQP and RsbTU. Dephosphorylated RsbV attacks the σ^B^-RsbW_2 complex to induce σ^B^ release, thereby activating the σ^B^ regulon expression (378, 379). Notably, genes encoding σ^B^ and its regulators lie within a σ^B^-controlled operon (380), resulting in positive and negative feedback loops.

Recently, it was shown that under energy stress conditions σ^B^ is activated in a series of transient pulses increasing in frequency as stress becomes more severe (103). It has also been shown that this pulsatile activation of σ^B^ is sensitive to the rate of stress-signal increase (381). As a result, σ^B^ pulse amplitudes can encode the rate of increase in stress inducing agents such as ethanol (381). While its clear, that the pulsatile activation of σ^B^ is routed in the complex architecture of its regulatory network (Fig. 9.1A) its mechanism is not fully understood. Previous models of the σ^B^ network either did not produce the pulsatile response (382) or make ad hoc simplifications of the network topology somewhat inconsistent with experimentally observed details (103). As a result, it
remains unclear which design features of the $\sigma^B$ network enable its functional properties.

To address these issues we develop a detailed model of $\sigma^B$ network and examine its dynamics to understand the mechanistic principles underlying the pulsatile response. By decoupling the post-translational and transcriptional components of the $\sigma^B$ network we show that an ultrasensitive negative feedback is the basis for $\sigma^B$ pulsing. Moreover we find that the relative stoichiometry of $\sigma^B$, RsbW and RsbV synthesis rates plays a critical role in determining the qualitative nature of the $\sigma^B$ response. We also use our model together with experimental data to explain how the $\sigma^B$ network is able to encode the rate of stress increase and the size of stochastic bursts of stress phosphatase into the amplitudes of $\sigma^B$ pulses.

![Figure 9.1. $\sigma^B$ general stress response network.](image)

A. Network diagram of the $\sigma^B$ general stress response. Energy and environmental stresses activate the stress-sensing phosphatases RsbQP (QP) and RsbTU (TU) which in turn activate $\sigma^B$ by dephosphorylating RsbV which releases $\sigma^B$ from the $\sigma^B$-RsbW$_2$ complex. Free $\sigma^B$ activates the production of many stress-response genes and the $\sigma^B$-operon that includes RsbV, RsbW and $\sigma^B$. 
B. Dynamics of free $\sigma^B$ and $B_T$ in response to a step-increase in phosphatase concentration.

C. Representation of the pulse response in (B) in the $\sigma^B$-$B_T$ phase plane. Blue and cyan lines show the decoupled post-translational response at high and low phosphatase respectively. The black line shows the transcriptional response ($B_T=F_T(\sigma^B)$; $B_T$ as a function of $\sigma^B$). The step increase in phosphatase causes a shift in the steady-state post-translational response (from low phosphatase-cyan to high phosphatase-blue). This shift triggers a response in which the system moves from the initial steady state (gray dot) to the new steady state (black dot) via pulsatile trajectory (green curve).

D. Dependence of $\sigma^B$ post-translational response on the stoichiometry of $\sigma^B$ operon expression. The parameter space of $\sigma^B$ operon stoichiometry defined by $\lambda_W$ and $\lambda_V$ is divided into three Regions (I, II, III) by $d\log\sigma^B/d\log B_T$, i.e. the sensitivity of free $\sigma^B$ to total $\sigma^B$ concentration ($B_T$). $d\log\sigma^B/d\log B_T>0$ in Region I, $d\log\sigma^B/d\log B_T<0$ in Region II and $d\log\sigma^B/d\log B_T\sim0$ in Region III. The blue area denotes a subsection of Region II where the post-translational response is ultrasensitive, $d\log\sigma^B/d\log B_T<-1$. The combined post-translational and transcriptional network forms a positive feedback in Region I, an ultrasensitive negative feedback that allows pulsing in Region II and a non-responsive system in Region III (free $\sigma^B$ is always low). Red circles mark the selected values of $(\lambda_W, \lambda_V)$ used in panels E-G.

E-G. Simplified view of the decoupled post-translational and transcriptional components of the $\sigma^B$ network in Regions I-III. Blue and black lines show the post-translational response ($\sigma^B=F_P(B_T,P_T)$; $\sigma^B$ as a function of $B_T$) and the transcriptional response ($B_T=F_T(\sigma^B)$; $B_T$ as a function of $\sigma^B$) respectively. Gray dots mark the steady states of the full system. Dashed yellow lines show linear asymptotic approximations of the post-translational response (see text). The dashed vertical line in (G) shows the critical level of operon expression beyond which free $\sigma^B$ decreases as a function of $B_T$.

We further develop this model to investigate how the network functions in the context of other sigma factors. As in many other bacteria, $\sigma^B$ is one of the many sigma-factor that complex with RNA-polymerase core and activate their respective regulons. When induced, alternative sigma factors therefore compete for RNA polymerase with one another the housekeeping sigma factor $\sigma^A$. We use
our model to investigate how the ultrasensitive negative feedback design enables the network to function even in the presence of competition from $\sigma^A$ which has a significantly higher affinity RNA polymerase (383). Lastly, we investigate how more multiple stress sigma factors compete or cooperate when cells are exposed to multiple stresses simultaneously. Using our model we identify design features that are ubiquitous in stress sigma factor regulation and critical to bacterial survival under diverse types of stresses.

9.2. Results

9.2.1. Biochemically accurate model of $\sigma^B$ pulsing mechanism

A recent study demonstrated that a step increase in energy stress results in pulsatile activation of $\sigma^B$ and that increasing stress resulted in higher pulse frequencies (103). The study also proposed a minimal mathematical model of the network which reproduced pulsing in $\sigma^B$. However this model included several ad hoc assumptions: (i) Michaelis-Menten kinetics for phosphorylation and dephosphorylation reactions, (ii) $\sigma^B$ and RsbV represented as single lumped variable rather than separate species and (iii) partner-switching, and the formation and dissociation of various RsbW$_2$ complexes were ignored. Though this minimal model was able to reproduce experimental observations, it depicts a biochemically inaccurate picture of the $\sigma^B$ network. Consequently it cannot be used to uncover the design features that enable the observed pulsatile $\sigma^B$ response.

To understand the response of the $\sigma^B$ network we built on our earlier study (382) to develop a detailed model that explicitly includes all known molecular interactions in the network. We note that we made one significant change to the model discussed in (382). The model in (382) assumed that the synthesis rates for total $\sigma^B$ and its network partners (RsbW and RsbV) follow their stoichiometric binding ratios (i.e. $\text{RsbW}_{\text{tot}}/\sigma^B=\lambda_W=2$ and $\text{RsbV}_{\text{tot}}/\sigma^B=\lambda_V=2$). In contrast, following
the experimental measurements of (384) we assumed that $\sigma^B$, RsbW and RsbV are produced in non-stoichiometric ratios ($\lambda_w=4$ and $\lambda_v=4.5$).

Simulations of this detailed model showed that it is able to reproduce the experimentally observed response of the $\sigma^B$ network. Specifically we found that a step-increase in phosphatase concentration led to an increase in free $\sigma^B$ followed by an increase in the concentration of total $\sigma^B$ concentration ($B_T$) (Fig. 9.1B). Thereafter free $\sigma^B$ levels decreased thus forming the pulsatile response observed in experiments.

To understand this pulsatile response, we chose to decouple the transcriptional and post-translational responses of the network in our model. By varying the $\sigma^B$ operon transcription rate, while keeping the ratio of operon component synthesis rates ($\lambda_w$, $\lambda_v$) fixed we were able to calculate the post-translational response of the $\sigma^B$ network $\sigma^B=F_P(B_T,P_T)$, which provided the free $\sigma^B$ concentration as a function of total $\sigma^B$ concentration ($B_T$) and total phosphatase concentration ($P_T$). In parallel, we calculated the transcriptional response $B_T=F_T(\sigma^B)$, by using the free $\sigma^B$ concentration as an independent variable to calculate the total concentrations of $\sigma^B$ (RsbW and RsbV concentrations are proportional to $B_T$). Under this analysis framework, the steady state of the complete closed loop network can be determined by simultaneously solving the post-translational and transcriptional functions, $\sigma^B=F_P(B_T,P_T)$ and $B_T=F_T(\sigma^B)$, for $\sigma^B$ and $B_T$.

Using this decoupling method we found that free $\sigma^B$ is a non-monotonic function of $B_T$ (Fig. 9.1C). At low $B_T$, free $\sigma^B$ increases as a function of $B_T$ because RsbW is sequestered in the $W_2V_2$ complex. However at higher $B_T$, the kinase flux dominates the phosphatase flux and the anti-sigma factor RsbW$_2$ dimer sequesters $\sigma^B$ in the $W_2\sigma^B$ complex such that free $\sigma^B$ decreases ultrasensitively as a function of $B_T$ which allows the formation of a negative feedback loop in the $\sigma^B$ network. Notably, negative feedback is one of the few network motifs that is known to be capable of producing adaption-like pulsatile
responses (334). This explains why a step increase in phosphatase concentration in our model simulations leads to a pulse of $\sigma^B$ activity. Plotting the trajectory of the $\sigma^B$ pulse on the $(\sigma^B, B_T)$ plane over the post-translational and transcriptional responses (Fig. 9.1C) illustrates the mechanism driving this pulsatile response. Starting at the initial steady state, an increase in phosphatase shifts the ultrasensitive post-translational response so that free $\sigma^B$ is rapidly released from the $W_2\sigma^B$ complex whereas $B_T$ remains relatively unchanged. Once the free $\sigma^B$ level matches the post-translational response it stops increasing. The increase in $\sigma^B$ operon transcription eventually causes accumulation of $B_T$ (and the anti-sigma factor RsbW). This in turn forces the $\sigma^B$ level to decrease following the post-translational response curve to the new steady state which has very little free $\sigma^B$ thereby completing the $\sigma^B$ pulse.

This graphical illustration of the $\sigma^B$ pulse using the decoupled transcriptional and post-translational responses offers mechanistic understanding of the response and facilitates the identification of critical design features. First, our simulation results explain why no pulsatile response occurs in strains in which $\sigma^B$ operon is transcribed constitutively (103). In this case, the $\sigma^B$ network lacks the negative feedback necessary to produce a pulsed response. A step increase in phosphatase still leads to an increase in free $\sigma^B$ due to the change in the post-translational response however this not followed by an increase in $B_T$ (Fig. A8.1A). Consequently an increase in phosphatase results in a monotonic increase in free $\sigma^B$ rather than a pulse (Fig. A8.1A).

Our decoupling method also enables the calculation of the dependence of $\sigma^B$ pulse amplitude on the phosphatase level and transcriptional and post-translational parameters. Specifically, we found that $\sigma^B$ pulse amplitude is a threshold linear function of phosphatase concentration (Fig. A8.2A). Below the phosphatase threshold, the post-translational response $\sigma^B=F_P(B_T, P_T)\approx0$ and is insensitive to $B_T$ (Fig. A8.2B). As a result, the full system lacks the negative feedback and $\sigma^B$ does not pulse. Above the threshold, $\sigma^B$ pulse amplitudes increase linearly as a function of phosphatase. Our results show that the
threshold value of phosphatase depends on the basal level of $\sigma^B$ operon transcription and the binding affinity of $\sigma^B$ and RsbV for the RsbW$_2$ dimer (Fig. A8.2B). This binding affinity controls the ultrasensitivity of the post-translational dependence of $\sigma^B$ on $B_T$ in a manner similar to the molecular titration effect (385). Consequently, a decrease in this binding affinity decreases the phosphatase threshold for activating $\sigma^B$ pulsing (Fig. A8.2C). The sensitivity of pulse amplitudes to phosphatase above the threshold is also determined by post-translational parameters of the $\sigma^B$ network. In particular, the ratio of the catalytic rate constants of the phosphatase and kinase ($k_p/k_k$) that controls the ultrasensitivity of the post-translational response $\sigma^B = F_P(B_T, P_T)$. An increase in this ratio increases the ultrasensitivity and thereby the slope of the linear dependence of $\sigma^B$ pulse amplitudes on phosphatase (Fig. A8.2D).

9.2.2. $\sigma^B$ response dynamics depend on operon relative stoichiometry

Next we further investigated the post-translational response of the $\sigma^B$ network and found that the ratio of operon component synthesis rates ($\lambda_W$, $\lambda_V$) plays a critical role in determining both qualitative and quantitative features of $\sigma^B = F_P(B_T, P_T)$. As shown in Fig. 9.1D, the ($\lambda_W$, $\lambda_V$) parameter space can be divided into three regions based on qualitative differences in the post-translational response. In Region I ($\lambda_W <= 2 + \lambda_V k_{deg}/k_k$), the sensitivity of free $\sigma^B$ concentrations to $B_T$ is positive (dlog($F_P$)/dlog($B_T$)>0) for all phosphatase concentrations, indicating that free $\sigma^B$ increases as a function of $B_T$ irrespective of $P_T$. In contrast, in Region II ($2 + \lambda_V k_{deg}/k_k < \lambda_W < \lambda_V + 2$), excepting very high phosphatase concentrations, free $\sigma^B$ actually decreases ultrasensitively as a function of $B_T$ (dlog($F_P$)/dlog($B_T$)<<-1). Finally in Region III ($\lambda_W > \lambda_V + 2$), free $\sigma^B$ concentration is very low and insensitive to $B_T$ (dlog($F_P$)/dlog($B_T$)~0) irrespective of $P_T$. Figures 9.1E-G illustrate the differences in the post-translational responses in these three regions for three selected combinations (red dots on Fig. 9.1D) of stoichiometric ratios ($\lambda_W$, $\lambda_V$).
For $\lambda_W=2$ and $\lambda_V=2$ (Region I; Fig. 9.1E), free $\sigma^B$ increases as a function of $B_T$ in biphasic fashion. At low $B_T$, most of the phosphatase flux dominates the kinase flux and most of the anti-sigma factor RsbW is sequestered in the $W_2V_2$ complex thereby freeing up $\sigma^B$ such that $\sigma^B = B_T$. However at high enough $B_T$ ($B_T > 2k_\rho P_T/k_w/\lambda_W + 2B_0(P_T)/\lambda_W$), the kinase flux dominates the phosphatase flux and the anti-sigma factor RsbW$_2$ dimer sequesters some of the in the $W_2\sigma^B$ complex such that free $\sigma^B = (1+k_{deg}\lambda_V/k_k-\lambda_W/2)B_T + B_0(P_T)$. Here $k_\rho$, $k_k$ and $k_{deg}$ are the phosphatase, kinase and dilution rate constants. Note that irrespective of $B_T$, there is never enough $W_2$ dimer to sequester all $\sigma^B$ since $W_T <= 2B_T$ in this region. As a result, $\sigma^B$ always increases as a function of $B_T$ and since $\sigma^B$ transcriptionally activates $B_T$ production, in this region a positive feedback loop exists in the $\sigma^B$ network.

For $\lambda_W=8$ and $\lambda_V=4.5$ (Region III; Fig. 9.1F), $W_T > 2B_T + V_T$ at all values of $B_T$. As a result, irrespective of whether phosphatase is present or not there is always enough RsbW$_2$ dimer to sequester all $\sigma^B$ in the $W_2\sigma^B$ complex such that $\sigma^B/B_T = 0$.

Finally for $\lambda_W=4$ and $\lambda_V=4.5$ (Region II; Fig. 9.1G), free $\sigma^B$ is a non-monotonic function of $B_T$. At low $B_T$ ($B_T < 2B_0(P_T)/\lambda_W$), similar to region I, RsbW is sequestered in the $W_2V_2$ complex thereby freeing up $\sigma^B$ such that $\sigma^B = B_T$. However at higher $B_T$ ($B_T > 2B_0(P_T)/\lambda_W$), the kinase flux dominates the phosphatase flux and the anti-sigma factor RsbW$_2$ dimer sequesters some of the in the $W_2\sigma^B$ complex such that free $\sigma^B = (1+k_{deg}\lambda_V/k_k-\lambda_W/2)B_T + B_0(P_T) = B_0(P_T) - (\lambda_W/2-1)B_T$. Finally at even higher $B_T$, ($B_T > B_0(P_T)/(\lambda_W/2-1)$) all $\sigma^B$ is sequestered in the $W_2\sigma^B$ complex such that $\sigma^B/B_T = 0$. Thus, since $W_T > 2B_T$ in region II, for a range of $B_T$, $\sigma^B$ decreases ultrasensitively as a function of $B_T$ which allows the formation of a negative feedback loop in the $\sigma^B$ network.

Notably simulations of response to phosphatase step increase in Regions I and III of the $(\lambda_W, \lambda_V)$ parameter space do not produce $\sigma^B$ pulse (Fig. A8.1BC). This absence of pulsing can be explained by the lack of negative feedback in
these regions. Thus our model shows that the stoichiometric ratio of $\sigma^B$ operon expression rates and post-translational sensitivity of $\sigma^B$ to operon expression levels play a critical role in determining the qualitative nature of the $\sigma^B$ response. In the following sections, we used our model to further investigate the design-function relationship of the $\sigma^B$ network.

9.2.3. $\sigma^B$ network encodes phosphatase burst size into pulse amplitudes

Next we tried to understand how the fluctuations in phosphatase affect $\sigma^B$ pulses. Based on previous theoretical (199, 386) and experimental (57) studies we assume that the phosphatase level follows a gamma distribution which is described by two parameters - burst size ($b$, average number of molecules produced per burst) and burst frequency ($a$, number of bursts per cell cycle). The mean phosphatase in this case is the product of burst size and burst frequency ($<P_T>=ab$). Thus, stress can increase mean phosphatase by either changing burst size or burst frequency.

From the results of (103) it is unclear whether the increase in phosphatase at high stress is the result of increased mean burst size or burst frequency. First, we performed stochastic simulations in which mean phosphatase concentration was varied by changing burst size. These simulations showed that it is able to reproduce all the experimentally observed features of the $\sigma^B$ pulsatile response. Specifically our results show that stochastic bursts in stress phosphatase levels lead to pulses of $\sigma^B$ activity (Fig. 9.2A). Moreover, consistent with the experimental observations of (103), this model shows that the amplitude of $\sigma^B$ pulses is a linear function of the stress phosphatase level (Fig. 9.2B). Finally we found that stress-mediated increases in phosphatase concentration lead to an increase in the frequency of $\sigma^B$ pulsing (Fig. 9.2C) and an ultrasensitive increase in the level of $\sigma^B$ target expression (Fig. 9.2D).
Figure 9.2. Pulsatile response of the σB network to stochastic phosphatase bursts.

A. Simulation results of the detailed σB network model show that stochastic bursts in stress-sensing phosphatases RsbQP (QP) levels lead to pulses of σB target promoter activity. Light and dark green lines show sample trajectories from stochastic simulation at high and low stress respectively.

B. σB pulse amplitudes increase linearly as a function of mean phosphatase level. Green dots and errorbars show mean and standard deviations calculated from stochastic simulations. Black line shows a linear fit.

C. σB pulse frequency increases ultrasensitively as a function of mean phosphatase level. Green dots show the mean pulse frequency calculated from stochastic simulations. Black line shows a Hill-equation fit with $n_{Hill}>2$.

D. Mean σB target expression increases ultrasensitively as a function of mean phosphatase level. Green dots show the mean pulse frequency calculated from stochastic simulations. Black line shows a Hill-equation fit with $n_{Hill}=2$.

Stochastic simulations the σB network where mean phosphatase concentration was varied by changing burst frequency also led to an increase σB pulsing. A key difference was that σB pulse amplitude increases ~5-fold for burst size modulation (Fig. 9.2B), whereas it remains constant for burst frequency modulation (Fig. A8.3A). Additionally, we found that in the case of burst size the increase in σB pulse frequency is non-linear, whereas in the case of burst frequency the increase is linear (compare Figs. 9.2C and A8.3B). Similar to burst size modulation burst frequency modulation led to a non-linear increase in the level of σB target expression (Figs. 9.2D and A8.3C). Comparison of these results to the experimental observations reported in (103) suggests that phosphatase
concentration increase at high stress are primarily the result of increasing burst size.

To definitively establish which one of mean burst size or burst frequency controls the pulse response we next examined the cumulative histograms of pulse amplitudes at different phosphatase concentrations. These histograms carry different signatures for burst size or burst frequency encoding. Our results show that since the distribution of phosphatase burst sizes does not change when only burst frequency is increased, neither does the distribution of the resulting pulse amplitudes (Fig. A8.3D). In contrast if phosphatase levels are controlled by changing mean burst size then the distribution of phosphatase burst size changes as mean phosphatase is increased and so does the distribution of pulse amplitudes (Fig. A8.3E). Consequently, for burst frequency encoding but not for burst size encoding, the cumulative histograms of pulse amplitudes overlap when normalized. Applying this test to the data from (103), we found that the normalized cumulative pulse amplitudes histograms do not in fact overlap (Fig. A8.3F). These results confirm that the σB network follows the mean burst size modulation strategy and encodes phosphatase burst size into σB pulse amplitudes.

9.2.4. σB network encodes phosphatase ramp rates into pulse amplitudes

We also used our model to study the response of the σB network to different rates of stress increase. Previously it has been shown that for ramped increases in stress, σB pulse amplitude depends on the rate of stress increase (381). To replicate this in silico we modeled ramped stress with ramped increase in phosphatase concentration. Our simulations showed that the detailed model is indeed able to capture the effect of rate of stress increase on σB pulse amplitudes (Fig. 9.3A). Specifically for a fixed phosphatase increase, the pulse amplitudes decreased non-linearly as a function of duration of the phosphatase ramp (Fig. 9.3B, E).
Figure 9.3. Rate sensitivity of the $\sigma^B$ pulsatile response.

A. Ramped increases in phosphatase concentration used as model input for $\sigma^B$ network to simulate different rates of stress increase.

B. $\sigma^B$ pulse amplitudes in the wildtype model ($k_{\text{deg}}=0.72\text{hr}^{-1}$) resulting from the ramped increases in phosphatase concentration shown in (A). Note that the pulse amplitude decreases as phosphatase ramp durations increase.

C,D. $\sigma^B$ pulse amplitudes resulting from the ramped increase in phosphatase concentration shown in (C) for various degradation/dilution rates (D). Note that the pulse amplitudes decrease as degradation/dilution rates ($k_{\text{deg}}$) increase.

E. $\sigma^B$ pulse amplitudes as a function of phosphatase ramp duration for various values of $k_{\text{deg}}$. The non-linear dependence of amplitudes on ramp durations is characterized by the half-maximal constant, $K_{\text{ramp}}$, which increases as $k_{\text{deg}}$ decreases. Points and solid lines represent simulation results and Hill-equation fits respectively. Colors of points represent different $k_{\text{deg}}$ values as in (D).

F. $K_{\text{ramp}}$, the half-maximal constant of the non-linear dependence of amplitudes on ramp durations as a function of $k_{\text{deg}}$. 
We hypothesized that this ramp rate encoding is the result of the timescale separation between the fast post-translational and the slow transcriptional responses of the σ^B network. During the pulsed σ^B activation, post-translational response is rate-limited by the phosphatase ramp. In contrast, the transcriptional response is slow and its rate is set by the dilution rate of σ^B operon proteins. For a step increase in phosphatase the fast post-translational response ensures that σ^B reaches its post-translational steady state before the slow increase in RsbW is able to sequester and turn off the pulse (Fig. 9.3AB). However for a ramped increase in phosphatase the post-translational increase in σ^B is limited by the rate of phosphatase increase which allows RsbW to catch up earlier and terminate the σ^B pulse thereby decreasing the pulse amplitude. To test this, we varied the dilution rate and proportionally changed the rate of transcription of the σ^B operon to ensure that the total concentrations of σ^B, RsbW and RsbV are kept fixed (Fig. 9.3CD). Our simulations showed that the dependence of pulse amplitudes on ramp duration was indeed sensitive to the dilution rate (Fig. 9.3EF). This suggests that the time-scale separation between the post-translational and transcriptional responses is in fact the basis of ramp rate encoding into pulse amplitudes.

9.2.5. The design of the σ^B network enables it to compete with σ^A for RNA polymerase

The results thus far indicate that σ^B network essentially working in the negative feedback regime in which increase in the operon expression would decrease σ^B activity. Noting, that in other systems such negative feedbacks were shown to increase the robustness of the system to perturbations we decide to investigate how the design of the σ^B network affects its performance when there is competition for RNA polymerase from other sigma factors like the housekeeping sigma factor σ^A. Since σ^A has a much higher affinity for RNA polymerase (383), a small increase in σ^A can dramatically increase the amount of σ^B necessary to activate the transcription of the σ^B regulon. Thus changes in σ^A can alter the effective input-output relationship of a stress-response sigma factor
like $\sigma^B$ (Fig. A8.4) and thereby adversely affect the survival ability cells under stress.

To understand how the design of the $\sigma^B$ network handles competition for RNA polymerase, we expanded our model of the network to explicitly include $\sigma^A$, RNA polymerase (RNApol) and $\sigma$–RNApol binding (Fig. 9.4A). Since post-translational reactions only depend on the concentrations of $\sigma^B$ operon components and phosphatase, this inclusion of $\sigma^A$ did not change the post-translational function relating free $\sigma^B$ to $B_T$ and $P_T$. In contrast, looking at the transcription function, our model showed that an increase in $\sigma^A$ decreased the effective affinity of $\sigma^B$ for RNApol and as a result, higher levels of free $\sigma^B$ are necessary to achieve the same production rate of $B_T$ or $\sigma^B$ target genes.

Next we looked at the response of the full model, including both transcriptional and post-translational components to changes in $\sigma^A$ in the presence of the stress signal i.e. stochastically fluctuating RsbQP phosphatase levels. Our simulation results showed that phosphatase bursts lead to pulses of free $\sigma^B$ and $\sigma^B$ target promoter activity (Fig. 9.4BC) similar to the results in Fig. 9.2. The pulsatile $\sigma^B$ response is present since the full system still includes a negative feedback loop between $\sigma^B$ and $B_T$ (Fig. 9.4A). Surprisingly our results also showed that the pulse amplitudes of $\sigma^B$ target promoter activity are not affected by ~30% increase in $\sigma^A$ (Fig. 9.4C). This despite the fact that since RNApol level was fixed at 10µM, the increased competition significantly increases the effective half-maximal constant of $\sigma^B$ target expression rate.

The surprising insensitivity of the phosphatase-$\sigma^B$ target dose response to RNApol competition is the result of the ultrasensitive negative feedback between $\sigma^B$ and $B_T$. Owing to the ultrasensitivity of this feedback, a small decrease in $B_T$ resulting from the increase in $\sigma^A$ causes a major increase in $\sigma^B$ pulse amplitude (Fig. 9.4B, D). This increased amplitude compensates for the increased competition for RNApol and insulates the network from the perturbations (Fig. 9.4DE).
Figure 9.4. Negative feedback insulates the σ^B response from competition with housekeeping sigma factor σ^A.

A. Simplified network diagrams of stress sigma factor σ^B competing with housekeeping sigma factor σ^A RNA polymerase. In all cases, a σ^B phosphatase controls the stress-signal driven activation of σ^B.

B. Trajectories of free σ^B response to stochastic phosphatase input for all three networks at two different levels of σ^A (σ^A=9µM-low competition for RNA polymerase; σ^A=12µM-high competition for RNA polymerase).

C. Output trajectories of σ^B target promoter activity in response to stochastic phosphatase input for all three networks at two different levels of σ^A.
**D, E.** Mean free $\sigma^B$ concentration (D) and mean $\sigma^B$ target promoter activity (E) as a function of $\sigma^A$ for all three networks at fixed mean phosphatase (mean $P_T=0.5$µM). Gray vertical line shows the total RNA polymerase levels which was fixed at 10 µM.

To test the importance of the negative feedback in insulating the network we compared the response of the wildtype network to two mutant networks: one that lacks any feedback between $\sigma^B$ and $B_T$ and one that has a positive feedback between $\sigma^B$ and $B_T$ (Fig. 9.4A). As shown in Fig. 9.4BD, unlike the wildtype response, increase in $\sigma^A$ did not affect or decreased the free $\sigma^B$ response of the no feedback and positive feedback networks respectively. Accordingly the increased competition for RNApol at higher $\sigma^A$ results in reduced $\sigma^B$ target promoter activity in these networks (Fig. 9.4CE). Thus fluctuations in $\sigma^A$ can lead to unwanted variability in the $\sigma^B$ stress-response of these network designs. In contrast, the wildtype $\sigma^B$ network with its ultrasensitive negative feedback design is ideally designed to compensate for competition effects (Fig. 9.4DE).

### 9.2.6. Negative feedback designs of stress-response sigma factor networks minimizes interference

The negative feedback design of the network discussed here is not unique to $\sigma^B$. Several stress sigma factors in *B. subtilis* as well as other bacteria are regulated in similar fashion (380, 387-391). For example $\sigma^W$, a sigma factor in *B. subtilis* that controls the response to alkaline shock (392) is co-transcribed with its anti-sigma factor RsiW. In the absence of stress, RsiW sequesters $\sigma^W$ in an inactive complex. $\sigma^W$ is activated by stress signals which trigger the cleavage and degradation of thereby releasing and activating target expression (393). Although it is unknown whether the $\sigma^W$ network functions in a negative feedback regime similar to or if it pulses, it is possible for this network to exhibit these design properties. If RsiW is expressed in stoichiometric excess of its binding partner $\sigma^W$ from the $\sigma^W$-regulated operon which they share (394), then similar to the $\sigma^B$ network, $\sigma^W$, would operate in a negative feedback regime.
To determine if negative feedback control of stress sigma factor activity offers any advantages, we built a new model that includes three sigma factors: \( \sigma^B, \sigma^W \) and \( \sigma^A \). The regulation of free \( \sigma^B \) and \( \sigma^W \) was modeled with simplified identical versions of the negative feedback design of the \( \sigma^B \) network (Fig. A8.5A). Under this simplification, free \( \sigma^B (\sigma^W) \) is a non-monotonic function of \( B_T (W_T - \text{total } \sigma^W) \) that is qualitatively similar to the post-translational response function \( \sigma^B = F_P(B_T, P_T) \). Similar to \( \sigma^B = F_P(B_T, P_T) \), the simplified post-translational response of \( \sigma^W \) depends on a signaling proteins \( P_B \) (or \( P_W \) for \( \sigma^W \)). Anti-sigma factors RsbW (RsiW) and other details of post-translational regulation were excluded for simplicity. Following the previous section, this model explicitly includes \( \sigma^A \), RNA polymerase (RNApol) and \( \sigma - \text{RNApol} \) binding such that the transcriptional responses for both \( \sigma^B \) and \( \sigma^W \) depend on \( \sigma^A \) and RNApol concentrations (see Appendix 8). All parameters of the simplified model were chosen to approximately match the full \( \sigma^B \) network model and ensure that both \( \sigma^B \) and \( \sigma^W \) operate in the negative feedback regime. Consequently for the chosen parameters this simplified model acts like our detailed model and responds to step increases in the phosphatase \( P_B \) (or \( P_W \)) by producing a pulse of \( \sigma^B \) (or \( \sigma^W \)) activity (Fig. A8.5BC).

We used this simple model to study the response when cells are simultaneously exposed to multiple stresses creating competition for RNApol. Surprisingly our results showed that increasing the stress signal for one sigma factor also led to the activation of the other sigma factor. For example increasing \( P_B \) while keeping \( P_W \) fixed leads to an increase in free \( \sigma^B \) but also results in a smaller increase in free \( \sigma^W \) (Fig. 9.5C). This response can be explained by the ultrasensitive negative feedbacks controlling the stress sigma factors. Increase in free by \( P_B \) leads to increased competition for RNApol resulting in a decrease in production of \( W_T \). But since is regulated by a negative feedback, a decrease in \( W_T \) actually frees up more \( \sigma^W \) thereby insulating \( \sigma^W \) target activity from the effects of RNApol competition (Fig. 9.5E). Similarly the dynamic response of the stress sigma factors is also insulated from competition and an increase in fixed \( P_W \)
levels increases the pulse amplitude of $\sigma^B$ in response to step changes in $P_B$ (Fig. A8.5D-F).

Figure 9.5. Negative feedback minimizes competition between stress $\sigma$ factors for RNA polymerase.

**A,B.** Simplified network diagrams of stress sigma factors $\sigma^B$ and $\sigma^W$ and housekeeping sigma factor $\sigma^A$ competing with for RNA polymerase. $\sigma^B$ and $\sigma^W$ activities are regulated by negative and positive feedbacks in (A) and (B) respectively. In both cases, phosphatases $P_B$ and $P_W$ control the stress-signal driven activation of $\sigma^B$ and $\sigma^W$ respectively.

**C,D.** Dependence of free $\sigma^B$ and $\sigma^W$ levels on $P_B$ at fixed $P_W(=2\mu M)$. In the wildtype negative feedback system (C), increase in $\sigma^B$ phosphatase leads to an increase in both free $\sigma^B$ (green curve) and free $\sigma^W$ (red curve). In the positive feedback system (D),

increase in $\sigma^B$ phosphatase leads to an increase in free $\sigma^B$ (green curve) and a decrease in free $\sigma^W$ (red curve).

**E, F.** $\sigma^B$ and $\sigma^W$ target promoter activities as a function of $P_B$ at fixed $P_W$ in the wildtype negative feedback system (E), and the positive feedback system (F).

**G, H.** RNA polymerase bound $\sigma^B$ ($\text{Rpol-}\sigma^B$) as a function of $P_B$ at fixed $P_W$ in the wildtype negative feedback system (G) and the positive feedback system (H). Increase in $\sigma^B$ phosphatase leads to an increase in $\text{Rpol-}\sigma^B$ (green curve) and corresponding decreases $\Delta\text{Rpol-}\sigma^W$ in $\text{Rpol-}\sigma^W$ (red area) and $\Delta\text{Rpol-}\sigma^A$ in $\text{Rpol-}\sigma^A$ (blue area).

Thus the two stress sigma factors are able function simultaneously despite the scarcity of RNApol. This cooperation between stress sigma factors becomes clearer when we track the changes in $\sigma$–RNApol complexes as a function of stress signal $P_B$. As $P_B$ increases, more free becomes available and binds to RNApol (Fig. 9.5G). However this RNApol must be accounted for by the RNApol lost by the other operating sigma factors $\sigma^W$ and $\sigma^A$. Comparing the contributions of each shows that despite the fact that $\sigma^A$ has a much higher affinity for RNApol, most of the RNApol in the $\sigma^B$-RNApol complex is drawn from the $\sigma^A$-RNApol pool rather than $\sigma^W$-RNApol pool (Fig. 9.5G). Thus the negative feedback design allows stress sigma factors to cooperate and avoid competing with each other at the expense of the housekeeping sigma factor $\sigma^A$.

