In silico modelling of CD8 T cell immune response links genetic regulation to population dynamics

Thi Nhu Thao Nguyen, Madge Martin, C Arpin, Samuel Bernard, Olivier Gandrillon, Fabien Crauste

In silico modelling of CD8 T cell immune response links genetic regulation to population dynamics

Thi Nhu Thao Nguyen[1,2], Madge Martin[3], Christophe Arpin[1,4], Samuel Bernard[2,5], Olivier Gandrillon[1,4,7], Fabien Crauste[6,7]

1 septembre 2024

[1] Univ Lyon, ENS de Lyon, Univ Claude Bernard, CNRS UMR 5239, INSERM U1210, Laboratory of Biology and Modelling of the Cell, 46 allée d'Italie Site Jacques Monod, F-69007, Lyon, France

[2] Inria, Villeurbanne, 69693, France.

[3] CNRS, Univ Paris Est Creteil, Univ Gustave Eiffel, UMR 8208, MSME, F-94010, Créteil, France

[4] Centre International de Recherche en Infectiologie, Université de Lyon, ENS de Lyon, Université Claude Bernard, CNRS UMR 5308, INSERM U1111, Lyon, France

[5] Univ Lyon, Université Lyon 1, CNRS UMR5208, Institut Camille Jordan, 43 Blvd du 11 Novembre 1918, Villeurbanne-Cedex, F-69622, France

[6] Université Paris Cité, CNRS, MAP5, F-75006, Paris, France

[7] Corresponding authors

Abstract

The CD8 T cell immune response operates at multiple temporal and spatial scales, including all the early complex biochemical and biomechanical processes, up to long term cell population behavior.

In order to model this response, we devised a multiscale agent-based approach using `Simuscale` software. Within each agent (cell) of our model, we introduced

a gene regulatory network (GRN) based upon a piecewise deterministic Markov process (PDMP) formalism. Cell fate – differentiation, proliferation, death – was coupled to the state of the GRN through rule-based mechanisms. Cells interact in a 3D computational domain and signal to each other via cell-cell contacts, influencing the GRN behavior.

Results show the ability of the model to correctly capture both population behaviour and molecular time-dependent evolution. We examined the impact of several parameters on molecular and population dynamics, and demonstrated the add-on value of using a multiscale approach by showing the influence of molecular parameters, particularly protein degradation rates, on the outcome of the response, such as effector and memory cell counts.

Keywords: Gene Regulatory Networks, Cell population dynamics, CD8 T cell immune response, Stochastic gene expression, Multiscale modeling

1 Introduction

CD8 T cells are important for immune responses against viruses and intracellular bacteria, as well as for tumor surveillance. A naive CD8 T cell gets activated when it recognizes an antigen presenting cell (APC), through the formation of an immunological synapse [1]. Activated CD8 T cells first give rise to proliferating memory precursor (MP) cells [2]. Such MP cells could represent bipotential cells that face a choice between two fates: the terminally differentiated effector fate that is associated with the repression of their self-renewing capacity and the activation of their effector function, and the memory precursor fate that maintains their self-renewing capacity. Effector T cells massively proliferate while acquiring their effector potential, which allows them to kill infected or antigen-bearing malignant cells, before dying during the contraction phase. Meanwhile, part of MP cells differentiate into memory cells, providing long-term protection against reinfection [3] The capacity to generate long-lived memory cells during a primary immune response forms the basis for vaccination. At the end of the response, the number of memory cells remains stable, since memory cells hardly ever proliferate or die [4, 5].

This entire process functions across various temporal and spatial scales, encompassing intricate early biochemical and biomechanical processes, extending to the long-term behavior of cell populations. Building a comprehensive computational model, where all relevant scales and their interactions would be represented, would hold the potential to introduce new and more robust principles for designing vaccines aimed at swiftly adapting pathogens. This perspective motivated the recent development of advanced mathematical and computational models to depict these multiscale phenomena, moving us towards a more comprehensive understanding of the CD8+ T cell response [6].

There is a rich ecosystem of software developed to describe cell dynamics systems at different scales. Software can be distributed into three main classes:

1. Description of cell population dynamics, such as Compucell3D [7], Physicell [8] or CellSys [9].

2. Description of both cellular and intracellular scale dynamics, such as Virtual Cell [10], COPASI [11] or Smoldyn [12].

3. Description of coupling at least two different scales, like Vivarium [13], ENISI-MSM [14], EpiLog [15] coupled with COPASI, MSM [16] or Tissue Forge [17]. One should also cite PhysiBoSS [18] which results from the coupling of Physicell (processing up to 10^6 cells, but needs to run for several days) with MaBoSS, a tool based on Boolean modeling [19].

These tools are limited in some aspects for example in the number of cells they can simulate, the explicit description of a molecular level or their ability to deal with different cell types. Their computation time is also often extremely slow, limiting the relevance of such computational models for performing parameter estimation and model fitting, at least when dealing with highly proliferating cells. Indeed, when the number of cells is only a few thousands, the calculation process becomes cumbersome and time-consuming.

The development of a multiscale model of the immune response requires the ability to simulate the molecular state of an expanding T cell population over several scales at

the single-cell level, using a realistic GRN that can reproduce stochastic gene expression behavior.

In [20, 21], CompuCell3D was coupled with a molecular network described by an ODE system. Although these latest studies have qualitatively captured expected cellular and molecular behaviors, enabling cellular decision-making, the molecular network was modeled as a fully deterministic system using ODEs, whereas it is now accepted that gene expression at the single cell level is a stochastic process [22, 23, 24, 25, 26, 27, 28].

We therefore describe here the use of `Simuscale` [29], which enabled us to simulate the dynamics of interactions from the molecular to the cell-population scale, in conjunction with the use of a biologically realistic mechanistic GRN based on a stochastic 2-state model for gene expression [30]. Noticeably, we introduce a GRN that does not include real genes (as in [20, 21, 31] for instance) since such approaches did not prove neither to allow better characterization of the biological process nor to provide access to relevant parameter values. We then decided to highlight the feasibility of implementing a multiscale computational model whose GRN would enable to reproduce the dynamics observed with single cell RNA-sequencing data by using an ad hoc GRN, that could be later replaced by an experimentally inferred GRN. `Simuscale` is particularly relevant to model and simulate differentiating cell populations, whose dynamics are not solely dependent on biophysical rules.

Thanks to appropriate parameter calibration, this study demonstrates the ability to capture the expected time-dependent evolution of CD8 T cell population dynamics. In addition, issues related to simulation time, variability, the dependence on the initial condition of the number of cells in the population and differentiation states, were also addressed in this study. Finally, we demonstrate the benefit of multiscale coupling by assessing the population behavior (time of the peak of the response) as a function of a molecular parameter (protein half-life).

2 Material, Methods and Models

2.1 Simuscale

`Simuscale` [29] is a multiscale individual-based modeling platform for performing numerical simulations of heterogeneous populations of individual cells. Cells are assumed to evolve in time and interact physically and biochemically with each others. Models are described at two levels: cellular and population level. The cellular level describes the dynamics of single cells, as defined by the user/modeler. Cells have an internal state that includes default properties such as cell size and position, and may also include cell-specific states (*e.g.*, gene or protein expression). The population level describes the mechanical constraints and biochemical interactions between cells. Cells evolve in a bounded 3D domain, and can divide or die. Cells are represented spatially as visco-elastic spheres with a rigid core.

`Simuscale` implements the physical simulator that manages the simulations at the population level. Details of cellular dynamics to each cell are to be defined by the user. This makes `Simuscale` modular, as it can accommodate any number of cell models within the same simulation, including models with different modelling formalisms, such as ordinary and stochastic differential equations and up to Piecewise Deterministic Markov Process (PDMPs, [30]). Biochemical interactions occur between cells that are in contact with each other, through intercellular signals. Intercellular signals can be known to all or to a subset of the cells only.

`Simuscale` expects an input file describing the initial cell population and numerical options. It runs a simulation over a specified time interval, updating the cell population at given time steps, and it generates an output file containing the state of each cell at each time step, and the tree of cell divisions and deaths. All details regarding `Simuscale` can be found in Bernard et al. [29]

2.2 Gene Regulatory Network

The first step in building a multiscale model of the CD8 T cell immune response is to build a GRN whose dynamics will drive each CD8 T cell fate (proliferation, death, differentiation).

We first build a simplified GRN, based on essential principles that allow each cell to be able to proliferate, die, and differentiate in each relevant cell type (mainly, effector or memory CD8 T cell), see Section 3.1. This GRN is made of 9 genes, which have no equivalent in a real-biological setting (yet analogies are drawn in Section 3.2), each gene dynamics being driven by the gene model introduced in Section 2.3.

