Stochastic modeling of a gene regulatory network driving B cell development in germinal centers

Alexey Koshkin, Ulysse Herbach, María Rodríguez Martínez, Olivier Gandrillon, Fabien Crauste

Stochastic modeling of a gene regulatory network driving B cell development in germinal centers

Alexey Koshkin[1,2], Ulysse Herbach[3], María Rodríguez Martínez[4], Olivier Gandrillon[1,2,*], Fabien Crauste[5,*]

1. Inria Dracula, Villeurbanne, France
2. Laboratory of Biology and Modelling of the Cell, Université de Lyon, ENS de Lyon, Université Claude Bernard, CNRS UMR 5239, INSERM U1210, 46 allée d'Italie, Site Jacques Monod, 69007 Lyon, France
3. Université de Lorraine, CNRS, Inria, IECL, F-54000 Nancy, France
4. IBM Research Zurich, Switzerland
5. Université Paris Cité, CNRS, MAP5, F-75006 Paris, France

* fabien.crauste@math.cnrs.fr
* olivier.gandrillon@ens-lyon.fr

Abstract

Germinal centers (GCs) are the key histological structures of the adaptive immune system, responsible for the development and selection of B cells producing high-affinity antibodies against antigens. Due to their level of complexity, unexpected malfunctioning may lead to a range of pathologies, including various malignant formations. One promising way to improve the understanding of malignant transformation is to study the underlying gene regulatory networks (GRNs) associated with cell development and differentiation. Evaluation and inference of the GRN structure from gene expression data is a challenging task in systems biology: recent achievements in single-cell (SC) transcriptomics allow the generation of SC gene expression data, which can be used to sharpen the knowledge on GRN structure. In order to understand whether a particular network of three key gene regulators (BCL6, IRF4, BLIMP1), influenced by two external stimuli signals (surface receptors BCR and CD40), is able to describe GC B cell differentiation, we used a stochastic model to fit SC transcriptomic data from a human lymphoid organ dataset. The model is defined mathematically as a piecewise-deterministic Markov process. We showed that after parameter tuning, the model qualitatively recapitulates mRNA distributions corresponding to GC and plasmablast stages of B cell differentiation. Thus, the model can assist in validating the GRN structure and, in the future, could lead to better understanding of the different types of dysfunction of the regulatory mechanisms.

1 Introduction

Adaptive immune response is a complex mechanism, relying on B and T lymphocytes, which protects the organism against a range of pathogens. Crucial elements of adaptive immune response, the germinal centers (GCs) are the structures in lymphoid organs where activated naive B cells are expanded (in a dark zone, DZ) and selected (in a light zone, LZ) and can have multiple exit fates, such as antibody production (plasmablasts and plasma cells, PB_PC), long term storage of antigen information (memory B cells, MC), or death via apoptosis [1, 2].

It is currently thought that B cell differentiation in GC is controlled by a small network of transcription factors (TFs) constituted by B-cell lymphoma 6 (BCL6), interferon regulatory factor 4 (IRF4) and PR domain zinc finger protein 1 (BLIMP1) [3]. BCL6 controls formation of GC, terminal differentiation of B cells and lymphomagenesis [4, 5]. BCL6 disturbance can be triggered by several mechanisms, including proteasome degradation by BCR, T-cell-mediated CD40-induced IRF4 repression of BCL6 [6, 5], or disruption of BCL6 autoregulation loop [5, 7]. Transcription factor IRF4 is involved in the termination of GC B cell differentiation, in immunoglobulin class switch recombination (CSR) and plasma cell development [8]. Impairment of IRF4 expression is tightly connected with the appearance of multiple malignancies [8]. BLIMP1 regulates pathways responsible for B cell lineage (e.g., PAX5) and GC proliferation and metabolism (e.g., MYC) [9, 10]. BLIMP1 is also involved in the induction of genes (e.g., XBP-1, ATF6, Ell2) facilitating antibody synthesis [11, 12, 13]. These three TFs interact, through various activation/inhibition processes: IRF4 represses BCL6 and activates BLIMP1 [14] (hence being essential for GC maturation and B cell differentiation into plasmablast), BLIMP1 and BCL6 mutually repress each other [15, 16, 17, 18].

Martinez et al. [3] developed a deterministic kinetic ODE model capable of simulating normal and malignant GC exits using a GRN based on these three transcription factors. For the normal differentiation of GC B cells towards PB_PC stage, the kinetic ODE model fits microarray data at two steady-states: the first one associated with the GC stage of B cell differentiation (with high levels of BCL6 and low levels of IRF4 and BLIMP1), and the second one associated with PB_PC stage (with low levels of BCL6 and high levels of IRF4 and BLIMP1).

Recently, multiple protocols for SC RNA-seq data generation have been developed and used to answer various questions in biology [19, 20]. At the same time, different groups showed that gene transcription in eukaryotes is a discontinuous process and follows bursting kinetics [21, 22, 23, 24]. Such results suggest that the stochastic nature of gene expression at the single cell (SC) level can be partly responsible for the phenotype variation in living organisms [25]. Thus, by gaining access to a stochastic behavior of gene expression, the SC viewpoint may lead to further improvement of the understanding of the biological systems and their variability.

Nevertheless, stochastic modeling of GRNs using SC gene expression data is still in its early stage [26, 27] and has never been studied for GC B cells. Here, we apply a particular class of stochastic models combining deterministic dynamics and random jumps, called

piecewise-deterministic Markov processes (PDMPs) [28], to the description of GC B cell differentiation. It is a two-state model of gene expression introduced in [29] that allows a description of the system's dynamics at the promoter, transcription and translation levels for a given GRN. We apply this model to the GRN made of the three key genes, BCL6, IRF4 and BLIMP1, and simulate single B cell mRNA data [30]. We show that the model can qualitatively simulate the SC mRNA patterns for normal B cell differentiation at GC and PB_PC stages.

2 Material, Methods and Models

2.1 Single-cell data

We used the B cell dataset from human lymphoid organs published by Milpied et al. [30]. The authors studied normal B cell subsets from germinal centers of the human spleen and tonsil and performed integrative SC analysis of gene expression. They used an adapted version of the integrative single-cell analysis protocol [31]. In short, the authors prepared cells for flow cytometry cell sorting. Then in every 96-well plate the authors sorted three to six ten-cell samples of the same phenotype as a single-cell. They performed multiplex qPCR analysis using the Biomark system (Fluidigm) with 96x96 microfluidic chips (fluidigm) and Taqmann assays (Thermofisher) [30]. They obtained results in the form of fixed fluorescence threshold to derive Ct values. We used Ct values to derive Expression threshold (Et) values: $Et = 30 - Ct$. When there was an unreliably low or undetected expression ($Ct > 30$), Et was set to zero [30]. Using SC gene expression analysis of a panel of 91 preselected genes and pseudotime analysis (based on the cartesian coordinates of SC on the first and second principal components of the PCA), the authors separated GC DZ cells, GC LZ cells, memory cells and PB_PC cells.
Here we focused on three genes, BCL6, IRF4 and BLIMP1. We selected the SC gene expression values for BCL6, IRF4 and BLIMP1 for GC DZ cells (317 SC) and for PB_PC (104 SC) (see Figure 5). The experimental dataset includes at the GC B cell stage 30 cells with zero BCL6 mRNA amount, 292 cells with zero IRF4 mRNA amount and 292 cells with zero BLIMP1 mRNA amount. For the end of the B cell differentiation (PB_PC), there were 25 cells with zero BCL6 mRNA amount, 79 cells with zero IRF4 and 5 cells with zero BLIMP1 mRNA amount.

2.2 Kinetic ODE model

Martinez et al. [3] derived an ODE model that simulates B cell differentiation from mature GC cells towards PB_PC. Dynamics of each protein (BCL6, IRF4 and BLIMP1) are defined by a production rate (μ), a degradation rate (λ), a dissociation constant (k) and a maximum transcription rate (σ). Dynamics are described by System (1)-(3), where p, b

Index	Gene/Stimulus
1	BCL6
2	IRF4
3	BLIMP1
4	BCR
5	CD40

Table 1. Correspondence between gene or stimulus names and model index.

and r account for proteins BLIMP1, BCL6 and IRF4, respectively:

$$\frac{dp}{dt} = \mu_p + \sigma_p \frac{k_b^2}{k_b^2 + b^2} + \sigma_p \frac{r^2}{k_r^2 + r^2} - \lambda_p p, \tag{1}$$

$$\frac{db}{dt} = \mu_b + \sigma_b \frac{k_p^2}{k_p^2 + p^2} \frac{k_b^2}{k_b^2 + b^2} \frac{k_r^2}{k_r^2 + r^2} - (\lambda_b + BCR)b, \tag{2}$$

$$\frac{dr}{dt} = \mu_r + \sigma_r \frac{r^2}{k_r^2 + r^2} + CD40 - \lambda_r r. \tag{3}$$

In this model, CD40 and BCR act as stimuli on genes: BCR temporary represses BCL6 and CD40 temporary activates IRF4.