The role of the negative feedback in producing this cooperative response was made clearer when we compared the response of a mutant network that has positive feedbacks between $\sigma^B$ and $B_T$ and $\sigma^W$ and $W_T$. As shown in Fig. 9.5D, unlike the wildtype response, increase in $P_B$ and the resulting increase in free $\sigma^B$ decreases the free $\sigma^W$ in the positive feedback network. As a result of the increased competition for RNApol and the decreased free $\sigma^W$, $\sigma^W$ target promoter activity in this network decreases as a function of $P_B$ (Fig. 9.5F). Moreover comparing changes in $\sigma$–RNApol complexes as a function of stress signal $P_B$ we found that most of the RNApol in the $\sigma^B$-RNApol complex is drawn from the $\sigma^W$-RNApol pool rather than $\sigma^A$-RNApol pool (Fig. 9.5H). Thus the negative feedback
designs are not only essential for stress sigma factors to tolerate competition from $\sigma^A$ but also necessary for them to cooperate and avoid competing with each other when the cell is simultaneously exposed to multiple types of stresses.

9.3. Discussion

The $\sigma^B$ network is one of the most well-studied model systems for stress response sigma factors in bacteria (68, 380, 395). Recent studies have added a new dimension to the functional properties of this network by demonstrating that it responds to stress in a highly dynamic pulsatile manner. Prompted by these developments, here we have provided a new mathematical model that avoids making ad hoc simplifications and instead accurately captures all the known molecular details of the network. Our model reproduces all reported features of the response including pulsatile activation in response to stress. By decoupling the post-translational and transcriptional responses in our model we were able to derive a simplified view of the network that illustrates how the pulsatile response is mechanistically based on the ultrasensitive negative feedback in the network. Using this method we were able to identify the relative stoichiometry of $\sigma^B$, RsbW and RsbV synthesis rates as the most critical design property that determines all qualitative features of the network response.

Specifically our results showed that $\sigma^B$ pulsing is only possible when there is a stoichiometric imbalance between the synthesis rates of various network proteins. The concentration of RsbW needs to be in a sweet-spot: greater than the concentration of $\sigma^B$, but less than the total concentration of its two binding partners. In our model, stoichiometric imbalance is achieved by choosing synthesis rates which result in protein concentrations away from their stoichiometric ratios. Not surprisingly a previous model with synthesis rates that maintain stoichiometric balance neither observed an ultrasensitive response, nor pulsing in $\sigma^B$ (382). This prediction can be further tested experimentally by measuring the relative concentration of $\sigma^B$ and its network partners in the
presence of several mutations that promote or curb pulsing. Notably stoichiometric balance between network proteins has been shown to play an important role in the dynamic response of other networks. Surprisingly, whereas it is detrimental for pulsing in the $\sigma^B$ network, it plays a key role in generating robust circadian rhythms in Drosophila (396).

The decoupling of the post-translational and transcriptional response greatly facilitated the identification of critical design features despite the complexity of the network. This separation reduces the dimensionality of the dynamical system by enabling an independent input–output analysis in the two modules. Similar methods have been applied in variety of studies to deduce core functional properties in a variety of studies and model systems. Interestingly our analysis revealed that the post-translational and transcriptional module structures of the network and the phosphorelay controlling B. subtilis sporulation are remarkably similar (350). Despite the differences in molecular details, both networks possess a non-monotonic post-translational dependence of the active transcription factor on the concentration of the total transcription factor. Combining this response with the transcriptional feedback produces an ultrasensitive negative feedback in both networks. The relevance of these similarities is evidenced by the fact that both networks produce dynamically similar pulsatile responses even though they are activated by entirely different stimuli.

Our model also showed how the timescale separation between the fast post-translational and the slow transcriptional responses in the network allow it to encode the rate of stress increase into $\sigma^B$ pulses. Further we showed that this network encodes the size of stochastic bursts of stress phosphatase into the amplitudes of $\sigma^B$ pulses. Still the following question remains unanswered: how do bacteria specifically convert increase in stress to increase in phosphatase burst size? A recent work, which studied gene expression dynamics across the human genome, showed that weaker expression loci modulate burst frequency whereas stronger expression loci modulate burst size to increase activity (397). Thus, it
would be interesting to investigate the strength of the loci in *B. subtilis* where phosphatase is encoded. This result also raises a question about the utility of frequency-modulated strategy to control $\sigma^B$ activity. It has been suggested that frequency modulation results in proportional expression of downstream genes (103, 105). This proportional control requires the distribution of pulse amplitudes to remain fixed even as stress levels increase. However under the burst encoding strategy, pulse amplitude distributions change as stress levels increase thereby negating the efficacy of a pulsed response in producing proportional expression of downstream genes.

We also used our model to understand the response when placed in the larger context of other sigma factor networks and competition for RNA polymerase. Our results show how the design of the network is uniquely suited to insulating its response from RNA polymerase competition from the housekeeping sigma factor. Finally we demonstrate how ultrasensitive negative feedback, a ubiquitous feature of stress sigma factor regulation enables different stress sigma factors to operate simultaneously without inhibiting each other by competing for RNA polymerase. These results are relevant not only for understanding the stress response of bacteria but also increasingly for the design of synthetic circuits. The movement towards the construction of larger genetic circuits has produced a number of recent designs that include multiple independent modules that rely on shared resources or actuators to function (398-400). Our results highlight how competition between modules for these shared resources can significantly affect the performance of these synthetic circuits. Inspired by the design of naturally occurring stress sigma factor networks we also provide new design rules that if incorporated can improve and lead to failure resistant synthetic networks.

9.4. Methods
Mathematical model of the $\sigma^B$ network

Our mathematical model of $\sigma^B$ network is based on a previous model proposed in (382). This ODE-based model explicitly includes all known molecular species, post-translational reactions and the transcriptional regulation of the $\sigma^B$ operon by $\sigma^B$. The stress signals were assumed to control the concentrations of stress phosphatases $\text{RsbTU}$ and $\text{RsbQP}$. For $\text{RsbQP}$, energy stress was assumed to regulate the transcription rate of the phosphatase and the phosphatase concentration was assumed to be subject to stochastic fluctuations resulting from gene expression noise. In contrast $\text{RsbTU}$ concentration is regulated by environmental stress post-translationally, consequently $\text{RsbTU}$ concentration was assumed to be stress-dependent but not subject to stochastic fluctuations. The details of all biochemical reactions in the model and the corresponding differential equations are described in the Appendix 8. All the parameters used in the model are summarized in Table A8.1.

To model the effects of competition for RNA polymerase, the model of the network was expanded by including reactions for $\sigma^A$, RNA polymerase (RNApol) and $\sigma-$RNApol binding (see Appendix 8). To model the competition between $\sigma^B$, $\sigma^W$ and $\sigma^A$, we used a phenomenological non-monotonic function to model the post-translational regulation of stress sigma factors ($\sigma^B$ and $\sigma^W$; see Appendix 8 for details).

Calculation of steady state post-translational and transcriptional responses

The decoupled transcriptional and post-translational responses of the network at steady state were calculated using the MATLAB bifurcation package MATCONT. The post-translational response $\sigma^B = F_p(B_T, P_T)$, was calculated by varying the rate of operon transcription while keeping the component synthesis rates ($\lambda_W$, $\lambda_V$) and the total phosphatase concentration ($P_T$) fixed. Similarly, the transcriptional response $B_T = F_T(\sigma^B)$, was calculated by varying the free $\sigma^B$ concentration as an independent variable to calculate the total concentrations of $\sigma^B$, RsbW and RsbV.
Simulations

In the deterministic set-up, the system of differential equations was solved using standard *ode15s* solver in MATLAB. For stochastic simulations the time-varying total phosphatase level $P_T (= P + V_P P)$ was pre-computed using a gamma distributed Ornstein-Uhlenbeck process as in (103). This gamma distributed Ornstein-Uhlenbeck process permits independent modulation of mean burst size ($b$) and frequency ($a$) (401). For each phosphatase level, 50 simulations were performed each lasting 10 hours. Pulses were detected by examining local maxima and minima of the simulated trajectories, and subsequently this information was used to compute pulse statistics for amplitude and frequency.
Appendix 1. Bistability and low-pass filtering in the network module determining blood stem cell fate

A1.1. Supplementary Figures

![Figure A1.1](image)

Figure A1.1. Scl+19 distant enhancer controls the probability not rate of gene expression.

A. Flow cytometry analysis of β-galactosidase expression in day 5 embryoid bodies. A single copy of each of the reporter constructs for the Scl+19 enhancer were targeted to the HPRT locus in embryonic stem (ES) cells. ES cells carrying either no transgene, the wild-type Scl+19 enhancer or the Scl+19 enhancer with the GATA site mutated were differentiated into embryoid bodies and analyzed at Day 5 of differentiation (onset of
haematopoiesis) for the expression of β-galactosidase. The percentage of β-galactosidase positive cells is shown along with the geometric mean fluorescence intensity (GMFI) of the β-galactosidase positive population. Note that the number of cells expressing β-galactosidase is reduced by the mutation in the enhancer but the mean level of expression in β-galactosidase positive cells is the same for both the wild type and the mutated enhancer. The ES cell data demonstrates a bi-modal distribution of β-galactosidase expression that remains in the presence of the mutated form of the enhancer showing that the Scl+19 enhancer determines the number of cells expressing the enhancer but not the level of expression observed.

B. 416B myeloid progenitor cells were co-transfected with a puromycin expressing selection plasmid and either SV/β-geo or SV/β-geo/Scl+19 reporter constructs. Pools of stably transfected cells were analysed for LacZ activity by flow cytometry. The dashed lines highlight the fact that the main consequence of including the enhancer is the loss of population I. By contrast, any increase in fluorescence in population II following inclusion of the enhancer is much less pronounced.
Figure A1.2. Schematic diagram of ratchet model of distal enhancer action using Scl+19 enhancer as an example.

The chromatin exists in an open and closed state as shown here. TRs and RNA polymerase can bind to DNA only in the open state. Gata2 and Fli1 bind to the enhancer and shift the equilibrium towards an open conformation. The increased probability of the open state results in increased gene expression.

Figure A1.3. Robustness of switchable bistable response to variability in chromatin equilibrium constants.

Switchable bistability in the triad response to Notch, Bmp4 and Gata1 is robust to large changes in chromatin equilibrium constants $K_S, K_G$ and $K_F$ when the free energies for the model are fixed at values given in Table A1.1. Threshold levels of Notch (A) and Bmp4 (B) for the OFF-ON switch and threshold levels of $f([\text{gata1}])$ (C) for the deactivation switch ON-OFF change with the values of the chromatin equilibrium constants $K_S$ (blue curve), $K_G$ (green curve) and $K_F$ (black curve). System response is
bistable if threshold levels for Notch and Bmp4 are finite and lie within the range $0 < f \left( [\text{gata1}] \right) < 1$ for Gata1. Therefore the threshold levels indicate the ranges of chromatin equilibrium constant values where the system response is bistable. Moreover there is only one threshold for each signal over the range of parameter values shown here. Thus the irreversible nature of the bistable response is also robust to large changes in $K_S, K_G$ and $K_F$. Solid dots indicate the values of $K_S, K_G$ and $K_F$ for which the dose response is given in Figures 2.2A and D.

**Figure A1.4. Dose-response to Notch and Bmp4 in Gata2, Scl and Fli1 heterozygotes.**

Scl+/- (A), Gata2+/- (B) and Fli1+/- (C) all show an irreversible bistable response to Notch (blue) and Bmp4 (black). Stable and unstable states are denoted by solid lines and dashed lines respectively. (D) Heterozygous deletion of two genes Gata2+/- Fli1+/- (results in reversible bistability in response to both Notch and Bmp4, so that at low levels of Notch and Bmp4, only the low steady state of triad proteins exists. The insets depict
the dose response in linear scale to show that all heterozygous mutants except Gata2+/- and Fli1+/- have two steady states in the absence of Notch and Bmp4. Note that in B-D the basin of attraction for the high steady state is very small compared to the wild type (Figure 2.2A) for low values of activators, especially Notch. The high steady state of heterozygous deletions can therefore be very sensitive to fluctuations in protein concentrations in the absence of Notch and Bmp4.

A1.2. Supplementary Tables

Table A1.1. Enhancer library reporter expression results
Data adapted from (118).

<table>
<thead>
<tr>
<th>Enhancers</th>
<th>Binding Sites</th>
<th>Fold Enhancement of Expression Rate $I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Scl+19$</td>
<td>Wild Type (wt)</td>
<td>820.51</td>
</tr>
<tr>
<td></td>
<td>Mutant 1 (mut1)</td>
<td>141.02</td>
</tr>
<tr>
<td></td>
<td>Mutant 1 (mut2)</td>
<td>235.89</td>
</tr>
<tr>
<td></td>
<td>Mutant 1 (mut3)</td>
<td>1</td>
</tr>
<tr>
<td>$Gata2$-3</td>
<td>Wild Type (wt)</td>
<td>234.38</td>
</tr>
<tr>
<td></td>
<td>Mutant 1 (mut1)</td>
<td>1.72</td>
</tr>
<tr>
<td></td>
<td>Mutant 2 (mut2)</td>
<td>1.72</td>
</tr>
<tr>
<td></td>
<td>Mutant 3 (mut3)</td>
<td>120.31</td>
</tr>
<tr>
<td></td>
<td>Mutant 4 (mut4)</td>
<td>1</td>
</tr>
<tr>
<td>$Fli1$+12</td>
<td>Wild Type (wt)</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Mutant 1 (mut1)</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Mutant 2 (mut2)</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Mutant 3 (mut3)</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Mutant 4 (mut4)</td>
<td>1</td>
</tr>
</tbody>
</table>

Table A1.2. Notation for various symbols used in the main text and supplementary information

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_i^{C}$</td>
<td>Closed chromatin conformation of enhancer $i$: $(i = s (Scl+19), g (Gata2-3), f (Fli1+12))$</td>
</tr>
<tr>
<td>$G_i^{Gata2}$</td>
<td>Gata2 bound to enhancer $i$: $(i = s (Scl+19), g)$</td>
</tr>
<tr>
<td>Interaction Type</td>
<td>Description</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>$G_i^{Fli}$</td>
<td>Fli1 bound to enhancer $i$: $(i = s (Scl+19), g (Gata2-3), f (Fli1+12))$</td>
</tr>
<tr>
<td>$G_i^{Fli1Gata2}$</td>
<td>Gata2&amp; Fli1 bound to enhancer $i$: $(i = s (Scl+19), g (Gata2-3), f (Fli1+12))$</td>
</tr>
<tr>
<td>$G_i^{SclGata2Fli1}$</td>
<td>Scl&amp;Gata2&amp;Fli1 bound to enhancer $i$: $(i = g (Gata2-3), f (Fli1+12))$</td>
</tr>
<tr>
<td>$G^N$</td>
<td>Notch bound to promoter</td>
</tr>
<tr>
<td>$G^B$</td>
<td>Bmp4 bound to promoter</td>
</tr>
<tr>
<td>$G^P$</td>
<td>Polymerase bound to promoter</td>
</tr>
<tr>
<td>$K_i$</td>
<td></td>
</tr>
<tr>
<td>Steady State Concentrations</td>
<td>([SCL])</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>([GAT])</td>
<td></td>
</tr>
<tr>
<td>([FLI])</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dimensionless Concentrations</th>
<th>([scl])</th>
<th>Concentration of Scl relative to concentration in high steady state. ([scl]=\frac{[SCL]}{[SCL]})</th>
</tr>
</thead>
<tbody>
<tr>
<td>([gat])</td>
<td></td>
<td>Concentration of Gata2 relative to concentration in high steady state. ([gat]=\frac{[GAT]}{[GAT]})</td>
</tr>
<tr>
<td>([fli])</td>
<td></td>
<td>Concentration of Fli1 relative to concentration in high steady state. ([fli]=\frac{[FLI]}{[FLI]})</td>
</tr>
<tr>
<td>([n])</td>
<td></td>
<td>Concentration of Notch relative to dissociation constant of binding to the promoter. ([n]=\frac{[N]}{e^{\Delta}})</td>
</tr>
<tr>
<td>([b])</td>
<td></td>
<td>Concentration of Scl relative to concentration in high steady state. ([b]=\frac{[B]}{e^{\Delta}})</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other symbols used</th>
<th>(p_{oi})</th>
<th>Probability of open chromatin conformation for enhancer (i): ((i = s (Scl+19), g (Gata2-3), f (Fli1+12)))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p' (R_p))</td>
<td></td>
<td>Probability of RNA polymerase being bound to promoter for enhancer-reporter construct with enhancer (i): ((i = s (Scl+19), g (Gata2-3), f (Fli1+12))). (p' (R_p) = [R_p] e^{-\Delta} p_{oi})</td>
</tr>
<tr>
<td>(I_o)</td>
<td></td>
<td>Maximum rate of transcription from SV promoter.</td>
</tr>
</tbody>
</table>
\( I_i = kR_pe^{-G_i} \)

Rate of transcription from SV promoter-enhancer construct for enhancer \( i \): \( i = s \) (Scl+19), \( g \) (Gata2-3), \( f \) (Fli1+12) \( I_i = I_o p_i \)

\( I_1 \)

Maximum rate of protein production from SV promoter. \( I_1 = k_1 I_o \)

\( k \)

Rate of isomerization of SV promoter bound RNA polymerase to open conformation

\( k_1 \)

Number of proteins produced per lifetime of mRNA

\( Z_i \)

Partition function for enhancer \( i \): \( i = s \) (Scl+19), \( g \) (Gata2-3), \( f \) (Fli1+12)

\( Z_i^E \)

Sum of Boltzmann weights for all open chromatin state enhancer configurations of enhancer \( i \). \( i = s \) (Scl+19), \( g \) (Gata2-3), \( f \) (Fli1+12).

\( k_{d i} \)

Degradation rate for TR \( i \): \( i = s \) (Scl), \( g \) (Gata2), \( f \) (Fli1).

---

**Table A1.3. Free energy values used in the model for Scl-Gata2-Fli1 triad**

<table>
<thead>
<tr>
<th>TR</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scl+19</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gata2</td>
<td>( G_{s\text{Gata2}} )</td>
<td>-5.1304</td>
</tr>
<tr>
<td>Fli1</td>
<td>( G_{s\text{Fli1}} )</td>
<td>-2.8991</td>
</tr>
<tr>
<td>Fli1-Gata2</td>
<td>( G_{s\text{Fli1Gata2}} )</td>
<td>-17.0239</td>
</tr>
<tr>
<td><strong>Gata2-3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gata2</td>
<td>( G_{g\text{Gata2}} )</td>
<td>0.3229</td>
</tr>
<tr>
<td>Fli1</td>
<td>( G_{g\text{Fli1}} )</td>
<td>0.1614</td>
</tr>
<tr>
<td>Fli1-Gata2</td>
<td>( G_{g\text{Fli1Gata2}} )</td>
<td>-5.4957</td>
</tr>
<tr>
<td>Scl-Fli1-Gata2</td>
<td>( G_{g\text{SclGata2Fli1}} )</td>
<td>-13.6429</td>
</tr>
</tbody>
</table>
### A1.3. Supplementary Methods

#### A1.3.1. All or none regulation of gene expression from distant enhancers

We propose that the Scl, Gata2 and Fli1 regulate gene expression by ratcheting the equilibrium between open (promoter site accessible) and closed (promoter site inaccessible) chromatin conformations. This dynamic equilibrium of wrapped and unwrapped nucleosomal DNA has also been discussed elsewhere (160). Under our hypothesis the binding of TRs to the enhancer site increases gene expression by stabilizing the open conformation and thereby shifting the equilibrium towards it. This mechanism therefore allows the TRs to ratchet the spontaneous unwrapping of nucleosomal DNA and trap it in a state accessible to the transcriptional machinery. We illustrate this mechanism of the regulation of Scl expression by Gata2 and Fli1 in Figure A1.2. According to this mechanism binding of TRs only modulates the probability of gene expression and not the rate of transcription. Therefore, mutations of binding sites in the enhancers would affect the number of cells expressing the gene but the level of expression would be unaffected. To experimentally verify this all-or-none type of gene regulation embryonic stem cells containing the Scl+19 enhancer-reporter constructs in Figure 2.1 were analyzed for reporter expression with flow cytometry (Figure A1.1A). Mutation of the binding sites significantly reduces the fraction of cells expressing β-galactosidase. However the mean β-galactosidase expression among cells that do show expression did not change thereby confirming our hypothesis about the mechanism of gene regulation by TR binding at distant enhancers. The bi-modality of the data is obscured by the presence of several different cell types in the culture of ES cells. To highlight the all-or-none

| Enhancer          | Gene |  
|-------------------|------|------|
| Fli1+12           | Gata2| $G_{Fli1+12}^{Gata2}$ | 0.6687 |
| Fli1              | Fli1 | $G_{Fli1}$ | 0.3343 |
| Fli1-Gata2        | Fli1| $G_{Fli1Gata2}$ | -3.2967 |
| Scl-Gata2-Fli1    | Scl | $G_{SclGata2Fli1}$ | -7.4961 |
nature of gene regulation by the triad enhancers a similar experiment was carried out in 416B myeloid progenitor cells that were transfected with either SV/β-geo or SV/β-geo/Scl+19 reporter constructs. Figure A1.2B shows that the inclusion of the Scl+19 enhancer significantly enhances the number of cells expressing lacZ while the level of lacZ expression is unaffected. These results confirm our hypothesis about the all-or-none nature of gene regulation by enhancers within the Scl/Gata2/Fli1 triad.

A1.3.1. Experimental Methods

The wild type and mutated Scl+19 enhancer elements previously described in Pimanda et al (118) and shown in Figure 1 were cloned downstream of an SV40LacZ reporter construct and inserted as a Not I blunt fragment into the HPRT targeting vector pMP8NEBΔLacZ (a kind gift from Stephen Duncan) (402). HPRT targeted ES cells were generated as previously described (144) using the HM-1 ES cell line (403) and differentiated in 90mm Petri dishes in IMDM supplemented with 15% FCS, 2mM L-Glutamine, 300ug/ml Transferrin, 4x10-4M MTG, 50ug/ml Ascorbic acid and 5% PFHM-II. At day 5 of differentiation embryoid bodies were disrupted using trypsin and the single-cell suspension was analysed for β-galactosidase activity using FDG (fluorescein di-β-D-galactoside) (404). Briefly, cells were incubated with 1mM FDG at 37°C for 1 min followed by the addition of 1ml ice cold PBS. The cells were then incubated on ice for 10 mins and the reaction stopped using 1mM phenylethyl-β-D-thiogalactoside. The fluorescent signal was then analysed on the FACS Calibur (Becton and Dickinson) in the FL-1 channel. The Gata2-3 and Fli1+12 enhancers have been characterized in (118). The results for reporter expression from wild type and mutated enhancer elements for all three triad enhancers are shown in Table A1.1.

A1.3.1. Modeling the effects of heterozygous deletions of triad genes
HSCs heterozygous for triad proteins have been shown to have defects in repopulation capacity (152, 153). We use our model to analyze the effect of heterozygous deletions of triad genes on triad response to Notch and Bmp4. *Scl+/-*, *Gata2+/-* and *Fli1+/-* heterozygotes all show bistable dose response to Notch and Bmp4 (Figure A1.4 A-C). Note that the basins of attraction for the high steady states of heterozygous mutants at low concentration of Notch/Bmp4 are very small. These mutants should therefore be sensitive to stochastic fluctuations in triad protein concentrations. We expect that the noise induced switching from *ON* to *OFF* state in these mutants would be much higher than that for wild type cells. If the triad is a master regulator of the maintenance of stem cell state an increased rate of switching from *ON* to *OFF* state would result in an increase in the fraction of stem cells that produce multipotent progenitor cells committed to differentiation and a decrease in stem cell renewal. The decreased rate of stem cell proliferation would explain the deficient repopulation capacity of *Scl+/-*, *Gata2+/-* and *Fli1+/-* heterozygotic mutants (152, 153). We also analyzed the effects of heterozygous deletion of both the Gata2 and Fli1 genes and find that this cell line can only display reversible bistability-switching back to *OFF* state at low Notch and Bmp4 concentrations. Therefore, we predict a depletion of HSC pool in this mutant.

**A1.3.1. Estimation of free energies from enhancer library expression results**

Our approach for the calculation of free energies of TR-DNA and TR-TR interactions is discussed in Methods. The thermodynamic model is used to match the experimental results for the fold enhancement of gene expression from each enhancer (see Table A1.1 for experimental results). The experimental results are then used to reduce the dimension of the parameter space of the model. The following equations were derived to calculate the free energies that are parameters in the model from values of the chromatin equilibrium constants and the experimentally observed fold enhancement in expression.
Free energies for *Scl*+19:

Gata2 binding:

\[
G_s^{\text{Gata2}} = \log([\text{GAT}]) - \log \left( \frac{(1 - I_{\text{mut1}}^s / I_{\text{mut3}}^s)(K_s + 1)}{(K_s + 1 - I_{\text{mut1}}^s / I_{\text{mut3}}^s)} \right) \tag{A1.1}
\]

Fli1 binding:

\[
G_s^{\text{Fli1}} = 2\log([\text{FLI}]) - \log \left( \frac{(1 - I_{\text{mut2}}^s / I_{\text{mut3}}^s)(K_s + 1)}{(K_s + 1 - I_{\text{mut2}}^s / I_{\text{mut3}}^s)} \right) \tag{A1.2}
\]

Gata2-Fli1 binding:

\[
G_s^{\text{Fli1Gata2}} = \log([\text{GAT}][\text{FLI}]^2) - \log \left( \frac{(1 - I_{\text{wt}}^s / I_{\text{mut3}}^s)(K_s + 1)}{(K_s + 1 - I_{\text{wt}}^s / I_{\text{mut3}}^s)} \right) - \log \left( \frac{(1 - I_{\text{mut2}}^s / I_{\text{mut3}}^s)(K_s + 1)}{(K_s + 1 - I_{\text{mut2}}^s / I_{\text{mut3}}^s)} \right) - \log \left( \frac{(1 - I_{\text{mut1}}^s / I_{\text{mut3}}^s)(K_s + 1)}{(K_s + 1 - I_{\text{mut1}}^s / I_{\text{mut3}}^s)} \right) \tag{A1.3}
\]

Free energies for Gata2-3:

Gata2 binding:

\[
G_g^{\text{Gata2}} = \log([\text{GAT}]) - \log \left( \frac{(1 - I_{\text{mut1}}^g / I_{\text{mut3}}^g)(K_g + 1)}{(K_g + 1 - I_{\text{mut1}}^g / I_{\text{mut3}}^g)} \right) \tag{A1.4}
\]

Fli1 binding:

\[
G_g^{\text{Fli1}} = 2\log([\text{FLI}]) - \log \left( \frac{(1 - I_{\text{mut2}}^g / I_{\text{mut3}}^g)(K_g + 1)}{(K_g + 1 - I_{\text{mut2}}^g / I_{\text{mut3}}^g)} \right) \tag{A1.5}
\]

Gata2-Fli1 binding:
\[ G_{g}^{\text{FliH} \text{Gata2}} = \log \left[ \left( \frac{I_{g}}{I_{\text{mut3}}} \right) \right]^2 \]

\[ -\log \left( \frac{(1 - I_{\text{wt}}/I_{\text{mut3}})(K_g + 1)}{(K_g + 1 - I_{\text{mut1}}/I_{\text{mut3}})} \right) \]

\[ (A1.6) \]

Scl-Gata2-Fli1 binding:

\[ G_{g}^{\text{Scl} \text{Gata2} \text{Fli1}} = \log \left[ \left( \frac{I_{g}}{I_{\text{mut3}}} \right) \right]^2 \]

\[ -\log \left( \frac{(1 - I_{\text{wt}}/I_{\text{mut3}})(K_g + 1)}{(K_g + 1 - I_{\text{mut1}}/I_{\text{mut3}})} \right) \]

\[ (A1.7) \]

Free energies for \( \text{Fli1} + 12 \):

Gata2 binding:

\[ G_{f}^{\text{Gata2}} = \log \left[ \left( \frac{I_{f}}{I_{\text{mut3}}} \right) \right] - \log \left( \frac{(1 - I_{\text{mut1}}/I_{\text{mut3}})(K_f + 1)}{(K_f + 1 - I_{\text{mut1}}/I_{\text{mut3}})} \right) \]

\[ (A1.8) \]

Fli1 binding:

\[ G_{f}^{\text{Fli1}} = 2 \log \left[ \left( \frac{I_{f}}{I_{\text{mut3}}} \right) \right] - \log \left( \frac{(1 - I_{\text{mut2}}/I_{\text{mut3}})(K_f + 1)}{(K_f + 1 - I_{\text{mut2}}/I_{\text{mut3}})} \right) \]

\[ (A1.9) \]

Gata2-Fli1 binding:

\[ G_{f}^{\text{Fli}H \text{Gata2}} = \log \left[ \left( \frac{I_{f}}{I_{\text{mut3}}} \right) \right]^2 \]

\[ -\log \left( \frac{(1 - I_{\text{wt}}/I_{\text{mut3}})(K_f + 1)}{(K_f + 1 - I_{\text{mut1}}/I_{\text{mut3}})} \right) \]

\[ (A1.10) \]

Scl-Gata2-Fli1 binding:

\[ G_{f}^{\text{Scl} \text{Gata2} \text{Fli1}} = \log \left[ \left( \frac{I_{f}}{I_{\text{mut3}}} \right) \right]^2 \]

\[ -\log \left( \frac{(1 - I_{\text{wt}}/I_{\text{mut3}})(K_f + 1)}{(K_f + 1 - I_{\text{mut2}}/I_{\text{mut4}})} \right) \]

\[ (A1.11) \]
All free energies can be determined using the values of chromatin rewrapping equilibrium constants in these equations. The dimensions of the parameter space are therefore reduced to just these equilibrium constants. We choose \( K_s, K_g \) and \( K_f \) from a range of values where the triad exhibits switchable bistability and determine free energies for the chosen set of parameter values. The free energies are shown in Table A1.3. The free energy of TR-DNA binding varies between enhancers due to the effect of sequences flanking the binding sites. Notably the TR-TR interaction (e.g. Gata2-Fli1) energies vary very little between enhancers.
Appendix 2. Thermodynamic models of combinatorial gene regulation by distant enhancers

A2.1. Supplementary Figures

Figure A2.1. Gene expression response functions for contact model logic gates.
Parameters and equations have been adapted from Ref. (186).

(a) Gene expression rates relative to the maximum level of expression \([A] = [B] = 10^3\)
for the AND logic with typical cellular concentrations of TFs \(A\) and \(B\) \((G_A = G_B = 1.25, G_{AB} = -2.99, G_{AP} = G_{BP} = -2.99, G_{ABP} = -6.68, e^{-G_T} = 0.029}\).

(b) Gene expression rates for the OR logic normalized relative to the maximum expression level at \([A] = [B] = 10^3\) \((G_A = G_B = -2.31, G_{AB} = 0, G_{AP} = G_{BP} = -2.99, G_{ABP} = -3.69, e^{-G_T} = 0.05}\).
(c) For the NAND logic transcription rates were normalized relative to the maximum expression level at \([A] = [B] = 0\) \( (G_A = G_B = -2.31, G_{AB} = -2.99, e^{-G_R} = 100) \).

(d) For the XOR logic the maximum expression level at \([A] = 10^7, [B] = 0\) was used to normalize the transcription rates
\[ (G_A^l = G_B^l = -1.62, G_{AB}^l = 0, G_{AP}^l = G_{BP}^l = -2.99, G_{ABP}^l = -3.69, G_A^2 = G_B^2 = -0.11, G_{AB}^2 = -2.99, e^{-G_R} = 0.1) \).

Figure A2.2. Comparison of the sensitivities to free energy values of the AND, NAND and XOR logic gate responses.

(a)-(d) The AND gate response is most sensitive to the free energies \(G_A\) and \(G_{AB}\) at high concentrations of TFs \(A\) and \(B\), respectively. The sensitivity of the AND gate response to these free energies is similar for the chromatin mechanism and the direct contact mechanism. (The sensitivity to \(G_B\) is symmetric to the sensitivity to \(G_A\)).
(e)-(h) Similar to the AND gate response, the NAND gate response is most sensitive to $G_A$ and $G_{AB}$ at high TF concentrations. The sensitivities of the contact and chromatin mechanisms are identical.

(i) In the contact model the XOR gate response is sensitive to $G_A^J$ only at intermediate concentrations of the TFs.

(j) The XOR gate response for the chromatin mechanism is less sensitive than the contact mechanism at intermediate concentrations of TFs and more sensitive at high concentrations of TFs.

(k),(l) The contact mechanism’s response is more sensitive to $G_{AB}^J$ at high TF concentrations than the chromatin mechanism’s response. Rather than parameters sensitivities these differences are largely a result of the difference in XOR response of the two mechanisms (see section 3.2.3). Parameters and equations for the contact model were abstracted from (186).

Figure A2.3. Sensitivity of the parameter estimation method to the choice of chromatin equilibrium constant $K$.

a,c. The Gata2 response function for $K = 250$ and $K = 500$ respectively, as determined using the expression data given in Figure 3.5A (the colorbars represent fold-change increase in gene expression relative to the expression rate at $[Gata2]=[Fli1]=0$). Note
that the response function within the range of wild-type TF concentrations
\(0 < [\text{Gata2}], [\text{Fli1}] \leq 1\); demarcated by white lines) does not change over this two-fold change in \(K\). However the response functions are drastically different when the TFs are over-expressed. In fact the maximum fold-change (at saturating TF concentrations) in each case are approximately equal to \(K\). Therefore the value of \(K\) can be determined by over-expressing the enhancer binding TFs.

**b, d.** Sensitivity of the Gata2 response to the value of \(K\) for \(K = 250\) and \(K = 500\) respectively. In the range of wild-type concentrations \(0 < [\text{Gata2}], [\text{Fli1}] \leq 1\); area bounded by white lines) the response functions are not sensitive to the value of \(K\), but outside this region the response function depends strongly on the chosen value of \(K\). Therefore the response function inside the region of wild-type concentrations can be reliably predicted even without the knowledge of \(K\).

**A2.2. Methods**

**A2.2.1. Calculation of waiting times in transcriptional machinery bound (ON) and unbound (OFF) states**

The methods discussed in (202) were used to calculate the probability distribution function (PDF) of the time spent in the transcriptional machinery bound (ON) and unbound (OFF) states for the contact and chromatin mechanism kinetic schemes. The dynamics for either mechanism can be described by a system of ordinary differential equations. Let \(H\) be a rate matrix such that each element \(H_{ij}\) represents the rate constant of \(j \rightarrow i\) transition and \(H_{ii} = -\sum_{j \neq i} H_{ij}\).