In order to obtain realistic dynamics, we also consider an augmented GRN based on the 9-genes GRN, to which three so-called 'decorating genes' are added to each main gene, resulting in a 36-genes GRN with the same properties. A "decorating gene" is activated by its main gene but does not act on any downstream gene and therefore has no influence on the GRN dynamics.

The GRNs are dynamical mathematical models, based on the coupling of deterministic and probabilistic formalisms, resulting in stochastic dynamics (see Section 2.3). In particular, genes of the GRN can act on other genes of the GRN, either by activating or inhibiting their expression, resulting in highly nonlinear dynamics.

It is important to note that each CD8 T cell will be embedded with the same GRN. Nonetheless, depending on their previous experiences and interactions with other cells, CD8 T cell molecular, intracellular states will differ from cell to cell and will be specific to each cell, thanks to individual values of each gene expression in a given cell.

2.3 Gene model

Given the stochastic nature of gene expression at the single cell level [24, 32], we chose to model the expression of each gene in the GRN (see Section 2.2) as a stochastic two-state model [33]. This stochastic process consists of three components for each gene i, $i \in \{1, ..., n\}$: the promoter state E_i, the mRNA level M_i and the protein level P_i.

The promoter can be in two states $E_i = 0$ or 1 (inactive or active). It opens with

a rate k_{on} and closes with a rate k_{off}. When the promoter is in the open state, mRNA gets synthetized at a rate s_0 and degraded at a rate d_0. From mRNA, proteins get synthetized at a rate s_1 and degraded at a rate d_1.

Such a model can be implemented in a variety of formalisms [34]. In the present work we consider the so-called bursty regime of the two-state model. It corresponds to the experimentally observed situation where active periods are short and characterised by a high transcription rate, thereby generating bursts of mRNA [28]. In such a regime, the promoter state no longer needs to be explicitly described since active periods can be considered as infinitely short: random jumps will instantaneously increase the amount of mRNA levels M_i.

The construction of a gene trajectory is as follows: starting from state $(M_i(t), P_i(t))$, the dynamics of mRNA and protein levels are given by

$$\begin{cases} M_i' = -d_{0,i} M_i, \\ P_i' = s_{1,i} M_i - d_{1,i} P_i. \end{cases} \tag{1}$$

until a burst occurs for an isolated gene i, at rate $k_{\text{on},i}$. Following the burst, the quantity M_i jumps by a random height according to an exponential law with rate $k_{\text{off},i}/s_{0,i}$ (see [34] or [35] for details of the model). This stochastic model constructs a piecewise-deterministic Markov process (PDMP).

To describe regulatory mechanisms within the GRN, we introduce the dependence of the burst frequency upon the proteomic field and the external signalling activity, i.e. a burst occurs for gene i, at a rate $k^{\theta}_{\text{on},i}(P, S)$, where $P = (P_1, \ldots, P_n)$ is the protein level vector of all genes in the network and $S = (S_1, \ldots, S_m)$ is the signalling state vector of the network. For every $i \in (1, \ldots, m)$, $S_i = 1$ or 0 (active or inactive signalling). More precisely, in [34, 35], the burst rate of each gene i is calculated using

$$k^{\theta}_{on,i}(P, S) = k_{0,i} + (k_{1,i} - k_{0,i}) \left(1 + \exp\left(-\sigma^{\theta}_i(P, S)\right)\right)^{-1}, \tag{2}$$

where $k_{0,i}$ and $k_{1,i}$ correspond respectively to the minimum and maximum burst fre-

quencies of gene i, and

$$\sigma_i^{\theta}(P,S) = \beta_i + \sum_{j=1}^{n} \theta_{ij} P_j + \sum_{j=1}^{m} \alpha S_j, \tag{3}$$

where β_i is the basal activity of gene i [34], α is a signalling parameter, and $\theta = \{\theta_{ij}\}_{i,j\in\{1,\dots,n\}}$ is the gene-gene interaction matrix, representing the considered GRN (see Supplementary Figure 1). Parameters θ_{ij} can be positive or negative corresponding either to the activation or the inhibition behaviour.

2.4 Cell types

When a simulation starts, there are two cell types in the population, APC and naive CD8 T cell. APC do not possess an internal state, whereas we recall that all CD8 T cells share the same GRN structure and CD8 cell types are defined as a function of the dynamical state of their GRN. A naive cell is activated as soon as it encounters an APC. This "activated" state will be transmitted throughout the descent of an activated cell. Activated CD8 T cells will then acquire specific cell type identity as a function of their position in the gene expression space. They will be considered bipotent if their protein levels in both genes 7 and 8 (see Section 3.1) are below certain threshold values P_7^* and P_8^*, respectively. When $P_7 \geq P_7^*$ while $P_8 < P_8^*$, they will be considered effector cells, and when $P_8 \geq P_8^*$, they will be considered memory cells. Noticeably, when $P_8 \geq P_8^*$ then, thanks to the toggle switch between genes 5 and 6, gene 7 will be inhibited and unlikely to satisfy $P_7 \geq P_7^*$ (if it happens, then a cell would opt for the memory phenotype, in order to avoid undefined phenotypes and to prevent the generation of immortal effector cells associated with $P_7 \geq P_7^*$ and a strong inhibition of gene 3). Table 1 summarizes all CD8 T cell differentiation states considered in this work.

2.5 Signalling

We consider two kinds of signals, namely APC signalling (*via* T cell-APC contact) and TCC signalling (effector-T cell contact), inducing either activation or apoptosis in a CD8 T cell after its encounter with an APC or an effector cell, respectively (see Figure 1).

Table 1: **Differentiation states of CD8 T cells.**

Cell type	Definition
Naive	No contact with an APC (inactivated)
Bipotent	Activated, protein levels $P_7 < P_7^*$, $P_8 < P_8^*$
Effector	Activated, protein level $P_7 \geq P_7^*$, $P_8 < P_8^*$
Memory	Activated, protein level $P_8 \geq P_8^*$

A CD8 T cell perceives a signal from another cell (either an APC or an effector T cell) if

$$R_i + R_j < d(i, j), \tag{4}$$

where R_i and R_j are the external radii of both cells and $d(i, j)$ is the Euclidean distance between the centers of the two cells. Condition (4) means that a CD8 T cell comes into contact with another cell if their spheres intersect. Once a contact is detected, the phenotype of the contacting cell is determined and the corresponding signalling (APC signalling or TCC signalling) is applied.

When a naive CD8 T cell encounters an APC (contacts on its surface), it receives an APC signalling, which impacts the GRN behaviour (see below). T cells integrate APC signalling in a cumulative way (that is, signals from exposure to any APC are cumulated). It has been shown that when a CD8 T cell encounters an APC, it adheres tightly to the APC for up to 20 hours in *vitro* before starting to proliferate [36]. Therefore, in the model, until the first proliferation occurs an activated T cell remains attached to an APC, through adjusting the coefficient σ_T (see Section 2.7). More precisely, when a naive CD8 T cell encounters an APC, it tends to stay with this APC for a while thanks to a modification of its velocity coefficient σ_T which takes a very low value. Then the APC signals to the T cell (APC signalling equals 1). If the activated CD8 T cell has not yet divided for the first time and the cumulative contact time with the APC is greater than 10 hours (i.e. it only takes 5 hours if the CD8 T cell encounters 2 APCs at the exact same time), gene number 1 is activated by the APC signaling (the level of expression of

this gene determines the first proliferation, see Section 3.1), with a signalling parameter $\alpha = 20$ (no unit) in order to strongly initiate gene expression. The effect of this value on model's outputs has been numerically investigated and shows very limited impact (Supplementary Figure 2). Noticeably, a CD8 T cell may leave the APC before its protein 1 level reaches the value required for first division (this is rare, but possible due to the probabilistic nature of the model), the level of proteins then remains at the level before contact breaking (save for protein degradation) and will increase upon new contact with APC. In this case, due to its fast mobility (σ_T goes back to its initial value upon contact breaking), the CD8 T cell will easily contact a new APC, its σ_T value will decrease again and its protein level will increase again rapidly.

In all cases, immediately after its first division, the CD8 T cell leaves the APC and moves randomly. Further contacts with APC for this cell may occur, yet subsequent divisions will no longer depend on the APC signalling. In particular, this will be the case for the daughter cells of activated cells, as we assume that both daughter cells remember their mother's first division, so they will not need to be activated by the APC signalling in order to proliferate [37]. In addition, both daughter cells dissociate from the APC, due to their fast motility (σ_T returns to its initial value after the first division).

Concerning apoptosis, we assume effector CD8 T cells' targets include all CD8 T cells, so fratricide killing may occur [38, 39]. In this case, an effector cell sends an apoptotic signal called TCC signalling to another CD8 T cell it contacts. This leads to an increase in gene 3 bursting frequency (see Section 3.1), and protein 3 level increases in the receiver cell, eventually resulting in its death. Only an effector cell can emit TCC signalling. When a CD8 T cell receives a signal emitted by an effector cell, its TCC signalling value equals 1 and its signalling parameter α equals 20 (no unit).

