HHS Public Access

Author manuscript

Proc IEEE Conf Decis Control. Author manuscript; available in PMC 2020 March 09.

Published in final edited form as:

Proc IEEE Conf Decis Control. 2018 December ; 2018: 1886–1892. doi:10.1109/cdc.2018.8619300.

Stochastic multistationarity in a model of the hematopoietic stem cell differentiation network

M. Ali Al-Radhawi[1,2], **Nithin S. Kumar**[1], **Eduardo D. Sontag**[2], **Domitilla Del Vecchio**[1]

[1]Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139.

[2]Department of Electrical and Computer Engineering and Department of Bioengineering, Northeastern University, 805 Columbus Ave, Boston, MA 02115, USA.

Abstract

A central issue in the analysis of multi-stable systems is that of controlling the relative size of the basins of attraction of alternative states through suitable choices of system parameters. We are interested here mainly in the stochastic version of this problem, that of shaping the stationary probability distribution of a Markov chain so that various alternative modes become more likely than others.

Although many of our results are more general, we were motivated by an important biological question, that of cell differentiation. In the mathematical modeling of cell differentiation, it is common to think of internal states of cells (quanfitied by activation levels of certain genes) as determining the different cell types. Specifically, we study here the “PU.1/GATA-1 circuit” which is involved in the control of the development of mature blood cells from hematopoietic stem cells (HSCs). All mature, specialized blood cells have been shown to be derived from multipotent HSCs.

Our first contribution is to introduce a rigorous chemical reaction network model of the PU.1/GATA-1 circuit, which incorporates current biological knowledge. We then find that the resulting ODE model of these biomolecular reactions is incapable of exhibiting multistability, contradicting the fact that differentiation networks have, by definition, alternative stable steady states. When considering instead the stochastic version of this chemical network, we analytically construct the stationary distribution, and are able to show that this distribution is indeed capable of admitting a multiplicity of modes. Finally, we study how a judicious choice of system parameters serves to bias the probabilities towards different stationary states. We remark that certain changes in system parameters can be physically implemented by a biological feedback mechanism; tuning this feedback gives extra degrees of freedom that allow one to assign higher likelihood to some cell types over others.

malirdwi@mit.edu.

I. INTRODUCTION

In cell-fate gene regulatory networks (GRNs), attractors are typically associated with biological phenotypes [1], [2]. Hence, a great issue of interest in the theory of multi-stable GRNs is that of shaping the relative size of the basins of attraction of the multiple attractors by suitable choices of system parameters, including constant inputs, as well as the use of time-varying inputs to drive the state from one region of attraction to another. For stochastic systems, the first of these questions can be translated into the shaping of the stationary probability distribution of an associated Markov chain. Also for stochastic systems, variations of parameters (which can be adjusted by means of appropriate feedback mechanisms, as discussed later) can be used in order to reshape this distribution, so that the time evolution of the distribution of states of this Markov chain will converge to the desired landscape. In this paper, we present a case study of this issue in an important biological context, that of differentiation of stem cells into different types of blood cells. We introduce a biochemical model that represents known biological data, and then remark that the deterministic model of this network does not exhibit multistationarity, which contradicts experimental data and hence suggests that one cannot ignore randomness in this model. We then proceed to analyze this model from the point of view of stochastic multistationarity and show how to shape the stationary distribution via controlling the parameters.

Deterministic models are usually justified under the assumptions of sufficiently large volume and sufficiently large number of molecules [3], or, under some conditions such as fast promoter kinetics [4], as will be discussed in the text. In such cases, an ODE model captures the system's dynamics, and it produces a similar qualitative behaviour to the one produced by the stochastic model. However, these assumptions are not usually satisfied in practice due to the fact that cell-fate GRNs have usually very low gene copy numbers. For instance, an investigation of the stochastic model of the toggle switch [5] has shown an increase in the number of admissible phenotypes. Furthermore, endogenous GRNs in eukaryotic cells can have the binding/unbinding rate of transcription factors (TFs) to promoters occurring at a slower rate than transcription and translation due to the complicating effects of chromatin structure and regulation [4],[6], [7]. Therefore, the qualitative behaviour produced by deterministic models can be erroneous.

The specific network studied in this paper is the PU.1-GATA-1 GRN, which is involved in lineage determination in hematopoietic stem cells (HSCs) [8], [9]. All mature, specialized blood cells have been shown to be derived from multipotent HSCs [10]-[11]. The classical hierarchical tree-like model of blood cells (hematopoiesis) describes the differentiation of HSCs into progenitor cells which have the ability to further differentiate into committed cells [12], [13], [14]. The decision to commit to a particular lineage is thought to depend on the relative expression levels of certain TFs [15], [16], [17]. Here we consider two particular TFs, PU.1 and GATA-1, which self-activate and interact antagonistically [8], [9], [18] as pictorially depicted in Fig. 1. These two TFs are thought to be involved in reinforcing the lineage commitment of a differentiating HSC, typically at the Common Myeloid Progenitor (CMP) state, to either the Megakaryocyte-Erythroid Progenitor (MEP) or the Granulocyte-Macrophage Progenitor (GMP). Commitment to the MEP or GMP phenotype results in the differentiation to the erythroid or myeloid/lymphoid lineages, respectively [19]. The

expression level of GATA-1 has been shown to increase down the lineage from HSC to MEP and the TF is a key regulator of erythroid genes [20], [21]. Similarly, PU.1 expression increases as cells differentiate to the myeloid/lymphoid lineage [19] and has been shown to be critical for myeloid cell regulation. Early progenitor cells around the CMP stage have been shown to have relatively low levels of GATA-1 and PU.1 compared to more differentiated cells [21], [22]. The PU.1-GATA-1 network has been extensively studied since the early 2000s [8], [9], [18]. Previous models of the PU.1-GATA-1 GRN in the literature make assumptions on the biological interaction between the TFs to explain the biologically expected bistable property of this GRN, in which one stable state corresponds to high levels of GATA-1 and low levels of PU.1 and vice-versa. However, some of these assumptions have not been experimentally validated. Specifically, the studies in [23] and [24] assume high cooperativity of TFs (n=2 and 4, respectively), though mutual repression and self-activation have been shown to occur primarily in their monomeric form [25], [26], [8]. The model presented in [27] introduces a gene X that transcriptionally represses PU.1 and is activated by GATA-1, which has not been experimentally identified. The model studied in [28] assumes both TFs can directly bind to each others' promoters and transcriptionally repress each other but this has not been shown. In this paper, we consider a set of biomolecular reactions for the system, in which none of these assumptions are made. Interestingly, we mathematically demonstrate that the corresponding ODE model is monostable, which does not agree with the fact that the network should be capable of exhibiting two phenotypes, high GATA-1, low PU.1 (MEP) and low GATA-1, high PU.1 (GMP). We therefore use the same biomolecular reactions to construct the CME model under the assumptions of low gene copy number and slow promoter kinetics and analytically demonstrate that the resulting stationary distribution can have multiple modes, each possibly corresponding to a cell phenotype, including the MEP and GMP phenotypes.