2.3 Stochastic model

The stochastic model that describes the coupled dynamics of gene i and the other genes of the GRN is defined by the series of equations:

$$\begin{cases} E_i(t) : 0 \xrightarrow{k_{\text{on},i}(P_1,P_2,P_3,Q_s)} 1, 1 \xrightarrow{k_{\text{off},i}(P_1,P_2,P_3,Q_s)} 0, \\ M_i'(t) = s_{0,i} E_i(t) - d_{0,i} M_i(t), \\ P_i'(t) = s_{1,i} M_i(t) - d_{1,i} P_i(t), \end{cases} \tag{4}$$

where $E_i(t)$, $M_i(t)$ and $P_i(t)$ are, respectively, the activation status of the promoter, the quantity of mRNA and the quantity of proteins of gene i, for $i \in \{1, 2, 3\}$. For $s \in \{4, 5\}$, Q_s accounts for external stimuli intensity. Each index i refers to one of the gene in the GRN, either BCL6, IRF4, or BLIMP1, and each index s to stimuli BCR and CD40 (see Table 1).

For each gene i, System (4) is defined by the promoter state switching rates $k_{\text{on},i}$ (h^{-1}) and $k_{\text{off},i}$ (h^{-1}), by a degradation rate of mRNA ($d_{0,i}$, h^{-1}), a protein degradation rate ($d_{1,i}$, h^{-1}), a transcription rate ($s_{0,i}$, mRNA $\times\, h^{-1}$), a translation rate ($s_{1,i}$, protein $\times$ mRNA^{-1} $\times$ h^{-1}), and interaction parameters $\theta_{j,i}$ with either gene or stimulus j. Interactions between genes are based on the assumption that $k_{\text{on},i}$ is a function of the proteins P_1, P_2, P_3 and stimuli Q_s and is given by:

$$k_{\text{on},i}(P_1, P_2, P_3, Q_s) = \frac{k_{\text{on},i}^{min} + k_{\text{on},i}^{max} \beta_i \Phi_i(P_1, P_2, P_3, Q_s)}{1 + \beta_i \Phi_i(P_1, P_2, P_3, Q_s)} \tag{5}$$

where

$$\Phi_i(P_1, P_2, P_3, Q_s) = \prod_{s=\mathrm{BCR}}^{\mathrm{CD40}} \frac{1 + e^{\theta_{s,i}} Q_s}{1 + Q_s} \prod_{j=1}^{3} \frac{1 + e^{\theta_{j,i}} (P_j / H_{j,i})^{\gamma}}{1 + (P_j / H_{j,i})^{\gamma}}. \tag{6}$$

In (6), $H_{j,i}$ represents an interaction threshold for the protein j on gene i , while β_i is a scaling parameter. The structure of System (4)-(6) for the particular network considered in this paper is illustrated in Figure 1.

A detailed derivation of the model is presented in the supplementary material of [29]. Starting from a simple biochemical model of gene expression, the authors described higher-order interactions and took into consideration possible auto-activations. After normalization and simplifcation steps, Herbach et al. [29] and Bonnaffoux et al. [32] described the promoter switching rates $k_{\mathrm{on},i}$ and $k_{\mathrm{off},i}$ in the form of (5) and (6) by introducing the scaling parameter β_i.

It can be noted that the promoter state evolution of gene i between time t and $t + \delta t$ in System (4)-(6) is defined, for small δt, as a Bernoulli-distributed random variable [29, 32]:

$$E_i(t + \delta t) \sim \mathrm{Bernoulli}(\pi_i(t)),$$

where probability $\pi_i(t)$, derived by solving the master equation [29, 33], is given by

$$\pi_i(t) = E_i(t) e^{-\delta t (k_{\mathrm{on},i} + k_{\mathrm{off},i})} + \frac{k_{\mathrm{on},i}}{k_{\mathrm{on},i} + k_{\mathrm{off},i}} \left(1 - e^{-\delta t (k_{\mathrm{on},i} + k_{\mathrm{off},i})}\right).$$

It follows that the promoter state of gene i averages to $k_{\mathrm{on},i}/(k_{\mathrm{on},i} + k_{\mathrm{off},i})$ in the fast promoter regime ($k_{\mathrm{on},i} + k_{\mathrm{off},i} \gg 1/\delta t$). This quantity will be used to reduce System (4)-(6) into an ordinary differential equation (ODE) system in Section 3.1.

2.4 Simulating the stochastic model

During B cell differentiation in GC, B cells first receive BCR signal, through follicular dendritic cells interaction, that represses BCL6. Then, B cells integrate CD40 signals, through T follicular helper, that activate IRF4 [3, 6, 34].

In order to simulate these interactions, we assumed that BCR was acting on BCL6 from 0h until 25h, and CD40 was acting on IRF4 from 35h until 60h. Stimuli were implemented in three steps: first a linear increase ($t_{\mathrm{BCR}} \in [0.5h; 1.5h]; t_{\mathrm{CD40}} \in [35h; 36h]$), then a stable stimulus ($t_{\mathrm{BCR}} \in [1.5h; 24h]; t_{\mathrm{CD40}} \in [36h; 60h]$), finally a linear decrease ($t_{\mathrm{BCR}} \in [24h; 25h]; t_{\mathrm{CD40}} \in [60h; 61h]$) (see Supplementary Figure S1).

In all simulations, the system evolves for 500h so it can reach a steady state before applying the stimuli (at time $t = 0$h). After the first stimulus (BCR) is applied, the system is simulated for an additional 500h. For each simulation, the amounts of mRNA counts have been collected every 0.5h.

The stochastic system (4)-(6) is defined by 40 parameters, whose values are given in Tables 2 to 5.

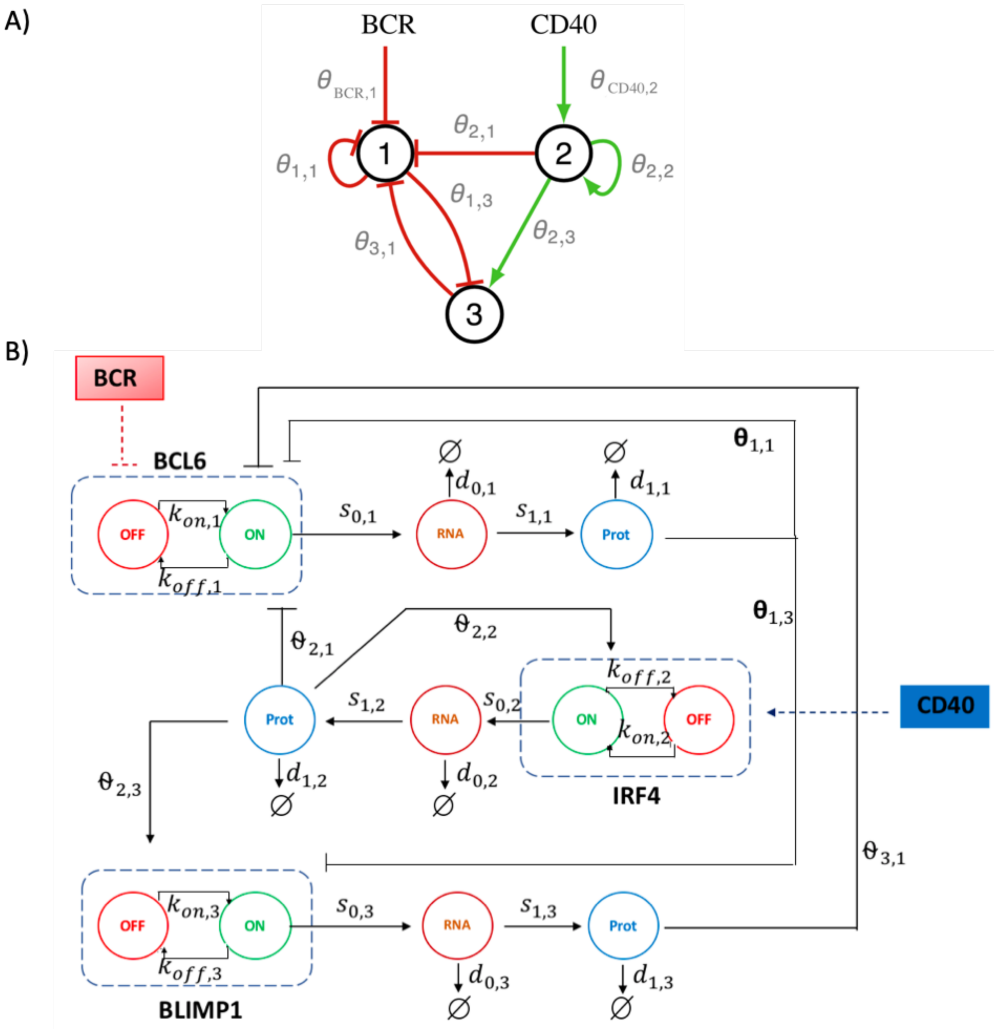

Figure 1. A) Schematic representation of the three-gene GRN involved in B cell differentiation. It consists of BCL6 (gene 1), IRF4 (gene 2) and BLIMP1 (gene 3), and with stimuli BCR and CD40 acting on the network. The interaction $j \to i$ between a regulating protein j and a target gene i is represented by the interaction parameter $\theta_{j,i}$. B) Schematic representation of the associated stochastic model. A gene is represented by its promoter state (dashed rectangle), which can switch randomly from on to off (and vice versa), with rates $k_{\text{on},i}$ ($k_{\text{off},i}$). When promoter state is on, mRNA molecules are continuously produced at $s_{0,i}$ rate. Proteins are constantly translated from mRNA at $s_{1,i}$ rate. Parameters $d_{0,i}$ and $d_{1,i}$ are degradation rates of mRNA and proteins. The interaction between a regulator gene j and a target gene i is defined by the dependence of $k_{\text{on},i}$ and $k_{\text{off},i}$ on the protein level P_j and the interaction parameter $\theta_{j,i}$. IRF4 gene exhibits an autoactivation loop ($\theta_{2,2}$). Additionally, two external stimuli, BCR and CD40, act on the GRN.