2.6 Initial conditions

Initial molecular content In order to obtain realistic values for the initial molecular content, we initialize each simulation with the values observed after 24 hours of simulating T cells in the absence of APCs. Specifically, we run the simulation with 980 CD8

Table 2: **CD8 T cell velocities.**

σ_T	Contact behaviour with respect to cell velocity
5	no contact with others/after first division upon encountering APC
0.02	contact with an APC
0.8	contact with an effector cell

T cells in the absence of APCs, so that they start to synthesize mRNAs and proteins without being activated. We then use simulated quantities of mRNA and proteins as initial data for the naive cells of the next simulation, where they are seeded in presence of APCs at time $t = 0$.

Initial positions in the computational domain Initial positions of APCs and CD8 T cells are uniformly randomized within the bounded computational domain (see Supplementary Figure 3).

2.7 Cell motion and fate

Cell movement T cells move randomly in the computational domain. The random movement of T cells was modelled through the use of a Gaussian distribution. More precisely, at each time step dt, new 3D coordinates of CD8 T cells are updated by an amount identically distributed in each direction, as follows:

$$\begin{cases} dx = \sqrt{dt}\sigma_T \mathcal{N}(0,1), \\ dy = \sqrt{dt}\sigma_T \mathcal{N}(0,1), \\ dz = \sqrt{dt}\sigma_T \mathcal{N}(0,1), \end{cases} \tag{5}$$

where σ_T characterizes the velocity of the CD8 T cell when in contact or non-contact with an effector cell or an APC (see Table 2).

Cell fate Regarding the proliferation process, if a CD8 T cell reaches its maximum volume (through the linear growth function of the current volume) and satisfies a condi-

tion on protein levels ($P_1 \geq P_1^*$ for the first division, $P_2 \geq P_2^*$ for subsequent divisions, see Section 3.1), then it divides into 2 daughter cells of equal volume. At division, the molecular content C of the mother cell is normally distributed between both daughter cells, and equals $(1 \pm 0.2\mathcal{N}(0, 0.2))C/2$ when C is the protein concentration and $(1 \pm 0.1\mathcal{N}(0, 0.2))C/2$ when C is the mRNA concentration. No proliferation program has been implemented, so cells can potentially divide without any limitation. However, due to differentiation (in memory cell) and apoptosis (see below), CD8 T cells either die or lose their ability to proliferate (memory phenotype). No more than 10 divisions have been observed for a single cell (lineage) in all simulations introduced in this manuscript.

Concerning apoptosis, if a CD8 T cell reaches the condition of death (protein level of gene 3 is above a threshold value P_3^*), then the cell disappears from the population immediately. We also set up an APC death mechanism to ensure that they will start dying after day 12 (i.e. after the expansion phase) and will disappear completely from the computational domain by day 20. This is a way to mimic the displacement of CD8 T cells out of the APC-containing organ [40], or the eradication of the antigen.

Cell differentiation is discussed in Section 2.4, see Table 1.

2.8 Simulation Parameters

All parameters used throughout this study are introduced in this section. Some parameters are needed for the functioning of the core code of `Simuscale`, other parameters are specific to the CD8 T cell simulation.

For running `Simuscale` simulations, we use an environment simulation domain inside a cube measuring 40x40x40 (same unit as the initialized radius, see [29]). This is what will be called 'computational domain' in this paper. Assuming cells are around 10 μm in diameter, a space unit (SU) would correspond to around 7 μm, for a domain side approximately 300 μm in length. Each cell is a sphere with a maximum volume of 2 SU^3 and a minimum volume of 1 SU^3. There is a linear growth function, where the initial volume of each cell is randomly chosen in the range $[1, 2]$, and the growth factor is fixed at $2^{dt/10}$, i.e. it takes 10 hours for a cell with a minimum volume to reach its maximum

Table 3: **Parameter values of the gene-gene interaction matrix.** Parameters $\theta_{i,j}$ give the value of the action of gene i onto gene j, while parameters $\theta_{i,i.j}$ represent the interaction between gene i and its three decorating genes $i.j$ ($j \in \{1, 2, 3\}$).

Parameter	$\theta_{1,2}$	$\theta_{2,2}$	$\theta_{2,3}$	$\theta_{2,4}$	$\theta_{3,3}$	$\theta_{4,5}$	$\theta_{4,6}$	$\theta_{5,5}$	$\theta_{5,6}$	$\theta_{5,7}$
Value	40	16	-10	16	60	16	16	90	-150	25
Parameter	$\theta_{6,5}$	$\theta_{6,6}$	$\theta_{6,8}$	$\theta_{8,2}$	$\theta_{8,9}$	$\theta_{9,3}$	$\theta_{i,i.j}$	All other $\theta_{i,j}$		
Value	-150	90	15	-800	530	-3000	15	0		

volume.

The time step of all simulations is set to $dt = 0.1$ hour, and the final simulation time is 960 hours, i.e. 40 days.

For T cell specific simulations, various parameters associated with cell fate, motion, and contacts must be defined. We will first focus on parameters related to gene dynamics.

The β_i parameter represents the basal activity of gene i, in a logarithmic scale, in the absence of any stimulation (see equation (3)). Its value was arbitrarily fixed to $\beta_i = -5$ for all genes i and allowed to reproduce consistent behaviors of gene dynamics. Values of the parameters of the gene-gene interaction matrix are shown in Table 3.

Values of protein degradation rates for all genes are given in Table 4. Since $d_1 \ll d_0$, we kept the degradation rate of mRNA equal to 1 for all genes, so $d_{0,i} = 1$ for $i \in \{1, \ldots, n\}$, and varied only the protein degradation rates. Those are chosen to progressively decrease along the main GRN, thus creating a delay in expressions and a wave-like behavior of signal propagation in the GRN [41]. Burst frequency of gene i is defined by $k_{\text{off},i}/s_{0,i}$, whose value has been chosen to correspond to an exponential distribution of bursts with mean $\mu = 50$. The effect of this value on model's outputs has been numerically investigated and shows limited impact (Supplementary Figure 2).

Threshold values for proteins defining the cell types are given in Table 5.

Finally, Table 2 presents the velocity coefficients of CD8 T cells in different cases. The default behavior of CD8 T cells is to move randomly and fast. When in contact

Table 4: **Gene kinetic parameters.** Values of: protein degradation rates $d_{1,i}$ of gene i ; protein degradation rates $d_{1,i.j}$ of decorating genes $i.j$, $j \in \{1,2,3\}$, of gene i ; burst frequency $k_{\text{off},i}/s_{0,i}$ of gene i ; synthesis rates $s_{1,i}$ of gene i proteins (translation rates).

Protein degradation rate	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{1,4}$	$d_{1,5}$	$d_{1,6}$
Value	0.03	0.021	0.021	0.02	0.019	0.009
Protein degradation rate	$d_{1,7}$	$d_{1,8}$	$d_{1,9}$	$d_{1,i.j}$		
Value	0.02	0.00014	0.00013	0.03		
Synthesis rate	$k_{\text{off},i}/s_{0,i}$	$s_{1,i}$				
Value	0.02	$0.01 d_{1,x}$				

Table 5: **Protein threshold values.** P_i^* is the threshold value for the protein of gene i, used to define cell fate (proliferation, apoptosis) and differentiation states, see Table 1.

Parameter	P_1^*	P_2^*	P_3^*	P_7^*	P_8^*
Value	0.5	0.4	0.75	0.01	0.02

with another T cell, they reduce their speed, and almost stick to an APC when they first encounter the APC, until their first proliferation occurs. Consequently, three different velocity parameters are used depending on the situation.

2.9 Model's outputs and simulations

All simulations are performed by running `Simuscale` code on a computation platform of the "Pôle Scientifique de Modélisation Numérique" (PSMN) of ENS Lyon. During each simulation, all cell population information is recorded, including the total number of cells in the population, cell identity, coordinates, volume, signalling, cell type and molecular content. This information is then post-processed on personal computers, using Python and R tools. All figures related to model's outputs presented in this paper have been generated from the outputs of the simulations and plotted with Python and R.

Simulations of the GRN have been performed through a C++ code embedded in `Simuscale` and adapted from the Python code in Herbach [42].

Typical simulation times range from a few minutes (35 CD8 T cells and 25 APCs initially) to a few hours (1,960 CD8 T cells and 1,400 APCs initially). It took only 2h to simulate the simulation associated to the largest initial cell counts in this study (average value over 10 simulations), executed on PSMN's partition Cascade with 96 cores and 384 GB RAM (for more details, see https://www.ens-lyon.fr/PSMN/Documentation/clusters_usage/partitions_overview.html). All codes are available at https://gitlab.inria.fr/thinnguy/Simuscale_Lymphocytes

The maximum number of cells simultaneously simulated is slightly over 32,000 cells (corresponding to the highest initial condition), while the maximum number of cells the computational domain can contain is 64,000 cells. With parameter values used and mentioned in this paper, no crowding effect can be observed.