This paper is organized as follows. In Section II, we present a system of biochemical reactions that describe the PU.1-GATA-1 GRN. In Section III, we derive a deterministic ODE model of this GRN and demonstrate that this system is unable to demonstrate bistability. In Section IV, a stochastic model with slow promoter kinetics is presented. Section V presents a numerical example of how one would use our techniques in order to shape the distribution. The conclusion is given in section VI. The appendix lists the biological modeling assumptions that have lead to the reaction model.

II. MODEL OF REACTIONS

The reactions that describe this model are given below. The biological justification for the model is included in the appendix. Here p_0 (shown in Fig. 2(a)) and g_0 are the unbound promoters of TFs PU.1 (P) and GATA-1 (G):

$$p_0 + P \underset{d_0}{\overset{a_0}{\rightleftharpoons}} p_1, g_0 + G \underset{d_0'}{\overset{a_0'}{\rightleftharpoons}} g_1, \tag{1}$$

$$p_1 + G \underset{d_1'}{\overset{a_1'}{\rightleftharpoons}} p_2, g_1 + P \underset{d_1}{\overset{a_1}{\rightleftharpoons}} g_2, \tag{2}$$

$$p_0 \xrightarrow{\alpha_{p0}} p_0 + P, g_0 \xrightarrow{\alpha_{g0}} g_0 + G, \tag{3}$$

$$p_1 \xrightarrow{\alpha_{p1}} p_1 + P, g_1 \xrightarrow{\alpha_{g1}} g_1 + G, \tag{4}$$

$$P \xrightarrow{\delta_P} \varnothing, G \xrightarrow{\delta_G} \varnothing . \tag{5}$$

The reversible binding reactions between the unbound promoters (p_0 and g_0) and their TFs (P and G) to form complexes p_1 (shown in Fig. 2(b)) and g_1, respectively are given by (1). Reactions (2) describe the formation of complexes p_2 and g_2 by reversible binding of G(P) with $p_1(g_1)$, respectively. These complexes represent the "off" state of the promoter wherein the gene is silenced as shown in Fig. 2(c) for p_2. Reactions (3) describe the leaky promoter one-step production of P and G with rates α_{p0} and α_{g0}, respectively. P(G) is produced at rate $\alpha_{p1}(\alpha_{g1})$ when it is bound to its promoter as shown in reactions (4). Lastly, reactions (5) describe the decay of transcription factors. Since the genes are self-activating, we have $\alpha_{p0} < \alpha_{p1}$ and $\alpha_{g0} < \alpha_{g1}$. To simplify the model, we assume that there is no expression from the repressed p_2 and g_2 configurations.

III. DETERMINISTIC MODEL

The most common model for reaction networks is a deterministic model [29]. It assigns to each species a state variable corresponding to its concentration. The time-evolution of species' concentrations is given by an ordinary differential equation of the form:

$$\dot{x} = \Gamma R(x), \tag{6}$$

where $x \in \mathbb{R}^n$ is the concentration vector. Γ is called the stoichiometry matrix of the network, and $R(x) = [R_1(x), \ldots, R_m(x)]^T$ is the reaction rate function which are to be defined below. Assume that the i^{th} reaction is given as:

$$\sum_{i=1}^{n} \alpha_{ij} X_i \rightarrow \sum_{i=1}^{n} \beta_{ij} X_i,$$

where α_{ij}, β_{ij} are the stoichiometric coefficients. Then the corresponding *stoichiometry vector* is: $\gamma_j = [\beta_{1j} - \alpha_{1j}, \ldots, \beta_{nj} - \alpha_{nj}]^T$, and the stoichiometry matrix is given as

$$\Gamma := [\gamma_1, \ldots, \gamma_m] . \tag{7}$$

A conservation law for a reaction network is a positive vector $v \in \mathbb{R}^n_{\geq 0}$ such that $v^T \Gamma = 0$. Multiplying both sides of (6) by v^T and integrating yields $\sum_{i=1}^n x_i(t) \equiv M$ where M is the conserved quantity.

The most commonly used form for reaction rate function is the *Mass-Action kinetics* [29], which is given as follows:

$$R_j(x) = k_j \prod_{i=1}^{n} x_i^{\alpha_{ij}}, \tag{8}$$

where k_j is the kinetic constant.

Setting the derivatives to zero for the mass-action ODE (6) and performing algebraic manipulations, we obtain a quintic equation. It is not easy to show directly that this equation has a unique solution for all possible choices of kinetic parameters, however. Thus, in order to determine the number of positive equilibria, we use the advanced deficiency algorithm developed in [30], [31],[32], and implemented in the “Chemical Reaction Network” toolbox [33]. When the algorithm is applied to our network (1)–(10), it shows that it cannot admit multiple positive equilibria for any combination of kinetic parameters. Hence, the deterministic model cannot explain the bistable behaviour observed experimentally [19], [20], [21].

IV. STOCHASTIC MODELING WITH SLOW PROMOTER KINETICS

The validity of the deterministic model rests on high copy numbers of the species in the GRN. However, this assumption is not usually satisfied in differentiation networks where TFs are expressed through genes located on the chromosome, which has one or two copies only. Nevertheless, the issue of low gene copy numbers has been overlooked in the literature by assuming high protein copy numbers and fast promoter kinetics [4], i.e. the binding and unbinding of TFs to the genes are assumed to be much faster than protein production and decay. The underlying justification for this approximation is that the fast promoter kinetics will “smooth out” the discrete effects of low gene copy numbers. However, this simplifying assumption does not generally hold in eukaryotic cells, which have more complex transcription machinery [4]. In such cells, transcriptional regulation is often mediated by an additional regulation layer dictated by DNA methylation and histone modifications, commonly referred to as chromatin dynamics. For example, the presence of nucleosomes makes binding sites less accessible to TFs and therefore TF-gene binding/unbinding is modulated by the stochastic process of chromatin opening [34], [35], [36]. Several experiments have confirmed the role of the aforementioned complex transcription processes in slow promoter kinetics [37], [38], [39], [40].