2.5 Model execution in a computational center

All models were established as part of the WASABI pipeline [32] and were implemented in Python 3. All computations were performed using the computational center of IN2P3 (Villeurbanne/France).

2.6 Tuning of the PDMP model

2.6.1 Parameters estimation for the ODE-reduced model

In Section 3.1, we use a reduced, deterministic version of System (4)-(6), namely System (11). Initial guess of each parameter has been chosen randomly in the same order of magnitude as in Bonnaffoux et al. [32]. Specifically, the initial value of k_{on} for IRF4 ($k_{on,\,\text{init},\,\text{IRF4}}$) has been estimated by comparison with values of the kinetic model from Martinez et al. [3]. Initial values of k_{on} for BCL6 and BLIMP1 were selected in the same order of magnitude as $k_{on,\,\text{init},\,\text{IRF4}}$.

2.6.2 Estimation of the parameters for the stochastic model: Automatized approach

After we have established the parameters for the reduced model (11), and we have shown that (11) has two steady states, we used these values as initial guess for the stochastic model (4)-(6). The goal was then to further tune parameter values so the stochastic model (4)-(6) fits the experimental SC data.

We investigated a possible effect of $H_{j,i}$ values, $\theta_{j,i}$ values and $k_{on,\,\text{init}}$ values on the quality of the fitting (for each parameter combination, simulation was performed for 200 SC). We tested the values of interaction threshold $H_{j,i}$ within the set $\{0.01, 0.1, 1\}$ for $j, i \in \{1, 2, 3\}$, $(i \neq j)$, and the set $\{0.0001, 0.001, 0.1, 1, 100\}$ for BCR repression stimuli on BCL6 node, CD40 activation stimuli on IRF4 and for $H_{2,2}$. We also tested the values of $\theta_{j,i}$ by multiplying by a factor $f_\theta \in \{1,\, 5\}$ for $j, i \in \{1, 2, 3\}$, $(i \neq j)$, and by multiplying by a factor $f_\theta \in \{1, 10\}$ for BCR repression stimuli on BCL6 node, CD40 activation stimuli on IRF4 and for $\theta_{2,2}$ (IRF4 autoactivation loop). In total we tested two different values of $\theta_{j,i}$ for 8 interactions (2^8), 3 values of $H_{j,i}$ for 5 interactions (3^5), and 5 values of $H_{j,i}$ for 3 interactions (5^3), generating $2^8 \times 3^5 \times 5^3 \approx 7.8 \times 10^6$ combinations of parameters.

During this automatized tuning procedure, we selected a set of parameter values that allows the system to provide the best fit of the experimental mRNA values for BCL6, IRF4 and BLIMP1 at the GC stage, based on a quality-of-fit criterion. This criterion was defined as a comparison between the average model-derived values (Υ) and the average experimental values (Ω), with an objective function (OF) to minimize for the set of genes $G = \{BCL6, IRF4, BLIMP1\}$ and stages $ST = \{GC, PB\_PC\}$ defined by

$$OF = \sum_{\delta'=1}^{|G|} \sum_{\delta''=1}^{|ST|} \left| \frac{\Omega_{\delta',\delta''} - \Upsilon_{\delta',\delta''}}{\Omega_{\delta',\delta''}} \right|. \tag{9}$$

Parameter	Version I, II, III
H_{12}	1
H_{32}	1
H_{33}	1
θ_{11}	-0.2
θ_{12}	0
θ_{32}	0
θ_{13}	-1
θ_{33}	0
$d_{0,\mathrm{BCL6}}$	0.05
$d_{0,\mathrm{IRF4}}$	0.05
$s_{1,\mathrm{BCL6}}$	100
$s_{1,\mathrm{IRF4}}$	160
$s_{1,\mathrm{BLIMP1}}$	40
$d_{1,\mathrm{BCL6}}$	0.138
$d_{1,\mathrm{IRF4}}$	0.173
$d_{1,\mathrm{BLIMP1}}$	0.173
$k_{\mathrm{off,\,init,\,BCL6}}$	1
$k_{\mathrm{off,\,init,\,IRF4}}$	1
$k_{\mathrm{off,\,init,\,BLIMP1}}$	1

Table 2. Parameter set of the stochastic model (4)-(6) and reduced model (11). Version I - initial parameter set. Version II - parameter set obtained from the automatized approach. Version III - parameter set obtained after semi-manual tuning. Parameters are defined in the text.

The quality-of-fit criterion is then

$$\min_{PS} OF, \tag{10}$$

where PS is the set of parameter values from Tables 2 to 5.

2.6.3 Estimation of the parameters for the stochastic model: Semi-manual tuning

The automatized estimation procedure was followed by a semi-manual tuning of the parameters of the stochastic model (4)-(6) to improve the quality of the fit.
Values of candidate parameters have been tested in an interval of interest and the rest of the parameter values have been fixed at this stage. After model execution, model-simulated SC values of gene expression were collected. Then we selected the values of the parameters that provided the best qualitative fitting (see Equation (10)) of the experimental SC data. Ranges of tested values are summarised in Table 6.

2.7 Evaluation of model variability using Kantorovich distance

To compare distributions and to evaluate model variability, we used the Kantorovich distance (KD, particular case of Wasserstein distance, WD), as defined by Baba et al. [35] and implemented in Python 3 by Bonnaffoux et al. [32].

Parameter	Version I	Version II	Version III
H_{11}	1	0.001	0.1
H_{13}	0.1	1	0.01
$H_{\mathrm{BCR},1}$	0.01	1	0.001
$H_{\mathrm{CD40},2}$	1	0.001	1
θ_{21}	-10	-100	-50
θ_{31}	-2	-20	-0.5
θ_{22}	8	5	11
$\theta_{\mathrm{BCR},1}$	-200	-20	-200
$\theta_{\mathrm{CD40},2}$	10	40	10
$s_{0,\mathrm{IRF4}}$	2	1	2.1
$s_{0,\mathrm{BLIMP1}}$	6.5	1	100

Table 3. Parameter set of the stochastic model (4)-(6) and reduced model (11), presented parameters are different between all versions. Version I - initial parameter set. Version II - parameter set obtained from the automatized approach. Version III - parameter set obtained after semi-manual tuning. Parameters are defined in the text.

Parameter	Version I, II	Version III
H_{22}	0.01	0.001
H_{23}	0.001	0.1
θ_{23}	40	50
$d_{0,\mathrm{BLIMP1}}$	0.1733	0.007

Table 4. Parameter set of the stochastic model (4)-(6) and reduced model (11), presented parameters are equal between versions I and II. Version I - initial parameter set. Version II - parameter set obtained from the automatized approach. Version III - parameter set obtained after semi-manual tuning. Parameters are defined in the text.

Parameter	Version I	Version II, III
H_{21}	0.1	0.01
H_{31}	1	0.01
$s_{0,\mathrm{BCL6}}$	6.5	100
$k_{on,\,\mathrm{init},\,\mathrm{BCL6}}$	0.1	0.15
$k_{on,\,\mathrm{init},\,\mathrm{IRF4}}$	0.0017	0.007
$k_{on,\,\mathrm{init},\,\mathrm{BLIMP1}}$	0.1	0.001

Table 5. Parameter set of the stochastic model (4)-(6) and reduced model (11), presented parameters are equal between versions II and III. Version I - initial parameter set. Version II - parameter set obtained from the automatized approach. Version III - parameter set obtained after semi-manual tuning. Parameters are defined in the text.