It is important to note that among the functionalities of `Simuscale` there is the possibility to color-code cells according to the quantitative value of any cellular variable. Coupled with a video display of the output, this is an invaluable feature of the model to:

1. Verify the correct functioning of the model. For example do divisions occur as planned when a certain value of P_2 is reached, as expected?

2. For many biologists "seeing is believing". Videos are therefore the perfect medium for interactions between modellers and biologists.

3. In the present work there is no resulting 3D structure since cell movement tends to homogenize spatial cellular localisation, but for modeling spatially constrained structures, this output would be critical.

As an illustrative example, a movie of the CD8 T cell population dynamics is available at: https://gitlab.inria.fr/thinnguy/Simuscale_Lymphocytes/ (Pop_T_cell_APC.mov). All relevant cellular variables are examined during the course of the response.

2.10 Sensitivity analysis

We perform a sensitivity analysis in Section 3.6. Variables of interest are: the number of CD8 T cells at the peak of the response; the ratio between the average number of CD8 T cells at the peak of the response and the initial number of CD8 T cells; the time of the response peak; the average number of memory CD8 T cells at the end of the simulation ($t = 40$ days); the ratio between the average number of memory CD8 T cells at the end of the simulation and the average number of CD8 T cells at the peak of the response.

First, we perform a local sensitivity analysis and consider two types of influence: either a sensitivity to initial conditions or to molecular parameters. In each case, we fix parameter values, perform 10 simulations, calculate the mean and standard deviations of the outputs, and compute variables of interest.

The influence of initial conditions is assessed by either using different CD8 T cell and APC initial cell counts, but preserving the ratio T cell counts/APC counts (ratio equals to 1.4), or by fixing the initial number of CD8 T cells (two values were used, 210 and 980 cells) and varying the initial number of APCs (with T cell/APC ratios ranging in $[0.3; 8.4]$).

The influence of molecular content is assessed by computing variables of interest for various values of some molecular parameters: the degradation rates of proteins 2, 3, and 6 that determine proliferation, apoptosis and memory T cell differentiation, respectively. These parameters have been selected based on the *a priori* essential roles played by genes 2, 3 and 6 in controling CD8 T cell fate.

Second, we perform a higher dimensional sensitivity analysis using the Random Balance Designs–Fourier Amplitude Sensitivity Test (RBD-FAST) implementation of the SALib Python library [43, 44, 45]. Briefly, RBD-FAST is a numerical method for computing the main (first-order) effects in global sensitivity analysis, measuring the contribution of each parameter to the model's output variance. RBD-FAST has the advantage that it does not require a specific sampling method nor a specific number of samples. This is useful because numerical simulations may not be necessarily successful for all sets of parameter values (see Section 3.6).

We performed this global sensitivity analysis on the nine degradation rates $d_{1,1}$ to $d_{1,9}$, on intervals ranging $\pm 20\%$ of their nominal values, with all other parameter values fixed. We sampled 1,000 degradation parameter sets using a Latin Hypercube Sampling (LHS) algorithm implemented in *scipy.stats.qmc.LatinHypercube*. Initial populations were set to 980 T cells and 700 APCs. For each parameter and each response variable, the first-order sensitivity and the 95% confidence interval were obtained from the RBD-FAST method.

3 Results

3.1 A model involving a gene regulatory network constructed by 9 genes

For the intracellular scale, we first build a GRN composed of 9 genes (see Figure 1.A), which triggers the proliferation, apoptosis, and differentiation of a single cell. All genes are hereafter denoted Gi, $i = 1, \ldots, 9$. From now on, we refer to this GRN as a 'principle-

based GRN', since it is built upon few principles required to reproduce the main features of a CD8 T cell immune response. The logic behind this GRN is as follows.

First, APC signalling duration (cumulative contact time with APC is greater than 10 hours [36]) activates $G1$, which leads to the first proliferation of CD8 T cells when the protein concentration of $G1$ reaches a threshold value P_1^* and the cell has doubled its volume (volume is equal to 2). After division, the activated CD8 T cell leaves the APC due to an increased velocity coefficient and the activation stops. As long as all APCs have not been eliminated, activated CD8 T cells can encounter APC again but they do not really attach to APC due to their high velocity.

In the meantime, $G1$ activates $G2$, which also activates itself in order to maintain a strong proliferation phase at the beginning of the immune response. The second and subsequent proliferation events are based on $G2$ dynamics: when the G_2 protein level exceeds a threshold value P_2^* then cells divide. Also, $G2$ inhibits $G3$, which induces CD8 T cell apoptosis above a threshold value P_3^*. Gene G_3 is activated through TCC signalling and $G3$ also activates itself.

Furthermore, $G2$ activates $G4$, which simultaneously activates two genes, $G5$ and $G6$. There is a toggle switch between these two genes which then creates the differentiation states. $G5$ activates $G7$, which will be the marker for the effector phenotype (when the protein of this gene exceeds a value P_7^*). In addition, $G6$ activates $G8$, which induces the memory phenotype (when its protein concentration reaches a threshold value P_8^*). Gene $G8$ inhibits $G2$ to mimic cell cycle exit of memory cells. Finally, $G8$ activates $G9$, which inhibits $G3$, leading to memory cell long-term survival.

All threshold values are shown in Table 5.

For application purposes that require many more genes in the network (see Section 3.5), we also added so-called "decorating genes" to the principle-based GRN, resulting in a second GRN model (hereafter referred to as the 'augmented model'). More precisely, each gene, from $G1$ to $G9$ of the principle-based GRN, simultaneously activates n redundant "decorating genes" that have no impact on the GRN dynamics. In the next sections, we illustrate the results with simulations of the augmented GRN, including

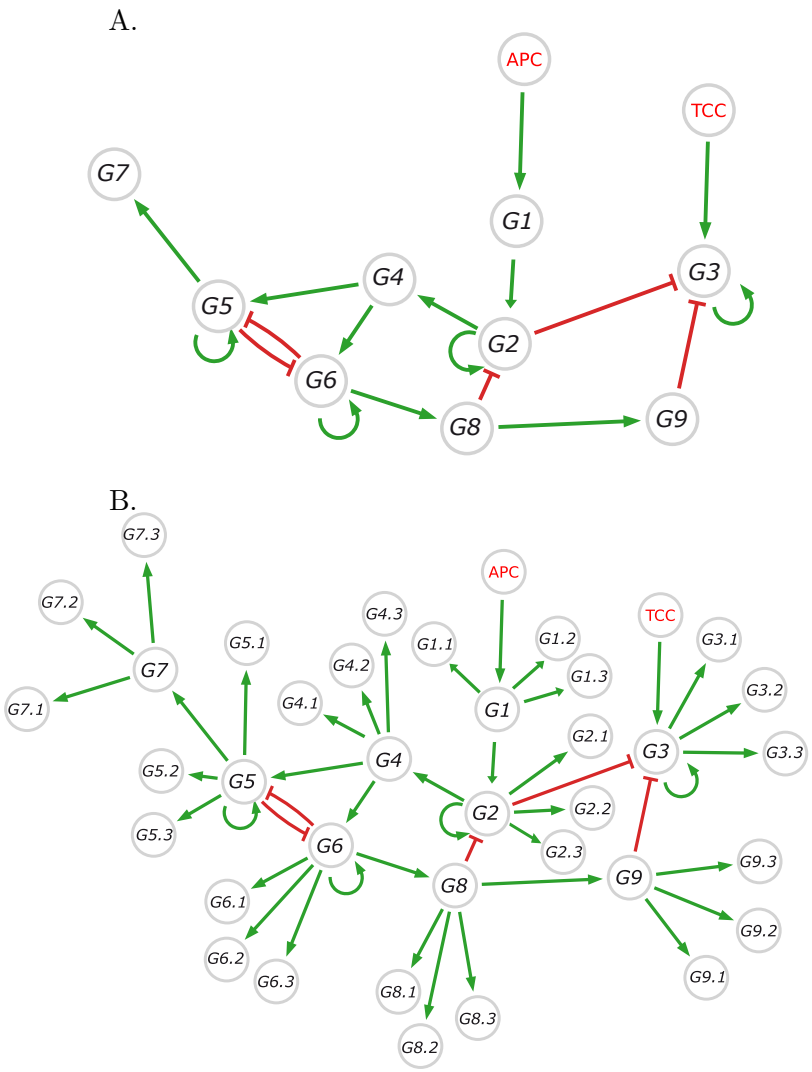


Figure 1: **Schematic representation of the GRN.** (A) The core circuitry of the principle-based GRN using 9 genes. The notation Gi corresponds to the gene number i. APC signalling and TCC signalling are perceived when CD8 T cells are in contact with APCs or effector cells, respectively. The green (resp. red) arrows represent activation (resp. inhibition). (B) The augmented GRN, made of 36 genes, incorporates the core principle-based GRN together with "decorating" genes. The gene $Gi.j$, $j \in \{1, 2, 3\}$, is the j-th gene activated by gene i. Networks have been drawn using the *plot_network* function of the Harissa package [42]

three ($n = 3$) "decorating genes" for each main gene (see Figure 1.B).