Since it is difficult to characterize the stationary distribution of the Chemical Master Equation (CME) [29] in general, our aim is to investigate the effect of slow promoter kinetics on the steady-state landscape of the network, and whether it can lead to the emergence of more than one phenotype in contrast to the unique one predicted by the deterministic model.

Let $X(t) \in \mathbb{Z}^n_{\geq 0}$ be the vector of copy numbers of all the species in the network and let the stoichiometry matrix Γ be defined as in (7). The CME employs propensity functions assigned to each reaction, and we assume that they follow the form of Mass-Action kinetics. However, the expression of the propensity functions is slightly different at low molecule number from the reaction rate functions given (8). Nevertheless, in the special case where the stoichiometric coefficients $\alpha_{ij} \leq 1$, then (8) can be used.

Let $p_x(t) = \Pr[X(t) = x]$. Then, the CME is given by [29]:

$$\dot{p}_x(t) = \sum_{j=1}^{m} R(x - \gamma_j) p_{x-\gamma_j}(t) - R(x) p_x(t). \tag{9}$$

Let x_0, x_1, x_2 be an enumeration of $\mathbb{Z}^n_{\geq 0}$, and let

$$p(t) = \left[p_{x_0}, p_{x_1}, ..\right]^T,$$

then the CME (6) can be written as:

$$\dot{p} = \Lambda p(t), \tag{10}$$

where Λ is the infinitisimal generator of the Markov chain defined elementwise as:

$$\lambda_{x\tilde{x}} := \begin{cases} R_j(x) & \text{if } \exists j \text{ such that } \tilde{x} = x - \gamma_j \\ -\sum_{j=1}^{m} R_j(z) & \text{if } \tilde{x} = x \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Considering the network (1)–(10), we assume that there is one copy number for each gene. Hence, the conservation laws imply that $\sum_{k=1}^{3} p_k(t) = \sum_{k=1}^{3} g_k(t) = 1$. Therefore, we can replace the six stochastic processes $p_k(t)$, $g_k(t)$ with two processes p, g defined as $p(t) = i$ iff $p_i(t) = 1$, and similarly for g.

We employ the results presented in [41] to analyze the time-scale separation scenario described above. The gene reactions (1)–(4) are assumed to be much slower than the protein reactions (5)–(10). To express this separation we start by decomposing $X(t) = [D(t)^T, Y(t)^T]^T$, where $D(t) = [p(t), g(t)]^T \in \{0, 1, 2\}^2$, $Y(t) = [P(t), G(t)]^T \in \mathbb{Z}^2_{\geq 0}$.

The probability distribution vector is decomposed correspondingly as:

$$p(t) = [p_{00}(t), p_{01}(t), \ldots, p_{22}(t)]^T,$$

where $p_{ij}(t) = [p_{y0ij}(t), p_{y1ij}(t), p_{y2ij}(t), ..]^T$, $p_{yij}(t) = \Pr[Y(t) = y, p(t) = i, g(t) = j]$, and $\{y_0, y_1, ..\}$ is an enumeration of $\mathbb{Z}^2_{\geq 0}$.

Hence, using the notation above we can decompose the CME (10) as [41]:

$$\dot{p}(t) = \Lambda p(t) = (\widetilde{\Lambda} + \varepsilon\widehat{\Lambda})p(t), \tag{12}$$

where

$$\widetilde{\Lambda} = \begin{bmatrix} \Lambda_{00} & & \\ & \ddots & \\ & & \Lambda_{22} \end{bmatrix}, \text{ and } p(t) = \begin{bmatrix} p_{00}(t) \\ \vdots \\ p_{22}(t) \end{bmatrix}, \tag{13}$$

where $\varepsilon > 0$ is assumed to be small. The slow matrix $\widehat{\Lambda}$ contains the reaction rates from the gene reactions (1)–(4), while the fast matrix $\widetilde{\Lambda}$ contains the reactions corresponding to the protein reactions (1)–(5). Each of the submatrices on the diagonal representation of $\widetilde{\Lambda}$ in (13) can be interpreted as representing an infinitesimal generator for the network conditioned on a certain *gene state*, i.e Λ_{ij} is the infinitesimal generator for the Markov chain conditioned on $p(t) = i$, $g(t) = j$.

The above representation allows us to decompose the state space into weakly coupled ergodic classes. The dynamics on each class consists of uncoupled birth-death processes which are known to have a steady state Poisson distribution. Hence, the overall stationary distribution is expected to be close to a mixture of Poisson distributions as $\varepsilon \to 0$.

Using singular perturbation techniques, the following theorem can be stated.

Theorem 1 ([41]): Given the network (1)–(10) and the CME (10), assume that $\boldsymbol{\pi}_\varepsilon$ is the marginal stationary distribution for P, G. Writing $\boldsymbol{\pi}_\varepsilon = \boldsymbol{\pi} + \varepsilon\boldsymbol{\pi}_1 + o(\varepsilon)$, we have:

$$\pi = \lim_{\varepsilon \to 0^+} \pi_\varepsilon = \sum_{i=1}^{3}\sum_{j=1}^{3} \rho_{ij}\mathbf{P}(P, G; \alpha_{Pi}/\delta_P, \alpha_{Gi}/\delta_G), \tag{14}$$

where $\mathbf{P}(x, y; a, b) := \frac{a^x}{x!}\frac{b^y}{y!}e^{-a-b}$, $\rho = [\rho_{00}, \rho_{01}, \ldots, \rho_{22}]^T$ is the normalized principal eigenvector of the following matrix:

$$\begin{bmatrix} \mathbf{1}^T & \mathbf{0}^T & \ldots & \mathbf{0}^T \\ \mathbf{0}^T & \mathbf{1}^T & \ldots & \mathbf{0}^T \\ & & \ddots & \\ 0^T & \mathbf{0}^T & \ldots & \mathbf{1}^T \end{bmatrix} \widehat{\Lambda}\left[\pi_{Y|00}\, \pi_{Y|01} \ldots \pi_{Y|22}\right], \tag{15}$$

where $\boldsymbol{\pi}_{Y|ij}(P, G) = \mathbf{P}(P, G; \alpha_{Pi}/\delta_P, \alpha_{Gi}/\delta_G)$.