Parameter	Definition	Tested values	Selected value
θ_{11}	Interaction parameter	$[-200; -10^{-2}]$	-0.2
θ_{21}	Interaction parameter	$[-200; -10^{-2}]$	-50
θ_{31}	Interaction parameter	$[-200; -10^{-2}]$	-0.5
θ_{22}	Interaction parameter	$[0.1; 200]$	11
θ_{13}	Interaction parameter	$[-200; -0.1]$	-1
θ_{23}	Interaction parameter	$[0.1; 200]$	50
$\theta_{\text{BCR},1}$	Interaction parameter	$[0.1; 200]$	200
$\theta_{\text{CD40},2}$	Interaction parameter	$[0.1; 200]$	10
$s_{0,\text{BCL6}}$	Transcription rate	$[0.1; 625]$	100
$s_{0,\text{IRF4}}$	Transcription rate	$[0.1; 625]$	2.1
$s_{0,\text{BLIMP1}}$	Transcription rate	$[0.1; 625]$	100
$d_{0,\text{BCL6}}$	Degradation rate of mRNA	$[10^{-3}; 10]$	0.05
$d_{0,\text{IRF4}}$	Degradation rate of mRNA	$[10^{-3}; 10]$	0.05
$d_{0,\text{BLIMP1}}$	Degradation rate of mRNA	$[10^{-3}; 10]$	0.007
$s_{1,\text{BCL6}}$	Translation rate	$[1; 1000]$	100
$s_{1,\text{IRF4}}$	Translation rate	$[1; 1000]$	160
$s_{1,\text{BLIMP1}}$	Translation rate	$[1; 1000]$	40
$d_{1,\text{BCL6}}$	Degradation rate of protein	$[0.1; 10]$	0.138
$d_{1,\text{IRF4}}$	Degradation rate of protein	$[0.1; 10]$	0.173
$d_{1,\text{BLIMP1}}$	Degradation rate of protein	$[0.1; 10]$	0.173
$k_{\text{on, init, BCL6}}$	Initial value of $k_{on,\text{BCL6}}$	$[10^{-5}; 10]$	0.15
$k_{\text{on, init, IRF4}}$	Initial value of $k_{\text{on, IRF4}}$	$[10^{-5}; 10]$	0.007
$k_{\text{on, init, BLIMP1}}$	Initial value of $k_{\text{on, BLIMP1}}$	$[10^{-5}; 10]$	0.001

Table 6. Parameters tested during the semi-manual tuning of the stochastic model.

Consider two discrete distributions p and q, defined on N bins of equal sizes, and denote by x_k the center of the k-th bin. Then the Kantorovich distance between p and q is given by

$$\mathrm{KD} = \sum_{n=1}^{N} \left| \sum_{k=1}^{n} p(x_k) - \sum_{k=1}^{n} q(x_k) \right|.$$

We chose WD because it suggested to be preferable over alternative methods such as Kullback-Leibler (KL) divergence or Jensen-Shannon (JS) divergence [36]. More specifically, WD does not require that distributions belong to the same probability space. At the same time, WD is more tractable and has higher performance compared to KL divergence [37]. JS divergence, in turn, does not provide a gradient for the distributions of non-overlapping domains, compared to WD [36]. Also, because WD is a metric and accounts both for the "cost" for the transfer (distance) and "the number of counts" to transfer, we selected its 1D case of WD (Kantorovich Distance, KD) for comparison of discrete experimental distributions versus model-derived distributions [38].

3 Results

3.1 Reduced model

In [3], Martinez et al. applied the kinetic ODE model (1)-(3) to the BCL6-IRF4-BLIMP1 GRN associated with B cell differentiation and successfully simulated GC B cell dynamics based on microarray data. Before using the complex, stochastic model (4)-(6) to fit SC data, we considered a reduced version of System (4)-(6) that can be compared to model (1)-(3), hence providing an initial guess for a key parameter of the model.
Since model (1)-(3) is deterministic, it is necessary to simplify the stochastic model (4)-(6) to perform a comparison of both models dynamics. We assume, in this section, that the stochastic process $E(t)$ (promoter status) in (4)-(6) equals its mean value, $\langle E(t) \rangle$, given by $k_{\mathrm{on}}/(k_{\mathrm{on}} + k_{\mathrm{off}})$. System (4)-(6) then reduces to

$$\begin{cases} \langle E(t) \rangle &= \dfrac{k_{\mathrm{on}}(t)}{k_{\mathrm{on}}(t) + k_{\mathrm{off}}(t)}, \\ \dfrac{dM}{dt} &= s_0 \langle E(t) \rangle - d_0 M(t), \\ \dfrac{dP}{dt} &= s_1 M(t) - d_1 P(t). \end{cases} \tag{11}$$

Comparing mathematical formulations of systems (1)-(3) and (11), one can see that it is possible to identifiy an initial value of the promoter state $E(t)$ for IRF4 gene in System (11) that will correspond to GC differentiation stage (see Supplementary Material A.1). Indeed, after rewriting System (11) in terms of System (1)-(3), we obtained the candidate value of $k_{on,\,\mathrm{init,\,IRF4}} = 1.7 \times 10^{-3}$. Using this value of $k_{on,\,\mathrm{init,\,IRF4}}$, System (11) successfully simulates two steady states for IRF4, i.e. it recapitulates the qualitative dynamics of System (1)-(3) (see Figure 2).

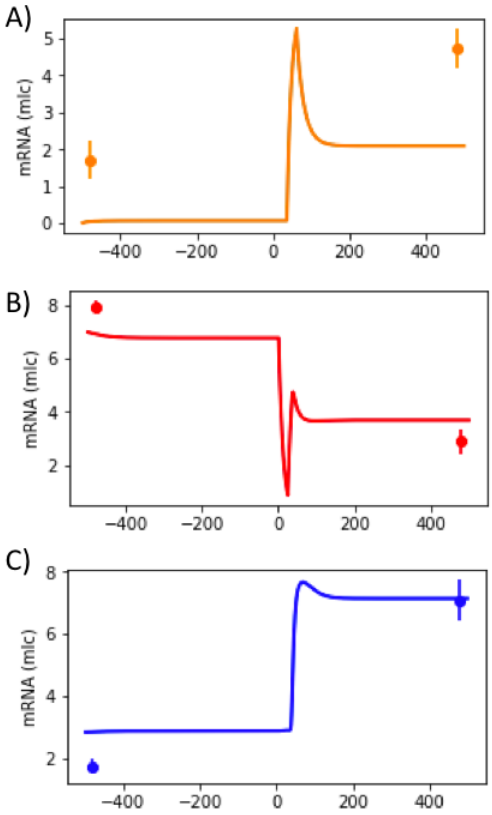

Figure 2. Temporal evolution of mRNA counts of IRF4 (A), BCL6 (B) and BLIMP1 (C) (see Figure 1), generated by the reduced model (11). BCR stimulus was applied from 0h until 25h and CD40 stimulus from 35h until 60h. Microarray gene expression dataset from GEO accession no. GSE12195 was used to estimate model's parameters (see Tables 2 to 5, version I) and are shown as dots with error bars.

Before application of BCR and CD40 stimuli, the system is at a steady state (simulating GC B cell stage) that corresponds to a low amount of IRF4 and BLIMP1 and a high amount of BCL6 mRNA molecules. After application of both stimuli, the system has transitioned to a second steady state that corresponds to a high number of IRF4 and BLIMP1 mRNA molecules and a low number of BCL6 mRNA molecules. However, it can be noted that for the current parameter set (see Tables 2-5, version I), System (11) underestimates the amount of IRF4 mRNA at both steady states (see Figure 2).

Dynamics of System (11) shows the existence of two steady-states for the parameter set from Tables 2-5, version I. Notably, if we test a random value of $k_{on,\,\mathrm{init},\,\mathrm{IRF4}}$ in combination with the parameters from Tables 2-5, version I (see Supplementary Table S1), System (11) has only one steady-state (see Supplementary Figure S2). To our knowledge, there may be more than one set of parameter values associated with two steady states of System (11).

We showed that for the parameter set from Tables 2-5, version I, the reduced model (11) is capable to qualitatively recapitulating the expected behavior of GC B cell differentiation

Figure 3. Model-to-model distributions of KD for GC and PB_PC stages and the three genes, BCL6, IRF4, BLIMP1. Model (4)-(6) was simulated with parameter values from Tables 2-5, version I. The violin plots show the shapes of the distributions, median value, interquartile range and 1.5x interquartile range of the KD values.

GRN (see Figure 2). Next we wanted to understand if the stochastic system (4)-(6) can fit the experimental SC data.

3.2 Stochastic modeling of B cell differentiation

3.2.1 Assessing the variability of the stochastic model

Due to the stochastic nature of the stochastic system (4)-(6), it is important to first evaluate the variability of the model-generated SC data, that is of model's outputs. Indeed, when one repeatedly simulates a finite number of cells from the stochastic system (4)-(6) for the same parameter value set (Tables 2-5, version I), the resulting model-derived empirical distributions are slightly different between each run due to the stochasticity of the model. We investigated how strongly shapes of distributions of simulated SC mRNA molecules vary for different executions of model (4)-(6).

We evaluated the level of variability of model (4)-(6) using the Kantorovich distance (KD, see Section 2.7). We simulated 200 datasets, each containing 200 single cells, of System (4)-(6) with a fixed parameter set (see Tables 2-5, version I). We estimated the KD between pairs of simulated datasets (mRNA counts for three genes at GC and PB_PC stages for 200 simulated cells), and obtained a distribution of all KD that we call the model-to-model (m-t-m) distribution (Figure 3). Shapes of m-t-m distributions are different for each gene and stage of differentiation. For instance, for BLIMP1, long tails are observed. We can also notice that the mean value of IRF4 at GC stage is low compared to other genes.