To obtain the PDF of the time spent in the ON state we first require the probability that the system does not release the bound transcriptional machinery for time \(t\) given that the transcriptional machinery binds to the regulatory region at \(t = 0\). Given that the system starts in the ON sub-state \(j\), then \(G_i(t)\), the probability that the system is in the ON sub-state \(i\) at time \(t\) without releasing the transcriptional machinery is given by the rate equation:
\[ \frac{d}{dt} G = (H - U)G \]  

(62)

with the initial conditions \( G_j(0) = 1, G_k(0) = 0 \) \( \forall k \neq j \). Here \( U \) is the matrix of rate constants such that \( U_{ij} = K_{ij} \) \( \forall i \neq j \) and the \( j \rightarrow i \) transition represents the dissociation of the transcriptional machinery from the regulatory region. All remaining elements of \( U \) are set to zero. Similarly we can define a matrix \( V \) such that \( V_{ij} = K_{ij} \) \( \forall i \neq j \) and the \( j \rightarrow i \) transition represents the binding of transcriptional machinery to the regulatory region. The solution of equation (62) is \( G(t) = \exp((K - U)t)G(0) \). The PDF of spending time \( \tau \) in the \( ON \) state depends on three factors: 1) the probability of entering the \( ON \) state at time \( t_0 \), 2) the probability of going from state \( j \rightarrow i \) in time \( \tau \) and 3) the probability of exiting the \( ON \) state from \( ON \) state at time \( t_0 + \tau \). This PDF can be calculated as:

\[ P_{ON}(\tau) = (UG(\tau))^\dagger p_{in} \]

where, \( p_{in} = \frac{Vp^{ss}}{1^T Vp^{ss}} \)  

(63)

Here \( p^{ss}(j) \), the vector of steady probability of each state is computed using equation (2) and the partition function for each mechanism (see section 3.2.5). The probability \( p_{in}(j) \) of entering the \( ON \) state at time \( t_0 \) is a weighing factor for the calculation of the \( ON \) state PDF (\( \dagger \) represents transpose and \( 1 \) represents a unit vector). Similarly the \( OFF \) state PDF can also be calculated:

\[ P_{OFF}(\tau) = (VG(\tau))^\dagger p_{in} \]

where, \( p_{in} = \frac{Up^{ss}}{1^T Up^{ss}} \)  

(64)

For the contact mechanism we use the following numbering of sub-states: O (1), OA (2), OR (3) and OAR (4). For Figure 3.6A, the matrices \( H, U \) and \( V \) for the contact mechanism are:
To find $P_{\text{ON}}(\tau)$, the PDF of waiting time in the ON state for the contact model equations, first the equations (62), (63) and (65)-(67) were used with the Laplace transform to solve for $\tilde{P}_{\text{ON}}(s)$:

$$\tilde{P}_{\text{ON}}(s) = \frac{k^A[A]\left(\frac{k^A k_d}{\omega} + \frac{k^R k_d}{\omega} (k^A[A] + k^R + s)\right) + k^A\left(\frac{k^A k_d}{\omega} + \frac{k^R k^A[A]}{\omega} + k^R\left(\frac{k^R}{\omega} + s\right)\right)}{(k^A + k^A[A])\left(\frac{k^A}{\omega}(k^R + s) + (k^A[A] + k^R + s)\left(\frac{k^R}{\omega} + s\right)\right)}$$

(68)

The inverse Laplace transform of $\tilde{P}_{\text{ON}}(s)$ shows that the PDF $P_{\text{ON}}(\tau)$ is a sum of two exponential terms:

$$P_{\text{ON}}(\tau) = w_1 r_1 e^{-r_1 \tau} + (1 - w_1) r_2 e^{-r_2 \tau}$$

$$r_{1,2} = \frac{1}{2} \left(\frac{k^A}{\omega} + k^A[A] + k^R + \frac{k^R}{\omega} \pm \sqrt{\left(\frac{k^A}{\omega} + k^A[A] + k^R + \frac{k^R}{\omega}\right)^2 - 4\left(\frac{k^R k^A[A]}{\omega} + \frac{k^R k^R}{\omega} + \frac{(k^R)^2}{\omega}\right)}\right)$$

(69)
where \( w_i \in (0,1) \) is a weighing factor. Equation (69) shows that the PDF has two timescales: \( 1/r_1 \) and \( 1/r_2 \). The point of separation of these two timescales \( \tau_s \) is the point at which the two exponential terms are equal.

\[
\tau_s = \log \left( \frac{w_i r_1}{(1-w_i) r_2} \right) \quad \frac{r_1 - r_2}{r_2} \tag{70}
\]

Using equations (62), (64) and (65)-(67) we can calculate \( \tilde{P}_{\text{OFF}}(s) \), the Laplace transform of the OFF state waiting time distribution:

\[
\tilde{P}_{\text{OFF}}(s) = \frac{k^R}{k^R + s} \tag{71}
\]

The inverse Laplace transform of \( \tilde{P}_{\text{OFF}}(s) \) shows that the PDF \( P_{\text{OFF}}(\tau) \) has a single exponential term:

\[
P_{\text{OFF}}(\tau) = k^R e^{-k^R \tau} \tag{72}
\]

Thus the waiting time distribution of the OFF state in the contact model has only one timescale: \((k^R)^{-1}\).

We use the same method to calculate the waiting time distributions for the chromatin mechanism. We use the following numbering of sub-states: C(1), O (2), OA (3), OR (4) and OAR (5). The matrices \( H, U \) and \( V \) for the chromatin mechanism are:
\[
H = \begin{bmatrix}
-k_o & k_c & 0 & 0 & 0 \\
k_o & -(k_c + k^R + k^A[A]) & k^A / \omega & k^R / \omega & 0 \\
0 & k^A[A] & -(k^R + k^A d) / \omega & 0 & k^A / \omega \\
0 & k^R & 0 & -(k^R d + k^A[A]) & k^A / \omega \\
0 & 0 & k^R & k^A[A] & -(k^R d + k^A d / \omega)
\end{bmatrix}
\]

Using equations (62), (63) and (73)-(75) we solve for the Laplace transform of the ON state waiting time distribution \( \tilde{P}_{ON}^{\text{chr}}(s) \):

\[
\tilde{P}_{ON}^{\text{chr}}(s) = \frac{k^R d / \omega}{k^R d / \omega + s}
\]  

The inverse Laplace transform of \( \tilde{P}_{ON}^{\text{chr}}(s) \) shows that the PDF \( P_{ON}^{\text{chr}}(\tau) \) has a single exponential term:

\[
P_{ON}^{\text{chr}}(\tau) = \frac{k^R d}{\omega} e^{-k^R d / \omega \tau}
\]  

Again using equations (62), (64) and (73)-(75) we can calculate \( \tilde{P}_{OFF}^{\text{chr}}(s) \) the Laplace transform of the OFF state waiting time distribution.
\[ \tilde{P}^\text{chr}_{\text{OFF}}(s) = \frac{N(s)}{D(s)} \]

\[ N(s) = k^R \left( 2 \left( \frac{k_d}{\omega} \right)^2 (k_o + s) + \frac{k_d}{\omega} (k_o + s) \left( k^A [A] + k^R + s \right) + \right) \]

\[ k^A [A] \left( k^A [A] (k_o + s) + k_o \left( k^R + s \right) + s \left( k_c + k^R + s \right) \right) \]

\[ D(s) = \left( \frac{k_d}{\omega} + k^A [A] \right) \left( \frac{k_d}{\omega} \left( k_o \left( k^R + s \right) + s \left( k_c + k^R + s \right) \right) + \left( k^R + s \right) \left( k^A [A] (k_o + s) + k_o \left( k^R + s \right) + s \left( k_c + k^R + s \right) \right) \right) \]

(78)

The inverse Laplace transform of \( \tilde{P}^\text{chr}_{\text{OFF}}(s) \) shows that the PDF \( P^\text{chr}_{\text{OFF}}(\tau) \) is a sum of three exponential terms:

\[ P^\text{chr}_{\text{OFF}}(\tau) = w_1 r_1 e^{-r_1 \tau} + w_2 r_2 e^{-r_2 \tau} + (1 - w_1 - w_2) r_3 e^{-r_3 \tau} \]

(79)

where \( w_1, w_2 \in (0, 1) \) are weighing factors that represent the probability of timescales \( r_1 \) and \( r_2 \) respectively. The moments of these waiting time distributions can be easily calculated from the PDFs.

For the calculations shown in the main text the rate of binding of TF \( A \) was assumed to be near the diffusion limit \( k^A = 0.001 \text{nM}^{-1} \text{s}^{-1} \) (405) and a typical value was chosen for the TF dissociation constant \( K_d = k_d / k^A = 1 \text{nM} \) (156, 157).

The binding rate constant for the transcriptional machinery was also assumed to be diffusion limited, \( k^R = 0.001 \text{s}^{-1} \) (note that this is a first order rate constant unlike \( k^A \)) (405-407). Binding of the transcriptional machinery to core promoters is typically weak (156, 196), so we assumed that \( k_d^R = 0.1 \text{s}^{-1} \). The chromatin equilibrium constant \( \omega \) and the strength of TF-transcriptional machinery interactions \( \omega \) are both known to be in the range 10-1000 (176, 408) so we assumed \( \omega = K + I = 20 \). The concentration of TF \( A \) was chosen to ensure that the response is not saturated \( [A] = 1 \text{nM} \).
A2.2.2. Construction of logic gates for the chromatin mechanism

Parameters for the contact mechanism logic gates were taken from Ref. (186). The parameters for each logic gate of the chromatin mechanism were chosen to ensure that the response of this design was as close as possible to the response of the corresponding contact mechanism logic gate.

The response function for the logic gates is given by

\[ f^i([A],[B]) = \frac{p_B([A],[B])}{\max(p_B)} \]  

where \( f^i \) is the normalized rate of gene expression in the presence of TFs \( A \) and \( B \) \((i = \text{con} \text{ for the contact mechanism; } i = \text{chr} \text{ for the chromatin mechanism})\) relative to the maximum rate of expression.

For the AND, OR and XOR gates the parameters were estimated numerically by minimizing the square of the difference between the response functions of each logic gate of the two mechanisms. The normalized rate of gene expression for the AND logic of the contact mechanism \( f^{\text{con}}_{\text{AND}} \) was calculated by using equation (4) for the probability of transcription \( p_B^{\text{con}} \). In a similar fashion, the normalized rate of gene expression for the AND logic of the chromatin mechanism \( f^{\text{chr}}_{\text{AND}} \) was calculated by substituting \( Z_{\text{ON}} \) and \( Z_{\text{OFF}} \) from equation (11) in equation (4) to find \( p_B^{\text{chr}} \).

\( f^{\text{con}}_{\text{AND}} \) and \( f^{\text{chr}}_{\text{AND}} \) were used to construct the objective function \( S \).

\[ S = \int \left( f^{\text{con}}_{\text{AND}}([A],[B]) - f^{\text{chr}}_{\text{AND}}([A],[B]) \right)^2 d\log[A] d\log[B] \]  

To approximate this integral, the range of concentrations \((1,10^4)\) was discretized into 40 log-uniform intervals for each TF. The square of the difference between the normalized transcription rates of the two mechanisms was
calculated at each combination of TF concentrations \(([A]_i, [B]_j ; i, j = 1, 2...40)\) and summed to construct the following discrete version of the objective function:

\[
S = \sum_{i,j}(f_{AND}^{con}([A]_i, [B]_j) - f_{AND}^{chr}([A]_i, [B]_j))^2
\]  

(82)

The parameters \(K, e^{-G_T}, G_A, G_B\) and \(G_{AB}\) for the chromatin mechanism were chosen to minimize \(S\). The \textit{fminimax} library routine of MATLAB was used to solve the nonlinear optimization. The parameters for the OR gate and XOR gate were estimated using similar objective functions. Normalized transcription rates for these gates can be easily calculated with the equations listed in sections 3.2.3.

Parameters of the NAND gate of the chromatin mechanism were derived analytically from the parameters of the NAND gate of the contact mechanism. The probability of transcription is highest when \([A] = [B] = 0\) in the case of NAND logic. Numerical methods were not necessary for estimating parameters of the NAND gate because there are no TF-transcriptional machinery interactions in the contact model of the NAND gate (see Ref (186)). The probabilities \(p_{B}^{con}\) and \(p_{B}^{chr}\) for the NAND gate of the contact mechanism and the NAND gate of chromatin mechanism are given by:

\[
p_{B}^{con} = \frac{e^{-G_{T}^{con}}}{e^{-G_{T}^{con}} + I + [A]e^{-G_A^{con}} + [B]e^{-G_B^{con}} + [A][B]e^{-G_{AB}^{con}}} 
\]  

(83)

\[
p_{B}^{chr} = \frac{e^{-G_{T}^{chr}}}{e^{-G_{T}^{chr}} + K + I + [A]e^{-G_A^{chr}} + [B]e^{-G_B^{chr}} + [A][B]e^{-G_{AB}^{chr}}}
\]  

(84)

Using the substitutions

\[
G_{T}^{con} = G_{T}^{chr} - G_{AB}^{con}, \quad G_{A}^{con} = G_{A}^{chr} - G_{AB}^{con}, \quad G_{B}^{con} = G_{B}^{chr} - G_{AB}^{con} and \quad e^{-G_{AB}^{con}} = K + I
\]

(85)
the response function for the contact mechanism in equation (83) can be rearranged to give the response function for the chromatin mechanism shown in equation (84). Clearly the two expressions are analytically identical and the parameters of the chromatin mechanism can be derived from the substitutions used above. Note that in the response function for the chromatin mechanism the TFs $A$ and $B$ do not interact. This implies that the cooperativity between the two TFs emerges from the equilibrium between open and closed chromatin states. The strength of this emergent cooperativity matches the free energy of the TF-TF interaction in equation (83).

$$G_{AB}^\text{con} = -\log(K + l)$$  \hspace{1cm} (86)
Appendix 3. Reconciling Effects of Genetic Perturbations on HSC Emergence

A3.1. Supplementary Figures

Figure A3.1. Enhancer and promoter binding sites for the cis-regulatory regions that control Smad1, Smad6 and Runx1 expression.

Figure A3.2. Runx1 requirement for triad activation depends on $\alpha$.
At low $\alpha$ (<0.75), the Bmp4 signal threshold $k_p^*$ approaches infinity at low Runx1 production rates ($v_i^*$). At higher $\alpha$, a finite value of $k_p^*$ exists even at $v_i^*=0$. 
Figure A3.3. Robustness of WT-Runx1\(^{+/−}\) activation time difference predictions.

A. The difference in activation timing between the WT and Runx1\(^{+/−}\) systems decreases with increase in the activating Bmp4 signal \(k_p\). The red dot shows the \(k_p\) used in Fig. 4.3D (=5\(k_p^*\) for WT). The inset shows the time to reach 90% of Scl steady state for WT (solid line) Runx1\(^{+/−}\) (dashed line) systems and as a function of \(k_p\). The grey bar marks \(k_p =5k_p^*\) for WT and the dots show the response times at that \(k_p\). Simulations were run with [Notch1]=1.

B. The difference in activation timing between the WT and Runx1\(^{+/−}\) systems is dependent upon the WT Runx1 maximum production value \(v_r^o\). For each \(v_r^o\), the Runx1 production rate was set to 0.5 \(v_r^o\) for the Runx1\(^{+/−}\) mutant and the \(k_p\) was set to 5 times the WT \(k_p^*\) at that particular \(v_r^o\). [Notch1] =1 for all simulations. The inset shows time to reach 90% of Scl steady state for WT as a function of \(v_r^o\).

Figure A3.4. Robustness of minimum Runx1 requirement time window.
The earliest time that Runx1 can be knocked down without affecting triad activation decreases with increasing $k_p$. Runx1 is still needed for $\sim 3$ days even for $k_p = 10k_p^*$, the Bmp4 signal threshold. Simulations were run at $[\text{Notch1}] = 1$. The red dot shows the $k_p$ used in Fig. 4.5AB.

**Figure A3.5.** WT-$Runx1^{+/\text{c}}$ activation time difference in the presence of Notch dynamics.

(A) and (B) show dynamics of triad activation in response to a step-increase in Bmp4 signal and an exponential decrease in $[\text{Notch}]$ levels (blue curves). The dynamics depend on $Runx1$ gene dosage and Bmp4 signal level.

A. Unlike the triad response with a fixed $[\text{Notch}]$ level (see Fig. 4.3D), only the haploid mutant ($Runx1^{+/\text{c}}$ - dashed red curve) and not the WT ($Runx1^{+/+}$ - solid red curve) is able to switch to the HIGH state when Bmp4 signal is increased from $k_p=0$ to $k_p= 50\text{hr}^{-1}$.

B. With a larger step-increase in Bmp4 signal from $k_p=0$ to $k_p= 100\text{hr}^{-1}$, both the haploid mutant and WT are able to switch to the HIGH state. Grey region represents $\Delta T_{90}^{[\text{Scl}]}$, the difference in time to reach 90% of $[\text{Scl}]$ level in HIGH state.

C. When $[\text{Notch}]$ decreases over time, the activation timing of both WT ($Runx1^{+/+}$ - solid curve) and haploid mutant ($Runx1^{+/\text{c}}$ - dashed curve) decreases with increasing Bmp4 signal similar to the response when $[\text{Notch}]$ is fixed (compare to Fig. A3.3A). Note that
the decrease in [Notch] shifts the WT activation timing curve towards higher Bmp4 levels.

D. The difference in activation timing between the WT and Runx1\(^{-}\) systems also decreases with increase in the activating Bmp4 signal \(k_p\). The red dots show the \(\Delta T_{90}^{[\text{Scl}]}\) at \(k_p=100\text{hr}^{-1}\).

**Table A3.1. Notation for species concentrations**

<table>
<thead>
<tr>
<th>Species</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scl</td>
<td>[S]</td>
</tr>
<tr>
<td>Gata2</td>
<td>[G]</td>
</tr>
<tr>
<td>Fli1</td>
<td>[F]</td>
</tr>
<tr>
<td>Smad1</td>
<td>[S1]</td>
</tr>
<tr>
<td>Phospho-Smad1</td>
<td>[pS1]</td>
</tr>
<tr>
<td>Runx1</td>
<td>[R]</td>
</tr>
<tr>
<td>Phospho-Smad1:Runx1 Complex</td>
<td>[pS1-R]</td>
</tr>
<tr>
<td>Smad6</td>
<td>[S6]</td>
</tr>
<tr>
<td>Phospho-Smad1:Smad6 Complex</td>
<td>[pS1-S6]</td>
</tr>
<tr>
<td>Smad6:Runx1 Complex</td>
<td>[R-S6]</td>
</tr>
</tbody>
</table>

**Table A3.2. Parameter values for the triad regulatory functions**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scl</td>
<td>(G_{g1})</td>
<td>-5.1301</td>
</tr>
<tr>
<td></td>
<td>(G_{f1})</td>
<td>-2.8989</td>
</tr>
<tr>
<td></td>
<td>(G_{fg1})</td>
<td>-12.5</td>
</tr>
<tr>
<td></td>
<td>(K_s)</td>
<td>821</td>
</tr>
<tr>
<td>Gata2</td>
<td>(G_{g2})</td>
<td>0.3231</td>
</tr>
<tr>
<td></td>
<td>(G_{f2})</td>
<td>0.1615</td>
</tr>
<tr>
<td></td>
<td>(G_{fg2})</td>
<td>-5.4672</td>
</tr>
<tr>
<td></td>
<td>(G_{s1})</td>
<td>-9.0181</td>
</tr>
<tr>
<td></td>
<td>(G_{bp})</td>
<td>-1.6094</td>
</tr>
<tr>
<td></td>
<td>(K_g)</td>
<td>233.44</td>
</tr>
<tr>
<td></td>
<td>(\alpha)</td>
<td>0.6</td>
</tr>
<tr>
<td>Fli1</td>
<td>(G_{g3})</td>
<td>0.6718</td>
</tr>
<tr>
<td></td>
<td>(G_{f3})</td>
<td>0.3359</td>
</tr>
<tr>
<td>Parameter</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>$G_{fg3}$</td>
<td>-3.2359</td>
<td></td>
</tr>
<tr>
<td>$G_{s2}$</td>
<td>-5.8703</td>
<td></td>
</tr>
<tr>
<td>$G_{bp}$</td>
<td>-1.6094</td>
<td></td>
</tr>
<tr>
<td>$K_f$</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.6</td>
<td></td>
</tr>
</tbody>
</table>

Table A3.3. Parameter values for the Smad1 regulatory function

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G'_{s1}$</td>
<td>-0.7587</td>
</tr>
<tr>
<td>$G''_{s1}$</td>
<td>0.1</td>
</tr>
<tr>
<td>$G^*_{s1}$</td>
<td>0.1054</td>
</tr>
<tr>
<td>$K_{s1}$</td>
<td>100</td>
</tr>
<tr>
<td>$v^o_{s1}$</td>
<td>25</td>
</tr>
</tbody>
</table>

Table A3.4. Parameter values for the Smad6 regulatory function

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G'_{s6p}$</td>
<td>-1.15</td>
</tr>
<tr>
<td>$G'_{s6p}$</td>
<td>-1.099</td>
</tr>
<tr>
<td>$G''_{s6p}$</td>
<td>-2.303</td>
</tr>
<tr>
<td>$K_{s6p}$</td>
<td>100</td>
</tr>
<tr>
<td>$G'_{s6e}$</td>
<td>-3.09</td>
</tr>
<tr>
<td>$G''_{s6e}$</td>
<td>0.105</td>
</tr>
<tr>
<td>$K_{s6e}$</td>
<td>100</td>
</tr>
<tr>
<td>$v^o_{s6}$</td>
<td>35</td>
</tr>
</tbody>
</table>

Table A3.5. Parameter values for the Runx1 regulatory function

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_{pl1}$</td>
<td>-1.099</td>
</tr>
<tr>
<td>$K_{rp1}$</td>
<td>10</td>
</tr>
<tr>
<td>$G'_{m2}$</td>
<td>-0.35</td>
</tr>
<tr>
<td>$K_{rp2}$</td>
<td>2</td>
</tr>
<tr>
<td>$G^g_{re}$</td>
<td>-0.5596</td>
</tr>
<tr>
<td>$G^r_{re}$</td>
<td>-0.003</td>
</tr>
</tbody>
</table>
Table A3.6. List of parameters values for post-translational reactions

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value (hr(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k_p)</td>
<td>Smad1 phosphorylation rate</td>
<td>Bmp4 signal</td>
</tr>
<tr>
<td>(k_b)</td>
<td>Rate of Smad6 binding</td>
<td>1</td>
</tr>
<tr>
<td>(k_u)</td>
<td>Rate of Smad6 unbinding</td>
<td>0.1</td>
</tr>
<tr>
<td>(k_{b1})</td>
<td>Rate of Runx1+pSmad1</td>
<td>10</td>
</tr>
<tr>
<td>(k_{u1})</td>
<td>Rate of pSmad1-Runx1</td>
<td>0.1</td>
</tr>
<tr>
<td>(k_{pro})</td>
<td>Rate of proteolytic</td>
<td>10</td>
</tr>
</tbody>
</table>

A3.3. Parameter robustness of model predictions

A3.3.1. Runx1 requirement for triad activation in response to Bmp4

In our model we have assumed that both pSmad1 and pSmad1:Runx1 complex can bind to and control transcription from the Gata2 and Fli1 promoters. Crucially, we have assumed that the maximum fold-change in transcription from these promoters in response to pSmad1 binding is smaller than the response to pSmad1:Runx1 complex binding (fold-change for pSmad1:Runx1 is \(f_1\) and fold-change for pSmad1 is \(\alpha f_1\) where \(\alpha=0.6\)). Our simulations show that whether or not Runx1 is essential for the irreversible switching of triad bistable switch in response to Bmp4 (cf. Fig. 4.3B) depends on the value of this factor \(\alpha\). Fig. A3.2 shows that at lower values of \(\alpha\) (< 0.75), the Bmp4 signal threshold at which the triad can switch to high state approaches infinity at low Runx1 production rates, \(\nu^*_r\). However as \(\alpha\) is increased, pSmad1 alone is sufficient to switch the triad, pSmad1:Runx1 complex is no longer necessary, and even at \(\nu^*_r=0\) a finite Bmp4 signal threshold for triad activation exists. Note that the non-monotonic dependence of this signal threshold around higher Runx1 expression rates
which is the result of Runx1’s activation of Smad6 expression, is not significantly affected by changes in $\alpha$.

### A3.3.2. Robustness of difference in WT-Runx1$^{+/\text{-}}$ activation timing predictions

The difference in activation timing between the WT and Runx1$^{+/\text{-}}$ systems is dependent upon the magnitude of the activating Bmp4 signal ($k_p$). As the signal strength is increased the time to reach 90% of the final steady state Scl concentration in the WT (see solid line, Fig. A3.3A inset) as well as Runx1$^{+/\text{-}}$-mutant (dashed line in Fig. A3.3A inset) decreases. The difference between these activation times for the two systems also decreases (solid line, Fig. A3.3A); however, the Runx1$^{+/\text{-}}$ system, with the lower Bmp4 signal threshold, consistently activates before the WT system across a large range of relevant signal levels.

The difference in activation timing between the WT and Runx1$^{+/\text{-}}$ systems is also dependent upon the Runx1 maximum production value ($v_r^o$). Fig. A3.3B shows the difference between response times for WT and Runx1$^{+/\text{-}}$ systems, in response to the same activation signal (5 times the WT activation threshold at the $v_r^o$—see Fig. 4.3B). Both the response time for the WT system and the difference in WT and Runx1$^{+/\text{-}}$ activation times increases with $v_r^o$ in the range where the Bmp4 signal threshold increases with $v_r^o$ (see Fig. 4.3B). Simulations were run at $[\text{Notch1}] = 1$.

### A3.3.3. Robustness of Runx1 requirement time-window

The earliest time that Runx1 can be knocked out through Cre-recombination mediated conditional deletion while still allowing the triad to be activated also depends on the strength of the Bmp4 signal ($k_p$) used in our simulations. As the signal strength is increased, the triad is activated faster and consequently Runx1 can be knocked down earlier without affecting triad
activation (Fig. A3.4). However our model predicts that even at high signal levels Runx1 has to be present for at least 3 days to allow for triad activation.

**A3.3.3. Qualitative robustness of WT-Runx1<sup>+/−</sup> activation timing predictions to changes in Notch dynamics**

Fig. A3.3A shows that the difference in WT-Runx1<sup>+/−</sup> activation timing is robust to changes in Bmp4 signal levels. However that comparison was made keeping [Notch1] levels at a fixed value ([Notch1]=1). Recently it has been show that [Notch1] activity decays during the emergence of HSCs. We used a phenomenological approach to determine if this dynamical change in [Notch1] would affect our results. We repeated our model simulations for response to step-increases Bmp4 signal while assuming that [Notch1] signal decreases exponentially over time ([Notch1]=exp(-t/τ); τ=6 days). We found that due to the decrease in [Notch1] over time the WT system requires a higher step-increase to switch the triad to HIGH expression state (compare Fig. A3.5A,B with Fig. 4.3C). However similar to our results with fixed [Notch1], the difference in activation timing between the WT and Runx1<sup>+/−</sup> systems decreases with the magnitude of the activating Bmp4 signal (compare Fig. A3.5C,D and Fig. A3.3). Therefore while changes in [Notch1] dynamics affect the quantitative behavior of our model, the qualitative prediction regarding the earlier activation of the triad in the Runx1<sup>+/−</sup> haploid mutant are robust to this type of dynamical perturbations in upstream signaling.
Appendix 4. Understanding the ultrasensitive dependence of sporulation decision on Spo0A~P

A4.1. Supplementary Figures

Figure A4.1. KinA induction in ASI.

A. Increase in KinA concentration with IPTG induction is not ultrasensitive. The synthesis of KinA from the Phy-spank promoter was induced by adding varying concentrations of IPTG as indicated. The KinA expression levels were measured by immunoblotting and normalized as described previously (129). The induction response can be fitted with a sigmoidal curve with Hill-coefficient of ~2 (red curve).

B. Measurement of noise in KinA expression at 4 and 10µM IPTG. KinA expression in single cells was tracked using a functional KinA-GFP fusion protein and processed by the published procedure (279). Mean KinA expression levels change ~two-fold from 4 and 10µM IPTG similar to A. The histograms of KinA expression are shown with bars and gamma distribution fits with solid curves of the same color. Note that distributions at 4 and 10µM IPTG show significant overlap. Probability densities were calculated by binning data and dividing the fraction of the population in each bin by bin size.

C. KinA concentration as a function of IPTG. The grey and colored dots indicate the measured expression level of KinA at various IPTG concentrations (data from A4.1A
normalized to absolute concentration measurements as in Ref. (279)). The black curve shows the prediction of mean KinA expression from Phy-spank as given by the model described in the methods above. Distributions of KinA at 4 and 10 μM IPTG (red and green, respectively) as computed from this model are shown on the right hand side.

D. Ultrasensitive response of the sporulation network to KinA (black curve, modeled as a Hill function with Hill coefficient of 20) converts overlapping unimodal distributions of KinA (same as panel A, shown along x-axis) to bimodal distributions of hypothetical output gene expression (distributions shown along y-axis). A threshold level of the output gene expression above which cells sporulate (grey line) is used to produce Panel E.

E. Predicted fraction of sporulating cells (solid curve) is in good agreement with measured spore fraction calculated from sporulation efficiency (blue dots) as a function of IPTG.

Figure A4.2. Ultrasensitivity of the phosphorelay increases response times.

A. Results of parameter sampling for the phosphorelay model show that a bistable phosphorelay response is associated with high ultrasensitivity (characterized by the Max. Log Gain of the IPTG-Spo0A~P response) and high response times (>10 hrs) that are physiologically unrealistic for sporulation.

B. No significant accumulation of essential downstream targets of Spo0A~P, like the spoliG operon product SpoIIGB, is seen even three hours after IPTG induction for the bistable phosphorelay response parameter sets shown in A. The parameter sets for which the phosphorelay response is bistable have high effective cooperativity of the positive feedback from Spo0A~P to Spo0F and Spo0A transcription (shown here are parameter sets for strong Spo0F feedback - Hill coefficient $n_{OF}$ > 5).
Figure A4.3. Ultrasensitive activation of $\sigma^F$ in the forespore at sub-threshold KinA concentrations.

A. Basic model of $\sigma^F$ regulation (see text).

B. SpoIIE threshold for $\sigma^F$ activation depends on the level of spoIIA operon (IIAA, IIAB, $\sigma^F$) expression. The contour diagram shows the concentration of free $\sigma^F$ predicted by the model at different SpoIIE and total $\sigma^F$ concentrations (total $\sigma^F$ is assumed to be representative of spoIIA operon expression). The numbers on each contour curve indicate the free $\sigma^F$ concentration on that contour. Note that for each given total $\sigma^F$ concentration there is a threshold concentration of SpoIIE at which free $\sigma^F$ concentration
changes dramatically from nano-molar to micromolar levels. The SpoIIE threshold increases with increasing total $\sigma^F$ concentration.

**C.** $\sigma^F$ is activated rapidly after septation. Despite the ultrasensitivity of the $\sigma^F$ activation response to SpoIIE, $\sigma^F$ is activated within ~20 minutes of septation induced SpoIIE increase through most of the range of total $\sigma^F$ and SpoIIE concentrations.

**D.** Effect of Spo0A~P up-regulation on free $\sigma^F$ concentration. An increase in KinA expression from 4$\mu$M to 10$\mu$M IPTG leads to an increase in Spo0A~P and thus increase in spollA operon and spoIIE gene expression. The change in spollA expression (total $\sigma^F$ concentration increases two-fold from 4$\mu$M IPTG to 10$\mu$M IPTG) shifts the ultrasensitive SpoIIE-free $\sigma^F$ response function (gray-4$\mu$M to black-10$\mu$M). A simultaneous increase in SpoIIE leads to a jump between response curves and a ten-fold increase in free $\sigma^F$ concentration (4$\mu$M IPTG-red dot to 10$\mu$M IPTG-green dot).

**E.** Simultaneous single cell measurements of a Spo0A~P and $\sigma^F$. Spo0A and $\sigma^F$ activity in single cells measured simultaneously using PspollG-gfp and PspollQ-mCherry reporters respectively. Fluorescence measurements from the two reporter strain at 4 (green squares) and 10 $\mu$M IPTG (orange dots) are shown here in scatter plot. Each point represents a single cell (data shown is pooled from measurements of 5 different colonies for PspollQ and 4$\mu$M IPTG different colonies for 10 $\mu$M IPTG; each colony had ~100 cells). The single variable distributions for PspollQ-mCherry and PspollG-gfp are shown alongside the y-axis and x-axis respectively. Note that the PspollG-gfp is unimodal and the PspollQ-mCherry distribution is bimodal at both at 4 (green lines) and 10 $\mu$M IPTG (orange lines). The grey bars represent thresholds used for cell-fate predictions. The PspollQ-mCherry distributions and threshold (horizontal gray bar) are the same as those shown in Fig. 5.4C. The fraction of cells above this threshold changes less than two-fold from 55% to 74% from 4 to 10 $\mu$M IPTG. The vertical gray bar represents a Spo0A~P threshold below which cells do not activate $\sigma^F$. Note that almost all cells to the left of this threshold have very little PspollQ-mCherry fluorescence indicating that $\sigma^F$ is not active.

**F.** Distributions of $\sigma^F$ activity using the PspollQ-mCherry reporter in 5 different colonies each for 4 $\mu$M (red curves) and 10 $\mu$M IPTG (green curves). Each colony had ~100 cells. Note that all distributions are bimodal around the same threshold level (grey dashed bar).
Figure A4.4. Ultrasensitive activation of $\sigma^E$ in the mother cell.

A. $\sigma^E$ activation network (see text). Vact and Vprot represent the activation and proteolytic degradation fluxes of $\sigma^E$.

B. Vprot is a saturating function of the active $\sigma^E$ concentration (red curve) whereas Vact the activation flux is independent of free $\sigma^E$. At lower Vact fluxes (blue and black horizontal lines), proteolytic degradation prevents the accumulation of $\sigma^E$ and free $\sigma^E$ levels at steady state are low (grey dots). As the activation flux increases beyond the maximum possible degradation flux, the imbalance between Vact (black line) and Vprot leads to accumulation of high levels of free $\sigma^E$.

C. Increase in KinA expression from 4µM to 10µM IPTG leads to an increase in Spo0A~P which directly increases spoIIG operon expression and indirectly increases spoIIR gene expression via up-regulation of $\sigma^F$. This simultaneous increase shifts the active $\sigma^E$-SpoIIR response curve to a lower threshold and higher maximum $\sigma^E$ value (compare 4µM (red curve) to 10µM (green curve)). In effect the change in IPTG concentration leads to a jump between the two $\sigma^E$ activation curves, which produces a
highly ultrasensitive increase in active $\sigma^E$ between 4 $\mu$M (grey dot) and 10 $\mu$M (black dot) IPTG.

D. $\sigma^F$ (or SpoIIIR) threshold for $\sigma^E$ activation depends on the level of spoIIG operon (SpoIIGA, pro-$\sigma^E$) expression. The contour diagram shows the concentration of free $\sigma^E$ predicted by the model at different $\sigma^F$ and SpoIIGA concentrations (SpoIIGA is representative of spoIIG operon expression rate). The numbers on each contour curve indicate the active $\sigma^E$ concentration on that contour. Note that for each given SpoIIGA concentration there is a threshold concentration of $\sigma^F$ at which free $\sigma^E$ concentration changes dramatically from nanomolar to micromolar levels. The $\sigma^F$ threshold decreases with increasing SpoIIGA concentration.