Remark 1: Theorem 1 presents a reduction of an infinite dimensional Markov chain into a finite dimensional Markov chain which has nine states only. The matrix in (15) is the infinitesimal generator for the reduced order Markov chain. Instead of computing a product of infinite dimensional matrices, we can use the procedure described in [41]. For every state (i, j), the procedure entails replacing a reaction of the form:

$$X + p_i \xrightarrow{a} p_{i'},$$

with

$$p_i \xrightarrow{\alpha \mathbb{E}[X \mid p = i, g = j]} p_{i'},$$

where $\mathbb{E}$ denotes conditional expectation.

Remark 2: If a mode is defined as a local maximum of a stationary distribution, then our theorem does not necessarily imply that the stationary distribution has nine modes since the peak values of two Poisson distributions can be very close to each other. In the remainder of the paper, we will call each Poisson distribution in the mixture a "mode" in the sense that it represents a component in the mixture distribution. The number of local maxima of a distribution can be found easily given the above expression.

V. Case Study: Controlling the distribution

While the deterministic model cannot explain the emergence of bistability, the stochastic modeling framework has shown the capacity to produce up to nine modes.

In order to estimate the probability landscape of the network we need to estimate the ratio of protein production rate in the active state to the leakage. We found that we can use the following ratio for PU.1 gene $a_{P1} : a_{P0} \approx 100 : 1$, and $a_{G1} : a_{G0} \approx 5 - 10 : 1$ for the GATA-1 gene [42],[43]. This implies that the production ratios of the gene states p_0, g_0 and p_2, g_2 are close to each other compared to the production ratio of p_1 and g_1. Hence, we can group the nine modes of the landscape into four groups as follows:

1. (low,low)$:= \{(a_{p0}/\delta_P, a_{g0}/\delta_G), (a_{p0}/\delta_P, 0), (0, a_{g0}/\delta_G), (0, 0)\}$.
2. (low,high)$:= \{(a_{p0}/\delta_P, a_{g1}/\delta_G), (0, a_{g1}/\delta_G)\}$.
3. (high,low)$:= \{(a_{p1}/\delta_P, a_{g0}/\delta_G), (a_{p1}/\delta_P, 0)\}$.
4. (high,high)$:= \{(a_{p1}/\delta_P, a_{g1}/\delta_G)\}$.

For convenience, will refer to weight corresponding to each mode as p_{ll}, p_{lh}, p_{hl}, p_{hh} respectively.

The weighting coefficients in (14) play a crucial role in the "visibility" of a certain mode in the landscape, since a very small coefficient implies that the corresponding mode can be ignored in the analysis as it is unlikely to be biologically observable. However, the expression of the weighting coefficients for this network is complicated by the fact that we obtain a high-order rational polynomial which is hard to optimize or to derive analytic bounds for. Therefore, we adopt a numerical approach as will be explained below.

In order to illustrate the results, consider the network with the following production ratios:

$$\frac{\alpha_{P0}}{\delta_P} = 25, \frac{\alpha_{P1}}{\delta_P} = 2500, \frac{\alpha_{G0}}{\delta_G} = 250, \frac{\alpha_{G1}}{\delta_G} = 2500.$$

Hence, the weighting coefficients in (14) depend on four association ratios, namely $d_0/a_0, d_0'/a_0', d_1/a_1, d_1'/a_1'$. We study two cases that correspond to biologically relevant phenotypes.

First, we are interested in the bistable phenotype where the network behaves as a toggle switch that has antagonistic modes. We assume that $d_0/a_0 = d_1/a_1 := k_1$, $d_0'/a_0' = d_1'/a_1' := k_2$. In this case, numerical computations show that $p_{hh} < 0.01$, which is negligible. Hence, we are interested in minimizing p_{ll} while keeping p_{hl}, p_{lh} relatively balanced. Figure 3-a shows the contour curves for intervals of interest. We can achieve low p_{ll} and balanced probabilities of the antagonistic modes by choosing $k_1, k_2 \leq 5$. Figure 3-b depicts the probability distribution for parameters in this range. We can clearly see that the system is bistable.

Second, we are interested in a tristable phenotype where we have a (high,high) phenotype in addition to the two antagonistic phenotypes. We assume that $d_0/a_0 = d_0'/a_0' := k_1$, $d_1'/a_1' = d_1/a_1 := k_2$. In this case, we are interested in keeping p_{ll} negligible. Hence, we choose to keep $p_{ll} < 0.07$, having p_{hh} with at least 0.2 probability, while keeping p_{hl}, p_{lh} relatively balanced. Figure 4-a shows the contour curves for intervals of interest. We can have the desired phenotype with $k_1 \leq 4$, $k_2 \geq 1250$. Figure 3-b depicts the probability distribution for parameters in this range. We see the system is tristable, where the probabilities $p_{hl} \approx p_{hh} \approx p_{lh} \approx 1/3$, $p_{ll} \approx 0$.

Next, we investigate the control objective of increasing the production and decay reaction while keeping them at a fixed ratio. This can be interpreted as adding the following reactions to the network:

$$\varnothing \underset{K}{\overset{Kp^*}{\rightleftharpoons}} P, \ \varnothing \underset{K}{\overset{Kg^*}{\rightleftharpoons}} G, \tag{16}$$

Note this implies that the stationary distribution at the limit of slow promoter kinetics is given as:

$$\pi = \sum_{i=1}^{3} \sum_{j=1}^{3} \rho_{ij} \mathbf{P}\left(P, G; \frac{\alpha_{Pi} + Kp^*}{\delta_P + K}, \frac{\alpha_{Gi} + Kg^*}{\delta_G + K}\right), \tag{17}$$

and for a sufficiently high K, the distribution becomes unimodal at a chosen mode (p^*, g^*). This procedure can be interpreted as a high-gain state-feedback to "stabilize" a chosen point in the space [44]. Figure 5 shows a numerical example where the nominal network has the bistable phenotype (as shown in Figure 3-b), and then the "controller" reactions (16) are included which produces a uni-modal distribution at a desired location.