In order to get a more accurate evaluation of the variability in model's outputs, we plotted distributions of the number of mRNA molecules (model's outputs) for each node of the GRN with the highest m-t-m distribution at both GC and PB_PC stages (Figure 4). Qualitatively, no difference is detected in the shapes of model-generated distributions. For all 6 nodes, the shapes of distributions are remarkably similar.

These results suggest that it may be sufficient to perform parameter tuning of the stochastic model (4)-(6) using only one simulation run for each parameter value set.

3.2.2 Initial estimation step based on an automatized approach

Variability of the stochastic model being assessed, and comparison of experimental data and a single model's output in order to assess their closeness being validated, we now focus on the estimation of parameter values. Model (4)-(6) comprises 40 parameters, so we first apply a straightforward strategy, that we call automatized approach, which consists in solving the stochastic system (4)-(6) for a number of fixed parameter values

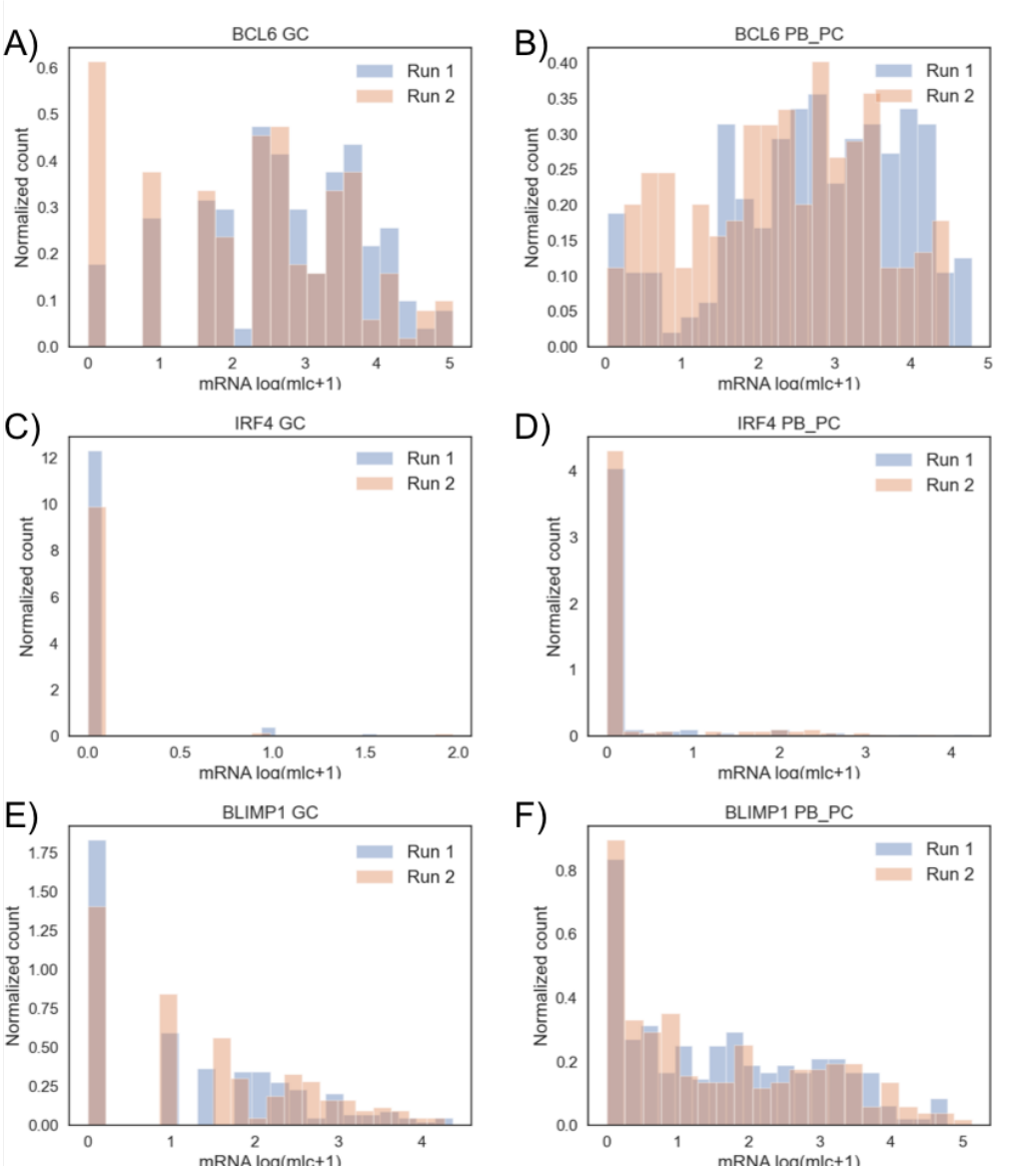

Figure 4. Histograms of two model-generated mRNA counts of BCL6, IRF4 and BLIMP1 at GC and PB_PC stages with the highest KD. The subgraphs A, C, E represent log2 (molecule+1) transformed values for BCL6, IRF4 and BLIMP1 at GC stage. The subgraphs B, D, F represent log (molecule+1) transformed values for BCL6, IRF4 and BLIMP1 at PB_PC stage (Parameters from Tables 2-5, version I).

and selecting the set of parameter values associated with the best fit (see Section 2.6.2) of experimental data [30].
Approximately 8×10^6 combinations of parameter values have been tested (see Section 2.6.2), then the best set of parameter values has been selected based on the quality of BCL6, IRF4 and BLIMP1 fitting at the GC and PB_PC stages (see Equations (9)-(10)). Numbers of mRNA molecules estimated by the stochastic model (4)-(6) are in a similar range of magnitude as the experimental SC data (see Supplementary Figure S3). However, the selected parameter values (Tables 2-5, version II) generate model-derived mRNA distributions that have sufficient overlap with experimental data for GC stage but insufficient overlap for PB_PC stage (see Supplementary Figure S3). Indeed, distributions of numbers of mRNA molecules at PB_PC stage mostly underestimate the experimental SC data (see Supplementary Figure S3B, D and F).
Implementing an automatized approach for estimating parameter values helped to establish a set of parameter values that allows System (4)-(6) to correctly estimate the number of mRNA molecules for 3 out of 6 nodes of the GRN. In order to improve the quality of the fit, a more directed and sensitive tuning of the parameter set is then performed (see Section 2.6.3).

3.2.3 Generation of simulated distributions of mRNA counts describing B cell differentiation

Due to the complexity of the stochastic model (4)-(6) that is made of 40 parameters, it is important to identifiy which parameters should be targeted to improve the quality of fit. To do so, we rely on the properties of the GRN (see Figure 1A). Thanks to the topological structure of the BCL6-IRF4-BLIMP1 GRN, where IRF4 activates BLIMP1 and autoactivates itself, we hypothesize that System (4)-(6) underestimates the experimental SC data at the PB_PC stage due to low values of the parameters responsible for IRF4 autoactivation (θ_{22}, and to a lesser extent $s_{0,\,\mathrm{IRF4}}$) and BLIMP1 activation by IRF4 (θ_{23}). Further, we improved the quality of the fit, in particular of BLIMP1 distribution, by focusing on BLIMP1-related interaction parameters (θ_{13}, θ_{31}).
Indeed, if IRF4 autoactivation reaction is not efficient enough, there are not enough IRF4 molecules to affect BCL6 and BLIMP1 activity at PB_PC stage. Because IRF4 activity is only impacted by its autoactivation loop, we first modulated values of the parameter related to this reaction (θ_{22}). During preliminary tests, we noticed that this reaction is crucial for the transition from GC towards PB_PC stage and that when interaction parameter θ_{22} and transcription rate $s_{0,\,\mathrm{IRF4}}$ have low absolute values then the system cannot reach PB_PC stage, even after application of the stimuli. It can be explained by the insufficient amount of IRF4 molecules produced (see Supplementary Figure S3, C and D). On the other hand, when parameters θ_{22} and $s_{0,\,\mathrm{IRF4}}$ have high values, model (4)-(6) transitions from GC towards PB_PC stage even before application of stimuli, exhibiting an abnormal behavior.
After comparison of the stochastic system (4)-(6) outputs for a range of different θ_{22} and $s_{0,\,\mathrm{IRF4}}$ values (described in Table 6), we selected the parameter set for which model (4)-(6)

correctly fits the IRF4 experimental data at both GC and PB_PC stages. Such model-derived SC pattern is obtained using the values ($\theta_{22} = 11$ and $s_{0,\,\mathrm{IRF4}} = 2.1$ molecule.h^{-1}) We additionally performed simulations to improve the quality of the fitting of BLIMP1 and BCL6 distributions by testing parameters that are directly responsible for the balance between BLIMP1 and BCL6, such as interaction parameters θ_{13}, θ_{31} and θ_{23}. We also tested parameters which can influence BCL6 and BLIMP1 indirectly, such as transcription rates ($s_{0,\,\mathrm{BCL6}}$ and $s_{0,\,\mathrm{BLIMP1}}$), and degradation rates of mRNA ($d_{0,\,\mathrm{BCL6}}$, $d_{0,\,\mathrm{IRF4}}$ and $d_{0,\,\mathrm{BLIMP1}}$).