E. Distributions of $\sigma^E$ activity using the PspoIID-GFP reporter in 4 different colonies each for 4 $\mu$M (red curves) and 10 $\mu$M IPTG (green curves). Each colony had ~100 cells. Note that all distributions are bimodal.
Figure A4.5. Low spolIG expression limits $\sigma^E$ activation and sporulation below the KinA threshold.

A. Stochastic simulations of the sporulation network model show that fraction of sporulating cells, based on a threshold level of $\sigma^E$ activity, increases at a lower IPTG level when spolIG expression is constitutive in the ASI strain harboring the Pspac$^C$-spolIG (MF4883) (solid curve) as opposed to when it is Spo0A activity dependent (dashed curve, same as in Fig. 5.4). A Hill-equation fit of the fraction of $\sigma^E$ active cells as a function of IPTG for this strain (Pspac$^C$-spolIG) has a half-maximal threshold of 4.35 µM IPTG (95% confidence interval: 4.232-4.458) and a Hill-exponent of nH~5. These values are comparable to those for the fraction of $\sigma^F$ active cells (dotted curve, same as in Fig. 5.3) but below those for the dashed curve (half-maximal threshold of 6.74 µM IPTG, n~7). These result indicate that constitutive expression of spolIG decreases the sensitivity and threshold of sporulation response to KinA induction.

B. Images showing $\sigma^E$ activity in single-cells of MF4883, tracked using the PspollD-gfp reporter at T3 for 4 µM (top) and 10 µM IPTG (bottom). Note that a significant fraction of cells exhibits $\sigma^E$ activity even at 4 µM IPTG (compare with Fig. 5.4E).

C. Bimodal distributions of $\sigma^E$ activity at 4 µM (red lines) and 10 µM IPTG (green lines) for MF4883. Histograms of $\sigma^E$ activity are shown for two independent experiments at each IPTG concentration. Fluorescence measurements shown are the mean fluorescence of each cell in arbitrary units (a.u.). The gray dashed line indicates the $\sigma^E$ activity threshold used to classify cells as $\sigma^E$ active. The fractions of $\sigma^E$ active cells computed using these distributions are in excellent agreement with computational predictions in (A) (4µM IPTG: red circle; 10µM IPTG: green square).
Figure A4.6. $\sigma^F$ activation does not ensure $\sigma^E$ activation or completion of sporulation in starving wild-type cells.

Triple reporter strain (MF4859) was used to simultaneously measure Spo0A and sigma factors activities in wild-type cells grown in starvation media.

A. $\sigma^F$ activity ($PspollQ$-mCherry reporter) in wild-type cells at 2.5 hours after suspension in sporulation media shows a bimodal distribution whereas Spo0A activity ($PspollA$-cfp) is broadly heterogeneous. A threshold level of $\sigma^F$ activity (grey dashed horizontal bar) can be used to divide the population into $\sigma^F$ active/inactive cells. Note that the number of $\sigma^F$ active cells changes dramatically upon crossing a Spo0A activity threshold (grey dashed vertical bar).

B. Similarly $\sigma^E$ activity ($PspollD$-yfp) is bimodally distributed at 2.5 hours after after suspension in sporulation media. Number of $\sigma^E$ active cells (YFP fluorescence derived from $PspollD$ greater than the threshold indicated by the grey dashed horizontal bar) also changes dramatically upon crossing a Spo0A activity threshold (grey dashed vertical bar).

C. Time-lapse microscopy was used to track $\sigma^F$ activity and cell-fate in single cells over 24 hours. Out of 400 cells tracked 79% of cells formed an asymmetric septum and activated $\sigma^F$. Among the $\sigma^F$ active cells 84% of cells went on to engulf the forespore and completed sporulation, some cells died post engulfment and ~2.5% of cells failed to engulf the forespore despite having activated $\sigma^F$.

D. Single-cell tracking of $\sigma^E$ activity. Trajectories of $\sigma^E$ activity in cells that sporulate (yellow curves) and cells that remain vegetative (grey curves) are clearly distinguishable. All cells that end up as spores cross a minimum threshold of $\sigma^E$ activation (dashed horizontal line).
E. Out of 280 cells tracked using time-lapse microscopy, 61% of cells activated $\sigma^E$. Among these $\sigma^E$ active cells 89% of cells completed sporulation while 21% cells died during the process of sporulating. All cells that activated $\sigma^E$ went on to engulf the forespore.

**A4.2. Supplementary Tables**

**Table A4.1. Strains used in this study**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype or description</th>
<th>Fig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MF1027</td>
<td>amyE::PspollQ-gfp spc</td>
<td>5.5C-F and Fig. A4.6C</td>
</tr>
<tr>
<td>MF248</td>
<td>amyE::PspollID-gfp spc</td>
<td>A4.6D,E</td>
</tr>
<tr>
<td>MF1917</td>
<td>kinAΩPhy-spank-kinA cm, amyE::PspollI-G-lacZ spc</td>
<td></td>
</tr>
<tr>
<td>MF1996</td>
<td>kinAΩPhy-spank-kinA-gfp spc kan</td>
<td></td>
</tr>
<tr>
<td>MF1929</td>
<td>kinAΩPhy-spank-kinA cm, amyE::PspollIG-gfp spc</td>
<td></td>
</tr>
<tr>
<td>MF1956</td>
<td>kinAΩPhy-spank-kinA cm, amyE::PspollIQ-gfp spc</td>
<td></td>
</tr>
<tr>
<td>MF1957</td>
<td>kinAΩPhy-spank-kinA cm, amyE::PspollID-gfp spc</td>
<td>5.4C-E</td>
</tr>
<tr>
<td>MF3799</td>
<td>kinAΩPhy-spank-kinA cm, thrC::PspollIQ-mCherry erm</td>
<td>5.3C-E and Fig. A4.3EF</td>
</tr>
<tr>
<td>MF3765</td>
<td>kinAΩPhy-spank-kinA cm, amyE::PspollIQ-gfp spc, thrC::PspollIQ-mCherry erm</td>
<td></td>
</tr>
<tr>
<td>MF4856</td>
<td>spoIIAΩPspollA-cfp cm, amyE::PspollIQ-yfp spc</td>
<td></td>
</tr>
<tr>
<td>MF4857</td>
<td>spoIIAΩPspollA-cfp cm, amyE::PspollID-yfp spc</td>
<td></td>
</tr>
<tr>
<td>MF4859</td>
<td>spoIIAΩPspollA-cfp cm, amyE::PspollIQ-mCherry spc, thrC::PspollID-yfp erm</td>
<td>5.5AB and Fig. A4.6AB</td>
</tr>
<tr>
<td>MF4883</td>
<td>kinAΩPhy-spank-kinA cm, ΔspollGAB::tet, amyE::PspacC-spoIIGAB spc, thrC::PspollID-gfp erm</td>
<td>A4.5A-C</td>
</tr>
<tr>
<td>Purpose</td>
<td>Name</td>
<td>Sequence</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>----------</td>
</tr>
<tr>
<td><em>Pspoll</em>Q</td>
<td>omf42</td>
<td>5'-gcggaatccggtatcgcgtttaccatt-3'</td>
</tr>
<tr>
<td><em>Pspoll</em>Q</td>
<td>omf43</td>
<td>5'-gcgaagcttagcaacattgaacacttt-3'</td>
</tr>
<tr>
<td><em>Pspoll</em>D</td>
<td>omf37</td>
<td>5'-gcggaatcgcagacatcgaacgtgtaat-3'</td>
</tr>
<tr>
<td><em>Pspoll</em>D</td>
<td>omf38</td>
<td>5'-gcgaagcttcgctcgggattgactcta-3'</td>
</tr>
<tr>
<td>cfp</td>
<td>omf39</td>
<td>5'-gccaagcctacataagggaggaactactatggtcagcaagggagaggaagat-3'</td>
</tr>
<tr>
<td>cfp</td>
<td>om50</td>
<td>5'-gccaagcctacataagggaggaactactatggttcaaagggcaagaactg-3'</td>
</tr>
<tr>
<td>cfp</td>
<td>om52</td>
<td>5'-gcgggatccctttacttataaagttgcgtccatgccagagtaagggagaagaact-3'</td>
</tr>
<tr>
<td>yfp</td>
<td>omf31</td>
<td>5'-gcgaagctacataagggaggaactactatggtcagcaagggagaggaagat-3'</td>
</tr>
<tr>
<td>yfp</td>
<td>omf39</td>
<td>5'-gcgaagctacataagggaggaactactatggtcagcaagggagaggaagat-3'</td>
</tr>
<tr>
<td>mCherry</td>
<td>om84</td>
<td>5'-gcgggatctttgtagttgtatcgcctgc-3'</td>
</tr>
<tr>
<td>mCherry</td>
<td>om87</td>
<td>5'-gcgaagctacataagggaggaactactatggtcagcaagggaggaagat-3'</td>
</tr>
<tr>
<td>mCherry</td>
<td>om88</td>
<td>5'-gcgggatctttgtagttgtatcgcctgc-3'</td>
</tr>
</tbody>
</table>
Table A4.3. Model parameters for transcriptional regulation

<table>
<thead>
<tr>
<th>Gene</th>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>spo0A</td>
<td>$k_v$</td>
<td>Maximum transcription rate for $P_v$</td>
<td>50 nM hr$^{-1}$</td>
<td>(263, 288)</td>
</tr>
<tr>
<td></td>
<td>$K_v$</td>
<td>Binding affinity for $P_v$ repression site</td>
<td>100 nM</td>
<td>(130)</td>
</tr>
<tr>
<td></td>
<td>$k_s$</td>
<td>Maximum transcription rate for $P_s$</td>
<td>500 nM hr$^{-1}$</td>
<td>(263, 288, 367)</td>
</tr>
<tr>
<td></td>
<td>$K_s$</td>
<td>Binding affinity for $P_s$ activation sites</td>
<td>100 nM</td>
<td>(130)</td>
</tr>
<tr>
<td>spo0F</td>
<td>$k_{vF}$</td>
<td>Maximum transcription rate for $P_v$</td>
<td>72 nM hr$^{-1}$</td>
<td>(279), (130)</td>
</tr>
<tr>
<td></td>
<td>$k_{sF}$</td>
<td>Maximum transcription rate in sporulation conditions</td>
<td>360 nM hr$^{-1}$</td>
<td>(130)</td>
</tr>
<tr>
<td></td>
<td>$K_{sF}$</td>
<td>Binding affinity for $P_s$ activation sites</td>
<td>50 nM</td>
<td>(130)</td>
</tr>
<tr>
<td>spoIIA</td>
<td>$v_{bIIA}$</td>
<td>Basal transcription rate</td>
<td>100 nM hr$^{-1}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$f_{IIA}$</td>
<td>Fold Change</td>
<td>15</td>
<td>(130)</td>
</tr>
<tr>
<td></td>
<td>$n_{IIA}$</td>
<td>Hill coefficient</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_{IIA}$</td>
<td>0A−P binding affinity</td>
<td>140 nM</td>
<td>(130)</td>
</tr>
<tr>
<td>spoIIE</td>
<td>$v_{bIIE}$</td>
<td>Basal transcription rate</td>
<td>75 nM hr$^{-1}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$f_{IIE}$</td>
<td>Fold Change</td>
<td>15</td>
<td>(130)</td>
</tr>
<tr>
<td></td>
<td>$n_{IIE}$</td>
<td>Hill coefficient</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_{IIE}$</td>
<td>0A−P binding affinity</td>
<td>230 nM</td>
<td>(130)</td>
</tr>
<tr>
<td>spoIIG</td>
<td>$v_{bIIG}$</td>
<td>Basal transcription rate</td>
<td>80 nM hr$^{-1}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$f_{IIG}$</td>
<td>Fold Change</td>
<td>16</td>
<td>(130)</td>
</tr>
<tr>
<td></td>
<td>$n_{IIG}$</td>
<td>Hill coefficient</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_{IIG}$</td>
<td>0A−P binding affinity</td>
<td>1700 nM</td>
<td>(130)</td>
</tr>
</tbody>
</table>
A4.3. Mathematical Modeling of the Sporulation Network

Our model of the sporulation network includes three modules: the phosphorelay, \( \sigma^F \) activation and \( \sigma^E \) activation. Each module was studied independently with deterministic and stochastic simulations to understand its functional role in the network. Thereafter a combined model was used to predict the cell fate response as described below.

*Modeling Induced KinA expression from the Phy-spank promoter*

We chose the following model for KinA expression from the IPTG ([I]) inducible \( P_{\text{hy-spank}} \) promoter.

\[
\begin{align*}
\text{Promoter}_{\text{off}} & \xrightarrow{k_{\text{on}}} \text{Promoter}_{\text{on}} \\
\text{Promoter}_{\text{on}} & \xrightarrow{k_1} \text{Promoter}_{\text{on}} + \text{mRNA} \\
\text{mRNA} & \xrightarrow{k_2} \text{mRNA} + \text{KinA} \\
\text{mRNA} & \xrightarrow{\gamma_1} \emptyset \\
\text{Protein} & \xrightarrow{\gamma_2} \emptyset
\end{align*}
\]

Based on this model the steady state concentration of KinA production is given by:

\[
[\text{KinA}] = \frac{k_{\text{on}} k_1 k_2}{k_{\text{on}} + k_{\text{off}} \gamma_1 \gamma_2}
\]  \[1\]

The Promoter ON to OFF transition depends on the binding of the Lac-repressor dimer which was assumed to have some fixed concentration \([R_T]\). We assumed that IPTG binds to Lac-repressor dimers and reduces the effective free repressor concentration. Following this assumption, the dependence of Promoter OFF-rate \( k_{\text{off}} \) on IPTG concentration \([I]\) is given by:

\[
k_{\text{off}} ([I]) = k_0 \frac{[R_T]}{1 + ([I]/K)^2}
\]  \[2\]
Here $K$ is the dissociation constant for Lac-dimer IPTG binding. Using equations [1] and [2], we can rewrite the steady state $[\text{KinA}]$ as a function of IPTG concentration [I]:

$$[\text{KinA}] = \frac{k_{on} k_{k_2}}{k_{on} + k_0[R_T]} \frac{k_2}{\gamma_1 \gamma_2} \left( 1 + \frac{[I]^2 (k_0[R_T]/k_{on})}{K^2 (1+k_0[R_T]/k_{on}) + [I]^2} \right) = \frac{v_b}{\gamma_2} \left( 1 + \frac{[I]^2 f}{K_m^2 + [I]^2} \right)$$  \[3\]

$$v_b = \frac{k_{on} k_2}{k_{on} + k_0[R_T]} \gamma_1$$  \[4\]

$$f = \frac{k_0[R_T]}{k_{on}}$$  \[5\]

$$K_m^2 = K^2 (1+k_0[R_T]/k_{on})$$  \[6\]

Here $v_b$, $f$ and $K_m$ represent the basal expression rate, fold-change and half-maximal constant for the IPTG-dependent promoter. We assumed that the combined degradation-dilution rate for KinA is $\gamma_2 = 1$ hr$^{-1}$. KinA expression data from (129) and (279), was fit with this IPTG-dose dependent equation to determine $v_b$, $f$ and $K_m$. Fig. A4.1A shows a fit with $f = 3.53$, $K_m = 6$µM and $v_b = 0.67$ µM hr$^{-1}$. Using these values in equations [5] and [6], we found $k_0[R_T]/k_{on} = 3.53$ and $K^2 = 7.95$ µM$^2$. These estimates are consistent with previously reported biochemical measurements for the LacI repressor (409). We also used a P$_{hy-spank}$-KinA-GFP fusion to quantify the distributions of $[\text{KinA}]$ at 4 and 10 µM IPTG (see Fig. A4.1B). Following the methods described in (57), these distributions were fit using gamma distributions (solid curves in Fig. A4.1B) to determine the mean number of KinA bursts-a and the mean burst size-b (µM). Fig. A4.1B shows gamma distribution fits with $a = 11$, $b = 0.11$ for 4µM IPTG and $a = 11$, $b = 0.218$ for 10µM IPTG. Note that IPTG dose does not affect the mean number of bursts. This is consistent with our model for KinA production where IPTG only affects $k_{off}$, the rate at which the P$_{hy-spank}$ promoter turns off. We followed (410) to find an expression for the burst size as a function of IPTG concentration [I]:

$$b([I]) = \frac{k_{k_2}}{\gamma_1} \left( 1 + \frac{k_1}{k_{off} ([I])} \right)$$  \[7\]
Using the measurements for burst size $b([I])$ at 4 and 10 µM IPTG in this equation we determined $k_1 = 30.4 \text{ nM hr}^{-1}$, $k_{0[R]} = 345.4 \text{ hr}^{-1}$ and $k_{on} = 97.8 \text{ hr}^{-1}$. Then using the fact that $[\text{KinA}] = a \cdot b$ we found that $k_2/\gamma_1 = 100$ protein molecules per mRNA. We assumed a half-life of 10 min for KinA mRNA which implied $k_2 = 10 \text{ hr}^{-1}$. These kinetic constants completely characterize the $P_{\text{hy-spank}}$ promoter and the production/degradation of KinA protein and mRNA for our model and accurately predict both the deterministic dynamics of these species and the stochastic fluctuations of [KinA] in the ASI system.

An ultrasensitive single-cell network response can explain a switch-like increase in sporulation efficiency

In the ASI system (Fig. 5.1B), induction of KinA expression beyond a certain extent leads to a massive increase in the number of spore-forming cells (Fig. 5.1D) (129). Surprisingly, this increase occurs in a narrow range between 4 and 10 µM IPTG. In this range the number of cells forming spores increases ~20-fold while KinA concentration increases less than two-fold (Fig. 5.1D and Fig. A4.1A,C), indicating that at the population level $B. subtilis$ cells show an ultrasensitive response to a threshold concentration of KinA. We hypothesized that this ultrasensitivity of the population-level response could be explained if the expression of essential sporulation genes increased abruptly above the threshold of KinA.

To test this hypothesis theoretically, we illustrate how the fraction of sporulating cells varies with KinA induction using a model ultrasensitive response network. For simplicity, we did not consider details of the sporulation network and assumed that increasing the synthesis of KinA results in an ultrasensitive increase in the expression of some key sporulation gene. KinA expression from the IPTG inducible promoter was modeled based on experimental characterization of the $P_{\text{hy-spank}}$ promoter as described in the previous section. The input-output relation between KinA and the hypothetical target gene was modeled as a Hill function similar to equation [3] above with a Hill-coefficient of
20, half-maximal constant 2.2 μM of [KinA] and basal rate 0.8 μM hr⁻¹ and fold-change 10. Gillespie simulations of this network at different IPTG concentrations between 0-20 μM were used to estimate the distributions of [KinA] and gene three hours after KinA induction. Fig. A4.1D illustrates that this ultrasensitive transfer function (black curve) translates the overlapping and noisy distributions of KinA expression at low (4µM) and high (10µM) IPTG into well separated bimodal distributions of output gene expression (red and green distributions along the y-axis). To determine the fraction of sporulating cells from these distributions we used a threshold concentration to define which cells have enough expression of the hypothetical gene and which cells do not. Only cells that had higher-than-threshold concentrations were counted as sporulating. Using this threshold, for output gene expression, we can generate the relationship between the IPTG concentration and the cell fate. The resulting mathematical prediction of cell fate (Fig. A4.1E, solid black curve) is in good agreement with the experimentally measured values (Fig. A4.1E, purple dots; calculated from observed values of the sporulation efficiency = spores/total cells; see Methods in main text) (7).

This simple model illustrates how a modest two-fold change in the mean level of KinA can lead to a much larger increase in the fraction of sporulating cells if the change in KinA level leads to an ultrasensitive change in the expression of a set of sporulation genes. Based on this, we hypothesize that the abrupt increase in the number of sporulating cells above the threshold level of KinA does indeed originate from ultrasensitivity in individual cells. In the following sections we test this hypothesis and uncover the mechanism for generating such an ultrasensitive transfer function.

The Phosphorelay

Several models have been proposed for the phosphorelay (263, 307, 333, 411). Our model is based on the same details of post-translational reactions and
transcriptional regulation as these earlier models but we have modified certain crucial aspects and extended these models as discussed here.

Our model of the phosphorelay includes all species shown in Fig. 5.1B of the main text. Unlike the model of Jabbari et al. (411), we have not explicitly included the effects of transition-state regulators AbrB, SinI, SinR. We have assumed, based on recent experimental reports, that these regulators are quickly repressed early on in sporulation and have little effect on the dynamics of Spo0A~P accumulation or the expression of its targets (263). σ^H expression and activity are not explicitly modeled either since it is activated when AbrB is repressed and is not expected to be rate-limiting (263).

Spo0E is explicitly included in our model; we have assumed that it accumulates over time starting from low levels and does have an important effect on the dynamics of the phosphorelay. As shown in Fig. 5.1B, spo0F and spo0A expression are directly regulated by Spo0A~P in the model (70, 331, 412). In addition, Spo0A~P up-regulates the transcription from spolla, spolle and spollig promoters. A single Rap phosphatase species was included and was assumed to be representative of all Spo0F phosphatases similar to the model of Bischofs et al. (307).

Post-translational interactions in the phosphorelay include the following reactions (70, 290):

\[
\begin{align*}
\text{KinA} & \xrightleftharpoons[k_{\text{br}}]{k_p} \text{KinA}_p \\
\text{KinA}_p + \text{Spo0F} & \xrightarrow[k_{\text{ab}}]{k_i} \text{KinA}_p::\text{Spo0F} \xrightarrow[k_i]{k_i} \text{KinA} + \text{Spo0F}_p \\
\text{KinA} + \text{Spo0F} & \xrightarrow[k_i]{k_i} \text{KinA}::\text{Spo0F} \\
\text{Spo0F}_p + \text{Spo0B} & \xrightarrow[k_i]{k_i} \text{Spo0F}_p::\text{Spo0B} \xrightarrow[k_i]{k_i} \text{Spo0F} + \text{Spo0B}_p \\
\text{Spo0B}_p + \text{Spo0A} & \xrightarrow[k_i]{k_i} \text{Spo0B}_p::\text{Spo0A} \xrightarrow[k_i]{k_i} \text{Spo0B} + \text{Spo0A}_p \\
\text{Spo0F}_p + \text{Rap} & \xrightarrow[k_i]{k_i} \text{Spo0F}_p::\text{Rap} \xrightarrow[k_i]{k_i} \text{Spo0F} + \text{Rap} \\
\text{Spo0A}_p + \text{Spo0E} & \xrightarrow[k_i]{k_i} \text{Spo0A}_p::\text{Spo0E} \xrightarrow[k_i]{k_i} \text{Spo0A} + \text{Spo0E}
\end{align*}
\]
Here subscript P marks phosphorylated forms of the proteins, colon denotes protein complexes and each rate constant is introduced above its corresponding reaction arrow. All post-translational reactions were modeled with mass-action kinetics. Binding rates $k_b$ for all reactions were assumed to be diffusion limited with rate constant $1\text{nM}^{-1}\text{hr}^{-1}$. Spo0E and Rap mediated dephosphorylation were assumed to have catalytic rate constants $20\text{hr}^{-1}$ based on (413). All other rate constants were estimated from *in vitro* measurements of phosphorelay kinetics (290). KinA dephosphorylation is assumed to be negligible (367). These parameter values have order of magnitude agreement with the parameters used in other models (333, 411).

Transcription and translation were modeled explicitly for all species in the model. For *spo0B*, *spo0E* and *rap* we assumed constant rates of transcription. The same rates of translation and mRNA degradation rates were used for all genes except in certain instances that are specifically indicated below. mRNA degradation rates were assumed to be $10\text{hr}^{-1}$ (based on a ~5 min half-life). Translation rates were assumed to be $40\text{mRNA}^{-1}\text{hr}^{-1}$ (based on 4 proteins synthesized per mRNA). These estimates are similar to those used in (411). Degradation and dilution were assumed to have a combined rate of $1\text{hr}^{-1}$ for all genes except where specifically noted.

In modeling all transcriptional regulatory we assumed that binding-unbinding of transcription factors from their promoters is fast and therefore we have modeled their rates of transcription with appropriate Hill-functions. In the ASI, Spo0A-P controls the transcription of the phosphorelay genes - *spo0A* and *spo0F* whereas *spo0B* is transcribed constitutively ($v_{spo0B} = 10\text{nMhr}^{-1}$; ref. (279)). *kinA* expression from the *Phy-spank* promoter is controlled by IPTG as described above. Spo0A transcription is controlled by two promoters, namely $P_v$ and $P_s$. Chastanet and Losick (288) have shown that the weak vegetative promoter $P_v$ is active during exponential phase but repressed by Spo0A-P. On the other hand, the sporulation specific promoter $P_s$, is activated by binding of Spo0A-P to a region 70-90 bp upstream of $P_s$. $P_s$ has also been shown to be transcribed in a
σ^H-dependent manner. Since *spo0A* expression from *P_v* is much weaker than that from *P_s*, overall a σ^H-mediated positive feedback regulates *spo0A* transcription. Based on these results we used the following Hill-type equation to model *spo0A* expression from *P_v* and *P_s*:

\[
v_{spo0A} = k_v \frac{K_v}{K_v + [A^P]^i} + k_s \frac{[A^P]^i}{K_s^i + [A^P]^i}
\]

Here the first and second terms represent that rate of transcription from *P_v* and *P_s* respectively. [A^P] represents the concentrations of Spo0A~P. We assume that σ^H concentrations do not change significantly during the course of the experiments and therefore the effect of σ^H on *P_s* transcription can be absorbed into the rate *k_s*. Table A4.3 describes the parameters and parameter values used for this model of *spo0A* transcriptional regulation.

*spo0F* transcription is controlled by Spo0A~P via a σ^H-dependent promoter. We use the following Hill-type equation to model Spo0A~P controlled transcription of *spo0F*:

\[
v_{spo0F} = k_v + k_s \frac{[A^P]^i}{K_s^i + [A^P]^i}
\]

Here the first and second terms represent the maximum rate of transcription of *spo0F* in vegetative and sporulation conditions respectively. Table A4.3 describes the parameters and parameter values used for this model of *spo0F* transcriptional regulation.

For *spollA*, *spollE* and *spollG* we assumed that binding-unbinding of Spo0A~P from their promoters is fast and model their rates of transcription with the following Hill-function (also see Table A4.3):

\[
v_i = vb \left( 1 + \frac{[A^P]^h_i}{K_h^i + [A^P]^h_i} \right) \quad i \in \{IIA, IIIE, IIG\}
\]
Here $v_{bi}$, $f_i$ and $n_i$ represent the basal transcription rate, fold-change and Hill-coefficient for the regulation of each gene. The details of these regulatory interactions (130, 288) are as follows. Operons *spollA* and *spollG* produce polycistronic mRNAs that are translated into multiple products. The *spollA* operon produces SpoIIAA, SpoIIAB and SpoIIAC ($\sigma^F$). Production of SpoIIAA and SpoIIAB is known to be translationally coupled (414) and based on measurements of protein concentration we assumed their translation rate is five-fold 5-fold higher than that of SpoIIAC (415, 416). In addition, SpoIIAB was assumed to have a half-life of 30 min based on observations that a C-terminal sequence targets it for degradation *in vivo* (417). This effectively leads to [SpoIIAA]>[SpoIIAB]>[SpoIIAC] (at steady state [SpoIIAA]~2[SpoIIAB] and ~4[SpoIIAC]) which is essential for keeping $\sigma^F$ sequestered in the cell before septation while allowing rapid activation after septation in the forespore (270). The *spollG* operon produces the proteins SpoIIGA and SpoIIGB (pro-$\sigma^E$). It has been shown that SpoIIGB concentration is 10-fold higher than SpoIIGA which has a half-life of 20 min (418). We accordingly assumed that relative to SpoIIGB, SpoIIGA has a ~3-fold lower translation rate (12 mRNA$^{-1}$ hr$^{-1}$) and a 3-fold higher degradation rate (3 hr$^{-1}$).

*spollG* expression from the *Pspac* promoter, which is independent of IPTG concentration, was found to be well described by a gamma distribution similar to the expression of *spollG* in the ASI at 10 $\mu$M IPTG. Thus to model constitutive *spollG* expression from the *PspacC* promoter, the transcription rate of the *spollG* operon was set to a constant rate of 800 nM hr$^{-1}$, which is comparable to the rate of transcription of *spollG* at 10 $\mu$M IPTG.

We used the *ode45* simulator of MATLAB (©MathWorks) to study the deterministic response of the phosphorelay network to different IPTG doses (Fig. 5.2A). Global parameter sampling for the phosphorelay was used to determine if steady state Spo0A~P concentrations can increase ultrasensitively between 4 and 10$\mu$M IPTG (see section on Logarithmic Gains for definition of ultrasensitivity). Spo0A~P activity was quantified in all simulations using the rate
of transcription from the \textit{PspollG} promoter to maintain consistency with experiments.

Stochastic simulations for different IPTG doses were performed using the SSA algorithm of the Stochkit package (292). We assumed that promoter-transcription factor binding/unbinding steps are fast such that transcriptional regulation can be modeled using Hill-equations even for stochastic simulations. To estimate average response times (Fig. A4.2), stochastic simulations at each IPTG concentration were run till the 20 hrs. after IPTG induction. Subsequently, the first time a trajectory crosses 50% of the maximum value of \([\text{SpoIIGB}]\) concentration was recorded as the response time. Response times for each parameter set at each IPTG dose correspond to the average of 200 trajectories. The Log-Gains of the phosphorelay response were computed using the average \([\text{SpoIIGB}]\) concentration at 10hrs. after induction as output and the IPTG concentration as input in equation [8] below. These Log-Gains were used to construct Fig. A4.2. Parameter sets that display high Log-Gains were examined with deterministic simulations for the steady state response to IPTG and most of them were found to display bistability as indicated in Fig. A4.2.

\textit{The } \sigma^F \textit{ activation module}

Our model for the \(\sigma^F\) activation module is based on the model of Igoshin (270). We refer the reader to this previous report for detailed descriptions of all the post-translational reactions involved in this module and summarize them here (also see Fig. A4.3A). SpoIIAA and SpoIIBAB are present in multiple states: phosphorylation of SpoIIAA by SpoIIBAB (serine kinase) results in free, phosphorylated (inactive) SpoIIAA-P and a SpoIIBAB-ADP complex. Free SpoIIBAB-ATP forms a complex with \(\sigma^F\) and prevents it from activating gene expression by associating with RNA polymerase. SpoIIE, a membrane associated phosphatase can dephosphorylate and activate SpoIIAA. If enough SpoIIAA is available it can bind SpoIIBAB-ADP and sequester it in a dead-end complex, thereby titrating away SpoIIBAB and freeing up \(\sigma^F\). However SpoIIE
concentrations in the pre-divisional cell are too low to dephosphorylate sufficient amounts of SpoIIAA~P. Upon asymmetric division SpoIIIE localizes to the septum and is concentrated on the forespore face of the septum. Due to the asymmetry of septum placement, the forespore volume is one-eighth of the cell volume. This effectively leads to a 8-fold increase in concentration of SpoIIIE in the forespore. The high SpoIIIE concentration activates SpoIIAA and leads to rapid release of free σ^F. Following the results of Iber (269), we assume that due to the allostery of SpoIIAB (AB), binding of one monomer of AA (k_{on1}=100\mu M^{-1} hr^{-1}, k_{off1}=2.5 s^{-1}) increases the affinity for a second AA monomer ~3000-fold (k_{on2}=100\mu M^{-1} hr^{-1}, k_{off2}=7.5e^{-3} s^{-1}). This increase in affinity allows AA to effectively bind to AB-ADP and sequester it thereby preventing the titration of σ^F (σ^F-AB binding rates: (k_{on}=0.4\mu M^{-1} s^{-1}, k_{off}=5e^{-5} s^{-1}). σ^F affinity for AA-bound AB is assumed to significantly lower due to conformational changes induced by AA binding (k_{off}=0.4 s^{-1}). We also include SpoIIIE (IIE) catalyzed dephosphorylation of phosphorylated AA explicitly in the model of Igoshin (270).

\[ \text{AA}_p + \text{IIE} \xrightleftharpoons{k_{on}}{k_{off}} \text{AA}_p: \text{SpoIIIE} \xrightarrow{k_{cat}} \text{AA} + \text{IIE} \]

In this scheme, IIE-AA binding and unbinding and catalytic rates based on the results of (269) are taken to be: k_{on}=0.36 nM^{-1} hr^{-1}, k_{off}=1800 hr^{-1} and k_{cat}=31.32 s^{-1}, which represent a binding affinity of 5µM.

Similar to the model setup of Igoshin (270), our model for the σ^F activation module takes [AA], [AB], [AC] (total [σ^F]) and [IIE] to predict the active (free) [σ^F] concentration. Following our model of spoIIA expression in the phosphorelay, [AC] was assumed to be representative of the level of spoIIA operon expression and [AA] and [AB] concentrations were assumed to be proportional to [AC] ([AA]=4[AC], [AB]=2[AC]). As a result, the output of σ^F activation module is a two variable function that uses [AC] and [IIE] concentrations to predict active [σ^F]. We define this output as [σ^F]=f([AC],[IIE]). We used MATALB ode45 solver to solve the equations in the σ^F activation module over long time-spans to find the steady state response of this module. Steady state was reached within 60 min for the
whole range of [IIE] (0-30 µM) and [AC] (0-12 µM) concentrations that we studied. These steady states were insensitive to whether or not [IIE] was available before [AC] could be sequestered by anti-sigma factor AB. Notably active [σF] increase dramatically at a threshold concentration of IIE and this threshold depends on the AC concentration (Fig. A4.3B). Using the trajectories of the module response observed in the ODE-solutions we determined the minimum time required to reach 50% of the steady state active σF concentration at different combinations of [IIE] and [AC]. As shown in Fig. A4.3C, these response times lie within the 20 min range showing that the σF is activated rapidly in response to IIE availability and quickly reaches steady state levels.