3.2 Biological rationale for our GRN

It is important to note that the principle-based GRN was built in order to satisfy minimal requirements imposed by a realistic description of the CD8 T cell immune response: it required genes able to induce proliferation, death and differentiation of CD8 T cells, therefore there is no reason any gene in this GRN has an equivalent specific gene in a real biological setting. Nevertheless, one could draw the following analogies between the principle-based GRN introduced previously and known genes.

1. $G1$ is activated by APC signalling and leads to the first cell division of activated cells. This has been shown to be a function carried by the *mTORC1* gene [46].

2. $G2$ drives the second and subsequent divisions. This role can be carried by the *IL-2R* gene [47], which also amplifies itself as IL-2 signalling promotes further IL-2R expression [48].

3. $G3$ leads to programmed cell death (apoptosis), which is known to be the activity of *Fas* [49]. In such a case, the TCC signalling could be seen as FasL signalling [50].

4. $G5$ and $G6$ are involved in a toggle switch, leading to the differentiation states. Two members of the T-box transcription factors have been demonstrated to play this critical role: *T-bet* ($G5$) [51] and *Eomesodermin* ($G6$) [52], which promote effector and memory differentiation, respectively. Similar tandems could be *Id2/Id3* and *Blimp1/Bcl-6* [3].

5. $G7$ is a marker for the effector phenotype. Many genes could fall in this category, but since in this model effector cells have the capacity to kill through the TCC signalling, *FasL* is one obvious candidate for this function [53].

6. $G8$ is a marker for the memory phenotype, limits growth and prevents death by apoptosis, a role for which *FOXO1* gene would be a relevant candidate [54, 55].

7. Finally, $G9$ is activated by $G8$ and inhibits death, a role known to be played by the antiapoptotic gene *Bcl-2* [56].

3.3 Population trajectory and differentiation states of a reference case

We first focus on the results obtained for a reference case, i.e. an initial population composed with 980 CD8 T cells and 700 APCs. All the parameters used to perform simulations can be found in Section 2.8.

Figure 2 shows the evolution of the CD8 T lymphocyte population and of each subpopulation of differentiated cells (Naive, Bipotent, Effector and Memory populations). Results represent average and standard deviations over 10 simulations (using different random generator seeds). The results capture the expected dynamics of the CD8 T cell population qualitatively, i.e. naive cells rapidly disappear, differentiating into bipotent cells that further generate effector or memory cells. In the effector state, cells proliferate rapidly, reaching a peak followed by a contraction phase, at which point memory cells appear and accumulate. In the last phase, only memory cells remain that form a stable population (see [2]).

From a more quantitative perspective, after 18 hours, all 980 naive cells initially present (green solid line) have become activated by contacting with APCs, 72% of the activated cells became bipotent cells (brown) while 28% differentiated into effector cells (blue). Bipotent cells gradually transform into effector cells, which proliferate very rapidly and reach a peak between days 5 and 10 post-immunization. On average, a total of $16,422$ T cells (red) were formed at the peak (a 16.8 fold expansion) and $1,637$ memory cells (yellow) remain in the population at the end of the simulation (40 days). For more details about cell numbers and standard deviations, see Table 6.

Noticeably, despite the stochasticity introduced at the molecular level with the mRNA bursting regime and at the cellular level with the random cell movement, the resulting output from 10 independent simulations is remarkably predictable, with little variation between two simulations of the model, characterized by small standard deviations (see Figure 2, where standard deviations are illustrated by narrow shaded areas

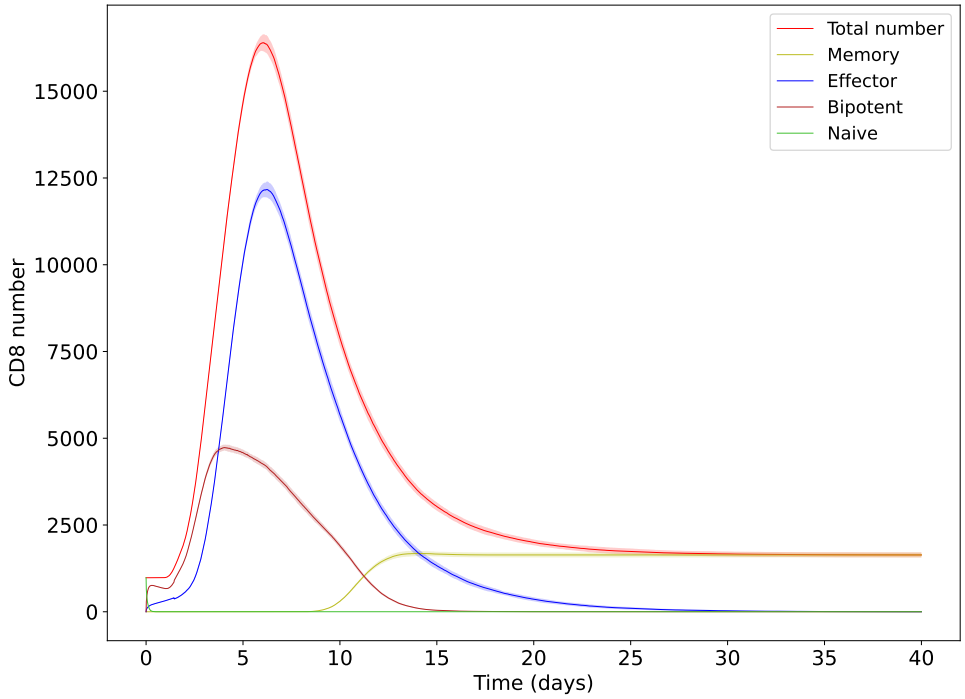


Figure 2: **Time-dependent evolution of differentiation states of the CD8 T cell population** over 10 simulations (mean (solid line) ± std (shaded area)). The green, brown, blue, and yellow lines present the total number of naive, bipotent, effector, and memory phenotype of CD8 T cells, respectively, and the red line the total number of CD8 T cells.

around the mean, and Table 6).

3.4 Time-dependent evolution of mRNA and protein

Having demonstrated a robust behavior of the model at the cellular level, we then assess its behavior at the molecular level. For this we plot in Figure 3 the histograms over time of mRNA expression for the 9 core genes of the augmented GRN through one simulation randomly selected from the 10 simulations (Figure 2). The times at which those distributions are shown were chosen to capture the early phase of expansion, the contraction phase and were more widely spaced for the final memory generation phase.

Altogether the genes display the expected dynamics. $G1$ shows a very brief period of activation, due to its early activation by APC signalling which stops after the first division. $G2$ shows a much more sustained period of activation during the expansion phase. $G3$ is mostly expressed during the contraction phase and both $G2$ and $G3$ expressions go towards 0 at the end of the simulation. The frequency of $G4$ mRNA expression follows the distribution of $G2$, because $G2$ activates $G4$. The expression of $G7$ and $G8$ mirrors that of $G5$ and $G6$, which are mutually exclusive as expected from their toggle switch connection.

It is important to note that those distributions harbour the characteristics known for patterns of gene expression at the single cell level, i.e. a strong zero component and a long tailed distribution. We have previously shown that the ability to generate realistic scRNA-seq datasets was a key asset of the bursty model [34].

We then explore the behaviour of the molecular model at the protein level. Figure 4 shows the time-dependent evolution of the mean expression of each protein across all cells, for the same simulation. Here too, the expected behaviour is observed.