After comparison of the stochastic system (4)-(6) outputs, we selected the parameters which allow the model to have a qualitative fit of the experimental data for all nodes at GC and PB_PC stages (see Figure 5, and Tables 2-5, version III). For this tuned parameter set, we see that the model (4)-(6) can have a good qualitative fitting of experimental data for all nodes. Results also show that for this parameter set (version III), the stochastic model (4)-(6) fits SC data at the GC stage for BCL6 (see Figure 5A). The model-derived empirical distribution of BLIMP1 was capable of showing overlap with experimental data at the PB_PC stage (see Figure 5F), but it overestimated the number of BLIMP1 mRNA molecules at the GC stage (see Figure 5E).

The current parameter set (Tables 2-5, version III) has difficulties to correctly evaluate the number of zero values. The model (4)-(6) tends to overestimate the number of BCL6 mRNA molecules at PB_PC stage, as well as the number of IRF4 mRNA molecules at GC stage and number of BLIMP1 mRNA molecules at GC stage (see Figure 5). Nevertheless, this parameter set allowed the model to generate SC data with a similar level of magnitude of the amount of mRNA as experimentally observed.

4 Discussion

In this work, we applied a particular class of stochastic models combining deterministic dynamics and random jumps to the simulation of SC data from two stages of B cell differentiation in germinal centers.

We first defined a reduced model (11) whose dynamics were compared to the ones of the kinetic model (1)-(3) and we established an initial parameter value for the key parameter $k_{on,\,\mathrm{init},\,\mathrm{IRF4}}$. We then showed that for a given parameter set (Table 2-5, Version I), the reduced model (11) admits two steady states. Secondly, we evaluated the effect of stochasticity on multiple independent generations of the number of mRNA molecules by the stochastic model (4)-(6) and we confirmed that for the same parameter set there is no noticeable difference between each model-generated outputs for BCL6-IRF4-BLIMP1 GRNs (see Figure 4). These results allow performing a combined parameter screening with the confidence that for each candidate parameter set, the algorithm needs to perform only one run of the model (4)-(6). Lastly, we showed that the model (4)-(6) can simulate distributions of the number of mRNA molecules for BCL6, IRF4, BLIMP1 at GC and PB_PC stages with the same order of magnitude as experimental data. However, as future scope of this work, a few strategies to improve the final parameter value set (Tables 2-5, version III) can be investigated.

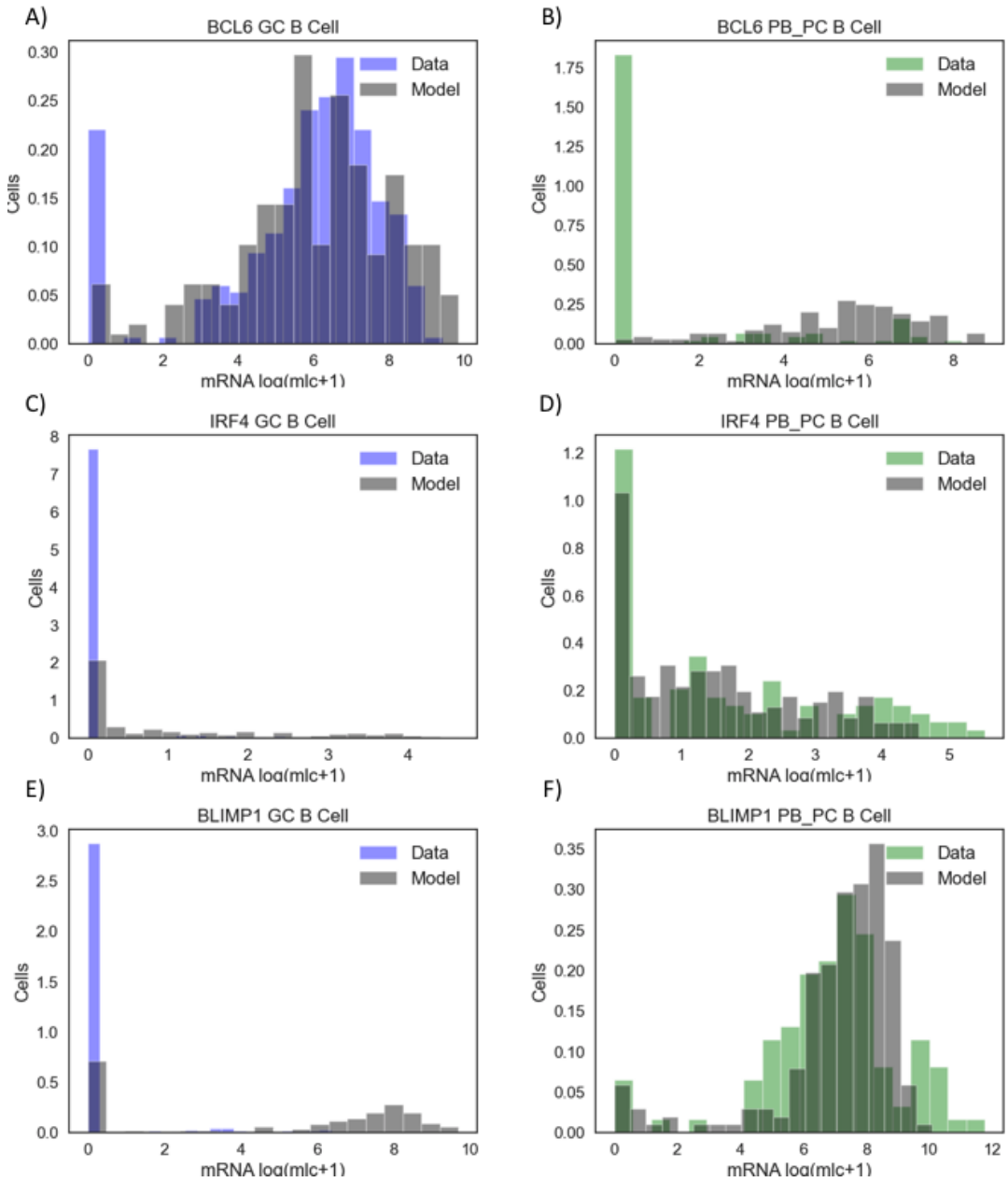

Figure 5. Histograms of model-generated and experimental mRNA counts of BCL6, IRF4, BLIMP1 at GC and PB_PC stages. The subgraphs A, C, E represent log (molecule+1) transformed SC values with BCL6, IRF4 and BLIMP1 compared between the model estimations at GC stage (grey) vs the experimental data from GC B cells (blue). The subgraphs B, D, F represent log (molecule+1) transformed SC values with BCL6, IRF4 and BLIMP1 compared between the model estimations at PB_PC stage (grey) vs the experimental data from PB_PC cells (green). Simulation of 200 single cells were used based on the parameter set, selected after semi-automatized parameter screening (see Tables 2-5, version III). Performed based on the dataset from Milpied et al. [30]

Since in BCL6-IRF4-BLIMP1 GRN, IRF4 activity depends only on its autoactivation reaction, we have only succeeded, by writing the reduced model (11) in terms of the kinetic model (1)-(3), in estimating the value of $k_{on,\,\mathrm{init},\,\mathrm{IRF4}}$. It would be advantageous to additionally estimate the values of $k_{on,\,\mathrm{init},\,\mathrm{BCL6}}$ and $k_{on,\,\mathrm{init},\,\mathrm{BLIMP1}}$, using the same logic. However, because BLIMP1 depends on BLIMP1, IRF4 and BCL6 (see Equation (1)) and BCL6 depends on both IRF4 and BLIMP1 (see Equation (2)), the rewriting of system (4)-(6) in terms of (1)-(3) would require additional calculations and simplifications.

The effect of mutual repression between BCL6 and BLIMP1 could be evaluated by performing a more extensive parameter value search. The current parameter value set (Tables 2-5, version III) makes model (4)-(6) overestimate the number of mRNA molecules of BLIMP1 at GC stage. Increasing BCL6 repression of BLIMP1 could potentially decrease the quantity of BLIMP1 at the GC stage.

The effect of the duration of the BCR and CD40 stimuli on the differentiation from GC B cells towards PB_PC could be investigated. Multiscale modeling of GCs performed by Tejero et al. [39] showed that CD40 signalling in combination with the asymmetric division of B cells results in a switch from memory B cells to plasmablasts. It would be relevant to evaluate a possible application of the stochastic model to study the effect of combined CD40 and BCR signaling with different intensities and durations at the SC level.