To integrate the phosphorelay and σF modules, we used AC and IIE concentrations provided by the phosphorelay module at 2 hours after IPTG induction as inputs for the σF activation module (AA-C and IIE concentrations reach steady state around 1.5-2 hours after IPTG induction). IIE concentrations were increased eight-fold for these simulations to account for the asymmetry of polar septation (419). spoIIR expression was included in the σF activation module for these simulations. SpoIIR transcription was modeled with a hyperbolic (Hill-function with exponent 1, i.e. Michaelis-Menten) dependence on active σF concentration. This type of non-cooperative effect is typical for sigma-factor regulated gene expression. We used a basal transcription rate 200nMhr⁻¹, fold-change=5 and half-maximal constant = 4µM σF. Translation rate and mRNA degradation rate were assumed to be 40 hr⁻¹ and 10 hr⁻¹ respectively. We note that since the forespore volume is 1/8th the volume of B. subtilis cells (~1fL), 1 molecule in the forespore corresponds to ~8nM. SpoIIR degradation rate has not been measured and was assumed to be 5 hr⁻¹ based on the rapid accumulation and degradation rates of GFP expressed from PspoIIR reported in (274). Both active σF and SpoIIR concentrations increase rapidly after septation and reach steady state within 30 minutes (Fig. A4.3C). We used the steady state level (~1 hr after septation or equivalently 3 hrs after induction of KinA) of active σF and SpoIIR as the outputs of this module.
At each IPTG concentration, distributions of active $\sigma^F$ and SpoIIR concentrations, one hour after septation, were calculated using stochastic simulations with SSA algorithm of the Stochkit package (292) using AC and IIE concentrations provided by the stochastic version of the phosphorelay module. Fig. 5.4A in the main text shows the average active $\sigma^F$ concentration obtained from 400 stochastic simulations at each IPTG dose one hour after septation. Variability of spoIIA and spoIIE expression combined with the ultrasensitive response of the $\sigma^F$ activation module produces bimodal distributions of activity for IPTG>3 µM (see Fig. 5.3). To predict cell fate based on these distributions we used a threshold of 4µM of $\sigma^F$ and counted all cells above the threshold as sporulating cells (Fig. 5.3D). It may seem surprising that although all cells were assumed to be capable of forming an asymmetric septum in our simulations we were still able to accurately match the fraction of cells that activate $\sigma^F$ rather than grossly overestimating it. This result is rationalized by experimental results from wild-type cells sporulating under starvation that show that during starvation many cells form an asymmetric septum without activating $\sigma^F$ (274, 282, 283) and therefore the SpoIIE requirement for $\sigma^F$ activation is higher than for asymmetric septum. Therefore, in our model, cells that are predicted to have enough IIE to activate are very likely to have a septum.

The $\sigma^E$ activation module

$\sigma^E$ exists in two forms in the cell: a membrane-tethered, stable, transcriptionally inactive form (pro-$\sigma^E$) and a cytosolic, unstable and transcriptionally active form ($\sigma^E$) (265, 278, 291). Prior to septation, transcription of the gene for $\sigma^E$ (spoIIGB) generates the inactive form pro-$\sigma^E$. The N-terminus pro-sequence of the inactive form prevents it from activating the transcription of $\sigma^E$ regulon while simultaneously protecting it from proteolytic degradation (420). After asymmetric septation, the pro- sequence is proteolytically cleaved by the aspartic protease SpoIIGA (hereafter IIGA) which is synthesized along with pro-$\sigma^E$ from the spoIIG operon. Notably SpoIIGA only acts as an efficient protease for pro-$\sigma^E$ when bound to the SpoIIR hereafter (IIR), a $\sigma^F$ target gene that is
specifically expressed in the forespore (5, 8). As a result, $\sigma^E$ is activated in a compartment-specific manner in the mother cell only after polar septation and $\sigma^F$ activation (265, 278, 291). The activated form of $\sigma^E$ is transcriptionally active but also proteolytically labile. Our model of the $\sigma^E$ activation module includes the following reactions (see also Fig. A4.4A):

$$2\text{IIIGA} \xrightleftharpoons[k_{ba}]{k_{ab}} \text{IIGA}_2$$

$$\text{IIGA}_2 + \text{pro-}\sigma^E \xrightarrow[k_{ba}]{k_{ab}} \text{IIGA}_2:\text{pro-}\sigma^E \xrightarrow[k_{s}]{} \text{IIGA}_2 + \sigma^E$$

$$\text{IIIGA}_2 + \text{IIR} \xrightarrow[k_{a}]{k_{s}} \text{IIGA}_2:\text{IIR}$$

$$\text{IIGA}_2:\text{IIR} + \text{pro-}\sigma^E \xrightarrow[k_{ba}]{k_{ab}} \text{IIGA}_2:\text{IIR}:\text{pro-}\sigma^E \xrightarrow[k_{s}]{k_{a}} \text{IIGA}_2:\text{IIR} + \sigma^E$$

$$\sigma^E + \text{Clp} \xrightarrow[k_{a}]{k_{s}} \sigma^E:\text{Clp} \xrightarrow[k_{c}]{k_{r}} \phi$$

IIIGA and pro-\(\sigma^E\) are membrane proteins and in our simulations all the \([\text{IIIGA}]\) and \([\text{pro-}\sigma^E]\) was assumed to be concentrated at the septum (421). Binding rates for all these reactions were assumed to be diffusion limited (\(k_b=1 \text{ nM}^{-1} \text{ hr}^{-1}\)). We assume that the protein complexes are stable with slow rates of dissociation (\(k_{aa}=k_{ab}=k_{ar}=k_{arb}=1 \text{ hr}^{-1}\)). Since the rates of proteolytic cleavage of pro-\(\sigma^E\) have not been measured we used a catalytic rate of \(k_2=1 \text{ min}^{-1}\) (based on the rate of RseA cleavage and \(\sigma^E\) activation in the \textit{E. coli} envelope stress response (422)). Since pro-\(\sigma^E\) is only efficiently cleaved by the SpoIIIGA$_2$-SpoIIR complex (265, 291), we therefore assumed that pro-\(\sigma^E\) cleavage by SpoIIIGA alone is negligible (\(k_1=0\)). In addition, our model includes the proteolytic degradation of $\sigma^E$ by proteolytic enzymes such as Clp which are active during sporulation (420, 423). Since $\sigma^E$ is highly susceptible to proteolytic degradation, we assumed that $\sigma^E$ has high affinity for the proteolytic machinery (\(k_d=1 \text{ hr}^{-1}\)), and the catalytic rate is high (\(k_3=10 \text{ hr}^{-1}\)). Crucially, in a manner similar to zero-order ultrasensitivity, as the activation flux of $\sigma^E$ increases beyond the maximum proteolytic rate, the active $\sigma^E$ concentration increases in a switch-like manner to high levels (Fig. A4.4B). As a result, active $\sigma^E$ concentration depends in a switch-like threshold manner on both determinants of this activation flux: total \textit{spolIG} expression rate and SpoIIR
concentrations (Fig. A4.4CD). Notably, change in total $\sigma^E$ concentration reduces the SpoIIR threshold at which $\sigma^E$ is activated while simultaneously increasing the rate at which active $\sigma^E$ accumulates. As a result, deterministic simulations with the ode45 solver showed that the simultaneous increase in spolIG and spolIR expression resulting from an increase in Spo0A activity leads to a dramatic increase in active $\sigma^E$ concentrations (see Fig. A4.4C).

Stochastic simulations of the full sporulation network used the SpoIIGA, pro-$\sigma^E$ and SpoIIR concentrations predicted by the phosphorelay and $\sigma^F$ activation modules at 3 hours after KinA induction to calculate the active $\sigma^E$ distributions. Figures 5.4A and B show the average and distribution of active $\sigma^E$ concentrations predicted by 400 stochastic simulations at each IPTG dose. To predict cell fate based on these distributions we use a threshold of 1µM of $\sigma^E$ and counted all cells above the threshold as sporulating cells (Fig. 5.5D).

**Quantification of sensitivity with effective Logarithmic Gains**

Throughout our analysis of the sporulation network we refer to effective Logarithmic Gains (LG) as a measure of the input-output sensitivity of a module or network's response. LG is simply the normalized slope of any input-output relationship. Although it is strictly a local measure of input-output sensitivity, we define an effective $\text{LG}_{\text{output}}^{\text{input}}$ that approximates the sensitivity of any output variable to a finite change in an input parameter like IPTG concentration.

$$\text{LG}_{\text{output}}^{\text{input}} = \frac{\partial \log(\text{output})}{\partial \log(\text{input})} = \frac{\text{input } \partial(\text{output})}{\text{output } \partial(\text{input})} \approx \frac{\text{input } \Delta(\text{output})}{\text{output } \Delta(\text{input})} \quad [8]$$

Based on this measure we define ultrasensitivity as $\text{LG}_{\text{output}}^{\text{input}} \sim 10$ so that it reflects the (to within an order of magnitude) the degree of sensitivity of the sporulation cell-fate response to IPTG induction. Averages of stochastic simulations of $\text{PspolIG}$ activity (Spo0A–P), active $\sigma^F$ concentration and active $\sigma^E$ concentration were used output for the phosphorelay, $\sigma^F$ and $\sigma^E$ modules.
respectively. IPTG was used as the input to determine overall effective $LG_{input}^{output}$. For the results of stochastic simulations, input-output data for each module was fitted with Hill-equations and the Hill-coefficient was used as a measure of the $LG_{input}^{output}$.

The AND-type coherent feed-forward loops that control $\sigma^F$ and $\sigma^E$ activation are known to significantly increase $LG_{input}^{output}$ or effective cooperativity. A simple calculation illustrates how the cooperativity is amplified in these feed-forward loops. We define a function $f([AC],[IIE])$ that relates the active $\sigma^F$ concentration in the forespore to the $spoIIA$ (IIA) and $spoIIE$ (IIE) expression levels at septation. Then the log-gains $LG_{c}^{\sigma^F}$ of $\sigma^F$ activation response to IPTG concentration ($c$) can be calculated from the log-gains $LG_{c}^{IIA}$ and $LG_{c}^{IIE}$ using the chain rule:

$$[\sigma^F] = f([IIA],[IIE])$$

$$LG_{c}^{\sigma^F} = \frac{c}{[\sigma^F]} \frac{\partial f([IIA],[IIE])}{\partial c}$$

$$= \frac{c}{[\sigma^F]} \left( \frac{\partial f([IIA],[IIE])}{\partial [IIA]} \frac{\partial [IIA]}{\partial c} + \frac{\partial f([IIA],[IIE])}{\partial [IIE]} \frac{\partial [IIE]}{\partial c} \right)$$

$$= \frac{[IIA] \partial f([IIA],[IIE])}{[\sigma^F]} \frac{\partial [IIA]}{\partial c} + \frac{[IIE] \partial f([IIA],[IIE])}{[\sigma^F]} \frac{\partial [IIE]}{\partial c}$$

$$= LG_{c}^{IIA}LG_{c}^{IIA} + LG_{c}^{IIE}LG_{c}^{IIE}$$

Since IIA and IIE expression levels depend on KinA and Spo0A~P, $LG_{c}^{IIA}$ and $LG_{c}^{IIE}$ implicitly include the cooperativity of these upstream regulators. Note that both $LG_{c}^{IIA}$ and $LG_{c}^{IIE}$ > 1 due to the multiple Spo0A~P binding sites in the promoters of both $spoIIA$ and $spoIIE$. This cooperativity when combined with the ultrasensitivity of the post-translational interactions that control $\sigma^F$ activation results in a very high $LG_{c}^{\sigma^F}$ in the 3-6 µM IPTG range. Defining a function that relates active concentration to $spoIIG$ (IIG) expression and $spoIIIR$ expression or
equivalently to the $\sigma^F$ concentration in the forespore, $[\sigma^F] = g([\text{IIG}],[\sigma^F])$, a similar calculation shows that the cooperativity is further amplified in the $\sigma^E$ activation module:

$$[\sigma^E] = g([\text{IIG}],[\sigma^F])$$

$$LG_{\text{c}}^{\sigma^E} = \frac{c}{[\sigma^E]} \frac{\partial g([\text{IIG}],[\sigma^F])}{\partial c} = LG_{\text{c}}^{\sigma^E} LG_{\text{c}}^{\text{IIG}} + LG_{\sigma^F}^{\sigma^E} LG_{\text{c}}^{\sigma^F}$$

Here again $LG_{\text{c}}^{\text{IIG}}$ and $LG_{\text{c}}^{\sigma^F}$ implicitly include the cooperativity of upstream regulators, since IIG and $\sigma^F$ levels depend on KinA and Spo0A~P. In fact, $LG_{\text{c}}^{\text{IIG}} > 1$ due to the multiple Spo0A~P binding sites in the promoters of both spoIIG and $LG_{\text{c}}^{\sigma^F} > 1$ as shown in the analysis of the $\sigma^F$ activation module. The combined effect of these cascaded feed-forward loops results in a very high $LG_{\text{c}}^{\sigma^E} > 10$ in the 4-10 µM IPTG range. These feed-forward loops are thus responsible for the
Appendix 5. The importance of proper Spo0A dynamics in sporulation

A5.1. Supplementary Figures

Figure A5.1. Protein levels of KinC and Spo0A in the IPTG- and xylose-inducible strain.

A. For the determination of KinC level, cells of the KinC (MF2734) and wild-type (MF2659) strains, both harboring kinC-gfp, were cultured and then processed for immunoblots as described in Materials and Methods. In order to make measurements in a non-saturation range, protein samples were diluted before processing for immunoblots (0.125μg total, KinC* panel).

B. For the determination of Spo0A level, cells of the KinC-Spo0A strain (MF4317) were cultured as above, and xylose was added at indicated concentrations (0, 0.04, and 1%). Samples of MF2659 were used for the detection of wild-type Spo0A protein. Total protein samples (2μg total, top panel) were diluted and processed for immunoblot (0.125μg total, 0A* panel). Protein levels were normalized to the constitutively expressed σ^A and then to the level in wildtype cells in SM conditions. Relative protein levels are shown in the bottom graphs. All experiments were performed at least three times.
independently. Gray bars and error bars represent the means and standard deviations under each condition.

![Fluorescence microscopy images](image)

**Figure A5.2. Protein expression from IPTG- and xylose-inducible promoters in a single cell.**

**A.** Cells expressing both CFP (cyan) and YFP (yellow) under the control of the IPTG (P<sub>hy-spank</sub>) and xylose (P<sub>xylA</sub>) inducible promoters, respectively, were collected after 2h of addition of IPTG (10µM) and xylose (0.04%). Fluorescence microscopy and quantification of color intensity were carried out as described in Experimental Procedures. Representative images are shown in a series of four panels: phase contrast (PC), CFP, merge (CFP and YFP), and YFP. Scale bar, 2µm.

**B-C.** Fluorescence intensities of CFP (panel B) and YFP (panel C) in individual cells in the presence of 0.04% Xylose and 10µM IPTG, respectively, were measured as indicated in Experimental Procedures and are shown as histograms. Gamma distributions fits (black curves) were determined by fitting the fluorescence data. The mean CFP and YFP fluorescence values are also shown along with the noise in gene expression (noise η is the ratio of standard deviation to mean).
D. The CFP and YFP expression levels from $P_{\text{hy-spank}}$ and $P_{\text{xylA}}$ in single cells are not correlated (Spearman correlation coefficient $\rho = -0.12$).

**Figure A5.3. Increase in xylose concentration speeds up Spo0A~P accumulation.**

A-C. Model simulations for the inducible KinC-Spo0A two-component system show that increasing xylose concentration results in increased Spo0A and consequently speeds up (A) the accumulation of Spo0A~P ([0A~P]), (B) SpoIIGB ([IIGB]) activation and (C) DivIVA ([DivIVA]) repression.

**Figure A5.4. Early onset of divIVA repression at high IPTG and xylose concentrations.**

Level of DivIVA was assayed using cells expressing an endogenous DivIVA-CFP fusion protein in the KinC-Spo0A strain (MF4812) at 0µM IPTG, 0% xylose, 10µM IPTG, 0.04% xylose and 500µM IPTG, and 1% xylose at indicated time-points (0, 1, 3, and 5 hours.
after induction). The upper panel shows representative immunoblots for each condition. The lower panel shows the quantification of these immunoblots. Gray bars show the mean values of three independent experiments with standard deviations. Values are normalized to the constitutively expressed σA and then those at time zero (T0) in each case.

![Immunoblots and quantification](image)

**Phase Contrast**

**DNA**

**DNA DivIVA**

**DNA Membrane**

**DivIVA Membrane**

**DivIVA Membrane**

**Membrane**
Figure A5.5. Fluorescence microscopy of KinC-Spo0A strain.

Cells of MF4812 expressing an endogenous DivIVA fused to CFP (cyan) in the KinC-Spo0A strain were harvested at two hours after induction with inducers.

**A-G.** 10µM IPTG and 0.04% xylose.

**H-N.** 500µM IPTG and 1% xylose. Membrane and DNA were visualized with FM4-64 (magenta) and DAPI (yellow), respectively. White arrows (IJK) show an empty forespore and red arrows (IJK) show an empty cell that have formed as a result of improper chromosome segregation. Scale bar: 5µm.

![Graph](image.png)

**Figure A5.6.** Overexpression of KinC and Spo0A leads to an increase in chromosome segregation defects.

Cells of MF4812 were harvested at two hours after induction with inducers and stained with membrane (FM4-64) and DNA (DAPI) dyes to quantify chromosome segregation defects. The fraction of cells with chromosome free compartments (empty cells and empty/incomplete forespores) was higher in high inducer conditions (500µM IPTG, 1% xylose - red bars) relative to the optimal inducer condition (10µM IPTG, 0.04% xylose - black bars). Total represents the sum of all chromosome segregation defects. Number of cells counted: 316 (10µM IPTG, 0.04% xylose), 128 (500µM IPTG, 1% xylose).
Figure A5.7. *spo0A* auto-regulation ensures proper target expression even at high *KinC* levels.

**A-C.** Model predictions for the concentrations of *Spo0A~P* (A), *SpoIIGB* (B) and *DivIVA* (C) at different IPTG concentrations in the *KinC* induction strain with *spo0A* under its native promoters and *spo0B* null mutation (MF4419). Black and red curves show the concentration of each species at $T(\sigma^H=0.8)$ and steady state, respectively. Solid and dashed curves represent the mean and standard deviations at each IPTG concentration calculated from 5000 stochastic simulation trajectories. Blue lines represent the thresholds used to predict cell-fate. Thresholds: $0.46 \mu M < [\text{Spo0A}~P] < 0.55 \mu M$ at $T(\sigma^H=0.8)$ and $0.91 \mu M < [\text{Spo0A}~P] < 1.1 \mu M$ at steady state; $2 \mu M < [\text{SpoIIGB}]$ at $T(\sigma^H=0.8)$ and $13.5 \mu M < [\text{SpoIIGB}] \mu M$ at steady state; $0.97 \mu M < [\text{DivIVA}]$ at $T(\sigma^H=0.8)$ and $0.84 \mu M < [\text{DivIVA}]$ at steady state. Note that most cells fall within the thresholds for $8 \mu M < \text{IPTG} < 1000 \mu M$.

**D.** Accumulation dynamics of directly- and indirectly-activated *Spo0A* target genes (*SpoIIGB*-solid curves, $\sigma^H$-dashed curves) in the *KinC* induction strain under different IPTG concentrations as predicted by our model. Higher IPTG concentrations shift the onset of both $[\text{SpoIIGB}]$ and $[\sigma^H]$ accumulation to earlier time-points. Note that $[\sigma^H]$ accumulation always precedes $[\text{SpoIIGB}]$ accumulation in this strain.
E. Cartoon of the native $spo0A$ regulatory region that involves multiple positive and negative feedback loops. $spo0A$ expression from promoter $P_v$ ($\sigma^V$ dependent) is repressed by the binding of Spo0A~P at $O_1$. Further $spo0A$ expression from promoter $P_s$ ($\sigma^H$ dependent) is repressed and activated by binding of Spo0A~P at $O_2$ and $O_3$ respectively. F. Model predictions for the effect of mutations of the $spo0A$ regulatory region on sporulation efficiency in the KinC induction strains. Removal of the negative regulatory regions ($\Delta O_{12}$) leads to overexpression of Spo0A which results in a non-monotonic dependence of sporulation efficiency on IPTG. Removal of all Spo0A~P binding sites ($\Delta O_{123}$) reduces the expression of Spo0A relative to $\Delta O_{12}$ and thereby ensures gradual accumulation of Spo0A~P and a monotonic dependence of sporulation efficiency on KinC induction.

A5.2. Supplementary Tables

Table A5.1. Sporulation in the KinC strain (MF4531) in the presence of different IPTG concentrations.

Cells of MF4531 were collected at 3h after IPTG addition and assayed for $\beta$-Galactosidase activity from $P_{spoIIG}$-lacZ as indicated in Materials and Methods. $\beta$-gal activities shown correspond to the average of three independent experiments.

<table>
<thead>
<tr>
<th>IPTG (μM)</th>
<th>CFU/ml</th>
<th>Efficiency</th>
<th>$\beta$-gal (Miller units)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Viable</td>
<td>Spores</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>$5.5 \times 10^8$</td>
<td>$1.7 \times 10^6$</td>
<td>$3.1 \times 10^{-3}$</td>
</tr>
<tr>
<td>2</td>
<td>$7.7 \times 10^8$</td>
<td>$9.8 \times 10^5$</td>
<td>$1.3 \times 10^{-3}$</td>
</tr>
<tr>
<td>4</td>
<td>$7.5 \times 10^8$</td>
<td>$5.1 \times 10^5$</td>
<td>$6.8 \times 10^{-4}$</td>
</tr>
<tr>
<td>6</td>
<td>$5.2 \times 10^8$</td>
<td>$3.3 \times 10^5$</td>
<td>$6.3 \times 10^{-4}$</td>
</tr>
<tr>
<td>8</td>
<td>$6.1 \times 10^8$</td>
<td>$5.4 \times 10^7$</td>
<td>$8.9 \times 10^{-2}$</td>
</tr>
<tr>
<td>10</td>
<td>$5.7 \times 10^8$</td>
<td>$3.2 \times 10^8$</td>
<td>$5.6 \times 10^{-1}$</td>
</tr>
<tr>
<td>wt</td>
<td>$5.3 \times 10^8$</td>
<td>$3.4 \times 10^8$</td>
<td>$6.4 \times 10^{-1}$</td>
</tr>
</tbody>
</table>

Table A5.2. Sporulation Efficiency and Spo0A activity in the KinC-Spo0A double induction (MF4317) strain.

Cells of MF4317 were collected at 3h after IPTG and Xylose addition and assayed for $\beta$-Galactosidase activity from $P_{spollIG}$-lacZ as indicated in Materials and Methods. $\beta$-gal
activities shown correspond to the average of three independent experiments. Sporulation efficiency was determined as indicated in Materials and Methods.

<table>
<thead>
<tr>
<th>IPTG (µM)</th>
<th>Xylose (%)</th>
<th>CFU/ml</th>
<th>Efficiency</th>
<th>β-gal (Miller units)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Viable cells</td>
<td>Spores</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.04</td>
<td>$4.4 \times 10^8$</td>
<td>1.3 x</td>
<td>$3.0 \times 10^{-1}$</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>$4.4 \times 10^8$</td>
<td>2.2 x</td>
<td>$5.0 \times 10^{-2}$</td>
</tr>
<tr>
<td>500</td>
<td>0.04</td>
<td>$5.7 \times 10^8$</td>
<td>5.2 x</td>
<td>$9.1 \times 10^{-2}$</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>$6.3 \times 10^8$</td>
<td>6.2 x</td>
<td>$9.8 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

Table A5.3. KinC can activate Spo0A and trigger sporulation in Δspo0F background.

After IPTG addition, cells of MF2406 were assayed for Sporulation efficiency as indicated in Materials and Methods.

<table>
<thead>
<tr>
<th>KinC Δspo0F (MF2406)</th>
<th>CFU/ml</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPTG (µM)</td>
<td>Viable cells</td>
<td>Spores</td>
</tr>
<tr>
<td>0</td>
<td>$6.1 \times 10^8$</td>
<td>$4.7 \times 10^4$</td>
</tr>
<tr>
<td>10</td>
<td>$5.1 \times 10^8$</td>
<td>$1.7 \times 10^8$</td>
</tr>
<tr>
<td>500</td>
<td>$1.1 \times 10^8$</td>
<td>$3.2 \times 10^7$</td>
</tr>
</tbody>
</table>
### Table A5.4. Strains used in this study

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype or description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MF1913</td>
<td>$kinA\Omega P_{hy-spank}kinA\ cm$</td>
<td>(131)</td>
</tr>
<tr>
<td>MF1915</td>
<td>$kinC\Omega P_{hy-spank}kinC\ cm$</td>
<td>(131)</td>
</tr>
<tr>
<td>MF2406</td>
<td>$kinC\Omega P_{hy-spank}kinC\ cm, \Delta spo0F::tet$</td>
<td>This study</td>
</tr>
<tr>
<td>MF4419</td>
<td>$kinC\Omega P_{hy-spank}kinC\ cm, \Delta spo0B::tet$</td>
<td>This study</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype or description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MF750</td>
<td>$thrC::P_{spo0G-lacZ}erm$</td>
<td>(296)</td>
</tr>
<tr>
<td>MF4317</td>
<td>$kinC\Omega P_{hy-spank}kinC\ cm, amyE::P_{xylA-spo0A spc, \Delta spo0F::tet, \Delta spo0A::kan, thrC::P_{spo0G-lacZ}erm}$</td>
<td>This study</td>
</tr>
<tr>
<td>MF4531</td>
<td>$amyE::P_{hy-spank}kinC\ spc, \Delta kinA::tet, \Delta kinB::kan, \Delta kinC::cm, thrC::P_{spo0G-lacZ}erm$</td>
<td>This study</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype or description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MF929</td>
<td>$kinA\Omega kinA-gfp kan$</td>
<td></td>
</tr>
<tr>
<td>MF2659</td>
<td>$kinC\Omega kinC-gfp spc$</td>
<td></td>
</tr>
<tr>
<td>MF1996</td>
<td>$kinA\Omega P_{hy-spank}kinA-gfp spc kan$</td>
<td></td>
</tr>
<tr>
<td>MF2734</td>
<td>$kinC\Omega P_{hy-spank}kinC-gfp spc kan$</td>
<td></td>
</tr>
<tr>
<td>MF4318</td>
<td>$amyE::P_{xylA-cfp spc, thrC::P_{hy-spank-yfp erm}}amyE::P_{xylA-spo0A spc, \Delta spo0F::tet, \Delta spo0A::kan, thrC::P_{spo0G-gfp erm}}$</td>
<td>This study</td>
</tr>
<tr>
<td>MF4581</td>
<td>$amyE::P_{xylA-cfp spc,\Delta spo0A::kan, \Delta spo0F::tet}$</td>
<td>This study</td>
</tr>
<tr>
<td>MF4812</td>
<td>$amyE::P_{xylA-cfp spc, \Delta spo0A::kan, \Delta spo0F::tet}$</td>
<td>This study</td>
</tr>
</tbody>
</table>

### Table A5.5. Plasmids used in this study

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMF311</td>
<td>$amyE::P_{xylA-spo0A}$</td>
<td>This study</td>
</tr>
<tr>
<td>pMF604</td>
<td>$amyE::P_{xylA-cfp spc}$</td>
<td>This study</td>
</tr>
<tr>
<td>pMF607</td>
<td>$thrC::P_{hy-spank-yfp erm}$</td>
<td>This study</td>
</tr>
</tbody>
</table>
### Table A5.6. Oligonucleotide primers used in this study

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>om24</td>
<td>5’-ccgctcgaggctgtctgtatattaaaggca-3’</td>
<td>(XhoI) kinC-gfp</td>
</tr>
<tr>
<td>om29</td>
<td>5’-gcggaattcacatcttgacggtttccctcag-3’</td>
<td>(EcoRI) kinC-gfp</td>
</tr>
<tr>
<td>omf288</td>
<td>5’-ggcgtcagacataaggagaaactactatggaataagtttgttgctgc-3’</td>
<td>(SalI) P$xylA$-sp00A</td>
</tr>
<tr>
<td>omf318</td>
<td>5’-gccgcatgccataagcccttagctctaaccctcagc-3’</td>
<td>(SphI) P$xylA$-sp00A</td>
</tr>
<tr>
<td>om138</td>
<td>5’-gccgcatgcactggcgaaaagcacttacagtttcaagggcagaacgcacactagtgc-3’</td>
<td>(SalI) P$xylA$-cfp</td>
</tr>
<tr>
<td>om139</td>
<td>5’-gccgcatgccataagcccttagctcataagggcagacgcaagcactgcagc-3’</td>
<td>(SphI) P$xylA$-cfp</td>
</tr>
<tr>
<td>om137</td>
<td>5’-gccgcatgcactggcgaaaagcacttacagtttcaagggcagacgcaagcactgcagc-3’</td>
<td>(SalI) P$hy$-spank-yfp</td>
</tr>
<tr>
<td>omf317</td>
<td>5’-gccgcatgccataagcccttagctcataagggcagacgcaagcactgcagc-3’</td>
<td>(SphI) P$hy$-spank-yfp</td>
</tr>
</tbody>
</table>

### Table A5.7. Parameter values used for the model of transcriptional regulation of the native spo0A promoter.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_v$</td>
<td>Maximum transcription rate for $P_v$</td>
<td>10 nM hr$^{-1}$</td>
<td>(263, 288)</td>
</tr>
<tr>
<td>$K_1$</td>
<td>Binding affinity for $P_v$ repression site</td>
<td>100 nM</td>
<td>(130)</td>
</tr>
<tr>
<td>$k_s$</td>
<td>Transcription rate for $P_s$</td>
<td>72 nM hr$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_H$</td>
<td>$\sigma^H$ binding affinity for $P_s$</td>
<td>250 nM</td>
<td>(263, 288, 367)</td>
</tr>
<tr>
<td>$K_2$</td>
<td>Binding affinity for $P_s$ repression site ($O_2$)</td>
<td>50 nM</td>
<td>(325)</td>
</tr>
<tr>
<td>$K_3$</td>
<td>Binding affinity for $P_s$ activation site ($O_3$)</td>
<td>100 nM</td>
<td>(130)</td>
</tr>
<tr>
<td>$f_a$</td>
<td>Fold-change of activation ($O_2$)</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
### Table A5.8. Parameter values used for the models of transcriptional regulation of Spo0A target genes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Transcriptional Regulator, x (+): activator (-): repressor</th>
<th>Hill-exponent/No. of binding sites (n)</th>
<th>$K_d$ (nM)</th>
<th>Fold-Induction (f)</th>
<th>Maximum transcription rate ($v_f$ or $v_m$) (nM/hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$abrB$</td>
<td>Spo0A~P (-)</td>
<td>4</td>
<td>200</td>
<td>N/A</td>
<td>75</td>
</tr>
<tr>
<td>$spo0H$</td>
<td>AbrB (-)</td>
<td>4</td>
<td>100</td>
<td>N/A</td>
<td>75</td>
</tr>
<tr>
<td>$spoIIG$</td>
<td>Spo0A~P (+)</td>
<td>4</td>
<td>800</td>
<td>15</td>
<td>500</td>
</tr>
<tr>
<td>$divIVA$</td>
<td>Spo0A~P (-)</td>
<td>4</td>
<td>1600</td>
<td>N/A</td>
<td>75</td>
</tr>
</tbody>
</table>

### Table A5.9. Parameter values used for post-translational interactions in the model of Spo0A activation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_p$</td>
<td>KinC auto-phosphorylation</td>
<td>10 hr$^{-1}$</td>
<td>Based on KinA auto-phosphorylation, dephosphorylation kinetics (367)</td>
</tr>
<tr>
<td>$k_{dp}$</td>
<td>KinC auto-dephosphorylation</td>
<td>50 hr$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$k_b$</td>
<td>Diffusion-limited binding rate constant</td>
<td>$0.1\text{nM}^{-1}\text{hr}^{-1}$</td>
<td>Based on SpoIIE-SpoIIAA binding rate (269)</td>
</tr>
<tr>
<td>$k_{ub}$</td>
<td>KinC~P:Spo0A unbinding</td>
<td>1000 hr$^{-1}$</td>
<td>Based on a typical dissociation constants (269, 290)</td>
</tr>
<tr>
<td>$k_i$</td>
<td>Phospho-transfer catalytic rate</td>
<td>1000 hr$^{-1}$</td>
<td>Based on the Spo0B$^{\gamma}$-Spo0A phosphotransfer rate (290)</td>
</tr>
<tr>
<td>$k_e$</td>
<td>Spo0A~P dephosphorylation</td>
<td>5 hr$^{-1}$</td>
<td>(424)</td>
</tr>
</tbody>
</table>
A5.3.1. KinC governs entry into sporulation in a concentration-dependent manner

We constructed a strain expressing KinC under an IPTG-inducible hyper-spank promoter (hereafter referred to as $P_{hy-spank}$). To determine the effective level of KinC for efficient entry into sporulation, the KinC-inducible strain (MF1915, referred as KinC strain) was cultured in LB in the presence of varying concentrations of IPTG. Sporulation efficiency (Table A5.1) was determined by measuring the fraction of heat-resistant colony forming units (CFU). We note that in these nutrient rich media, the wild-type strain fails to sporulate (129). We found that, at IPTG concentrations of 10μM, sporulation was efficiently induced to the level comparable to that in the wild-type strain (i.e. without IPTG-inducible construct) under normal sporulation conditions (Table A5.1).

Second, to determine the effect of KinC induction on Spo0A activity we constructed a reporter system consisting of lacZ gene expression driven by the Spo0A-dependent spoIIG promoter (130, 296). We then introduced this reporter system into the KinC strain (MF4531) to examine β-galactosidase activity under the same conditions as above. We confirmed that Spo0A was efficiently activated similarly to that seen in the sporulating wild-type strain (Table A5.1). We note that, without IPTG, none of the kinases can activate Spo0A to a level that is sufficient to trigger sporulation in LB medium (129, 306, 425). These results show that artificially induced expression of KinC can be used to control Spo0A activation and entry into sporulation.

A5.3.2. Increase in KinC triggers sporulation in a phosphorelay-independent manner

We determined whether inducing the synthesis of KinC would bypass the requirement of the phosphorelay intermediate phosphotransferases for sporulation as suggested previously (306, 426). To demonstrate this, a null mutation of spo0F (MF2406) or spo0B (MF4419) was introduced into the KinC strain, and the resulting strains were tested for sporulation efficiency in LB in the
presence of varying concentrations of IPTG. We found that in the presence of 10μM or higher concentrations of IPTG, sporulation efficiency (10^{-1}) of the resulting strain was on the same order as that of the wild-type strain under normal sporulation conditions (compare Tables A5.1 and A5.3). In contrast, inducing the synthesis of KinA triggered massive induction of sporulation in wild-type background (MF1913), but no sporulation initiation was observed in the spo0F or spo0B null genetic background as shown previously (131, 367). Taken together the results from our phosphorelay bypass systems (Δspo0F (MF2406) and Δspo0B (MF4419)) show that direct phosphotransfer from KinC to Spo0A can produce sufficient Spo0A~P for efficient sporulation.

A5.3.3. Mathematical Modeling of the KinC-Spo0A two-component system

Our models focus on the KinC-Spo0A two-component system. This two-component system is relevant to strains MF4531 and MF4419 (see Table A5.4). In each of these strains the interaction of Spo0A with the phosphorelay has been eliminated by deletion of spo0B, and phosphorylation of Spo0A depends solely on KinC.