VI. Conclusion

We have studied the PU.1-GATA-1 differentiation GRN which is involved in lineage determination by committing to either the myeloid/lymhoid or erythroid lineage. Literature search was used to identify key reactions, which include self-activation, mutual repression and monomeric TF binding. A deterministic model consistent with these findings and faithful to biological reality does not admit bistability, a defining behavior of this differentiation network. We have shown that adopting a stochastic model and assuming realistic slow promoter kinetics enables the network to exhibit multimodality, and we can control the shape of the distribution.

Acknowledgments

This work was partially supported by AFOSR Grant FA9550-14-1-0060.

VII.: Appendix: Modeling the PU.1/GATA.1 network

Here we seek to model the differentiation of the PU.1-GATA-1 GRN at the bistable CMP stage, where overexpression of either TF leads to a different, positive stable steady state. The list of modeling assumptions we use to derive both deterministic and stochastic models are:

- PU.1 and GATA-1 transcriptionally self-activate their respective production [45], [46].
- PU.1 represses GATA-1 production by binding to the complex formed by GATA-1 and its promoter. This complex forms repressive chromatin structure effectively silencing transcription (the PU.1-GATA-1 complex has been shown to be present at repressed GATA-1 target genes) [47], [48].
- Similarly, GATA-1 represses production of PU.1 by binding to it on its target genes and prevents the recruitment of co-activators (such as cJun), which are critical for PU.1-mediated transcriptional activation [48], [8].
- PU.1 cannot directly bind to the GATA-1 promoter. [8] reports that PU.1 blocks GATA-1 activation without affecting GATA-1 mRNA, protein expression, or nuclear translocation.
- GATA-1 cannot directly bind to the PU.1 promoter. [49] reports two GATA-1 binding sites on the PU.1 locus, a −18 kb site, which has not been shown to have a functional regulatory role, and a −17 bp site that potentially transcriptionally represses PU.1 production. However, [45] reports that the −14 kb PU.1 URE (to which GATA-1 cannot directly bind) is significantly more critical than the proximal promoter (which includes the −17 bp GATA-1 binding site) in myeloid cell line 416B for PU.1 expression. Therefore, here we only consider the contribution of the PU.1 URE (either activation or repression when bound with GATA-1).
- All TF interaction occurs in their monomeric form. ETS TFs, such as PU.1 typically bind as monomers (both to DNA and other proteins) [25]. Though there is evidence of GATA-1 dimerization [26], it exists primarily in its monomeric

form. In particular, GATA-1 self-activation and binding to PU.1 occurs only in its monomeric form [26], [8].

- The promoters of both TFs are leaky [45], [46].
- Protein production occurs in a one-step process (no intermediary mRNA dynamics).

REFERENCES

[1]. Huang S, Eichler G, Bar-Yam Y, and Ingber DE, "Cell fates as high-dimensional attractor states of a complex gene regulatory network," Phys. Rev. Lett, vol. 94, p. 128701, 4 2005 [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.94.128701 [PubMed: 15903968]

[2]. Wu M, Su R-Q, Li X, Ellis T, Lai Y-C, and Wang X, "Engineering of regulated stochastic cell fate determination," Proceedings of the National Academy of Sciences, vol. 110, no. 26, pp. 10 610–10 615, 2013 [Online]. Available: http://www.pnas.org/content/110/26/10610.abstract

[3]. Kurtz TG, "The relationship between stochastic and deterministic models for chemical reactions," The Journal of Chemical Physics, vol. 57, no. 7, pp. 2976–2978, 1972.

[4]. Kærn M, Elston TC, Blake WJ, and Collins JJ, "Stochasticity in gene expression: from theories to phenotypes," Nature Reviews Genetics, vol. 6, no. 6, pp. 451–464, 2005.

[5]. Strasser M, Theis FJ, and Marr C, "Stability and multiattractor dynamics of a toggle switch based on a two-stage model of stochastic gene expression," Biophysical journal, vol. 102, no. 1, pp. 19–29, 2012. [PubMed: 22225794]

[6]. Voss TC and Hager GL, "Dynamic regulation of transcriptional states by chromatin and transcription factors," Nat Rev Genet, vol. 15, no. 2, pp. 69–81, 2 2014, review. [Online]. Available: 10.1038/nrg3623 [PubMed: 24342920]

[7]. Paldi A, "Stochastic gene expression during cell differentiation: order from disorder?" Cellular and Molecular Life Sciences CMLS, vol. 60, no. 9, pp. 1775–1778, 2003 [Online]. Available: 10.1007/s00018-003-23147-z [PubMed: 14523542]

[8]. Zhang P, Zhang X, Iwama A, Yu C, Smith KA, Mueller BU, Narravula S, Torbett BE, Orkin SH, and Tenen DG, "Pu. 1 inhibits gata-1 function and erythroid differentiation by blocking gata-1 dna binding," Blood, vol. 96, no. 8, pp. 2641–2648, 2000. [PubMed: 11023493]

[9]. Nerlov C and Graf T, "Pu. 1 induces myeloid lineage commitment in multipotent hematopoietic progenitors," Genes & development, vol. 12, no. 15, pp. 2403–2412, 1998. [PubMed: 9694804]

[10]. Birbrair A and Frenette PS, "Niche heterogeneity in the bone marrow," Annals of the New York Academy of Sciences, vol. 1370, no. 1, pp. 82–96, 2016. [PubMed: 27015419]

[11]. Rieger MA and Schroeder T, "Hematopoiesis," Cold Spring Harbor perspectives in biology, vol. 4, no. 12, p. a008250, 2012. [PubMed: 23209149]

[12]. Rosenbauer F and Tenen DG, "Transcription factors in myeloid development: balancing differentiation with transformation," Nature reviews. Immunology, vol. 7, no. 2, p. 105, 2007.