Additionally, one can evaluate the impact of including additional genes into the BCL6-IRF4-BLIMP1 GRN on the quality of data fitting by the stochastic model. One of the possible candidates to incorporation in the GRN is PAX5, which plays an important role in directing lymphoid progenitors towards B cell development [40]. PAX5 positively regulates IRF8 and BACH2, which indirectly positively regulate IRF4 and negatively regulate BLIMP1 at an early stage of B cell differentiation. During further development, BLIMP1 starts to repress PAX5, consequently decreasing the expression of IRF8 and BACH2. The correct orchestration of PAX5-IRF8-BACH2 during B cell differentiation is important for the successful differentiation towards antigen producing cells (PB_PC), while its malfunction can cause aberration in GC B cell development [41].

CD40 stimulation of B cells initiates NF-κB signaling which is associated with cellular proliferation. In B cells, NF-κB activates IRF4, negatively regulates BACH2, what leads to positive regulation of BLIMP1 and consecutive repression of BCL6 [5, 34].

Another important transcription factor in GC development is MYC, which regulates B cell proliferation [42] and the DZ B cell phenotype [43]. MYC indirectly activates the histone methyltransferase enhancer of zeste homologue 2 (EZH2), which is responsible for the repression of IRF4 and BLIMP1 [44, 45, 46, 47].

The transcription factors mentioned above are present in the SC RT-qPCR dataset from Milpied et al. [30] that we used and could be used to extend the current GRN. Inclusion of additional transcription factors may have both positive and negative effects on the application of model (4)-(6). On one side, it can increase the computational time and the number of parameters required for simulating System (4)-(6). On the other side, because the inclusion of transcription factors can more precisely describe the biological system it could improve the quality of the fitting. However, any inclusion of new nodes to GRN

should be carefully evaluated and only essential transcription factors should be added. For instance, there are no advantages in adding a transcrption factor that would only have one downstream output. As an example, MYC activates E2F1 and further activates EZH2. For this reason, incorporation of the chain MYC-E2F1-EZH2 should have a similar outcome, as the incorporation of the simplified MYC-EZH2 reaction. This is expected because in the modeling, intermediate elements of one-to-one redundant reactions can be omitted without significant changes in the quality of the simulations.

To further continue our study, we could also use SC RNA-seq dataset from Milpied et. al [30]. The authors have produced SC RNA-seq dataset from GC B cells and analysed the similarities between SC RNA-seq and SC RT-qPCR dataset. Even though the gene-gene correlation levels were lower in SC RNA-seq compared to SC RT-qPCR, SC RNA-seq analysis confirmed the observation obtained by SC RT-qPCR [30]. From the stochastic modeling perspective, combining the data from SC RT-qPCR and SC RNA-seq should improve our understanding of the SC dataset variability and the quality of the fitting.

To summarise, the stochastic model (4)-(6) is capable of qualitatively simulating and depicting the stochasticity of experimental SC gene expression data of human B cell differentiation at the GC and PB_PC stages using a GRN made of three-key genes (BCL6, IRF4, BLIMP1). These results are encouraging, and suggest that our model may be used to test the different B cell exits from GC. Future steps may include testing of the model (4)-(6) on alternative SC datasets [48, 49, 50] and investigating the malignant formations, by evaluating differences of the associated GRN compared to the normal B cell differentiation from GC towards PB_PC.

Funding

This work was supported by the COSMIC grant (www.cosmic-h2020.eu) which has received funding from European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 765158.

Acknowledgments

We thank Arnaud Bonnaffoux and Matteo Bouvier for their help with the WASABI framework and their critical reading of the manuscript. We thank the computational center of IN2P3 (Villeurbanne/France), especially Gino Marchetti and Renaud Vernet. We also thank Aurelien Pélissier and Elias Ventre for their scientific discussions.

References

[1] Luka Mesin, Jonatan Ersching, and Gabriel D Victora. "Germinal center B cell dynamics". In: *Immunity* 45.3 (2016), pp. 471–482.

[2] Tomohiro Kurosaki, Kohei Kometani, and Wataru Ise. "Memory B cells". In: *Nature Reviews Immunology* 15.3 (2015), pp. 149–159.

[3] María Rodríguez Martínez, Alberto Corradin, Ulf Klein, Mariano Javier Álvarez, Gianna M Toffolo, Barbara di Camillo, Andrea Califano, and Gustavo A Stolovitzky. "Quantitative modeling of the terminal differentiation of B cells and mechanisms of lymphomagenesis". In: *Proceedings of the National Academy of Sciences* 109.7 (2012), pp. 2672–2677.

[4] Shuang Song and Patrick D Matthias. "The transcriptional regulation of germinal center formation". In: *Frontiers in immunology* 9 (2018), p. 2026.

[5] Masumichi Saito, Jie Gao, Katia Basso, Yukiko Kitagawa, Paula M Smith, Govind Bhagat, Alessandra Pernis, Laura Pasqualucci, and Riccardo Dalla-Favera. "A signaling pathway mediating downregulation of BCL6 in germinal center B cells is blocked by BCL6 gene alterations in B cell lymphoma". In: *Cancer cell* 12.3 (2007), pp. 280–292.

[6] Hans-Willi Mittrücker, Toshifumi Matsuyama, Alex Grossman, Thomas M Kündig, Julia Potter, Arda Shahinian, Andrew Wakeham, Bruce Patterson, Pamela S Ohashi, and Tak W Mak. "Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function". In: *Science* 275.5299 (1997), pp. 540–543.

[7] Laura Pasqualucci, Anna Migliazza, Katia Basso, Jane Houldsworth, RSK Chaganti, and Riccardo Dalla-Favera. "Mutations of the BCL6 proto-oncogene disrupt its negative autoregulation in diffuse large B-cell lymphoma". In: *Blood, The Journal of the American Society of Hematology* 101.8 (2003), pp. 2914–2923.

[8] Nilushi S De Silva, Giorgia Simonetti, Nicole Heise, and Ulf Klein. "The diverse roles of IRF4 in late germinal center B-cell differentiation". In: *Immunological reviews* 247.1 (2012), pp. 73–92.

[9] Stephen L Nutt, Kirsten A Fairfax, and Axel Kallies. "BLIMP1 guides the fate of effector B and T cells". In: *Nature Reviews Immunology* 7.12 (2007), pp. 923–927.

[10] Yi Lin, Kwok-kin Wong, and Kathryn Calame. "Repression of c-myc transcription by Blimp-1, an inducer of terminal B cell differentiation". In: *Science* 276.5312 (1997), pp. 596–599.

[11] Daniel Radtke and Oliver Bannard. "Expression of the plasma cell transcriptional regulator Blimp-1 by dark zone germinal center B cells during periods of proliferation". In: *Frontiers in immunology* 9 (2019), p. 3106.

[12] Julie Tellier, Wei Shi, Martina Minnich, Yang Liao, Simon Crawford, Gordon K Smyth, Axel Kallies, Meinrad Busslinger, and Stephen L Nutt. "Blimp-1 controls plasma cell function through the regulation of immunoglobulin secretion and the unfolded protein response". In: *Nature immunology* 17.3 (2016), pp. 323–330.

[13] Martina Minnich, Hiromi Tagoh, Peter Bönelt, Elin Axelsson, Maria Fischer, Beatriz Cebolla, Alexander Tarakhovsky, Stephen L Nutt, Markus Jaritz, and Meinrad Busslinger. "Multifunctional role of the transcription factor Blimp-1 in coordinating plasma cell differentiation". In: *Nature immunology* 17.3 (2016), pp. 331–343.

[14] Roger Sciammas, AL Shaffer, Jonathan H Schatz, Hong Zhao, Louis M Staudt, and Harinder Singh. "Graded expression of interferon regulatory factor-4 coordinates isotype switching with plasma cell differentiation". In: *Immunity* 25.2 (2006), pp. 225–236.

[15] AL Shaffer, Xin Yu, Yunsheng He, Jennifer Boldrick, Erick P Chan, and Louis M Staudt. "BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control". In: *Immunity* 13.2 (2000), pp. 199–212.

[16] Laura Pasqualucci, David Dominguez-Sola, Annalisa Chiarenza, Giulia Fabbri, Adina Grunn, Vladimir Trifonov, Lawryn H Kasper, Stephanie Lerach, Hongyan Tang, Jing Ma, et al. "Inactivating mutations of acetyltransferase genes in B-cell lymphoma". In: *Nature* 471.7337 (2011), pp. 189–195.

[17] Oksana R Bereshchenko, Wei Gu, and Riccardo Dalla-Favera. "Acetylation inactivates the transcriptional repressor BCL6". In: *Nature genetics* 32.4 (2002), pp. 606–613.

[18] AL Shaffer, Kuo-I Lin, Tracy C Kuo, Xin Yu, Elaine M Hurt, Andreas Rosenwald, Jena M Giltnane, Liming Yang, Hong Zhao, Kathryn Calame, et al. "Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program". In: *Immunity* 17.1 (2002), pp. 51–62.

[19] Geng Chen, Baitang Ning, and Tieliu Shi. "Single-cell RNA-seq technologies and related computational data analysis". In: *Frontiers in genetics* 10 (2019), p. 317.