**Modeling KinC expression from the IPTG-inducible \(P_{hy-spank}\) promoter**

We chose the following model for KinC expression from the IPTG ([I]) inducible \(P_{hy-spank}\) promoter:

\[
\begin{align*}
\text{Promoter}_{\text{off}} & \xrightarrow{k_1} \text{Promoter}_{\text{on}} \\
\text{Promoter}_{\text{on}} & \xrightarrow{k_2} \text{Promoter}_{\text{on}} + \text{mRNA} \\
\text{mRNA} & \xrightarrow{\gamma_1} \text{mRNA} + \text{KinC} \\
\text{mRNA} & \xrightarrow{\gamma_2} \emptyset \\
\text{KinC} & \xrightarrow{\gamma_3} \emptyset
\end{align*}
\]

Based on this model the steady state concentration of KinA production is given by:
\[ [\text{KinC}] = \frac{k_{on} k_2}{k_{on} + k_{off}} \frac{k_1}{\gamma_1} \]  

\[ k_{off} ([I]) = k_{off} \frac{[R_T]}{1 + ([I]/K)^2} \]  

The Promoter ON to OFF transition depends on the binding of the Lac-repressor dimer which was assumed to have some fixed concentration \([R_T]\). We assumed that IPTG binds to Lac-repressor dimers and reduces the effective free repressor concentration. Following this assumption, the dependence of Promoter OFF-rate \(k_{off}\) on IPTG concentration \([I]\) is given by (427):

\[ k_{off} ([I]) = k_{off} \frac{[R_T]}{1 + ([I]/K)^2} \]  

Here \(K\) is the dissociation constant for Lac-dimer IPTG binding. Using equations [1] and [2], we can rewrite the steady state \([\text{KinC}]\) as a function of the IPTG concentration \([I]\):

\[ [\text{KinC}] = \frac{k_{on} k_2}{k_{on} + k_{off}[R_T]} \frac{k_1}{\gamma_1} \frac{[I]^2(k_0[R_T]/k_{on})}{K_2(1+k_0[R_T]/k_{on})+[I]^2} \]  

\[ v_b = \frac{k_{on}}{k_{on} + k_0[R_T]} \]  

\[ f = \frac{k_0[R_T]}{k_{on}} \]  

\[ K_m^2 = K_2(1+k_0[R_T]/k_{on}) \]

\(v_b, f,\) and \(K_m\) represent the basal expression rate, fold-change, and half-maximal constant for the IPTG-dependent promoter. We assumed \(\text{KinC}\) is a stable protein such that its combined degradation-dilution rate is \(\gamma_2=1\text{ hr}^{-1}\). Fitting the KinC expression data from the immunoblotting experiments with the IPTG-dose dependent equation [3] (Fig. 6.4A; fitting was done with the MATLAB curve fitting toolbox function \textit{cftool}) we estimate that \(f=19\) and \(K_m=9\mu M\). Using these values in equations [5] and [6] we found \(k_0[R_T]/k_{on}=19\) and \(K_2=4.05\mu M^2\). Based on the results of (279) we estimated \(v_b=0.1\mu M\text{hr}^{-1}\). We also used a P\_hy-spank-YFP fusion to quantify the stochastic properties of gene expression from the P\_hy-spank
promoter at 10 and 500µM IPTG by assuming that fluorescence is proportional to protein numbers in the cell (see Supplementary Fig. A5.2C). Following the methods described in (57), these distributions were fit using gamma distributions (using the gamfit function of MATLAB - see solid curves in Fig. A5.2C) to determine the mean number of bursts-a and the mean burst size-b (µM). Fig. A5.2B shows gamma distribution fits with a=19.9, b = 8 for 10µM IPTG. We use the following relationship to relate the mean size of bursts b with the inducer concentration [I]:

\[ b([I]) = \frac{k_2}{\gamma_1} \frac{k_{on}}{k_{on} + k_{off} ([I])} \]  

From [1] and [7] we note that \( a=k_1/\gamma_2 \), and since a for YFP (division rate \( \gamma_2 \sim 2 \text{ hr}^{-1} \)) is 19.9, for KinC, a stable membrane protein with a dilution rate \( \gamma_2 \sim 0.5 \text{ hr}^{-1} \), we estimate that \( a=80 \). This implies that \( k_1=40 \text{ hr}^{-1} \). Now with the values for \( v_b \), f, and \( K_m \) that we have calculated above, we estimate \( k_1 k_2/\gamma_1 = 2 \). For an mRNA degradation rate \( \gamma_1=4\text{hr}^{-1} (~10 \text{ min half-life}) \), this implies \( k_2=0.2\text{hr}^{-1} \). Finally, previously reported biochemical measurements for the LacI repressor were used to estimate \( k_{off}[R_T]=345.4\text{hr}^{-1} \) and \( k_{on}=18\text{hr}^{-1} \) (409). These kinetic constants completely characterize KinC expression from the P_hy-spank promoter for our models.

**Modeling Spo0A expression from the Xylose-inducible P_xyl promoter**

We used a simple model to describe the expression of spo0A from the xylose-inducible promoter.

\[
\begin{align*}
\text{Promoter}_{on} & \xrightarrow{k_1} \text{Promoter}_{on} + \text{mRNA} \\
\text{mRNA} & \xrightarrow{k_2} \text{mRNA} + \text{Spo0A} \\
\text{mRNA} & \xrightarrow{n} \emptyset \\
\text{Spo0A} & \xrightarrow{n} \emptyset 
\end{align*}
\]

The mean concentration of Spo0A based on this model is given by the following equation:
\[ \text{[Spo0A]} = ab = \frac{k_1 k_2}{\gamma_1 \gamma_2} \quad [8] \]

Where \( a(=k_1/\gamma_2) \) and \( b(=k_2/\gamma_1) \) are the mean number of bursts and the mean burst size of gene expression for this promoter. These parameters also characterize the gamma distribution of gene expression from the \( P_{xyl} \) promoter. We determined these parameters for expression of CFP from \( P_{xyl} \) by fitting a gamma distribution to the single-cell fluorescence histogram. For \( P_{xyl}\)-CFP, \( a=10.46, b=14.17 \) at 0.04\% Xylose. Similar to YFP, CFP division rate is \( \gamma_2 \sim 2 \text{hr}^{-1} \) (half-life \( \sim 30 \text{ min} \)) whereas Spo0A is assumed to have a dilution rate \( \gamma_2 \sim 1 \text{hr}^{-1} \). Accordingly, we estimate that \( a=20 \) for Spo0A at 0.04\% Xylose.

We assume that the transcription rate \( k_1 \) from the promoter depends on the Xylose concentration \([xyl]\) according to the following equation:

\[ k_1([xyl]) = v_0 \left( 1 + f_x \frac{[xyl]^n}{K_x^n + [xyl]^n} \right) \quad [9] \]

The other rate constants are independent of changes in Xylose concentration. We combined equations [8] and [9] to fit the Spo0A immunoblotting results (Fig. 6.4B). The results of fitting show that \( v_0 k_2/(\gamma_1 \gamma_2)=0.5\mu\text{M}, f_x=4.5, n=2, \) and \( K_x=0.048\% \). Assuming an mRNA degradation rate of \( \gamma_1=4 \text{hr}^{-1} \) (based on a half-life of \( \sim 10 \text{ min} \)) and a protein degradation rate of \( \gamma_2=1 \text{hr}^{-1} \), this implies that \( v_0 k_2=2\mu\text{Mhr}^{-2} \). Further using \( a=k_1([xyl]=0.04\%)/\gamma_2=20 \) as calculated above in equation [9] shows that \( v_0=14\mu\text{Mhr}^{-1} \) and therefore \( k_2=0.14\text{hr}^{-1} \).

**Modeling Spo0A expression from the native spo0A promoter**

Spo0A transcription is controlled by two promoters, namely \( P_v \) and \( P_s \). Chastanet and Losick (288) have shown that the weak vegetative promoter \( P_v \) is active during the exponential phase but repressed by Spo0A~P. On the other hand, the sporulation specific promoter \( P_s \) is both activated and repressed by binding of Spo0A~P to a region 70-90 bp upstream of \( P_s \). \( P_s \) has also been
shown to be transcribed in a $\sigma^H$-dependent manner. Thus the regulation of spo0A expression from its native promoter involves multiple positive and negative auto-regulatory feedback loops. Based on these results we used the following Hill-type equation to model spo0A expression from $P_V$ and $P_S$:

$$v_{spo0A} = k_v \frac{1}{1+[0A-P] / K_1} + k_s \frac{[\sigma^H]}{K_{\sigma^H} + [\sigma^H]} \frac{1+f_s [0A-P]^2 / K_s^2}{1+[0A-P]^2 / K_s^2}$$

Here the first and second terms represent that rate of transcription from $P_V$ and $P_S$ respectively. [0A~P] and $[\sigma^H]$ represent the concentrations of Spo0A~P and $\sigma^H$, respectively. Supplementary Table A5.7 describes the parameters and parameter values used for this model of spo0A transcriptional regulation.

We assumed an mRNA degradation rate of 8 hr$^{-1}$ based on a half-life of $\sim$5 min. We also assumed a typical translational burst size of 30 protein molecules per mRNA. Since Spo0A is a stable protein, we assumed that the combined rate constant of dilution and degradation is $\sim$1 hr$^{-1}$ based on a cell doubling time of 1 hour. For simulations of regulatory mutants (see Fig. A5.7) the appropriate [0A~P] dependent terms were excluded from $v_{spo0A}$.

Transcriptional regulation of target genes

Our models included the following Spo0A~P target genes: abrB, spo0H, spollG and divIVA. For these genes we assumed that binding-unbinding of transcriptional regulators from their promoters is fast. We model the rate of transcription, $v$, with appropriate Hill-functions:

Activation: $v = v_b \left( 1 + f \frac{x^n}{K_d^n + x^n} \right)$

Repression: $v = v_m \frac{K_d^n}{K_d^n + x^n}$
Where \( x \) stands for the concentration of the appropriate transcription factor. Table A5.8 describes the details of the transcription regulation for these genes (parameter values were derived from -(130, 289)).

In view of the lack of specific parameter values for these genes, we assumed an mRNA degradation rate of 8 hr\(^{-1}\) based on a half-life of \(~5\) min. We also assumed a typical translational burst size of 40 protein molecules per mRNA. Since proteins are typically stable, we assumed that the combined rate constant of dilution and degradation is \(~1\) hr\(^{-1}\) based on a cell doubling time of 1 hour.

**Post-translational interactions in the KinC-Spo0A two-component system**

Post-translational interactions in the phosphorelay include the following reactions:

\[
\text{KinC} \xrightarrow{k_p} \text{KinC} \sim P \\
\text{KinC} \sim P + \text{Spo0A} \xrightarrow{k_h} \text{KinC} \sim P : \text{Spo0A} \xrightarrow{k_i} \text{KinC} + \text{Spo0A} \sim P \\
\text{Spo0A} \sim P \xrightarrow{k_o} \text{Spo0A}
\]

Here P marks phosphorylated forms of the proteins, colon denotes protein complexes, and each rate constant is introduced above its corresponding reaction arrow. All post-translational reactions were modeled with mass-action kinetics. Table A5.9 lists the parameter values used to model these post-translational reactions.

**A5.3.4. Simulations**

We used the ode45 simulator of MATLAB (©MathWorks) to study the deterministic response of the KinC-Spo0A two-component system to different inducer concentrations. For the \( P_{\text{hy-sparl}} \)-KinC, \( P_{\text{xyl}} \)-Spo0A double inducible strain, KinC transcription rate was set with IPTG concentrations (10 or 500\(\mu\)M), and transcription rates for Spo0A were set based on our estimates for the \( P_{\text{xyl}} \)
promoter at 0.04 and 1% xylose. For the model where spo0A expression is controlled by its native promoter, we used IPTG concentrations in the range 0-1000µM.

Stochastic simulations for the two models were performed using the Gillespie algorithm of the COPASI package. We assumed that promoter-transcription factor binding/unbinding steps are fast, such that transcriptional regulation can be modeled using Hill-equations even for stochastic simulations. To allow comparison with experimental results, for the P_{hy-spank}^~ KinC, P_{xyl}^~ Spo0A double inducible strain we focused on the inducer concentrations shown in Table A5.2. For each of the four conditions, stochastic simulations were used to collect 5000 trajectories of the two-component system response. For the model with native spo0A expression, we used IPTG concentrations in the range 0-1000µM and computed 5000 trajectories for each inducer concentration. The concentrations of σ^H, SpoIIGB and DivIVA from each trajectory were used to compute the fraction of sporulating cells.

**A5.3.5. Calculation of Fraction of Sporulating cells**

Trajectories generated from the stochastic simulations of both models were used with the three threshold mechanisms described in the main text to calculate the fraction of sporulating cells at different inducer concentrations. Specifically, a trajectory was counted as sporulating if it satisfied all the thresholds for the particular mechanism:

- **Single Threshold**: [SpoIIGB]>c_1 at T5 (steady state)
- Spo0A~P Range: [SpoIIGB]>c_1 and [DivIVA]>c_2 at T5
- **Dynamic Thresholds**: [SpoIIGB]>c_1 and [DivIVA]>c_2 at T5 and [SpoIIGB]>d_1 and [DivIVA]>d_2 when [σ^H]=s_0.

s_0 was fixed to 0.8µM which corresponds to ~T2 in our simulations. The values of the thresholds c_1, c_2, d_1 and d_2 were chosen to by optimizing the fit of
the Dynamic Thresholds mechanism to the experimentally measured sporulation efficiencies using the \textit{fmincon} function in MATLAB. The threshold values used were $c_1=13.5$, $c_2=0.84$, $d_1=2$, $d_2=0.97$ (all values are in µM).

For the strain where \textit{spo0A} is expressed from its native promoter (see Fig. 6.7B), fraction of sporulating cells was computed using the same thresholds as above.
Appendix 6. Chromosomal locations of phosphorelay genes controls the Bacillus subtilis response to starvation

A6.1. Supplementary Figures

**Figure A6.1. Pulsing from $P_{0F}$ promoters.**

**A-B.** Fluorescence measurements for $P_{0F}$-yfp reporters inserted at $amyE$ (A) and $gltA$ (B) loci in WT background. Solid lines and shaded regions show one example lineage trace and the mean expression level over multiple traces respectively.

**C.** Measurements of maximum $P_{0F}$-yfp fluorescence. Empty circles show maximum YFP fluorescence levels achieved by individual cell lineages over 25 hours in starvation conditions. Filled circles and errorbars indicate the mean and standard deviations of measurements at each integration location.

**D-E.** Measurements for $P_{0F}$ promoter activity for $P_{0F}$-yfp reporters inserted at $amyE$ (D) and $gltA$ (E) loci in WT background. Solid lines and shaded regions show one example lineage trace and the mean expression level over multiple traces respectively. Note that $P_{0F}$ promoter activity pulses similar to $P_{0A}$ promoter activity (see Fig. 7.1C).
F. Measurements of maximum $P_{OF}$ promoter activity. Empty circles show maximum $P_{OF}$ promoter activity levels achieved by individual cell lineages over 25 hours in starvation conditions. Filled circles and errorbars indicate the mean and standard deviations of measurements at each integration location. Both maximum $P_{OF}$-yfp fluorescence and maximum $P_{OF}$ promoter activity are comparable for the amyE and gltA locations.

Figure A6.2. Mathematical modeling of the post-translational interactions shows the ultrasensitive dependence of [0A~P] on the [0F]$_T$/[KinA]$_T$ ratio.

Post-translational interactions of the phosphorelay included the substrate-inhibition of KinA by OF. To isolate the post-translational interactions, the rate of 0A transcription was fixed at 0.4µMhr$^{-1}$ and the rate of OF and kinA transcription was varied to calculate their effect on the steady state levels of phosphorylated 0A ([0A~P]).

A. Contour diagram showing steady state levels 0A~P as a function of [0F]$_T$ and [KinA]$_T$ concentrations. Steady state [0A~P] levels always increase with increasing [KinA]$_T$. However due to the substrate inhibition of KinA by OF, steady state [0A~P] levels depended non-monotonically on [0F]$_T$ concentrations and decreased at high OF levels.

B. Modeling of post-translational interactions show that 0A~P concentration decreases ultrasensitively for [0F]$_T$/[KinA]$_T$>1. Note that the ultrasensitivity of this decrease increases at higher KinA concentrations.
Figure A6.3. Modeling predictions for the effects of 0F translocation and inducible 0F expression.

Changes in kinA and 0F gene dosages and KinA autophosphorylation rate were used as inputs to the mathematical model to study the response of the phosphorelay. Yellow shaded regions and dashed vertical lines in panels E-X represent DNA replication periods and cell divisions respectively.

A-D. Chromosomal arrangements of 0F and kinA in (A) Wildtype (WT) B. subtilis and synthetic mutant strains: (B) Trans-0Fθglta, (C) iOФamyE, (D) iOФθglta. 0F is located close to the origin of replication in WT and iOФamyE strains and close to the terminus in the Trans-0Fθglta and iOФθglta strains. kinA is located close to the terminus in all strains. Note that 0F is expressed from the IPTG-inducible Pnap promoter, rather than the native 0A-P regulated P0F promoter in the inducible iOФamyE and iOФθglta strains.

E-H. KinA autophosphorylation rate was increased over multiple cell-cycles to mimic the effect of increasing starvation.

I-L. Changes in 0F gene dosage ratio in the WT and mutant strains. In WT (I) and iOФamyE (K), 0F is replicated soon after cell division due to its position close to the chromosome origin in these strains. The translocation of 0F close to the terminus in...
Trans-OF\textsuperscript{gltA} (J) and i0F\textsuperscript{gltA} (L) strains delays its replication so that it is replicated at the end of the DNA replication period in each cell-cycle.

**M-P.** Changes in \textit{kinA} gene dosage ratio in the WT and mutant strains. \textit{kinA} is replicated at the end of the DNA replication period due to its position close to the chromosome terminus in all strains. Note that \textit{kinA} and 0F gene dosage trajectories are identical in the \textit{Trans-OF\textsuperscript{gltA}} and \textit{i0F\textsuperscript{gltA}} strains.

**Q-T.** Model simulation results showing the time course of 0A~P concentrations in WT and mutant strains. Note that 0A~P pulses in WT (Q) but does not pulse or increase significantly in the fluctuates in the \textit{Trans-OF\textsuperscript{gltA}} mutant (R) that lacks the transient change in \textit{kinA:0F} gene dosage ratio. 0A~P does not pulse in the inducible strains \textit{i0F\textsuperscript{amyE}} (S) \textit{i0F\textsuperscript{gltA}} (T) but is fluctuates in \textit{i0F\textsuperscript{amyE}} due to the decrease in \textit{kinA:0F} gene dosage ratio during DNA replication. The pulse amplitude of 0A~P in WT and the level of 0A~P in the mutants increases with increasing KinA autophosphorylation rate (compare with experimental data from Fig. 7.4).

**U-X.** Model simulation results similar to Q-T showing the time course of \textit{P\textsubscript{0A}} (blue curves) and \textit{P\textsubscript{0F}} (pink curves) promoter activities in WT and mutant strains.

---

**Figure A6.4.** \textit{OF} and \textit{kinA} chromosomal locations are evolutionarily conserved in spore forming bacteria.

**A.** Radial histogram of chromosomal locations of \textit{kinA} (green) and \textit{OF} (purple) in 46 species of spore forming bacteria (see also Table A6.4). The histograms show that the \textit{0F} is located close to the origin \textit{oriC} and \textit{kinA} is located close to the terminus in all
Bacillus species. *kinA* and *0F* locations in *B. subtilis* 168 (*0F*-326° and *kinA*-126°) are marked by solid lines.

**B.** Radial histogram of chromosomal locations of *lacI* (blue) in the same 46 species of spore forming bacteria as (A). The histogram shows that unlike *kinA* and *0F*, *lacI* location relative to *oriC* varies widely over different Bacillus species.

**Figure A6.5. Comparison of pulsing and non-pulsing responses.**

**A-B.** Pulsing dynamics (A) and Non-pulsing dynamics (B) of *P_{0A}* promoter activity can be distinguished by comparing the maximum change in promoter activity (ΔMax) and the change in the promoter activity at the end of the cell-cycle (ΔEnd). Both changes are measured relative to the promoter activity at the start of the cell-cycle. The response can be classified as Pulsing dynamics only if ΔMax and ΔEnd differ significantly.

**C-G.** Comparison of ΔMax and ΔEnd in WT(C), *Trans-0F^{gltA}*(D), *Trans-0F^{amyE}*(E), *i0F^{amyE}*(F) and *i0F^{gltA}*(G). Note that ΔMax and ΔEnd differ significantly only in WT and *Trans-0F^{amyE}* strains.
## A6.2. Supplementary Tables

### Table A6.1. *B. subtilis* strains used in this study.

<table>
<thead>
<tr>
<th><em>B. subtilis</em> strains as referred in the article</th>
<th><em>B. subtilis</em> strain number</th>
<th>Genotype</th>
<th>Used in Figures</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Wildtype”, WT</td>
<td>AK151</td>
<td>AmyE::P_{spo0A-yfp}, P_{comG-mCherry} (Sp^R)&lt;br&gt;SacA::P_{spo11r-cfp} (Cm^R)</td>
<td>Fig. 7.4E; Fig. 7.5; Fig. A6.5C</td>
</tr>
<tr>
<td></td>
<td>AK456</td>
<td>AmyE::P_{spo0F-yfp} (Sp^R)</td>
<td>Fig. A6.1</td>
</tr>
<tr>
<td></td>
<td>AK2111</td>
<td>GltA::P_{spo0F-yfp} (Kan^R)</td>
<td>Fig. A6.1</td>
</tr>
<tr>
<td></td>
<td>AK2261</td>
<td>AmyE::P_{hsp-DnaN-YFP} (Sp^R)&lt;br&gt;pHP13- P_{spo0A-cfp}, P_{comG-mCherry} (Erm^R)&lt;br&gt;Spo0F:: kan (Kan^R)</td>
<td>Fig. 7.1B-D</td>
</tr>
<tr>
<td>i0F^{amyE}</td>
<td>AK2084</td>
<td>AmyE::P_{hsp-Spo0F} (Sp^R)&lt;br&gt;SacA:: P_{spo0F-yfp} (Cm^R)&lt;br&gt;pHP13- P_{spo0A-cfp} (Erm^R)&lt;br&gt;Spo0F:: erm (Erm^R)</td>
<td>Fig. 7.4M,O; Fig. 7.5A; Fig. A6.5G</td>
</tr>
<tr>
<td>Trans-i0F^{gltA}</td>
<td>AK2271</td>
<td>AmyE::P_{spo0A-yfp}, P_{comG-mCherry} (Sp^R)&lt;br&gt;SacA::P_{spo11r-cfp} (Cm^R)&lt;br&gt;GltA::P_{spo0F-Spo0F} (Pm^R)&lt;br&gt;Spo0F:: erm (Erm^R)</td>
<td>Fig. 7.4F; Fig. 7.5A; Fig. A6.5D</td>
</tr>
<tr>
<td>Trans-i0F^{amyE}</td>
<td>AK2301</td>
<td>AmyE:: P_{spo0F-Spo0F} (Sp^R)&lt;br&gt;SacA::P_{spo0A-yfp/P_{comG-cfp}} (Cm^R)&lt;br&gt;Spo0F:: erm (Erm^R)</td>
<td>Fig. 7.4G; Fig. A6.5E</td>
</tr>
<tr>
<td>i0F^{gltA}</td>
<td>AK2253</td>
<td>AmyE::P_{spo0A-yfp}, P_{comG-mCherry} (Sp^R)&lt;br&gt;SacA:: P_{spo11r-cfp} (Cm^R)&lt;br&gt;GltA::P_{hsp-Spo0F} (Pm^R)&lt;br&gt;Spo0F:: erm (Erm^R)</td>
<td>Fig. 7.4N,P; Fig. 7.5; Fig. A6.5F</td>
</tr>
<tr>
<td>iTrans-0F</td>
<td>AK2092</td>
<td>AmyE::P_{hsp-Spo0F} (Sp^R)&lt;br&gt;SacA::P_{spo11r-yfp} (Cm^R)&lt;br&gt;GltA::P_{spo0F-Spo0F} (Pm^R)&lt;br&gt;pHP13- P_{spo0A-cfp} (Erm^R)&lt;br&gt;Spo0F:: kan (Kan^R)</td>
<td>Fig. 7.4H</td>
</tr>
</tbody>
</table>
Table A6.2. Parameter values used for post-translational interactions in the model of sporulation phosphorelay.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Reaction</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_p$</td>
<td>$K_{\text{inA}} \rightarrow K_{\text{inA}}$</td>
<td>KinA auto-phosphorylation</td>
<td>8 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_{dp}$</td>
<td>$K_{\text{inA}} \rightarrow K_{\text{inA}}$</td>
<td>KinA dephosphorylation</td>
<td>2 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_b$</td>
<td>$K_{\text{inA}}.0F \rightarrow K_{\text{inA}}.0F$</td>
<td>Diffusion-limited binding rate constant</td>
<td>0.5nM$^{-1}$hr$^{-1}$</td>
</tr>
<tr>
<td>$k_i$</td>
<td>$K_{\text{inA}}.0F \rightarrow K_{\text{inA}}.0F$</td>
<td>KinA:0F complex dissociation</td>
<td>120 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_1$</td>
<td>$K_{\text{inA}}.0F \rightarrow K_{\text{inA}}.0F$</td>
<td>KinA$_P$.0F complex dissociation</td>
<td>300 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_2$</td>
<td>$K_{\text{inA}}.0F \rightarrow K_{\text{inA}}.0F_p$</td>
<td>0F phosphorylation</td>
<td>300 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_3$</td>
<td>$0F_p.0B_p \rightarrow 0F_p.0B$</td>
<td>0F$_P$.0B complex dissociation</td>
<td>500 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_4$</td>
<td>$0F_p.0B_p \rightarrow 0F.0B_p$</td>
<td>0F to 0B phosphotransfer</td>
<td>800 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_5$</td>
<td>$0B_p.0A_p \rightarrow 0B_p.0A$</td>
<td>0B$_P$.0A complex dissociation</td>
<td>200 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_6$</td>
<td>$0B_p.0A_p \rightarrow 0B_p.0A$</td>
<td>0B to 0A phosphotransfer</td>
<td>800 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_{ubr}$</td>
<td>$0F_p.0B \rightarrow 0F_p.0B_p$</td>
<td>0F$_P$.Rap complex dissociation</td>
<td>100 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_r$</td>
<td>$0F_p.0B \rightarrow 0F_p.0B_p$</td>
<td>0F$_P$ dephosphorylation</td>
<td>20 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_{ube}$</td>
<td>$0A_p.0E \rightarrow 0A_p.0E$</td>
<td>0A$_P$.0E complex dissociation</td>
<td>100 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_e$</td>
<td>$0A_p.0E \rightarrow 0A_p.0E$</td>
<td>0A$_P$ dephosphorylation</td>
<td>20 hr$^{-1}$</td>
</tr>
</tbody>
</table>
Table A6.3. Parameter values used for gene regulatory interactions in the model of sporulation phosphorelay.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Rate</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rightarrow m_{\text{KinA}}$</td>
<td>$v_{\text{kinA}} = v_{\text{kinA}}^0 + v_{\text{kinA}}^{\text{max}} \frac{[S]^m}{K_{\text{kinA}} + [S]^m}$</td>
<td>$v_{\text{kinA}}^{\text{max}} = 0.05 \text{ M hr}^{-1}$, $v_{\text{kinA}}^0 = 0.5 \text{ M hr}^{-1}$</td>
</tr>
<tr>
<td>$\rightarrow m_{0F}$</td>
<td>$v_{0F} = v_{0F}^0 + v_{0F}^{\text{max}} \frac{[S]^m}{K_{0F} + [S]^m}$</td>
<td>$v_{0F}^{\text{max}} = 0.1 \text{ M hr}^{-1}$, $v_{0F}^0 = 0.65 \text{ M hr}^{-2}$</td>
</tr>
<tr>
<td>$\rightarrow m_{0B}$</td>
<td>$v_{0B}$</td>
<td>$v_{0B} = 0.1 \text{ M hr}^{-1}$</td>
</tr>
<tr>
<td>$\rightarrow m_{0A}$</td>
<td>$v_{0A} = v_{0A}^0 + v_{0A}^{\text{max}} \frac{[S]^m}{K_{0A} + [S]^m}$</td>
<td>$v_{0A}^{\text{max}} = 0.2 \text{ M hr}^{-1}$, $v_{0A}^0 = 1 \text{ M hr}^{-1}$</td>
</tr>
<tr>
<td>$\rightarrow m_{\text{Rap}}$</td>
<td>$v_{\text{Rap}}$</td>
<td>$v_{\text{Rap}} = 0.015 \text{ M hr}^{-1}$</td>
</tr>
<tr>
<td>$\rightarrow m_{0E}$</td>
<td>$v_{0E}$</td>
<td>$v_{0E} = 5 \text{ nM hr}^{-1}$</td>
</tr>
<tr>
<td>Translation: mRNA $\rightarrow$ mRNA+Protein</td>
<td>$k_j[mRNA]$</td>
<td>$k_j = 30 \text{ hr}^{-1}$</td>
</tr>
<tr>
<td>mRNA degradation : mRNA $\rightarrow$</td>
<td>$k_m[mRNA]$</td>
<td>$k_m = 3 \text{ hr}^{-1}$</td>
</tr>
<tr>
<td>$\rightarrow S$</td>
<td>$k_s[0A_p]$</td>
<td>$k_s = 0.68 \text{ hr}^{-1}$</td>
</tr>
</tbody>
</table>
Table A6.4. Genome sizes and chromosomal locations of phosphorelay genes spo0F and kinA in 46 spore forming bacteria.

Data was drawn from the Pubmed Gene database. Only species which have both kinA and spo0F were selected. Gene location in degrees is given by the ratio: $360 \times \frac{\text{pos}_x}{G_s}$, where $\text{pos}_x$ is the position of the transcriptional start site of gene $x$ (=kinA or spo0F) in species $s$ (measured clockwise from the origin) and $G_s$ is the genome size of species $s$. ($0^\circ$=360$^\circ$=chromosome origin; 180$^\circ$=chromosome terminus).

<table>
<thead>
<tr>
<th>Species Name</th>
<th>Genome Size</th>
<th>0F Location (degrees/360)</th>
<th>kinA Location (degrees/360)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacillus subtilis subsp. subtilis str. 168</td>
<td>4215606</td>
<td>325.324</td>
<td>125.5358</td>
</tr>
<tr>
<td>Bacillus halodurans C-125</td>
<td>4202352</td>
<td>334.6104</td>
<td>160.656</td>
</tr>
<tr>
<td>Bacillus megaterium QM B1551</td>
<td>5097129</td>
<td>351.7039</td>
<td>165.9983</td>
</tr>
<tr>
<td>Bacillus amyloliquefaciens DSM 7</td>
<td>3980199</td>
<td>325.8056</td>
<td>134.0655</td>
</tr>
<tr>
<td>Bacillus subtilis subsp. spizizenii str. W23</td>
<td>4027676</td>
<td>323.492</td>
<td>127.4498</td>
</tr>
<tr>
<td>Bacillus pseudofirmus OF4</td>
<td>3858997</td>
<td>342.5017</td>
<td>184.0939</td>
</tr>
<tr>
<td>Bacillus pumilus SAFR-032</td>
<td>3704465</td>
<td>323.9976</td>
<td>131.4467</td>
</tr>
<tr>
<td>Bacillus amyloliquefaciens FZB42</td>
<td>3918589</td>
<td>325.016</td>
<td>123.1567</td>
</tr>
<tr>
<td>Bacillus thuringiensis str. Al Hakam</td>
<td>5257091</td>
<td>348.7712</td>
<td>202.84</td>
</tr>
<tr>
<td>Bacillus licheniformis DSM 13 = ATCC 14580</td>
<td>4222645</td>
<td>322.7962</td>
<td>134.3266</td>
</tr>
<tr>
<td>Bacillus cereus biovar anthracis str. Cl</td>
<td>5196054</td>
<td>347.6662</td>
<td>147.0065</td>
</tr>
<tr>
<td>Bacillus thuringiensis BMB171</td>
<td>5330088</td>
<td>348.2211</td>
<td>238.7468</td>
</tr>
<tr>
<td>Bacillus cereus Q1</td>
<td>5214195</td>
<td>348.1544</td>
<td>151.8708</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td><em>Bacillus anthracis</em> strain 'Ames Ancestor'</td>
<td>5227419</td>
<td>348.789</td>
<td>234.2406</td>
</tr>
<tr>
<td>Bacillus cereus ATCC 10987</td>
<td>5224283</td>
<td>347.5709</td>
<td>235.4239</td>
</tr>
<tr>
<td>Bacillus anthracis strain Ames</td>
<td>5227293</td>
<td>348.7887</td>
<td>234.2375</td>
</tr>
<tr>
<td>Bacillus megaterium DSM 319</td>
<td>5097447</td>
<td>351.4806</td>
<td>163.0007</td>
</tr>
<tr>
<td>Bacillus cereus E33L</td>
<td>5300915</td>
<td>348.7168</td>
<td>236.0028</td>
</tr>
<tr>
<td><em>Bacillus thuringiensis</em> serovar konkukian strain 97-27</td>
<td>5237682</td>
<td>348.5111</td>
<td>148.329</td>
</tr>
<tr>
<td>Bacillus anthracis strain A0248</td>
<td>5227419</td>
<td>348.7821</td>
<td>234.2337</td>
</tr>
<tr>
<td>Bacillus anthracis strain CDC 684</td>
<td>5230115</td>
<td>348.7835</td>
<td>57.5361</td>
</tr>
<tr>
<td>Bacillus cereus 03BB102</td>
<td>5269628</td>
<td>348.79</td>
<td>238.2591</td>
</tr>
<tr>
<td>Bacillus cereus AH820</td>
<td>5302683</td>
<td>347.8941</td>
<td>236.5036</td>
</tr>
<tr>
<td>Bacillus cereus G9842</td>
<td>5387334</td>
<td>347.8314</td>
<td>238.9148</td>
</tr>
<tr>
<td>Bacillus cereus B4264</td>
<td>5419036</td>
<td>348.8427</td>
<td>243.252</td>
</tr>
<tr>
<td>Bacillus cereus AH187</td>
<td>5269030</td>
<td>348.3469</td>
<td>233.8018</td>
</tr>
<tr>
<td>Alicyclobacillus acidocaldarius subsp. acidocaldarius Tc-4-1</td>
<td>3124048</td>
<td>339.1574</td>
<td>227.7419</td>
</tr>
<tr>
<td>Bacillus licheniformis 9945A</td>
<td>4376305</td>
<td>326.2416</td>
<td>132.4264</td>
</tr>
<tr>
<td>Bacillus subtilis subsp. subtilis 6051-HGW</td>
<td>4215610</td>
<td>325.3242</td>
<td>125.5355</td>
</tr>
<tr>
<td>Bacillus subtilis XF-1</td>
<td>4061186</td>
<td>323.4117</td>
<td>130.4759</td>
</tr>
<tr>
<td>Bacillus subtilis subsp. natto BEST195</td>
<td>4091591</td>
<td>323.2848</td>
<td>130.1218</td>
</tr>
<tr>
<td>Bacillus subtilis QB928</td>
<td>4146839</td>
<td>324.7672</td>
<td>125.8378</td>
</tr>
<tr>
<td>Bacillus amyloliquefaciens Y2</td>
<td>4238624</td>
<td>326.7711</td>
<td>115.8335</td>
</tr>
<tr>
<td>Species</td>
<td>CCNA</td>
<td>Mn</td>
<td>MZ</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td><em>Paenibacillus polymyxa M1</em></td>
<td>5864546</td>
<td>10.39716</td>
<td>52.97136</td>
</tr>
<tr>
<td><em>Bacillus amyloliquefaciens subsp. plantarum YAU B9601-Y2</em></td>
<td>4242774</td>
<td>326.7803</td>
<td>115.9104</td>
</tr>
<tr>
<td><em>Bacillus amyloliquefaciens XH7</em></td>
<td>3939203</td>
<td>326.5166</td>
<td>182.7644</td>
</tr>
<tr>
<td><em>Bacillus amyloliquefaciens TA208</em></td>
<td>3937511</td>
<td>326.5049</td>
<td>182.7237</td>
</tr>
<tr>
<td><em>Bacillus megaterium WSH-002</em></td>
<td>4983975</td>
<td>8.047817</td>
<td>339.2598</td>
</tr>
<tr>
<td><em>Bacillus amyloliquefaciens subsp. plantarum CAU B946</em></td>
<td>4019861</td>
<td>327.6388</td>
<td>124.9527</td>
</tr>
<tr>
<td><em>Bacillus thuringiensis serovar chinensis CT-43</em></td>
<td>5486830</td>
<td>348.9355</td>
<td>236.8537</td>
</tr>
<tr>
<td><em>Corallococcus coralloides DSM 2259</em></td>
<td>10080619</td>
<td>37.49527</td>
<td>25.68442</td>
</tr>
<tr>
<td><em>Paenibacillus larvae subsp. larvae DSM 25430</em></td>
<td>4046337</td>
<td>339.7109</td>
<td>319.8833</td>
</tr>
<tr>
<td><em>Bacillus amyloliquefaciens subsp. plantarum NAU-B3</em></td>
<td>4196170</td>
<td>327.0585</td>
<td>222.1804</td>
</tr>
<tr>
<td><em>Bacillus amyloliquefaciens subsp. plantarum UCMB5113</em></td>
<td>3889532</td>
<td>326.1037</td>
<td>124.4399</td>
</tr>
<tr>
<td><em>Bacillus amyloliquefaciens subsp. plantarum UCMB5033</em></td>
<td>4071167</td>
<td>326.4765</td>
<td>122.2481</td>
</tr>
<tr>
<td><em>Bacillus amyloliquefaciens subsp. plantarum UCMB5036</em></td>
<td>3910324</td>
<td>325.163</td>
<td>122.767</td>
</tr>
</tbody>
</table>
A6.3. Supplementary Methods

A6.3.1. Mathematical Model of the Sporulation Phosphorelay

We extended a previous mathematical model of sporulation phosphorelay network (304) to uncover the mechanism of pulsatile 0A activation. For this purpose, we employed a deterministic model of the phosphorelay network with ordinary-differential equations describing concentration of the phosphorelay proteins and their complexes as a function of time.