[13]. Cvejic A, "Mechanisms of fate decision and lineage commitment during haematopoiesis," Immunology and cell biology, vol. 94, no. 3, p. 230, 2016. [PubMed: 26526619]

[14]. Moignard V, Macaulay IC, Swiers G, Buettner F, Schütte J, Calero-Nieto FJ, Kinston S, Joshi A, Hannah R, Theis FJ, et al., "Characterisation of transcriptional networks in blood stem and progenitor cells using high-throughput single cell gene expression analysis," Nature cell biology, vol. 15, no. 4, p. 363, 2013. [PubMed: 23524953]

[15]. Orkin SH, "Transcription factors and hematopoietic development," Journal of Biological Chemistry, vol. 270, no. 10, pp. 4955–4958, 1995. [PubMed: 7890597]

[16]. Krause DS, "Regulation of hematopoietic stem cell fate," Oncogene, vol. 21, no. 21, p. 3262, 2002. [PubMed: 12032767]

[17]. Sieweke MH and Graf T, "A transcription factor party during blood cell differentiation," Current opinion in genetics & development, vol. 8, no. 5, pp. 545–551, 1998. [PubMed: 9794826]

[18]. Zhang P, Behre G, Pan J, Iwama A, Wara-Aswapati N, Radomska HS, Auron PE, Tenen DG, and Sun Z, "Negative cross-talk between hematopoietic regulators: Gata proteins repress pu. 1," Proceedings of the National Academy of Sciences, vol. 96, no. 15, pp. 8705–8710, 1999.

[19]. Nutt SL, Metcalf D, D'Amico A, Polli M, and Wu L, "Dynamic regulation of pu. 1 expression in multipotent hematopoietic progenitors," Journal of Experimental Medicine, vol. 201, no. 2, pp. 221–231, 2005. [PubMed: 15657291]

[20]. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, Snyder MP, Pritchard JK, Kundaje A, Greenleaf WJ, et al., "Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution," Nature genetics, vol. 48, no. 10, pp. 1193–1203, 2016. [PubMed: 27526324]

[21]. Akashi K, Traver D, Miyamoto T, and Weissman IL, "A clonogenic common myeloid progenitor that gives rise to all myeloid lineages," Nature, vol. 404, no. 6774, p. 193, 2000. [PubMed: 10724173]

[22]. Iwasaki H and Akashi K, "Myeloid lineage commitment from the hematopoietic stem cell," Immunity, vol. 26, no. 6, pp. 726–740, 2007. [PubMed: 17582345]

[23]. Roeder I and Glauche I, "Towards an understanding of lineage specification in hematopoietic stem cells: a mathematical model for the interaction of transcription factors gata-1 and pu. 1," Journal of theoretical biology, vol. 241, no. 4, pp. 852–865, 2006. [PubMed: 16510158]

[24]. Huang S, Guo Y-P, May G, and Enver T, "Bifurcation dynamics in lineage-commitment in bipotent progenitor cells," Developmental biology, vol. 305, no. 2, pp. 695–713, 2007. [PubMed: 17412320]

[25]. Hollenhorst PC, McIntosh LP, and Graves BJ, "Genomic and biochemical insights into the specificity of ets transcription factors," Annual review of biochemistry, vol. 80, pp. 437–471, 2011.

[26]. Crossley M, Merika M, and Orkin SH, "Self-association of the erythroid transcription factor gata-1 mediated by its zinc finger domains." Molecular and Cellular Biology, vol. 15, no. 5, pp. 2448–2456, 1995. [PubMed: 7739529]

[27]. Chickarmane V, Enver T, and Peterson C, "Computational modeling of the hematopoietic erythroid-myeloid switch reveals insights into cooperativity, priming, and irreversibility," PLoS computational biology, vol. 5, no. 1, p. e1000268, 2009. [PubMed: 19165316]

[28]. Tian T and Smith-Miles K, "Mathematical modeling of gata-switching for regulating the differentiation of hematopoietic stem cell," BMC systems biology, vol. 8, no. 1, p. S8, 2014.

[29]. Érdi P and Tóth J, Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models. Manchester University Press, 1989.

[30]. Ellison P and Feinberg M, "How catalytic mechanisms reveal them-selves in multiple steady-state data: I. basic principles," Journal of Molecular Catalysis A: Chemical, vol. 154, no. 1, pp. 155–167, 2000.

[31]. Ellison PR, "The advanced deficiency algorithm and its applications to mechanism discrimination," Ph.D. dissertation, The University of Rochester, 1998.

[32]. Ji H, "Uniqueness of equilibria for complex chemical reaction networks," Ph.D. dissertation, The Ohio State University, 2011.

[33]. Ellison P, Feinberg M, Ji H, and Knight D, "Chemical reaction network toolbox v2.3," Available online at http://www.crnt.osu.edu/CRNTWin, 2011.

[34]. Paldi A, "Stochastic gene expression during cell differentiation: order from disorder?" Cellular and molecular life sciences, vol. 60, no. 9, pp. 1775–1778, 2003. [PubMed: 14523542]

[35]. Miller-Jensen K, Dey SS, Schaffer DV, and Arkin AP, "Varying virulence: epigenetic control of expression noise and disease processes," Trends in biotechnology, vol. 29, no. 10, pp. 517–525, 2011. [PubMed: 21700350]

[36]. Voss TC and Hager GL, "Dynamic regulation of transcriptional states by chromatin and transcription factors," Nature Reviews Genetics, vol. 15, no. 2, pp. 69–81, 2014.

[37]. Raj A, Peskin CS, Tranchina D, Vargas DY, and Tyagi S, "Stochastic mrna synthesis in mammalian cells," PLoS Biol, vol. 4, no. 10, p. e309, 2006. [PubMed: 17048983]

[38]. To T-L and Maheshri N, "Noise can induce bimodality in positive transcriptional feedback loops without bistability," Science, vol. 327, no. 5969, pp. 1142–1145, 2010. [PubMed: 20185727]

[39]. Mariani L, Schulz EG, Lexberg MH, Helmstetter C, Radbruch A, Löhning M, and Höfer T, "Short-term memory in gene induction reveals the regulatory principle behind stochastic il-4 expression," Molecular systems biology, vol. 6, no. 1, p. 359, 2010. [PubMed: 20393579]

[40]. Yuan L, Chan GC, Beeler D, Janes L, Spokes KC, Dharaneeswaran H, Mojiri A, Adams WJ, Sciuto T, Garcia-Cardeña G, et al., "A role of stochastic phenotype switching in generating mosaic endothelial cell heterogeneity," Nature communications, vol. 7, 2016.

[41]. Al-Radhawi MA, Del Vecchio D, and Sontag ED, "Multi-modality in gene regulatory networks with slow promoter kinetics," arXiv preprint arXiv:1705.02330, 2017.