The mean protein expression of $G1$ (blue line) first increases sharply, then decreases after 1 or 2 days, correctly representing the expected action of $G1$ in the GRN. While $P2$ increases and quickly reaches a peak (orange line), $P3$ slowly increases (green line) due to the inhibition of $G3$ by $G2$. When $P2$ decreases then $P3$ increases sharply and reaches a peak around the time of the response peak. Since $G4$ is activated by $G2$, the

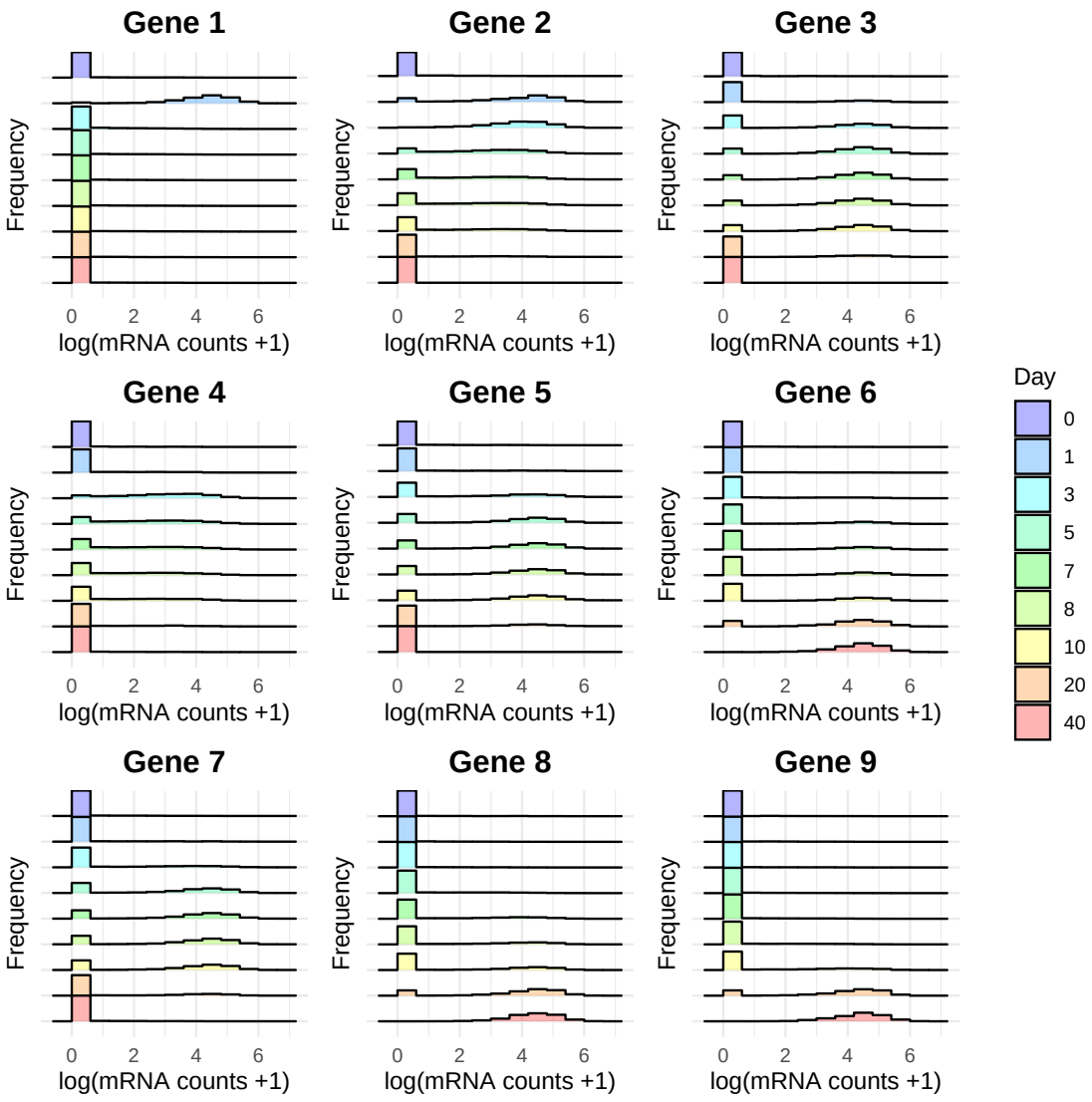


Figure 3: **Time dependent evolution of mRNA distributions of 9 genes** (see Figure 1.A): x-axis shows the value of mRNA+1 in logarithmic scale, while y-axis shows the frequency occurrence in the cell population observed at different time points.

growth pattern of $P4$ is similar to that of $P2$. The toggle switch between $G5$ and $G6$ is visible in the mean protein expression between $P5$ (purple) and $P6$ (brown). This is mirrored in changes between $P7$ and $P8$, with the expression of $G7$ (pink line) increasing initially and decreasing as $G8$ increases. Similarly, $G8$ activates $G9$, then the yellow line of $G9$ rises following the increase of $G8$.

One should also notice that the time-dependent evolution of the average protein amount is associated with small standard deviations (represented by very narrow shaded areas around the mean), hence highlighting a pretty reproducible behaviour of the model.

3.5 UMAP analysis

The highly dimensional nature of scRNAseq data has called for the development of suitable dimensionality reduction techniques. Among those, the UMAP representation (Uniform Manifold Approximation and Projection [57]) has established itself as one of the most popular. We therefore assess the ability of the model to produce a relevant UMAP representation of the CD8 differentiation sequence (Figure 5).

Our initial attempt at obtaining a UMAP representation of the model output based on the principle-based GRN gave rise to not very realistic images (see Supplementary Figure 4). At that stage we reasoned that the amount of information provided to the UMAP algorithm might have been too sparse. We therefore decided to add "decorating genes" (see Section 3.1) and used the augmented GRN instead. The decorating genes do not participate in the dynamics of the network, but add some redundant information that proved to be required to obtain the final correct UMAP representation observed on Figure 5.

On Figure 5.A, one can clearly see the time-dependent trajectory evolution of the CD8 T cell population, starting at day 0 on the right and going to day 40 to the top left. The UMAP representation therefore reveals a correct temporal arrangement of the cells.

Figure 5.B presents the trajectory evolution of the CD8 T cell population as a function of their differentiation states. We can see that at day 0 most cells are naive cells

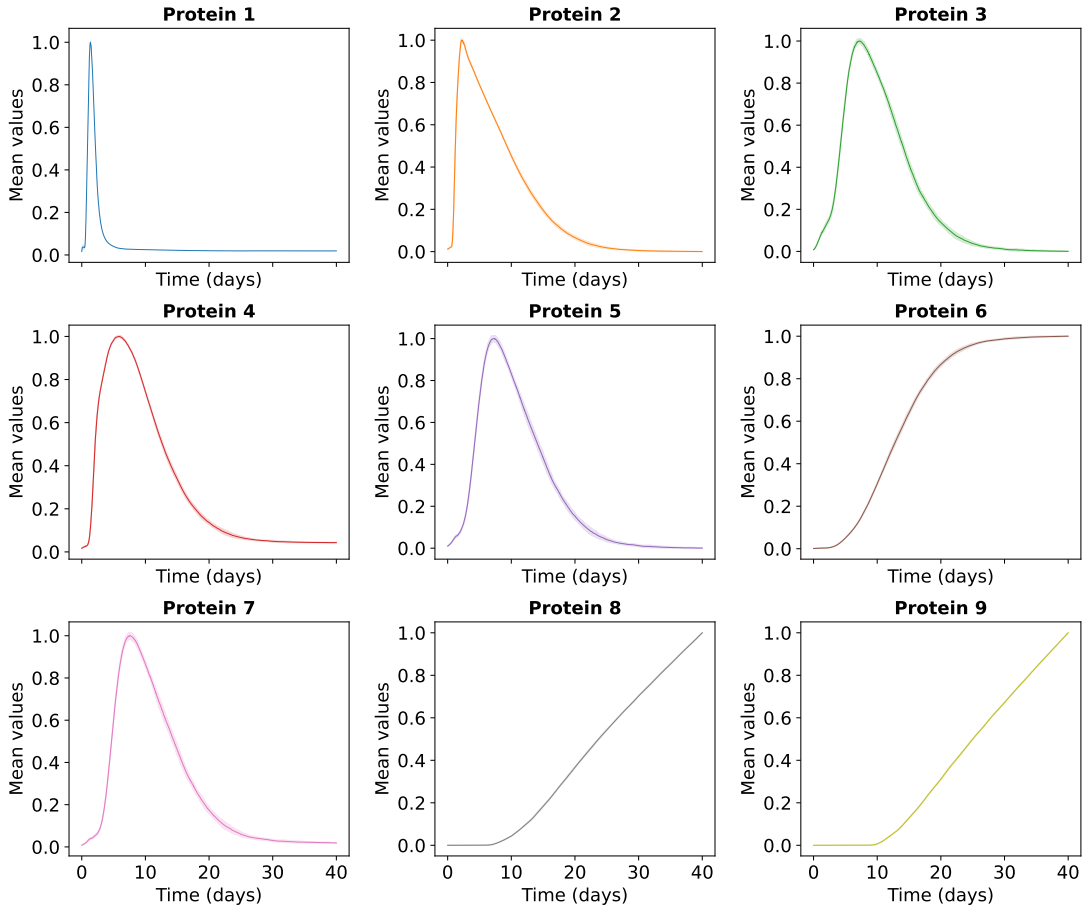
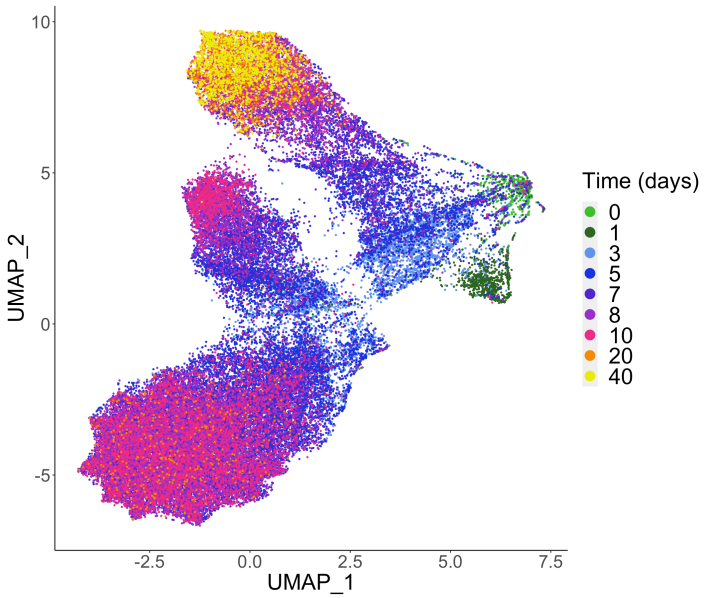