Our model can be subdivided into the following two parts: (i) the post-translational interactions that describe the phosphorylation/dephosphorylation of phosphorelay species and (ii) the transcriptional feedback interactions that control the expression of the phosphorelay proteins. Below we describe the reactions and associated assumptions and parameter values for each of these parts separately.

Post-translational interactions in the phosphorelay

The activity of the sporulation master regulator is controlled by the sporulation phosphorelay through post-translational phosphorylation and dephosphorylation reactions. Specifically, phosphoryl groups are transferred from the major sporulation kinase KinA to Spo0A (0A) via the phosphotransferases Spo0B (0B) and Spo0F (0F) (129, 330). Phosphorylated 0F and 0A are subject to negative regulation by phosphatases Rap and Spo0E (0E), respectively. Our model of these post-translational interactions in the phosphorelay includes the following reactions (R1-R7):
\[
\text{KinA} \xrightleftharpoons[k_{b}]{k_{p}} \text{KinA}_{p} \quad \text{(R1)}
\]
\[
\text{KinA}_{p} + 0F \xrightleftharpoons[k_{i}]{k_{h}} \text{KinA}_{p}:0F \xrightleftharpoons[k_{2}]{k_{h}} \text{KinA} + 0F_{p} \quad \text{(R2)}
\]
\[
\text{KinA} + 0F \xrightleftharpoons[k_{i}]{k_{h}} \text{KinA}:0F \quad \text{(R3)}
\]
\[
0F_{p} + 0B \xrightleftharpoons[k_{i}]{k_{h}} 0F_{p}:0B \xrightleftharpoons[k_{h}]{k_{i}} 0F + 0B_{p} \quad \text{(R4)}
\]
\[
0B_{p} + 0A \xrightleftharpoons[k_{i}]{k_{h}} 0B_{p}:0A \xrightleftharpoons[k_{h}]{k_{i}} 0B + 0A_{p} \quad \text{(R5)}
\]
\[
0F_{p} + \text{Rap} \xrightleftharpoons[k_{i}]{k_{h}} 0F_{p}:\text{Rap} \xrightleftharpoons[k_{i}]{k_{h}} 0F + \text{Rap} \quad \text{(R6)}
\]
\[
0A_{p} + 0E \xrightleftharpoons[k_{i}]{k_{h}} 0A_{p}:0E \xrightleftharpoons[k_{h}]{k_{i}} 0A + 0E \quad \text{(R7)}
\]

Here $0F$, $0B$, $0A$ and $0E$ refer to the proteins Spo0F, Spo0B, Spo0A and Spo0E respectively. Subscript $P$ marks phosphorylated forms of the proteins, colon denotes protein complexes and each rate constant is introduced above its corresponding reaction arrow. Reaction R3 was included in our model to account for the substrate inhibition effect of Spo0F on the kinase KinA. All post-translational reactions were modeled with mass-action kinetics. Table A6.2 shows the parameter values used in this model. Binding rates $k_b$ for all reactions were assumed to be diffusion-limited with rate constant $0.5 \text{nM}^{-1}\text{hr}^{-1}$. Dephosphorylation by $0E$ and Rap were assumed to have catalytic rate constants $100 \text{ hr}^{-1}$ based on (413). All other rate constants were estimated from the in vitro measurements of phosphorelay kinetics (290). These parameter values have order of magnitude agreement with the parameters used in other models (307, 333).

**Protein production and degradation in the phosphorelay**

For modeling the expression of phosphorelay proteins, we assumed that binding-unbinding of transcription factors from their promoters is fast and therefore rates of transcription can be modeled with appropriate Hill-functions.

The transcription of the phosphorelay genes $\text{kinA}$, $0F$, and $0A$ is regulated by $0A_p$ both directly and indirectly (via $\sigma^H$), thereby forming multiple feedback loops (289, 428). To model the delay induced by indirect feedback we assumed
that $0A_p$ levels control the expression an intermediate regulator $S$ which in turn controls the transcription of $kinA$, $0F$, and $0A$ (similar to (102)). The regulation of $kinA$, $0F$, and $0A$ transcription by the intermediate regulator was modeled with the generic Hill-function:

$$v = v^0 + v^{\max} \frac{[S]^m}{K^m + [S]^m}$$

Here $v^0$ and $v^{\max}$ represent the basal and maximal rate of transcription, respectively. K and m represent the half-maximal binding constant and the Hill-exponent, respectively. For simplicity, the rate of expression of the intermediate regulator was assumed to be linearly dependent on $0A_p$. For $spo0B$, $spo0E$ and $rap$ we assumed constant rates of transcription. The specific rate expressions and parameter values used are described in Table A6.3.

For the simulations of the $i0F$ strains (Figs. 7.3CD), the $0F$ expression rate was fixed at 0.6µMhr$^{-1}$ and 0.8 µMhr$^{-1}$ for comparisons to 5µM IPTG induction and 20µM IPTG induction respectively.

The rate of transcription of all genes in the model were also assumed to be proportional to the gene copy number and cell growth rate according to the following equation: $v = gv_p$. Where $v$ is the actual rate of gene expression, $v_p$ represents the expressions described for each gene in Supplementary Table A6.3, g is the gene copy number.

All mRNAs and proteins were assumed to be degraded/diluted with first-order kinetics. Degradation rate constant for all mRNAs was fixed at 3hr$^{-1}$ assuming a 10 min half-life for mRNAs. mRNA translation rate was fixed at 30hr$^{-1}$ for all mRNAs. Degradation rate constants for all proteins was fixed at 0.6hr$^{-1}$. Dilution rate constant for all proteins was set by the cell growth rate.

*Steady state response of phosphorelay*
To determine the effect of the substrate inhibition of KinA by 0F on our model results, we decoupled the transcriptional feedback from 0A_P to kinA, 0F, and 0A in our model and studied the response post-translational interactions in the phosphorelay in isolation. We kept the rate of 0A transcription fixed at 0.4µMhr⁻¹ for these simulations and calculated the steady state 0A_P concentrations at different combinations of total KinA and 0F concentrations ([KinA]_T and [0F]_T). As shown in the contour plot of Fig. A6.2A, steady state 0A_P levels always increased with increasing [KinA]_T. However when the substrate inhibition of KinA by 0F (reaction R3) was included in the model, steady state 0A_P levels depended non-monotonically on [0F]_T concentrations and decreased at high 0F levels. Plotting the steady state 0A_P concentrations as a function of the [0F]_T/[KinA]_T ratio (Fig. A6.2B) at different fixed levels of [KinA]_T showed that [0A_P] decreases ultrasensitively as this ratio becomes greater than one. Moreover, the ultrasensitivity of this decrease (measured by Hill-exponent) increases at higher [KinA]_T. These results suggest that the substrate inhibition of KinA by 0F makes the post-translational response of the phosphorelay highly sensitive to the ratio of 0F and kinA transcription rates. All steady state responses were calculated using the CL_MATCONT bifurcation package for MATLAB.

A similar procedure was used for Fig. 7.2B, to demonstrate the effect of substrate inhibition on the phosphorelay response in a mutant where 0F expression is externally induced. For this simulation, we decoupled the transcriptional feedback from 0A_P to 0F and independently varied the 0F expression rate to calculate the steady state 0A_P concentrations as a function of increasing 0F concentrations. As shown in Fig. 7.2B, when the substrate inhibition of KinA by 0F (reaction R3) was included in the model, steady state 0A_P levels and the P_0A promoter activity decreased ultrasensitively as 0F expression was increased. In contrast, when reaction R3 was excluded, 0A_P levels and P_0A promoter activity were far less sensitive to high 0F concentrations. These modeling results were compared to measurements of P_0A promoter activity in the
$i0F^{amyE}$ strain where 0F is expressed solely from an IPTG inducible promoter. As shown in Fig. 7.2B, $P_{0A}$ promoter activity in the $i0F^{amyE}$ strain decreases ultrasensitively at high IPTG levels thus confirming the substrate inhibition effect. These steady state calculations were all done with the CL_MATCONT bifurcation package.

**Time-Course Simulations**

All simulations of the phosphorelay response (Figs. 7.3 and Figs. A6.4) were done using the *ode15s* solver of MATLAB. The cell-cycle duration, $T_d$ was fixed at 8hrs and used to determine the growth/protein-dilution rate $\mu$ ($\mu = \log(2)/T_d$). DNA replication periods were modeled with a fixed 2 hour time duration. All origin proximal genes ($P_{0A}$ and $P_{0F}$ reporters, 0F in WT and $P_{hsp-0F}$ in $i0F^{amyE}$) were assumed to be replicated at the start of the DNA replication period and all terminus proximal genes ($kinA$, 0B, 0A, $P_{hsp-0F}$ in $i0F^{gltA}$, 0F in *Trans-OF*amyE) were assumed to be replicated at the end of the DNA replication time-window. For simplicity, replication was assumed to start immediately after cell-division.

In Fig. A6.4, we mimic the effect of increasingly severe nutrient deprivation over multiple cell-cycles spent in starvation conditions by varying the KinA autophosphorylation rate, $k_p$. As shown in Fig. A6.4, increasing $k_p$ leads to increasing $0A_P$ levels and $P_{0A}$ promoter activity similar to the increases seen in experimental results of Fig. 7.4. $k_p$ is not the only parameter that can affect $0A_P$ levels in this fashion. Changes in other parameters including $0A$ dephosphorylation rate, protein dilution rate etc. can have a similar effect (not shown). Notably however whereas changes in these parameters can affect quantitative aspects of our results such as the $0A_P$ pulse amplitude, the qualitative features of our results (i.e. pulsing in WT; no pulsing in *Trans-OF*gltA and $i0F^{gltA}$ etc.) are robust to such parameter variations.
Appendix 7. Slowdown of growth controls cellular differentiation

A7.1. Supplementary Figures

Figure A7.1. Effect of growth slowdown on pulsing from $P_{0F}$ promoter, cell volume and DNA replication.

A. Single cell time-lapse microscopy measurements of cell length (green) and cell growth rate (gray), over multiple cell-cycles in starvation media. In (A-C) vertical dashed lines indicate cell divisions.

B. Fluorescence measurements for $P_{0F}$-yfp reporters in WT background show that the expression level of $P_{0F}$-yfp increases in non-monotonic fashion.
C. Promoter activity of $P_{0F}$-yfp reporter (defined as production rate, an indicator of OAP level) shows pulses once every cell-cycle. Promoter activity pulse amplitudes increase as growth rate decreases.

D. Measurements of $P_{0F}$-yfp promoter activity show that the pulse amplitudes and growth rates are anti-correlated. Each dot corresponds to ranked measurements of the $P_{0A}$-yfp promoter activity pulse amplitude and growth rate of an individual cell-cycle. Red and gray dots indicate cell-cycle that end in sporulation and vegetative division respectively. The resulting Spearman's rank correlation $\rho=-0.56$, p-value<$10^{-60}$.

E. Measurements of cell length at division show that cell size decreases with decreasing growth rate. Gray circles show cell lengths at division over 25 hours of growth slowdown in starvation conditions. Blue circles and errorbars show the mean and standard deviations of division length measurements binned according to growth rate. Solid line indicates the phenomenological fit ($L(\mu)=3.466*\exp(-0.689/\mu)+3.743 \, \mu m$) for cell length dependence on growth rate that we use in our models.

F. Measurements of Trep, the duration of DNA replication period at different cell growth rates. Gray circles show Trep measured by detecting DnaN-YFP foci over 25 hours of growth slowdown in starvation conditions. Blue circles and errorbars show the mean and standard deviations of Trep measurements binned according to growth rate. Solid line indicates the phenomenological fit ($\text{Trep}(\mu)=0.15/\mu+0.78 \, \text{hrs}$) for DNA replication period dependence on growth rate that we use in our models.
Figure A7.2. Dependence of sporulation cell-fate on $P_{0A}$ promoter activity and growth rate.

**A.** Probability of sporulation as a function of peak $P_{0A-yfp}$ promoter activity calculated using logistic regression on the data in Fig. 8.2D. Solid lines and blue areas indicate the logistic regression curves and the 95% confidence intervals. Red dots show the fraction of cells sporulating calculated by binning data in linearly spaced bins of $P_{0A-yfp}$ promoter activity.

**B.** Probability of sporulation as a function of growth rate calculated using logistic regression on the data in Fig. 8.2D. Solid lines and blue areas indicate the logistic regression curves and the 95% confidence intervals. Red dots show the fraction of cells sporulating calculated by binning data in linearly spaced bins of growth rate.
Figure A7.3. Effect of nutrient availability on growth slowdown and sporulation dynamics.

A. Dynamics of growth slowdown in normal starvation media (RM). Gray circles and errorbars show the mean and standard deviations respectively of the growth rate of a colony of cells in RM. In RM substrate availability is low from the start of the experiment. As a result, cells grow slowly and accordingly nutrient levels are depleted gradually until cells cross the growth threshold for sporulation (green line) around 10 hours into the experiment. Fitting this data (black curve) shows that these growth rate dynamics can be explained by a simple population dynamics model for substrate amount and number of cells (see Appendix 7 Methods for details).

B. Dynamics of growth slowdown with increased initial nutrients (RM+0.025% glucose). The population dynamics model predicts that growth rate dynamics are sensitive to initial nutrient availability. Increased nutrients at the start of the experiment due to glucose addition increases the initial growth rate and postpones starvation but subsequently the increased number of cells results in rapid depletion of nutrients and decrease in growth rate (black curve). As a result, cells cross the growth threshold (green line) for sporulation earlier around 7 hours into the experiment. Experimental measurements of growth dynamics (blue circles) in RM+0.025% glucose confirm the model predictions. Blue circles and errorbars show the mean and standard deviations respectively.
Figure A7.4. Model predictions for 0A–P pulsing in an inducible KinA strain.

A. Model time course for an inducible KinA strain showing cell-cycle coordinated 0A–P pulsing. Upper and lower panels show the growth rate (model input) and 0A–P response (simulation result) respectively. Red, green and blue curves show the 0A–P response at increasing rates of KinA production ($V_{\text{kinA}}$) from an inducible promoter. Yellow bars and dashed lines represent DNA replication periods and cell division respectively. The simulation predicts that 0A–P pulses in the inducible KinA strain and that similar to WT,
pulse amplitudes increase with decreasing cell growth rate and with increase in KinA induction.

**B, C.** The two alternative hypotheses for the relationship between starvation, 0A~P and sporulation cell-fate. (B) Starvation controls 0A activation by affecting cell growth and by increasing KinA activity via an unknown signal. (C) Starvation controls 0A activation by affecting only cell growth but not KinA activity.

**D, E.** Model predictions for the dependence of the 0A~P pulse amplitudes on growth rate under the signal-dependent (D) and signal-independent (E) KinA-activity. The green line shows the 0A~P threshold used in model simulations to predict cell-fate. Note that the growth rate threshold (growth rate at which the 0A~P threshold is reached) depends on the KinA production rate ($V_{\text{kinA}}$) under both hypotheses.

**F, G.** Model predictions for the dependence of the 0A~P pulse amplitudes on KinA under the signal dependent (F) and signal independent (G) hypotheses. The green line shows the 0A~P threshold used in model simulations to predict cell-fate. Note that the KinA threshold (KinA level at which the 0A~P threshold is reached) depends on the KinA production rate ($V_{\text{kinA}}$) under the signal-dependent (F) but not the signal-independent (G) KinA activity.
Figure A7.5. Experimental measurement of the sporulation cell-fate outcomes as a function of KinA level and growth rate.
A. KinA-GFP levels in the IPTG-inducible $P_{hsp}$-kinA:gfp (at different IPTG concentrations) and WT ($P_{kinA}$-kinA:gfp) as a function of cell-cycle growth rate. Measurements confirm that 0A activity KinA-GFP levels increase as growth rate decreases during starvation. Each dot corresponds to a single cell-cycle. Black and red dots correspond to cell-cycles that end in vegetative division and spores respectively. Blue solid lines show the model predictions for KinA dependence on growth rate (same as that used in Fig. 8.2A). Horizontal and vertical green lines show the thresholds that can be used to predict cell fate as a function of KinA level and growth rate respectively.

B. Probability of sporulation as a function of KinA-GFP levels calculated using logistic regression on the data in A. Solid lines and blue areas indicate the logistic regression curves and the 95% confidence intervals. Red dots show the fraction of cells sporulating calculated by binning data in linearly spaced bins of KinA-GFP.

C. Probability of sporulation as a function of growth rate calculated using logistic regression on the data in A. Solid lines and blue areas indicate the logistic regression curves and the 95% confidence intervals. Red dots show the fraction of cells sporulating calculated by binning data in linearly spaced bins of growth rate.

D. Receiver Operating Characteristic (ROC) curves for KinA-GFP based and growth-based cell-fate prediction. Green and blue lines shows the relation between False positive rate and True positive rate for different values of KinA-GFP and growth rate threshold.

E. KinA-GFP thresholds calculated using the results of the logistic regression in B. Errorbars show standard errors. KinA threshold does not depend on IPTG level (One-way ANOVA: p-val=0.48).

F. Growth rate thresholds calculated using the results of the logistic regression in C. Errorbars show standard errors. Growth rate threshold increases with IPTG level (One-way ANOVA: p-val<1e-3).
A7.2. Supplementary Tables

Table A7.1. *B. subtilis* strains used in this study

<table>
<thead>
<tr>
<th><em>B. subtilis</em> strains as referred in the article</th>
<th>B. subtilis strain number</th>
<th>Genotype</th>
<th>Used in Figures</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Wildtype”, WT</td>
<td>AK151</td>
<td>AmyE::spo0A-yfp, P_comG-mCherry (Sp&lt;sup&gt;R&lt;/sup&gt;) SacA::spo0IIR-cfp (Cm&lt;sup&gt;R&lt;/sup&gt;)</td>
<td>Fig. 8.2D; Fig. A7.2</td>
</tr>
<tr>
<td>TC669</td>
<td>AmyE::p&lt;sub&gt;hap&lt;/sub&gt;-yfp (Sp&lt;sup&gt;R&lt;/sup&gt;)</td>
<td>Fig. 8.2A</td>
<td></td>
</tr>
<tr>
<td>AK2161</td>
<td>P&lt;sub&gt;spo0IIR&lt;/sub&gt;-YFP, P&lt;sub&gt;spo0A-CPFP&lt;/sub&gt;, pDG148-P&lt;sub&gt;rpsD&lt;/sub&gt;-mCherry</td>
<td>Fig. 8.1C-F; Fig. 3; Fig. A7.1E</td>
<td></td>
</tr>
<tr>
<td>AK456</td>
<td>AmyE::P&lt;sub&gt;spo0IIR&lt;/sub&gt;-yfp (Sp&lt;sup&gt;R&lt;/sup&gt;)</td>
<td>Fig. A7.1A-D</td>
<td></td>
</tr>
<tr>
<td>MF929</td>
<td>KinA::P&lt;sub&gt;kinA&lt;/sub&gt;-kinA-gfp (Kan&lt;sup&gt;R&lt;/sup&gt;)</td>
<td>Fig. 8.5CDE; Fig. A7.5</td>
<td></td>
</tr>
<tr>
<td>AK2261</td>
<td>AmyE:: p&lt;sub&gt;hap&lt;/sub&gt;-DnaN-YFP (Sp&lt;sup&gt;R&lt;/sup&gt;) pH13- P&lt;sub&gt;spo0A&lt;/sub&gt;-cfp, P&lt;sub&gt;comG&lt;/sub&gt;-mCherry (Erm&lt;sup&gt;R&lt;/sup&gt;)</td>
<td>Fig. A7.1F</td>
<td></td>
</tr>
<tr>
<td>iTrans-OF</td>
<td>AK2092</td>
<td>AmyE::P&lt;sub&gt;hap&lt;/sub&gt;-Spo0F (Sp&lt;sup&gt;R&lt;/sup&gt;) SacA:: P&lt;sub&gt;spo0IIR&lt;/sub&gt;-yfp (Cm&lt;sup&gt;R&lt;/sup&gt;) GltA::P&lt;sub&gt;spo0F&lt;/sub&gt;-Spo0F (Pm&lt;sup&gt;R&lt;/sup&gt;) pH13- P&lt;sub&gt;spo0A&lt;/sub&gt;-cfp (Erm&lt;sup&gt;R&lt;/sup&gt;) Spo0F:: kan (Kan&lt;sup&gt;R&lt;/sup&gt;)</td>
<td>Fig. 8.4D-I</td>
</tr>
<tr>
<td>Inducible KinA</td>
<td>MF2840</td>
<td>KinA::P&lt;sub&gt;hap&lt;/sub&gt;-kinA-gfp (Cm&lt;sup&gt;R&lt;/sup&gt;) (Kan&lt;sup&gt;R&lt;/sup&gt;)</td>
<td>Fig. 8.5CDE; Fig. A7.5</td>
</tr>
</tbody>
</table>

Table A7.2. Parameter values used for gene regulatory interactions in the model of sporulation phosphorelay.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Rate</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ KinA</td>
<td>$v_{\text{kinA}} = v_{\text{kinA}}^0 + v_{\text{kinA}}^\text{max} \frac{[S]^m}{K_{\text{kinA}} + [S]^m}$</td>
<td>$v_{\text{kinA}}^0 = 0.9 \mu M hr^{-1}$, $v_{\text{kinA}}^\text{max} = 1.5 \mu M hr^{-1}$</td>
</tr>
<tr>
<td>→ 0F</td>
<td>$v_{0F} = v_{0F}^0 + v_{0F}^\text{max} \frac{[S]^m}{K_{0F} + [S]^m}$</td>
<td>$v_{0F}^0 = 0.15 \mu M hr^{-1}$, $v_{0F}^\text{max} = 3 \mu M hr^{-1}$</td>
</tr>
<tr>
<td>→ 0B</td>
<td>$v_{0B}$</td>
<td>$v_{0B} = 0.3 \mu M hr^{-1}$</td>
</tr>
<tr>
<td>→ 0A</td>
<td>$v_{0A} = v_{0A}^0 + v_{0A}^\text{max} \frac{[S]^m}{K_{0A} + [S]^m}$</td>
<td>$v_{0A}^0 = 1.5 \mu M hr^{-1}$, $v_{0A}^\text{max} = 6 \mu M hr^{-1}$</td>
</tr>
<tr>
<td>→ Rap</td>
<td>$v_{\text{Rap}}$</td>
<td>$v_{\text{Rap}} = 0.075 \mu M hr^{-1}$</td>
</tr>
<tr>
<td>→ 0E</td>
<td>$v_{0E}$</td>
<td>$v_{0E} = 0.03 \mu M hr^{-1}$</td>
</tr>
<tr>
<td>→ S</td>
<td>$k_s[0A_p]$</td>
<td>$k_s = (\mu + 0.3) hr^{-1}$</td>
</tr>
</tbody>
</table>
Table A7.3. Parameter values used for the population dynamics model of growth dynamics during starvation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_m$</td>
<td>Maximum growth Rate</td>
<td>1.13 hr$^{-1}$</td>
</tr>
<tr>
<td>$K_m$</td>
<td>Half-maximal substrate concentration</td>
<td>0.82</td>
</tr>
<tr>
<td>$k_d$</td>
<td>Maximum sporulation rate</td>
<td>0.1 hr$^{-1}$</td>
</tr>
<tr>
<td>$K_s$</td>
<td>Half-maximal substrate concentration for sporulation</td>
<td>0.12 hr$^{-1}$</td>
</tr>
<tr>
<td>$m$</td>
<td>Hill-exponent for sporulation rate</td>
<td>4</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Substrate yield</td>
<td>0.02 substrate amt./cell</td>
</tr>
<tr>
<td>$N_0$</td>
<td>Initial Cell Number</td>
<td>SM: 5, SM+0.025% Glucose: 5</td>
</tr>
<tr>
<td>$S_0$</td>
<td>Initial Substrate Concentration</td>
<td>SM: 0.44, SM+0.025% Glucose: 0.72</td>
</tr>
</tbody>
</table>

A7.3. Supplementary Methods

A7.3.1 Mathematical Modeling Methods

Derivation of the dependence of protein concentrations on growth rate (Eq. [1])

To derive the Eq. [1] from the main text describing the dependence of protein concentration ($C$), on growth rate ($\mu$) we started with differential equations for protein molecule number ($N$) including production and degradation terms and for exponential growth of cell volume ($V$):

$$\frac{dN}{dt} = P - k_{\text{deg}} N$$

$$\frac{dV}{dt} = \mu V$$

In the first equation, we include protein production (rate $P$) and degradation (rate $k_{\text{deg}}$). The second equation describes the exponential increase in cell volume with
growth rate $\mu$. Using these equations and definition for concentration $C=\frac{N}{V}$, we can derive the rate of change of $C$:

$$\frac{dC}{dt} = \frac{d}{dt} \left( \frac{N}{V} \right) = \frac{1}{V} \frac{dN}{dt} - \frac{N}{V^2} \frac{dV}{dt} = \frac{1}{V} \left( P - k_{\text{deg}} N \right) - \frac{N}{V} \mu V = \frac{P}{V} - \left( \mu + k_{\text{deg}} \right) C$$

At steady state this equation result in the postulated dependence of $C$ on growth rate (Eq. [1] in the main text)

**Phosphorelay network model**

To investigate the dependence of $0A$ activity on cell growth rates we extended a previous mathematical model of sporulation phosphorelay network (350). This model used ordinary-differential equations describing concentration of the phosphorelay proteins and their complexes as a function of time to provide a deterministic description of the phosphorelay network response.

Our model can be subdivided into the following two parts: (i) the post-translational interactions that describe the phosphorylation/dephosphorylation of phosphorelay species and (ii) the transcriptional feedback interactions that control the expression of the phosphorelay proteins.

Post-translationally, the activity of the sporulation master regulator is controlled by the sporulation phosphorelay through phosphorylation and dephosphorylation reactions (Fig. 8.1B). Specifically, phosphoryl groups are transferred from the major sporulation kinase $\text{KinA}$ to $\text{Spo0A}$ ($0A$) via the phosphotransferases $\text{Spo0B}$ ($0B$) and $\text{Spo0F}$ ($0F$) (129, 330). Phosphorylated $0F$ ($0F\sim P$) and $0A$ ($0A\sim P$) are subject to negative regulation by phosphatases $\text{Rap}$ and $\text{Spo0E}$ ($0E$), respectively. All post-translational reactions were modeled exactly as in (350) with mass-action kinetics and the rate constants that were estimated from the in vitro measurements of phosphorelay kinetics (290).

Transcriptionally, the production of the phosphorelay genes $\text{kinA}$, $0F$, and $0A$ is regulated by $0A\sim P$ (Fig. 8.1B) both directly and indirectly (via $\sigma^H$), thereby forming multiple feedback loops (289, 428). For modeling the expression of
phosphorelay proteins, we again followed (350) and assumed that rates of transcription can be modeled with appropriate Hill-functions. To model the delay induced by indirect feedback we assumed that $0A\sim P$ levels control the expression an intermediate regulator $S$ which in turn controls the transcription of $\text{kinA}$, $0F$, and $0A$ (similar to (102)). The regulation of $\text{kinA}$, $0F$, and $0A$ transcription by the intermediate regulator was modeled with the generic Hill-function:

$$v_p = v^0 + v^{\text{max}} \frac{[S]^m}{K^m + [S]^m}$$

Here $v^0$ and $v^{\text{max}}$ represent the basal and maximal rate of transcription, respectively. $K$ and $m$ represent the half-maximal binding constant and the Hill-exponent, respectively. For simplicity, the rate of expression of the intermediate regulator was assumed to be linearly dependent on $0A\sim P$. For spo0B, spo0E and rap we assumed constant rates of transcription. The specific rate expressions and parameter values used are described in Table A7.2.

For the simulations of the inducible KinA strains (Figures 8.3CD), the $\text{kinA}$ expression rate $V_{\text{kinA}}$ was independent of $0A\sim P$ and varied between 0 µMhr$^{-1}$ and 5 µMhr$^{-1}$. For the simulations of the iTrans-$0F$ strain (Figures 8.4BC), the expression rate for the origin-proximal $P_{\text{hsp}}\sim 0F$ was independent of $0A\sim P$ and fixed at 0.7µMhr$^{-1}$. The protein degradation rate was constant for all proteins and was fixed at 0.3 hr$^{-1}$.

**Growth and gene copy number dependence of transcription rate**

The rate of expression of all genes in the model were also assumed to be proportional to the gene copy number and cell growth rate according to the following equation:

$$v = g \cdot v_p / F(\mu)$$

Where $v$ is the actual rate of gene expression, $v_p$ represents the expressions described for each gene in Table A7.2, $g$ is the gene copy number and $F(\mu)$ is a proportionality factor that models the effect of changes in cell size depending on the growth rate $\mu$. $F(\mu)$ is normalized such that...
F=1 for cells with doubling time of 1 hour (µ=log(2) hr⁻¹). We used the following phenomenological expression for F(µ): F(µ)=a*exp(b*µ)+c. The values for a, b and c were determined by fitting the data for change in cell length at division as a function of growth rate (Fig. A7.1E). We found that a=0.690, b=0.689 and c=0.745.

Simulations

All simulations of the phosphorelay response (Figs. 8.2, 8.5 and A7.4) were done using the ode15s solver of MATLAB and a decreasing series of cell-cycle growth rates (µ) to mimic the starvation response in the experiments (compare Figs. 8.1 and 8.2B).

The cell-cycle durations, T_{cyc} were fixed based on the growth rates:

T_{cyc}=\log(2)/\mu \text{ hrs}

The DNA replication period duration T_{rep}, was also assumed to be growth rate dependent and we used the following phenomenological expression for T_{rep}:

T_{rep}=0.78+0.15/\mu \text{ hrs}

The values of the coefficients in the above equation were determined by fitting the data for change in DNA replication periods as a function of growth rate (Fig. A7.1F). To identify DNA replication windows in time-lapse experiments we expressed a fluorescent DnaN-YFP fusion protein from the IPTG inducible P_{hsp} promoter and used the same quantification procedure as that described in (350).

For simplicity, replication was assumed to start immediately after cell-division.

All origin proximal genes (0F, P_{hsp}-0F in iTrans-0F and P_{0A} reporters) were assumed to be replicated at the start of the DNA replication period and all terminus proximal genes (kinA, 0B, 0A, 0F in iTrans-0F and P_{hsp}-kinA-gfp in the inducible KinA strain) were assumed to be replicated at the end of the DNA replication time-window.
For the signal dependent KinA activity hypothesis (Fig. 8.5A and Fig. A7.4BDF), the KinA autophosphorylation rate, $k_p$, was assumed to depend on the growth rate:

$$k_p = 1 + 8/((6.8 * \mu)^4 + 1) \text{ hr}^{-1}$$

For the signal independent KinA activity hypothesis (Fig. 8.5B and Fig. A7.4CEG), the KinA autophosphorylation rate, $k_p$, was fixed at $8 \text{ hr}^{-1}$ and assumed to be independent of growth rate.

**Dose responses of OA~P pulse amplitudes**

Under both signal-dependent and signal-independent KinA activity hypotheses, the OA~P pulse amplitude, growth rate and KinA concentration during each cell-cycle were calculated from these simulations to determine the OA~P pulse amplitude vs growth rate (Fig. 8.2D and Fig. A7.4DE) and OA~P pulse amplitude vs KinA concentration (Fig. A7.4FG) dose response relationships.

To calculate the growth rate and KinA thresholds the OA~P threshold was fixed at 0.9 µM and the dose response relationships were used to find the corresponding growth rate and KinA level. In the inducible KinA strain, the KinA and growth thresholds were calculated at different kinA production rates ($V_{\text{kinA}}$) to determine the interdependence of KinA and growth thresholds under the signal dependent and signal independent KinA activity hypotheses (Fig. 8.5AB).