[42]. Nishikawa K, Kobayashi M, Masumi A, Lyons SE, Weinstein BM, Liu PP, and Yamamoto M, "Self-association of gata1 enhances transcriptional activity in vivo in zebra fish embryos," Molecular and cellular biology, vol. 23, no. 22, pp. 8295–8305, 2003. [PubMed: 14585986]

[43]. Okuno Y, Huang G, Rosenbauer F, Evans EK, Radomska HS, Iwasaki H, Akashi K, Moreau-Gachelin F, Li Y, Zhang P, et al., "Potential autoregulation of transcription factor pu. 1 by an upstream regulatory element," Molecular and cellular biology, vol. 25, no. 7, pp. 2832–2845, 2005. [PubMed: 15767686]

[44]. Del Vecchio D, Abdallah H, Qian Y, and Collins JJ, "A blueprint for a synthetic genetic feedback controller to reprogram cell fate," Cell systems, vol. 4, no. 1, pp. 109–120, 2017. [PubMed: 28065574]

[45]. Okuno Y, Huang G, Rosenbauer F, Evans EK, Radomska HS, Iwasaki H, Akashi K, Moreau-Gachelin F, Li Y, Zhang P, et al., "Potential autoregulation of transcription factor pu. 1 by an upstream regulatory element," Molecular and cellular biology, vol. 25, no. 7, pp. 2832–2845, 2005. [PubMed: 15767686]

[46]. Nishikawa K, Kobayashi M, Masumi A, Lyons SE, Weinstein BM, Liu PP, and Yamamoto M, "Self-association of gata1 enhances transcriptional activity in vivo in zebra fish embryos," Molecular and cellular biology, vol. 23, no. 22, pp. 8295–8305, 2003. [PubMed: 14585986]

[47]. Stopka T, Amanatullah DF, Papetti M, and Skoultchi AI, "Pu. 1 inhibits the erythroid program by binding to gata-1 on dna and creating a repressive chromatin structure," The EMBO journal, vol. 24, no. 21, pp. 3712–3723, 2005. [PubMed: 16222338]

[48]. Liew CW, Rand KD, Simpson RJ, Yung WW, Mansfield RE, Crossley M, Proetorius-Ibba M, Nerlov C, Poulsen FM, and Mackay JP, "Molecular analysis of the interaction between the hematopoietic master transcription factors gata-1 and pu. 1," Journal of Biological Chemistry, vol. 281, no. 38, pp. 28 296–28 306, 2006.

[49]. Chou ST, Khandros E, Bailey LC, Nichols KE, Vakoc CR, Yao Y, Huang Z, Crispino JD, Hardison RC, Blobel GA, et al., "Graded repression of pu. 1/sfpi1 gene transcription by gata factors regulates hematopoietic cell fate," Blood, vol. 114, no. 5, pp. 983–994, 2009. [PubMed: 19491391]

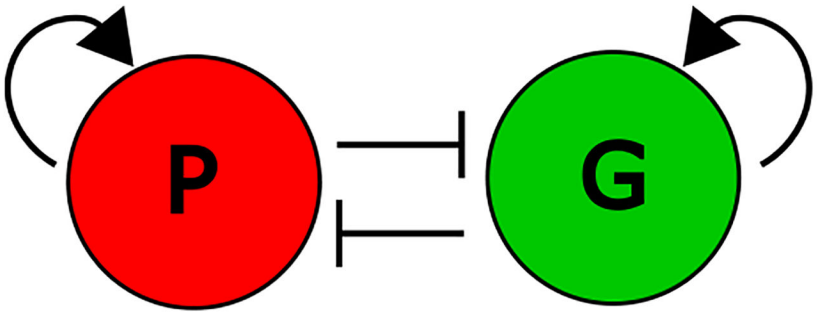


Fig. 1.
PU.1-GATA-1 GRN demonstrating self-activation and mutual repression.

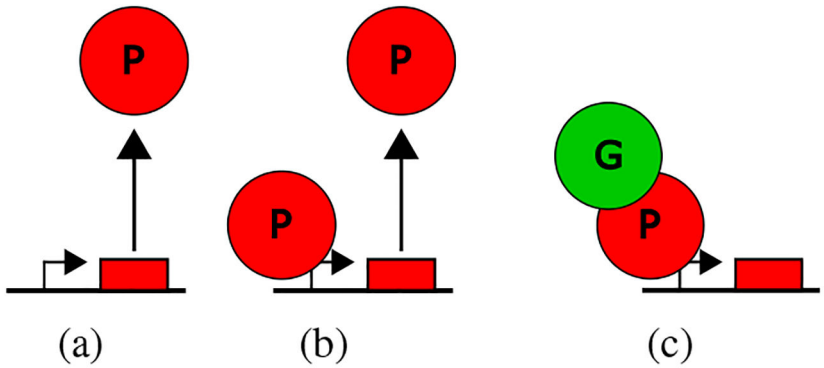
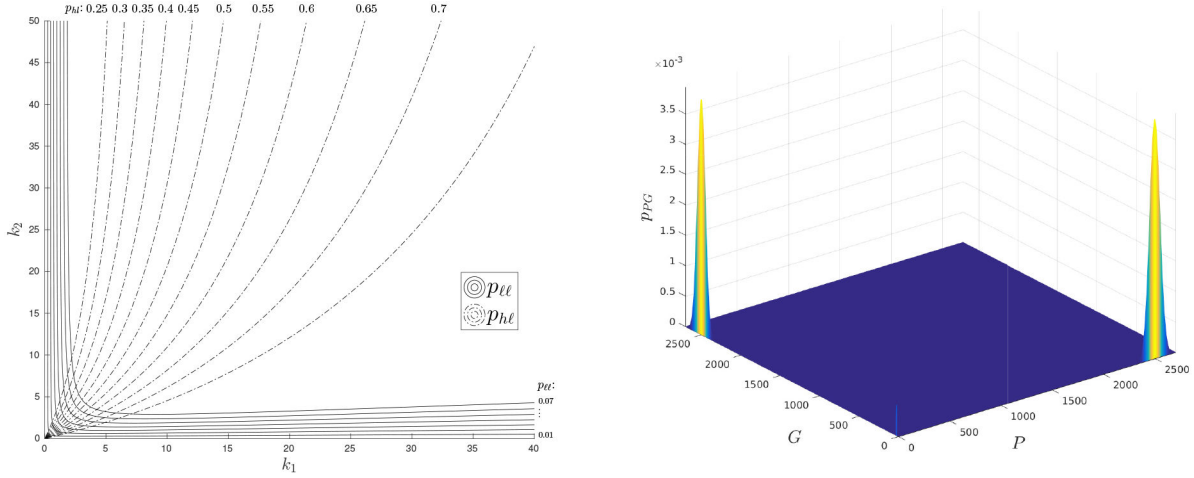


Fig. 2.