Figure 4: **Time-dependent normalized mean protein expressions over 10 simulations across all cells.** Shown is the mean (solid line) ± std (shaded area) for the 9 main genes of the augmented GRN.

A.



B.

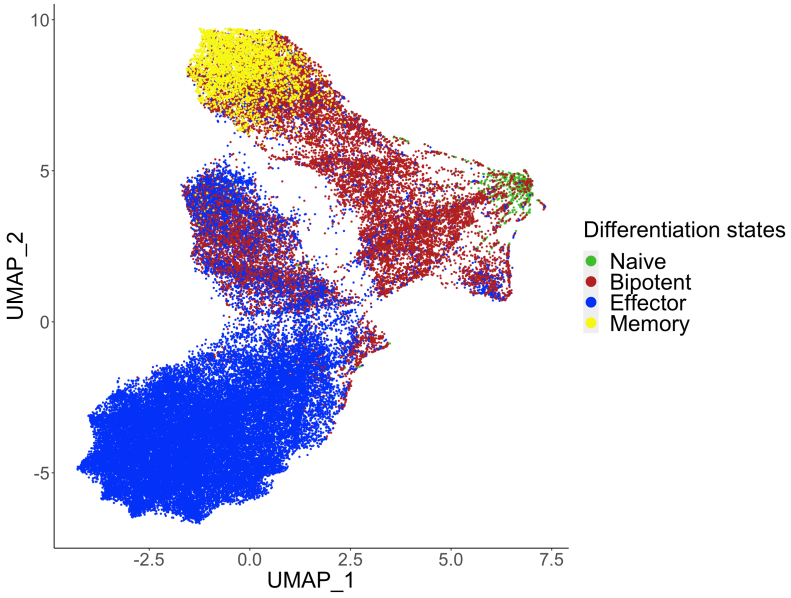


Figure 5: **UMAP representation of CD8 T cell population dynamics over time.** (A) Cells are color-coded as a function of the time they were observed. The same time points are displayed as in Figure 3. (B) The same graph, where the cells are now color-coded according to their phenotype.

(green in both figures) clustered close together, since their initial molecule content is almost 0. Bipotent cells appear at an early stage (brown in Figure 5.B)) and then become effector or memory cells, giving rise to two branches of cells. This clearly shows the correct behavior of the $G5/G6$ toggle switch pushing cells out of their bipotent state and forcing them into an effector or memory fate. The vast majority of cells are effector cells (blue in Figure 5.B) around day 6, but there is no effector cell left by day 40, suggesting that they have gone through the expected contraction phase. Indeed, at the end of the simulation, i.e. on day 40 (yellow color code in 5.A), only memory cells remain.

3.6 Parameter sensitivity analysis

To analyze the impact of some parameters on the model behavior, we first assess the role played by the initial number of cells. Table 6 recapitulates the results of the local sensitivity analysis.

We first explore the impact of modifying initial cell counts while keeping a constant ratio of T cells to APCs (equal to 1.4) on the overall amount of cells. One can see in Table 6 that there is a steady monotonous increase in the number of cells at the peak. Nevertheless if one examines the ratio between the average number of T cells at the peak of the response and the initial number of T cells, the situation is very different as illustrated in Figure 6.A. There is initially an increase in this variable, which then stabilizes in a very narrow range between 16.2 and 16.8. This amplification ratio therefore appears as a relatively robust emerging property of the model and suggests a minimal initial cell density is required for the optimal expansion of the effector population, as observed *ex vivo* [58].

One should note that the maximum number of cells generated was more than $32,000$ cells, highlighting the ability of the model to generate very large amounts of cells. The computational time was only 2h in this case (see Section 2.9).

We then explore the impact of varying the initial number of APCs while keeping constant the initial number of T cells (210 or 980). Larger ratios yield weaker cellular expansions in both situations (Figure 6.B), highlighting a positive effect of the antigen

Table 6: **Comparative values of different initial data, where protein degradation rate of $G3$ equals** 0.021**.** All ratios are computed using mean values of specified quantities.

T cell	APC	Ratio T cell/ APC	T cell at the peak (mean ± std)	Ratio T cell at the peak/ initial T cell	Time of the peak (days) (mean ± std)	Memory cell at the end (mean ± std)	Ratio Memory/ T cell at the peak (%)
35	25	1.4	500 ± 37	14.3	7.2 ± 0.5	61 ± 11	12
70	50	1.4	1,135 ± 29	16.2	6.7 ± 0.2	121 ± 10	11
140	100	1.4	2,292 ± 48	16.4	6.4 ± 0.2	228 ± 26	10
210	150	1.4	3,469 ± 62	16.5	6.2 ± 0.1	357 ± 37	10
980	700	1.4	16,422 ± 235	16.8	6.0 ± 0.1	1,637 ± 72	10
1,960	1,400	1.4	32,802 ± 165	16.7	6.0 ± 0.1	3,150 ± 95	10
210	25	8.4	2,177 ± 128	10.4	7.4 ± 0.2	257 ± 19	12
210	50	4.2	2,924 ± 101	13.9	6.8 ± 0.2	286± 32	10
210	210	1.0	3,583 ± 45	17.1	6.2 ± 0.1	351 ± 44	10
210	420	0.5	3,596 ± 78	17.1	6.0 ± 0.1	395 ± 42	11
210	630	0.4	3,577 ± 57	17.0	6.0 ± 0.1	368 ± 24	10
980	117	8.4	12,724 ± 472	13.0	7.1 ± 0.1	1,337 ± 87	11
980	175	5.6	14,592 ± 426	14.9	6.7 ± 0.1	1,472 ± 79	10
980	1,050	0.9	16,502 ± 274	16.8	6.0 ± 0.1	1,686 ± 106	10
980	1,960	0.5	16,750 ± 229	17.1	6.0 ± 0.1	1,738 ± 76	10
980	2,800	0.4	16,744 ± 284	17.1	6.0 ± 0.1	1,733 ± 115	10