**Sensitivity of OA~P pulse amplitudes**

To calculate the sensitivity of OA~P pulse amplitudes to variations in the phosphorelay protein levels (Fig. 8.2C) we tested the effect of increasing the production rate of the proteins on the OA~P pulse amplitude. For each phosphorelay protein $p$, the production rate $v_p$, was increased by $\Delta=10\%$ and then the OA~P pulse amplitude $[\text{OA~P}]_\Delta$, was calculated at the growth rate $\mu=0.15\text{ hr}^{-1}$ (corresponds to the $[\text{OA~P}]_{\text{WT}}=0.9\mu\text{M}$ – the sporulation threshold in our wildtype simulations). The normalized sensitivity of OA~P pulse amplitudes to each phosphorelay protein $p$ was then calculated using the following equation:
A7.3.2. Population Dynamics Model of Growth and Sporulation

To understand the effect of glucose addition on cell growth rate dynamics and thereby sporulation (Fig. 8.3C-F), we built a simple population dynamics model. We assumed that cell growth rate during starvation follows Monod kinetics (Kovarova-Kovar and Egli, 1998). Based on the observation of (Veening et al., 2008), we also assumed that cell death/sporulation releases nutrients that can be reused for cell growth. Our model is given by two equations for the number of cells (N) and amount of substrate (S):

\[
\frac{dN}{dt} = \left( v_m \frac{S}{K_m + S} \cdot \frac{k_d}{1 + (S/K_S)^m} \right) N
\]

\[
\frac{dS}{dt} = -\gamma \left( v_m \frac{S}{K_m + S} \cdot \frac{k_d}{1 + (S/K_S)^m} \right) N
\]

Here \( v_m \) and \( K_m \) are the maximum growth rate and the half-maximal substrate concentration for the Monod growth kinetics, respectively. To model the growth threshold-based sporulation decision, we assumed that sporulation rate is a non-linear function of the available substrate concentrations. \( k_d, K_s \) and \( m \) are the maximum sporulation/death rate, half-maximal concentration and the Hill-exponent, respectively. Parameter \( \gamma \) is the substrate yield. These model parameters along with the initial substrate concentration (arbitrary units: amt. substrate) were determined by fitting the model to data for cell growth from time-lapse experiments in Resuspension Media. The Levenberg-Marquardt algorithm of MATLAB fsolve function was used for fitting. The initial cell number was fixed to 5. The parameter values determined from fitting are shown in Supplementary Table A7.4. Using these same values for parameters, the model was used to explain the effect of increased nutrient availability at the start of the experiment.
As shown in Fig. A7.3, this model shows that the dynamics of cell growth are sensitive to the initial substrate availability. In regular sporulation media, substrate availability is low from the start of the experiment. As a result, cells grow slowly and the nutrient levels are accordingly depleted gradually until the cells cross the growth threshold for sporulation around 10 hours into the experiment (Fig. A7.3A). In contrast, the addition of 0.025% glucose at the start of the experiment increases the initial substrate availability and postpones starvation (Fig. A7.3B). Under these conditions cells grow rapidly and multiply. The increased number of cells at the onset of starvation results in a rapid depletion of nutrients and a decrease in growth rate (Fig. A7.3B). Consequently, in these conditions the cells start to sporulate earlier: around 7 hours into the experiment (Fig. 8.3B).
Appendix 8. Design and functional properties of the general stress response network in *Bacillus subtilis*

A8.1. Supplementary Figures

Figure A8.1. $\sigma^B$ does not pulse for networks that lack negative feedback.

A-C. Decoupled post-translational and transcriptional components of $\sigma^B$ networks that lack negative feedback. (A) $\lambda_W=2$, $\lambda_V=2$ (Region I) - positive feedback system; (B) $\lambda_W=8$, $\lambda_V=4.5$ (Region III) a non-responsive system; (C) $\lambda_W=4$, $\lambda_V=4.5$ with no transcriptional
feedback – no feedback system. In (A)-(C) cyan and blue lines show the post-translational response at low and high phosphatase respectively. Black line shows the and the transcriptional responses. Gray and black dots mark the steady states of the full system. The step increase in phosphatase causes a shift in the steady-state post-translational response (from low phosphatase-cyan to high phosphatase-blue) and leads to an increase in $\sigma^B$ (green curve) in all three systems.

**D-F.** Time-course representations of the green trajectories described in (A-C). Note that $\sigma^B$ does not pulse in any of the three systems.

**Figure A8.2. Dependence of $\sigma^B$ pulse amplitudes on post-translational parameters.**

**A.** $\sigma^B$ pulse amplitudes are a threshold-linear function of phosphatase concentration. The function is characterized by a phosphatase threshold (gray vertical line) and slope.

**B.** Effect of changes in the ratio of the phosphatase and kinase rate constants ($k_p/k_k$) on the threshold-linear dependence of $\sigma^B$ pulse amplitudes on phosphatase concentrations.

**C.** Changes in the binding rate constant ($k_b$) control the phosphatase threshold of the threshold-linear dependence of $\sigma^B$ pulse amplitudes on phosphatase concentrations.
Figure A8.3. Pulsatile response of the $\sigma^B$ network encodes stochastic phosphatase burst sizes not burst frequency.

**A-C.** Simulation results for the response of the $\sigma^B$ network model to stochastic fluctuations in stress-sensing phosphatase RsbQP (QP) levels for fixed mean burst size and varying burst frequency. Mean $\sigma^B$ pulse amplitudes (A) do not increase as a function of mean phosphatase level when this increase is driven by changes in burst frequency. Green dots and errorbars show mean and standard deviations calculated from stochastic simulations. Black line shows a linear fit. $\sigma^B$ pulse frequency (B) increases linearly as a function of mean phosphatase level. Green dots show the mean pulse frequency calculated from stochastic simulations. Black line shows a linear fit. Mean $\sigma^B$ target expression (C) increases non-linearly as a function of mean phosphatase level. Green dots show the mean pulse frequency calculated from stochastic simulations. Black line shows a Hill-equation fit.

**D-F.** Pulse amplitude cumulative histograms for stochastic simulations with (D) burst frequency modulation, (E) burst size modulation and (F) experimental data taken from (103). Different colors represent varying levels of mean phosphatase in the model or mycophenolic acid (MPA, energy stress) in experiments.
Figure A8.4. Sensitivity of the $\sigma^B$ target expression to $\sigma^A$ and competition for RNA polymerase.

A. Steady state dependence of the concentration of $B_{\text{target}}$, a $\sigma^B$ target, on the level of free $\sigma^B$ for different levels of the housekeeping sigma factor $\sigma^B_{\text{tot}}$.

B. $K_{\text{sigB}}$, the half-maximal constant of the dependence of target expression on $\sigma^B$ as a function of the concentration of the housekeeping sigma factor $\sigma^B$.

Figure A8.5. Pulsatile response of the multiple stress sigma factor model.

A. Decoupled post-translational and transcriptional components of the simplified model for the competition of stress sigma factors. Cyan and blue lines show the post-translational response at low and high $\sigma^B$ phosphatase respectively. Black line shows the and the transcriptional responses. A step increase in $\sigma^B$ phosphatase $P_B$ that
activates causes a shift in the steady-state post-translational response (from low phosphatase-cyan to high phosphatase-blue) and leads to a pulsatile $\sigma^B$ response trajectory (green curve). The $\sigma^W$ phosphatase concentration, $P_w$ was kept fixed at 0.1$\mu$M.

B-D. Time-course representations of the green trajectories described in (A) showing $\sigma^B$ (B), $\sigma^B$ promoter activity (B) and $B_T$ (D) respectively.

E-H. Same as (A-D) but with $\sigma^W$ phosphatase concentration, $P_w$ kept fixed at 2$\mu$M. Note that the higher $P_w$ in (E-H) leads to a higher $\sigma^B$ pulse amplitude which compensates for RNA polymerase competition and as a result $\sigma^B$ promoter activity and $B_T$ responses are unaffected.

A8.2. Supplementary Tables

Table A8.1. List of parameters values used in the model for $\sigma^B$ network

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{bw}$</td>
<td>72 $\mu$M$^{-1}$hr$^{-1}$</td>
</tr>
<tr>
<td>$k_{dw}$</td>
<td>18 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_{b1}$, $k_{b2}$, $k_{b3}$, $k_{b5}$, $k_{bb}$, $k_{ba}$, $k_{bpb}$</td>
<td>144 $\mu$M$^{-1}$hr$^{-1}$</td>
</tr>
<tr>
<td>$k_{d1}$, $k_{d2}$, $k_{d3}$, $k_{d5}$</td>
<td>18 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_{b4}$</td>
<td>7.2 $\mu$M$^{-1}$hr$^{-1}$</td>
</tr>
<tr>
<td>$k_{d4}$</td>
<td>7.2 $\mu$M$^{-1}$hr$^{-1}$</td>
</tr>
<tr>
<td>$k_{db}$</td>
<td>172.8 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_{da}$</td>
<td>2.88 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_{dpb}$</td>
<td>14.4 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_{k1}$, $k_{k2}$</td>
<td>36 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_{p}$</td>
<td>180 hr$^{-1}$</td>
</tr>
<tr>
<td>$k_{deg}$</td>
<td>0.72 hr$^{-1}$</td>
</tr>
<tr>
<td>$v_0$</td>
<td>0.4 $\mu$Mhr$^{-1}$</td>
</tr>
<tr>
<td>$f$</td>
<td>30</td>
</tr>
<tr>
<td>$K$</td>
<td>0.2 $\mu$M</td>
</tr>
<tr>
<td>$\lambda_W$</td>
<td>4</td>
</tr>
<tr>
<td>$\lambda_V$</td>
<td>4.5</td>
</tr>
<tr>
<td>$[p_3]_{tot}$</td>
<td>0.05 $\mu$M</td>
</tr>
<tr>
<td>RNApol$_tot$</td>
<td>10 $\mu$M</td>
</tr>
</tbody>
</table>
A8.3. Supplementary Methods

A8.3.1. Mathematical model of $\sigma^B$ stress-response network

We extend a previous mathematical model of $\sigma^B$ network (382) to determine the mechanism responsible for pulsatile activation of $\sigma^B$. Below we formulate the set of reactions and associated differential equations.

*Model reactions*

The events shown in Figure 9.1A can be described by the following set of biochemical reactions:

**Dimerization of anti-sigma factor RsbW**

$$2\text{RsbW} \xrightleftharpoons[k_{d2}]{k_{21}} \text{RsbW}_2$$  \hspace{1cm} (87)

**Reversible binding of the anti-anti-sigma factor RsbV to anti-sigma factor dimer RsbW$_2$ to form the complexes RsbW$_2$-RsbV and RsbW$_2$-RsbV$_2$**

$$\text{RsbW}_2 + \text{RsbV} \xrightleftharpoons[k_{d1}]{k_{21}} \text{RsbW}_2 - \text{RsbV}$$  \hspace{1cm} (88)

$$\text{RsbW}_2 - \text{RsbV} + \text{RsbV} \xrightleftharpoons[k_{d2}]{k_{22}} \text{RsbW}_2 - \text{RsbV}_2$$  \hspace{1cm} (89)

**Phosphorylation of the anti-anti-sigma factor RsbV by RsbW$_2$**

$$\text{RsbW}_2 - \text{RsbV} \xrightarrow{k_{s1}} \text{RsbW}_2 + \text{RsbV}_p$$  \hspace{1cm} (90)

$$\text{RsbW}_2 - \text{RsbV}_2 \xrightarrow{k_{s2}} \text{RsbW}_2 - \text{RsbV} + \text{RsbV}_p$$  \hspace{1cm} (91)

**Reversible binding of $\sigma^B$ to RsbW$_2$ to form the complex RsbW$_2$-$\sigma^B$**

$$\text{RsbW}_2 + \sigma^B \xrightleftharpoons[k_{s3}]{k_{d3}} \text{RsbW}_2 - \sigma^B$$  \hspace{1cm} (92)

**Reversible displacement of $\sigma^B$ by RsbV in the complex RsbW$_2$**

$$\text{RsbW}_2 - \sigma^B + \text{RsbV} \xrightleftharpoons[k_{d4}]{k_{44}} \text{RsbW}_2 - \text{RsbV} + \sigma^B$$  \hspace{1cm} (93)

**Dephosphorylation of phosphorylated anti-anti-sigma factor RsbV$_p$**

$$\text{RsbV}_p + \text{RsbP} \xrightleftharpoons[k_{d5}]{k_{55}} \text{RsbV}_p - \text{RsbP} \xrightarrow{k_p} \text{RsbV} + \text{RsbP}$$  \hspace{1cm} (94)
Protein degradation/dilution due to cell growth

\[ \text{deg} \xrightarrow{k_{\text{aug}}} X \rightarrow 0 \]  

(95)

where \( X \) is any protein or protein complex in the \( \sigma^B \) network.

Production of \( \sigma^B \), RsbW and RsbV

\[ \nu_B \rightarrow \sigma^B + \lambda_W \text{RsbW} + \lambda_V \text{RsbV} \]  

(96)

\( \sigma^B \), RsbW and RsbV were assumed to be synthesized proportionally as all three are part of the same operon. Synthesis was modeled using Hill equation which includes the effect of positive feedback due to \( \sigma^B \) autoregulation: \( \nu_B = \nu_0 (1 + f \sigma^B/(K + \sigma^B)) \). \( \nu_0 \) is the basal synthesis rate, \( f \) is the fold change in protein synthesis due to positive autoregulation and \( K \) is the equilibrium dissociation constant for the binding of \( \sigma^B \) to DNA. \( \lambda_W \) and \( \lambda_V \) are the proportionality constants denoting differential synthesis of operon genes.

**Model equations**

We assume mass-action kinetics for all the above reactions (equations 1-10) to obtain the following set of equations that describe network dynamics:
\[
\begin{align*}
\frac{d[B_T]}{dt} &= v_B - k_{\text{deg}}[B_T] \\
\frac{d[W_T]}{dt} &= \lambda_W v_B - k_{\text{deg}}[W_T] \\
\frac{d[V_T]}{dt} &= \lambda_V v_B - k_{\text{deg}}[V_T] \\
\frac{d[\sigma^B]}{dt} &= v_B - k_{b3}[W_2][\sigma^B] + k_{d3}[W_2\sigma^B] + k_{p4}[W_2\sigma^B][V] - k_{d4}[W_2V][\sigma^B] - k_{\text{deg}}[\sigma^B] \\
\frac{d[W_2]}{dt} &= k_d[W]^2 + (k_{d1} + k_{k1})[W_2][V] + k_{d3}[W_2\sigma^B] - (k_{b1}[V] + k_{b3}[\sigma^B] + k_{\text{deg}})[W_2] \\
\frac{d[V_P]}{dt} &= k_{k1}[W_2][V] + k_{k2}[W_2V_2] + k_{d6}[V_P] - (k_{b5}[P] + k_{\text{deg}})[V_P] \\
\frac{d[W_2V]}{dt} &= k_{b1}[W_2][V] + (k_{d2} + k_{k2})[W_2V_2] + k_{p4}[W_2\sigma^B][V] - (k_{d4} + k_{k1} + k_{b2}[V] + k_{d4}[\sigma^B] + k_{\text{deg}})[W_2V] \\
\frac{d[W_2V_2]}{dt} &= k_{b2}[W_2][V] - (k_{d2} + k_{k2} + k_{\text{deg}})[W_2V_2] \\
\frac{d[V_PP]}{dt} &= k_{b5}[V_P][P] - (k_{d5} + k_{p} + k_{\text{deg}})[V_PP] \\
[B_T] &= v_B - k_{\text{deg}}[B_T] \\
[W_T] &= \lambda_W v_B - k_{\text{deg}}[W_T] \\
[V_T] &= \lambda_V v_B - k_{\text{deg}}[V_T] \\
[W_2\sigma^B] &= [W_2][\sigma^B] \\
[W_2V] &= [W_2][V] \\
[W_2V_2] &= [W_2][V_2] \\
[V_P] &= [V_P][P] \\
[P_P] &= [P_P][P] \\
\end{align*}
\]

where \([B_T], [W_T], [V_T]\) and \([P_T]\) are the concentrations of total \(\sigma^B\), RsbW, RsbV and phosphatase; \(\sigma^B\) is the concentration of free \(\sigma^B\), \([W], [W_2]\) and are the concentrations of monomeric and dimeric RsbW; \([V]\) and \([V_P]\) are the concentrations of unphosphorylated and phosphorylated RsbV; \([W_2\sigma^B], [W_2V], [W_2V_2]\) and \([V_P]\) are the concentrations of complexes.

**Derivation of steady state asymptotes for the \(\sigma^B\) post-translational response**

To understand the steady state post-translational response of \(\sigma^B\) network (Fig. 9.1D-G) we used the mass balance for the operon components RsbW, RsbV and \(\sigma^B\) together with the phosphate flux balance to derive approximate
dependence of free $\sigma^B$ on $B_T$. We found that the post-translational response of the network varies depending on whether the concentration of operon components is lower or higher than a threshold level defined by the concentration of the stress phosphatase $P_T$.

For low $B_T$ ($B_T < 2P_Tk_p/k_w/\lambda_W$), the RsbW kinase is dominated by the phosphatase and as a result, $V_P \approx 0$. In addition most of the anti-anti-sigma factor $V$ is in the $W_2V_2$ complex. Taking this into account and applying the mass balance for RsbV,

$$V_T = V + V_P + V_pP + W_2V + 2W_2V_2 \approx V_pP + 2W_2V_2$$

Next applying the balance for kinase and phosphatase fluxes,

$$k_k(W_2V + W_2V_2) = k_{deg}V_P + (k_p+k_{deg})V_P$$

or, $k_kW_2V_2 \approx (k_p+k_{deg})V_P$

$$V_T = (2 + k_k/(k_p+k_{deg}))W_2V_2$$

Using the above equation in the mass balance for $W$ we solved for $W_2B$ and thereby $\sigma^B$,

$$2W_2B = W_T - 2W_2V_2 = W_T - 2V_T/(2+k_k/(k_p+k_{deg}))$$

$$B_T = \sigma^B + W_2B = \sigma^B + W_T/2 - V_T/(2+k_k/(k_p+k_{deg}))$$

$$\sigma^B = B_T + V_T/(2+k_k/(k_p+k_{deg})) - W_T/2 = B_T(1 + \lambda_W/(2+k_k/(k_p+k_{deg})) - \lambda_W/2)$$

$$\approx B_T(1 + \lambda_W/2 - \lambda_W/2)$$

In contrast, for higher $B_T$ ($B_T > 2P_Tk_p/k_w/\lambda_W$), where the RsbW kinase dominates the phosphatase, $V_P$ is not negligible and the phosphatase is saturated ($V_P \approx P_T$). Again using this in the mass balance for RsbV,

$$V_T = V + V_P + V_pP + W_2V + 2W_2V_2 \approx V_P + V_P + 2W_2V_2$$
And applying the balance for kinase and phosphatase fluxes,

\[ k_k W_2 V_2 = k_{\text{deg}} V P + (k_{\text{deg}} + k_p)V P \approx k_{\text{deg}} V P + k_{\text{deg}}(k_p + k_p) P_T \]

\[ V_T = (k_k W_2 V_2 - (k_{\text{deg}} + k_p)P_T)/k_{\text{deg}} + P_T + 2W_2 V_2 \]

\[ W_2 V_2 = (V_T + P_T k_p/k_{\text{deg}})/(2 + k_v/k_{\text{deg}}) \approx V_T k_{\text{deg}}/k_k + P_T k_p/k_k \]

\[ 2W_2 B = W_T - 2W_2 V_2 = W_T - 2V_T k_{\text{deg}}/k_k - 2P_T k_p/k_k \]

\[ B_T = \sigma^B + W_2 B = \sigma^B + W_T/2 - V_T k_{\text{deg}}/k_k - P_T k_p/k_k \]

\[ \sigma^B = B_T + V_T k_{\text{deg}}/k_k + P_T k_p/k_k - W_T \]

\[ \sigma^B = B_T(1 + \lambda_v k_{\text{deg}}/k_k - \lambda_W/2) + P_T k_p/k_k = B_T(1 + \lambda_v k_{\text{deg}}/k_k - \lambda_W/2) + B_0(P_T) \]

where \( B_0(P_T) = P_T k_p/k_k \)

Taken together the dependence of \( \sigma^B \) on \( B_T \) can be described by the following system of equations:

\[
\sigma^B = \begin{cases} 
B_T (1 + \lambda_v/2 - \lambda_W/2) & B_T < 2P_T k_p/k_k / \lambda_W \\
B_T \left(1 + \lambda_v k_{\text{deg}}/k_k - \lambda_W/2\right) + B_0(P_T) & B_T \geq 2P_T k_p/k_k / \lambda_W 
\end{cases}
\]

Based on the above equation, the sensitivity of the \( \sigma^B \) post-translational response depends on \( (\lambda_W, \lambda_v) \), i.e. the stoichiometry of operon components:

\[
\frac{\partial \sigma^B}{\partial B_T} = \begin{cases} 
(1 + \lambda_v/2 - \lambda_W/2) & B_T < 2P_T k_p/k_k / \lambda_W \\
(1 + \lambda_v k_{\text{deg}}/k_k - \lambda_W/2) & B_T \geq 2P_T k_p/k_k / \lambda_W 
\end{cases}
\]

Which shows that the \( (\lambda_W, \lambda_v) \) parameter space can be divided into three regions based on qualitative differences in the post-translational response.

**Region I** \( \lambda_W < 2 + \lambda_v k_{\text{deg}}/k_k \): \( d\log(\sigma^B)/d\log(B_T) > 0 \) and free \( \sigma^B \) increases as a function of \( B_T \) irrespective of \( P_T \).
Region II \((2+\lambda_W k_{\text{deg}} / k_k < \lambda_W < 2+\lambda_V)\): \(\frac{d\log(\sigma^B)}{d\log(B_T)} > 0\) for \(B_T < 2P_T k_p / k_k / \lambda_W\) and \(\frac{d\log(\sigma^B)}{d\log(B_T)} < 0\) for \(B_T > 2P_T k_p / k_k / \lambda_W\). Thus free \(\sigma^B\) is a non-monotonic function of \(B_T\).

Region III \((\lambda_W > 2+\lambda_V)\): \(\frac{d\log(\sigma^B)}{d\log(B_T)} < 0\) and free \(\sigma^B\) decreases as a function of \(B_T\) irrespective of \(P_T\).

A8.3.2. Mathematical model of competition between \(\sigma^B\) and \(\sigma^A\)

Additional reactions for the model of competition between \(\sigma^B\) and \(\sigma^A\)

To model the competition for RNA polymerase between \(\sigma^B\) and the housekeeping sigma factor \(\sigma^A\) (Figs. 9.4 and A8.4), we extended the model described above and supplemented reactions (1-9) with the following reactions:

Reversible binding of sigma factors and RNA polymerase

\[
\begin{align*}
\sigma^B + \text{RNApol} & \stackrel{k_{bb}}{\rightleftharpoons} k_{bb} \stackrel{k_{bb}}{\leftarrow} \text{RNApol} - \sigma^B \\
\sigma^A + \text{RNApol} & \stackrel{k_{aa}}{\rightleftharpoons} k_{aa} \stackrel{k_{aa}}{\leftarrow} \text{RNApol} - \sigma^A 
\end{align*}
\]

Reversible binding of RNApol-\(\sigma^B\) complexes to target promoters

\[
\text{RNApol} - \sigma^B + p_B \stackrel{k_{bb}}{\rightleftharpoons} k_{bb} \stackrel{k_{bb}}{\leftarrow} \text{RNApol} - \sigma^B - p_B
\]

Production of \(\sigma^B\), RsbW and RsbV

\[
\text{RNApol} - \sigma^B - p_B \xrightarrow{v_0} \text{RNApol} - \sigma^B + p_B + \sigma^B + \lambda_W \text{RsbW} + \lambda_V \text{RsbV}
\]

Where \(v_0\) is the basal synthesis rate and \(v_B = v_0 * f / [p_B]_{\text{tot}}\) is the maximal rate. \(f\) is the fold change in protein synthesis due to positive autoregulation and \([p_B]_{\text{tot}}\) is the total concentration of the \(\sigma^B\) promoter.
Model equations

The following set of equations that describe network dynamics of this extended model:

\[
\frac{dB_T}{dt} = v_0 + v_B - k_{deg}[B_T]
\]

\[
\frac{d[W_T]}{dt} = \lambda_W v_0 + \lambda_W v_B - k_{deg}[W_T]
\]

\[
\frac{d[V_T]}{dt} = \lambda_V v_0 + \lambda_V v_B - k_{deg}[V_T]
\]

\[
\frac{d[\sigma^B]}{dt} = v_0 + v_B - k_{b3}[W_2][\sigma^B] + k_{d3}[W_2\sigma^B] + k_{b4}[W_2\sigma^B][V] - k_{dt}[W_2V][\sigma^B] - k_{deg}[\sigma^B]
\]

\[
- k_{bb}[\sigma^B][RNApol] + k_{db}[RNApol - \sigma^B]
\]

\[
\frac{d[W_2]}{dt} = k_d[W]^2 + (k_{d1} + k_{k1})[W_2V] + k_{d3}[W_2\sigma^B] - (k_{b1}[V] + k_{b3}[\sigma^B] + k_{deg})[W_2]
\]

\[
\frac{d[V_p]}{dt} = k_{k1}[W_2V] + k_{k2}[W_2V_2] + k_{ds}[V_pP] - (k_{b5}[P] + k_{deg})[V_p]
\]

\[
\frac{d[W_2V]}{dt} = k_{b1}[W_2][V] + (k_{d2} + k_{k2})[W_2V_2] + k_{b4}[W_2\sigma^B][V]
\]

\[
- (k_{d1} + k_{k1} + k_{b2}[V] + k_{d4}[\sigma^B] + k_{deg})[W_2V]
\]

\[
\frac{d[W_2V_2]}{dt} = k_{b2}[W_2][V] - (k_{d2} + k_{k2} + k_{deg})[W_2V_2]
\]

\[
\frac{d[V_pP]}{dt} = k_{b5}[V_p][P] - (k_{d5} + k_{p} + k_{deg},)[V_pP]
\]

\[
\frac{d[RNApol - \sigma^B]}{dt} = k_{bb}[\sigma^B][RNApol] - k_{db}[RNApol - \sigma^B] - k_{bpb}[p_B][RNApol - \sigma^B]
\]

\[
+ k_{dpb}[RNApol - \sigma^B - p_B] + v_B[RNApol - \sigma^B - p_B] - k_{deg}[RNApol - \sigma^B - p_B]
\]

\[
\frac{d[RNApol - \sigma^B - p_B]}{dt} = k_{bpb}[p_B][RNApol - \sigma^B] - k_{dpb}[RNApol - \sigma^B - p_B] - k_{deg}[RNApol - \sigma^B - p_B]
\]

\[
\frac{d[RNApol - \sigma^A]}{dt} = k_{ba}[\sigma^A][RNApol] - k_{da}[RNApol - \sigma^A] - k_{deg}[RNApol - \sigma^A]
\]
Model equations for the model of competition between $\sigma^B$, $\sigma^W$ and $\sigma^A$

To model the competition for RNA polymerase between $\sigma^B$ the housekeeping sigma factor $\sigma^A$ and the alkaline stress response sigma factor (Figs. 9.5 and A8.5), we simplified the model for the post-translational control of stress sigma factors while explicitly including reactions for the binding/unbinding of RNA polymerase, sigma factors and target promoters. This model included the following set of equations:

\[
[W_2\sigma^B] = [B_T] - [\sigma^B] - [\text{RNApol} - \sigma^B] - [\text{RNApol} - \sigma^B - p_b] \\
[V] = [V_T] - [W_2V] - 2[W_2V_2] - [V_p] - [V_pP] \\
[P] = [P_T] - [V_pP] \\
[p_b] = [p_b]_T - [\text{RNApol} - \sigma^B - p_b] \\
[\sigma^A] = [\sigma^A]_T - [\text{RNApol} - \sigma^A] \\
[\text{RNApol}] = [\text{RNApol}_T] - [\text{RNApol} - \sigma^B] - [\text{RNApol} - \sigma^B - p_b] - [\text{RNApol} - \sigma^A]
\]
\[
\begin{align*}
\frac{d[B_T]}{dt} &= v_{B0} + v_B - k_{\text{deg}}[B_T] \\
\frac{d[W_T]}{dt} &= v_{W0} + v_W - k_{\text{deg}}[W_T] \\
\frac{d[\sigma^B]_{\text{free}}}{dt} &= v_{B0} + v_B - k_{\text{deg}} \left(1 + \frac{([B_T]/K_{\beta})_{\text{mb}}}{[P_B]_{\text{mb}}}ight) [\sigma^B]_{\text{free}} \\
\frac{d[\sigma^W]_{\text{free}}}{dt} &= v_{W0} + v_W - k_{\text{deg}} \left(1 + \frac{([W_T]/K_{\gamma})_{\text{nw}}}{[P_W]_{\text{nw}}}ight) [\sigma^W]_{\text{free}} \\
\frac{d[\text{RNApol}-\sigma^B]}{dt} &= k_{\beta} [\sigma^B][\text{RNApol}] - k_{\beta_{\text{bb}}} [\text{RNApol}-\sigma^B] - k_{\beta_{\text{dpb}}} [\text{RNApol}-\sigma^B] - k_{\beta_{\text{dpb}}} [\text{RNApol}-\sigma^B] + v_B [\text{RNApol}-\sigma^B] - k_{\text{deg}} [\text{RNApol}-\sigma^B - p_B] \\
\frac{d[\text{RNApol}-\sigma^B - p_B]}{dt} &= k_{\beta_{\text{dpb}}} [\text{RNApol}-\sigma^B - p_B] + v_B [\text{RNApol}-\sigma^B - p_B] - k_{\text{deg}} [\text{RNApol}-\sigma^B - p_B] \\
\frac{d[\text{RNApol}-\sigma^W]}{dt} &= k_{\omega_{\beta}} [\sigma^W][\text{RNApol}] - k_{\omega_{\text{bw}}} [\text{RNApol}-\sigma^W] - k_{\beta_{\omega}} [\text{RNApol}-\sigma^W] + v_W [\text{RNApol}-\sigma^W] - k_{\text{deg}} [\text{RNApol}-\sigma^W - p_W] \\
\frac{d[\text{RNApol}-\sigma^W - p_W]}{dt} &= k_{\beta_{\omega}} [p_W] [\text{RNApol}-\sigma^W] - k_{\omega_{\text{dpw}}} [\text{RNApol}-\sigma^W - p_W] - k_{\omega_{\text{dpw}}} [\text{RNApol}-\sigma^W - p_W] - k_{\text{deg}} [\text{RNApol}-\sigma^W - p_W] \\
\frac{d[\text{RNApol}-\sigma^A]}{dt} &= k_{\alpha_{\beta}} [\sigma^A][\text{RNApol}] - k_{\alpha_{\text{ba}}} [\text{RNApol}-\sigma^A] - k_{\text{deg}} [\text{RNApol}-\sigma^A] \\

[\sigma^B] &= [\sigma^B]_{\text{free}} - [\text{RNApol}-\sigma^B] - [\text{RNApol}-\sigma^B - p_B] \\
[\sigma^W] &= [\sigma^W]_{\text{free}} - [\text{RNApol}-\sigma^W] - [\text{RNApol}-\sigma^W - p_W] \\
[p_B] &= [p_B]_{\text{T}} - [\text{RNApol}-\sigma^B - p_B] \\
[p_W] &= [p_W]_{\text{T}} - [\text{RNApol}-\sigma^W - p_W] \\
[\sigma^A] &= [\sigma^A]_{\text{T}} - [\text{RNApol}-\sigma^A] \\
[\text{RNApol}] &= [\text{RNApol}_{\text{T}}] - [\text{RNApol}-\sigma^B] - [\text{RNApol}-\sigma^B - p_B] \\
&\quad - [\text{RNApol}-\sigma^W] - [\text{RNApol}-\sigma^W - p_W] - [\text{RNApol}-\sigma^A] \\

\textbf{A8.3.3. Simulations}

The parameter values for reversible binding and phosphorylation reactions were taken from (382) or were analysis driven to obtain pulsing in \(\sigma^B\). All the parameters used in the model are summarized in Table A8.1. In the deterministic
set-up (Figs. 9.1B-G, 9.3, 9.5, A8.1, A8.2, A8.4 and A8.5) the system of
differential equations was solved using standard ode15s solver in MATLAB. For
stochastic simulations in Figs. 9.2, 9.4 and A8.3, the time-varying total
phosphatase level \( P_T = P + V_P \) was pre-computed using a gamma distributed
Ornstein-Uhlenbeck process as in (103). This gamma distributed Ornstein-
Uhlenbeck process permits independent modulation of mean burst size \( (b) \) and
frequency \( (a) \) (401). For each phosphatase level, 100 simulations were
performed each lasting 10000 minutes. Pulses were detected by examining local
maxima and minima of the simulated trajectories, and subsequently this
information was used to compute pulse statistics namely duration, amplitude and
frequency. The pulse amplitude and duration were obtained by calculating the
pulse height and width respectively.

For the simulations of the effect of competition for RNA polymerase (Figs.
9.4 and A8.4), the total housekeeping sigma factor concentration was varied
between 5 and 15 µM. In these simulations we used \( (\lambda_W=4, \lambda_V=4.5) \) and \( (\lambda_W=2, \lambda_V=2) \) to simulate the wildtype (negative feedback) and positive feedback
networks respectively. For the simulations of the no feedback network we used
\( (\lambda_W=4, \lambda_V=4.5) \) and \( f=0 \) and \( v_0=8.64 \, \text{µM/hr}^{-1} \) to model the \( \sigma^B \)–independent
constitutive production of operon components.

For the simulations of the competition between \( \sigma^B, \sigma^W \) and \( \sigma^A \) (Figs. 9.5
and A8.5), the total housekeeping sigma factor concentration was kept fixed at
12 µM. We used \( (n_b=7, m_b=5) \) and \( (n_b=0, m_b=3) \) to simulate the wildtype
(negative feedback) and positive feedback networks respectively. \( K_B \) and \( K_W \)
were fixed at 5µM for simulations of both networks.
References


131. Fujita M & Losick R (2005) Evidence that entry into sporulation in Bacillus subtilis is governed by a gradual increase in the level and activity of the master regulator Spo0A. *Genes Dev* 19(18):2236-2244.


259. Shen R, et al. (2006) Smad6 interacts with Runx2 and mediates Smad ubiquitin regulatory factor 1-induced Runx2 degradation. *J Biol Chem* 281(6):3569-3576.


418.


408.


409.


410.


411.


412.


413.


414.


415.


416.


417.


418.


419.


420.


421.


422.

Chaba R, Grigorova IL, Flynn JM, Baker TA, & Gross CA (2007) Design principles of the proteolytic cascade governing the sigmaE-mediated