Promoter states: (a) p_0 representing leaky production, (b) p_1 with self-activated production, and (c) p_2 when the promoter is fully repressed.

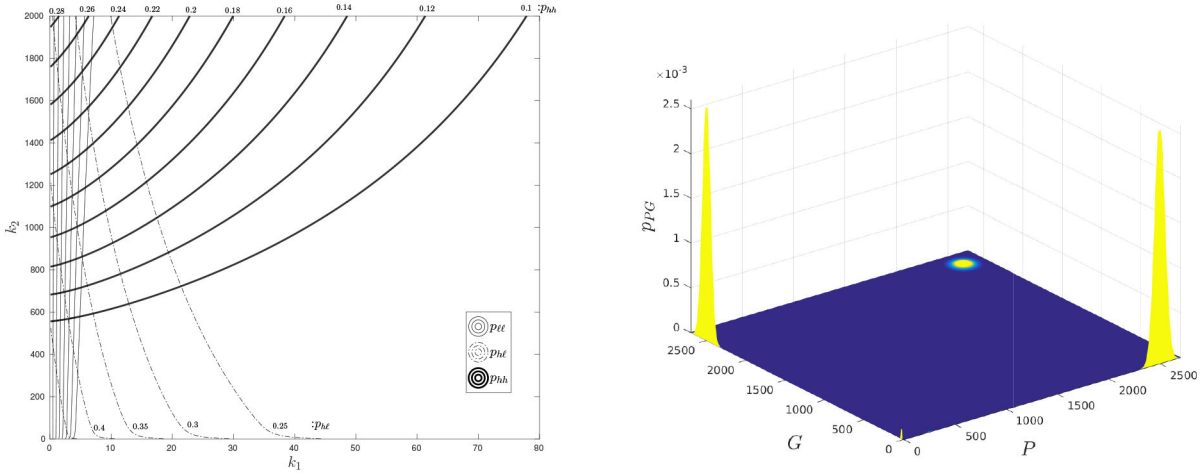


(a) Contour Curves

(b) The probability distribution

Fig. 3. The association ratios can be chosen to have a bistable system.

(a) Contour curves for the probabilities $p_{\ell\ell}$, $p_{h\ell}$. The curves are plotted for the intervals $0 \leq p_{\ell\ell} \leq 0.07$, $0.25 \leq p_{hl} \leq 0.75$. The variables on the axes are $k_1 := d_0/a_0 = d_1/a_1$, $k_2 := d_0'/a_0' = d_1'/a_1'$. (b) The probability distribution for $k_1 = k_2 = 1$ computed using (14). The probability of the corresponding modes are $p_{\ell\ell} = 0.021$, $p_{h\ell} = 0.485$, $p_{\ell h} = 0.494$, $p_{hh} \approx 0$.



(a) (b)

Fig. 4. The association ratios can be chosen to have a tristable system.

(a) Contour curves for the probabilities $p_{\ell\ell}$, $p_{h\ell}$, p_{hh}. The curves are plotted for the intervals $0 \le p_{\ell\ell} \le 0.07$, $0.25 \le p_{hl} \le 0.45$, $0.1 \le p_{hh} \le 0.3$. The variables on the axes are $k_1 := d_0/a_0 = d_0'/a_0'$, $k_2 := d_1'/a_1' = d_1/a_1$. The values for the contours for $p_{\ell\ell}$ are 0.01,0.02,..,0.07 from left to right, (b) The probability distribution for $k_1 = 0.1$, $k_2 = 2500$ computed using (14). The probability of the corresponding modes are $p_{hh} = 0.333$, $p_{h\ell} = 0.332$, $p_{\ell h} = 0.334$, $p_{\ell\ell} \approx 0.001$. Note that the (high,high) mode has a third of the probability despite having a small height.

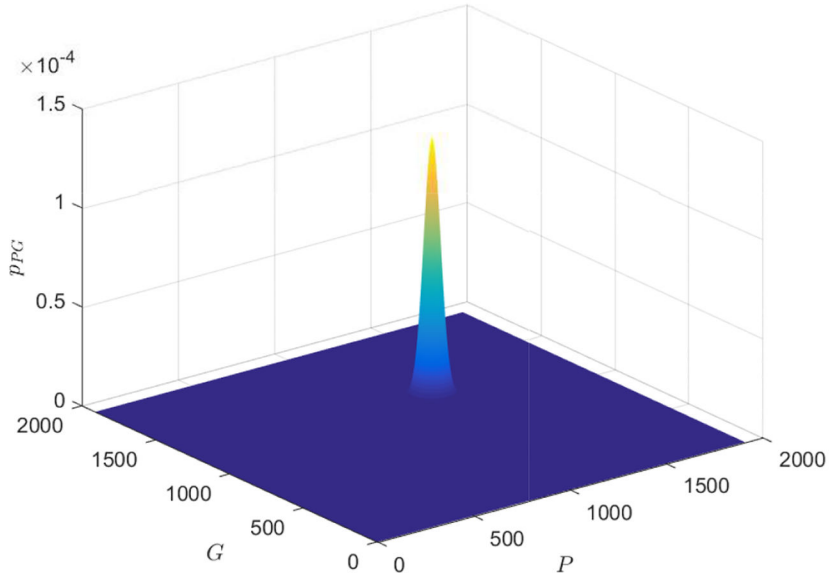


Fig. 5. Increasing production and decay rates simultaneously produces a uni-modal distribution at a desired location

The probability distribution for $k_1 = 0.1$, $k_2 = 1$, $K = 50$, $p^* = g^* = 1250$ computed using (17), where $k_1 := d_0/a_0 = d_1/a_1$, $k_2 := d_0'/a_0' = d_1'/a_1'$. Note that the distribution is uni-modal.