[20] Seitaro Nomura. "Single-cell genomics to understand disease pathogenesis". In: *Journal of Human Genetics* 66.1 (2021), pp. 75–84.

[21] Arjun Raj, Scott A Rifkin, Erik Andersen, and Alexander Van Oudenaarden. "Variability in gene expression underlies incomplete penetrance". In: *Nature* 463.7283 (2010), pp. 913–918.

[22] Arjun Raj, Charles S Peskin, Daniel Tranchina, Diana Y Vargas, and Sanjay Tyagi. "Stochastic mRNA synthesis in mammalian cells". In: *PLoS Biol* 4.10 (2006), e309.

[23] Jonathan R Chubb, Tatjana Trcek, Shailesh M Shenoy, and Robert H Singer. "Transcriptional pulsing of a developmental gene". In: *Current biology* 16.10 (2006), pp. 1018–1025.

[24] David M Suter, Nacho Molina, David Gatfield, Kim Schneider, Ueli Schibler, and Felix Naef. "Mammalian genes are transcribed with widely different bursting kinetics". In: *science* 332.6028 (2011), pp. 472–474.

[25] Arjun Raj and Alexander Van Oudenaarden. "Nature, nurture, or chance: stochastic gene expression and its consequences". In: *Cell* 135.2 (2008), pp. 216–226.

[26] Cameron P Gallivan, Honglei Ren, and Elizabeth L Read. "Analysis of single-cell gene pair coexpression landscapes by stochastic kinetic modeling reveals gene-pair interactions in development". In: *Frontiers in genetics* 10 (2020), p. 1387.

[27] Payam Dibaeinia and Saurabh Sinha. "Sergio: a single-cell expression simulator guided by gene regulatory networks". In: *Cell Systems* 11.3 (2020), pp. 252–271.

[28] Mark HA Davis. "Piecewise-deterministic Markov processes: a general class of non-diffusion stochastic models". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 46.3 (1984), pp. 353–376.

[29] Ulysse Herbach, Arnaud Bonnaffoux, Thibault Espinasse, and Olivier Gandrillon. "Inferring gene regulatory networks from single-cell data: a mechanistic approach". In: *BMC Systems Biology* 11.1 (2017), pp. 1–15.

[30] Pierre Milpied, Inaki Cervera-Marzal, Marie-Laure Mollichella, Bruno Tesson, Gabriel Brisou, Alexandra Traverse-Glehen, Gilles Salles, Lionel Spinelli, and Bertrand Nadel. "Human germinal center transcriptional programs are de-synchronized in B cell lymphoma". In: *Nature immunology* 19.9 (2018), pp. 1013–1024.

[31] Louise J McHeyzer-Williams, Pierre J Milpied, Shinji L Okitsu, and Michael G McHeyzer-Williams. "Class-switched memory B cells remodel BCRs within secondary germinal centers". In: *Nature immunology* 16.3 (2015), pp. 296–305.

[32] Arnaud Bonnaffoux, Ulysse Herbach, Angélique Richard, Anissa Guillemin, Sandrine Gonin-Giraud, Pierre-Alexis Gros, and Olivier Gandrillon. "WASABI: a dynamic iterative framework for gene regulatory network inference". In: *BMC Bioinformatics* 20.1 (2019), pp. 1–19.

[33] Jean Peccoud and Bernard Ycart. "Markovian modeling of gene-product synthesis". In: *Theoretical population biology* 48.2 (1995), pp. 222–234.

[34] Brian J Laidlaw and Jason G Cyster. "Transcriptional regulation of memory B cell differentiation". In: *Nature Reviews Immunology* 21.4 (2021), pp. 209–220.

[35] Akinori Baba and Tamiki Komatsuzaki. "Construction of effective free energy landscape from single-molecule time series". In: *Proceedings of the National Academy of Sciences* 104.49 (2007), pp. 19297–19302.

[36] Soheil Kolouri, Phillip E Pope, Charles E Martin, and Gustavo K Rohde. "Sliced Wasserstein auto-encoders". In: *International Conference on Learning Representations*. 2018.

[37] Alain Rakotomamonjy, Abraham Traoré, Maxime Berar, Rémi Flamary, and Nicolas Courty. "Distance measure machines". In: *arXiv preprint arXiv:1803.00250* (2018).

[38] Gabriel Peyré, Marco Cuturi, et al. "Computational optimal transport: With applications to data science". In: *Foundations and Trends® in Machine Learning* 11.5-6 (2019), pp. 355–607.

[39] Elena Merino Tejero, Danial Lashgari, Rodrigo García-Valiente, Xuefeng Gao, Fabien Crauste, Philippe A Robert, Michael Meyer-Hermann, María Rodríguez Martínez, S Marieke van Ham, Jeroen EJ Guikema, et al. "Multiscale Modeling of Germinal Center Recapitulates the Temporal Transition From Memory B Cells to Plasma Cells Differentiation as Regulated by Antigen Affinity-Based Tfh Cell Help". In: *Frontiers in immunology* 11 (2020).

[40] Jasna Medvedovic, Anja Ebert, Hiromi Tagoh, and Meinrad Busslinger. "Pax5: a master regulator of B cell development and leukemogenesis". In: *Advances in immunology* 111 (2011), pp. 179–206.

[41] Stephen L Nutt, Philip D Hodgkin, David M Tarlinton, and Lynn M Corcoran. "The generation of antibody-secreting plasma cells". In: *Nature Reviews Immunology* 15.3 (2015), pp. 160–171.

[42] Dinis Pedro Calado, Yoshiteru Sasaki, Susana A Godinho, Alex Pellerin, Karl Köchert, Barry P Sleckman, Ignacio Moreno De Alborán, Martin Janz, Scott Rodig, and Klaus Rajewsky. "The cell-cycle regulator c-Myc is essential for the formation and maintenance of germinal centers". In: *Nature immunology* 13.11 (2012), pp. 1092–1100.

[43] Hirokazu Tanaka, Itaru Matsumura, Sachiko Ezoe, Yusuke Satoh, Toshiyuki Sakamaki, Chris Albanese, Takashi Machii, Richard G Pestell, and Yuzuru Kanakura. "E2F1 and c-Myc potentiate apoptosis through inhibition of NF-κB activity that facilitates MnSOD-mediated ROS elimination". In: *Molecular cell* 9.5 (2002), pp. 1017–1029.

[44] Wendy Béguelin, Martín A Rivas, María T Calvo Fernández, Matt Teater, Alberto Purwada, David Redmond, Hao Shen, Matt F Challman, Olivier Elemento, Ankur Singh, et al. "EZH2 enables germinal centre formation through epigenetic silencing of CDKN1A and an Rb-E2F1 feedback loop". In: *Nature communications* 8.1 (2017), pp. 1–16.

[45] Frank M Raaphorst, Folkert J van Kemenade, Elly Fieret, Karien M Hamer, David PE Satijn, Arie P Otte, and Chris JLM Meijer. "Cutting edge: polycomb gene expression patterns reflect distinct B cell differentiation stages in human germinal centers". In: *The Journal of Immunology* 164.1 (2000), pp. 1–4.

[46] Irina Velichutina, Rita Shaknovich, Huimin Geng, Nathalie A Johnson, Randy D Gascoyne, Ari M Melnick, and Olivier Elemento. "EZH2-mediated epigenetic silencing in germinal center B cells contributes to proliferation and lymphomagenesis". In: *Blood, The Journal of the American Society of Hematology* 116.24 (2010), pp. 5247–5255.

[47] Laurie Herviou, Michel Jourdan, Anne-Marie Martinez, Giacomo Cavalli, and Jerome Moreaux. "EZH2 is overexpressed in transitional preplasmablasts and is involved in human plasma cell differentiation". In: *Leukemia* 33.8 (2019), pp. 2047–2060.

[48] Hamish W King, Nara Orban, John C Riches, Andrew J Clear, Gary Warnes, Sarah A Teichmann, and Louisa K James. "Single-cell analysis of human B cell maturation predicts how antibody class switching shapes selection dynamics". In: *Science Immunology* 6.56 (2021), eabe6291.

[49] Noemi Andor, Erin F Simonds, Debra K Czerwinski, Jiamin Chen, Susan M Grimes, Christina Wood-Bouwens, Grace XY Zheng, Matthew A Kubit, Stephanie Greer, William A Weiss, et al. "Single-cell RNA-Seq of follicular lymphoma reveals malignant B-cell types and coexpression of T-cell immune checkpoints". In: *Blood, The Journal of the American Society of Hematology* 133.10 (2019), pp. 1119–1129.

[50] Antony B Holmes, Clarissa Corinaldesi, Qiong Shen, Rahul Kumar, Nicolo Compagno, Zhong Wang, Mor Nitzan, Eli Grunstein, Laura Pasqualucci, Riccardo Dalla-Favera, et al. "Single-cell analysis of germinal-center B cells informs on lymphoma cell of origin and outcome". In: *Journal of Experimental Medicine* 217.10 (2020).