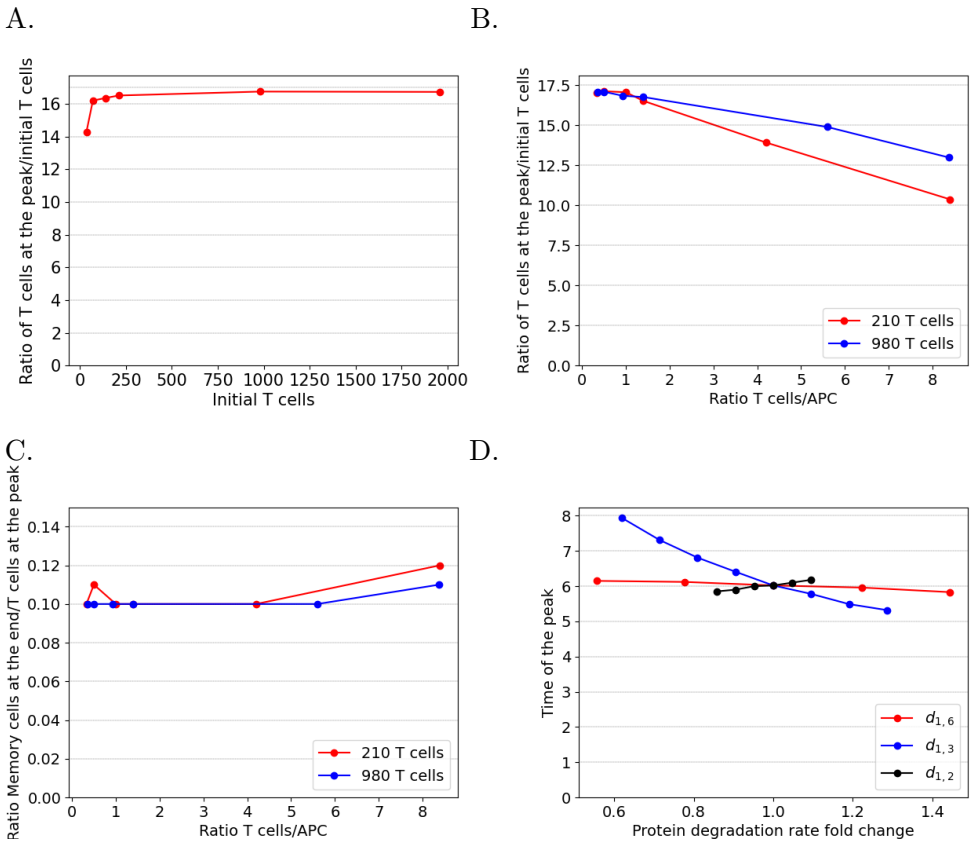


Figure 6: **Comparison of variables of interest when varying initial cell counts and protein degradation rates.** (A) Average cell expansion at the peak of the response as a function of initial cell numbers over 10 simulations. (B) Average cell expansion at the peak of the response as a function of the ratio of initial T cells to APCs counts, in two cases comprising 210 and 980 T cells. (C) Ratio of the average memory cell counts at the end of the simulation to average T cells count at the peak, as a function of the ratio of initial T cells to APCs counts, in two cases comprising 210 and 980 T cells. (D) Effect of proteins 2, 3 and 6 degradation rates upon the time of the peak (for a Ratio T cells/APC equals 1.4). Reference values of protein degradation rates are given in Table 4.

density on CD8 T cell expansion [59]. The range of values we used indicates there may be an optimal ratio of 1 T cell to 2 APC, when 980 CD8 T cells are initially seeded, since the highest total number of CD8 T cells is observed with this ratio (see Table 6).

We finally can demonstrate that the ratio between the average number of memory T cells at the end of the simulation and the average number of T cells at the peak of the response is quite insensitive to the initial APCs to T cells ratio (Figure 6.C), highlighting the initial antigen density does not influence the extent of cellular contraction after the peak of the response, as demonstrated *in vivo* [60]. It turns out that the value of this variable proved to be extremely robust, around $10\% - 12\%$ in all of the situations tested (see Table 6), which corresponds to *in vivo* situations [5].

Investigating the impact of molecular processes on the dynamics of the immune response through variations of different parameters (see Section 2.10) showed, as expected, that variations of Protein 2 degradation rate mostly impact the production of T cells at the peak of the response while variations of Protein 6 degradation rate mostly impact the generation of memory cells (Supplementary Figure 5). However, although the time at which the peak is observed is around 6-7 days and does not seem sensitive to initial cell counts, we could demonstrate the role played by the degradation rate of Protein 3 in setting the time of the response peak (Figure 6.D and Table 7). Indeed the time at which the expansion phase reaches its peak value is monotonically decreasing as a function of the degradation rate.

In order to investigate the robustness of the model to parameter changes, we next performed a higher-dimensional sensitivity analysis. We focused on the contribution of protein degradation rates, $d_{1,i}$, $i = 1, ..., 9$, to the variations of the model's outputs by computing first-order sensitivity indices over four cell population characteristics. Out of the $1,000$ sampled protein degradation parameter sets, 996 simulations (99.6%) were successful; the remaining four simulations either ran out of memory or were aborted after 10.5 hours of execution.

Figure 7 shows the first-order index and the 95% confidence intervals obtained for each protein degradation rate. As a result, the T cell count at the peak mostly depends

Table 7: **Effects of protein degradation rates $d_{1,\cdot}$ of genes 2, 3 and 6 on various variables of interest.** In each case, initial cell counts are: 980 T cells and 700 APC (ratio equals 1.4).

Value	T cell at the peak (mean ± std)	Ratio T cell at the peak/ initial T cell	Time of the peak (days) (mean ± std)	Memory cell at the end (mean ± std)	Ratio Memory/ T cell at the peak (%)
$d_{1,2}$					
0.023	19,821± 190	20	6.2± 0.1	1,918± 82	10
0.022	18,029± 150	18.4	6.1± 0.1	1,783± 68	10
0.021	16,422 ± 235	16.8	6.0 ± 0.1	1,637 ± 72	10
0.02	14,944± 150	15.3	6.0± 0.1	1,493± 93	10
0.019	13,453± 138	13.7	5.9± 0.1	1,371± 64	10
0.018	12,125± 101	12.4	5.9± 0.1	1,205± 51	10
$d_{1,3}$					
0.027	12,941 ± 120	13.2	5.3±0.1	1,582 ± 46	12
0.025	13,899 ± 110	14.2	5.5± 0.1	1,601± 57	12
0.023	15,105 ± 173	15.4	5.8 ± 0.1	1,632 ± 47	11
0.021	16,422 ± 235	16.8	6.0 ± 0.1	1,637 ± 72	10
0.019	17,400 ± 684	17.7	6.4 ± 0.1	1,592 ± 76	9
0.017	19,107 ± 600	19.5	6.8 ± 0.1	1,562 ± 137	8
0.015	21,305 ± 466	21.7	7.3 ± 0.1	1,672 ± 84	8
0.013	23,857 ± 258	24.3	7.9 ± 0.1	1,746 ± 87	7
$d_{1,6}$					
0.013	15,433± 162	15.8	5.8± 0.1	1,983± 98	13
0.011	15,990± 88	16.3	6.0±0.1	1,857± 40	12
0.009	16,422 ± 235	16.8	6.0 ± 0.1	1,637 ± 72	10
0.07	16,620± 155	17.0	6.1 ±0.1	1,312± 81	8
0.05	16,881± 109	17.2	6.2± 0.1	785± 43	5

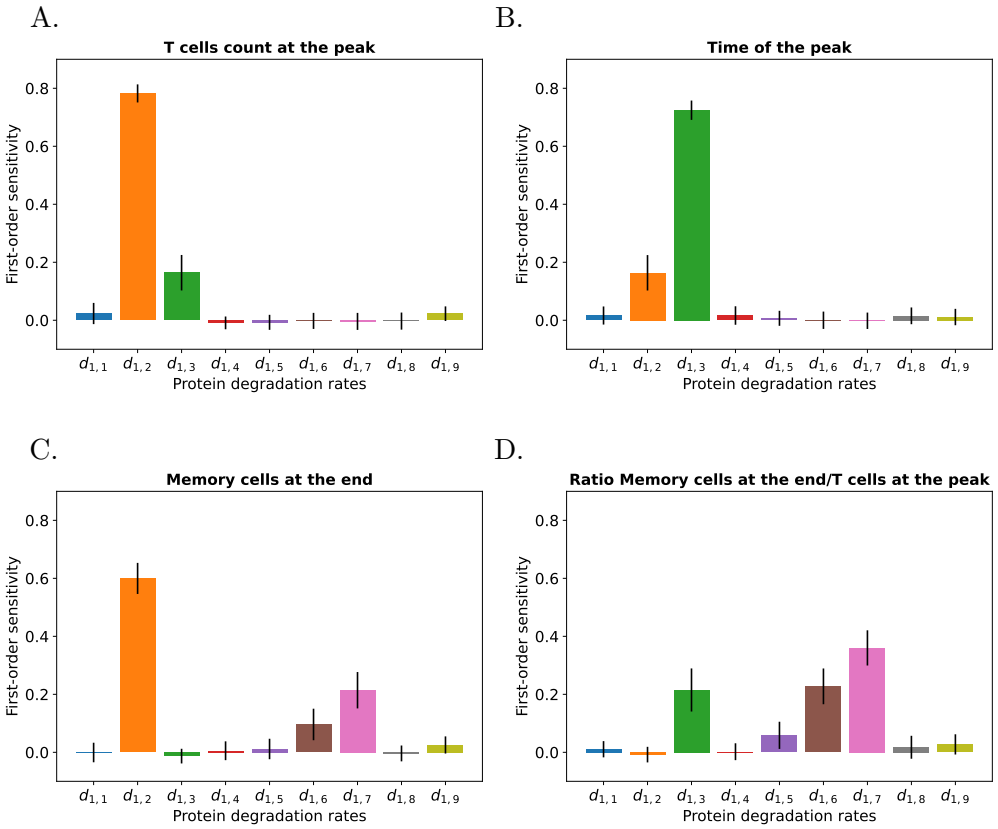


Figure 7: First-order sensitivity results for the protein degradation rates. First-order sensitivity index and 95% confidence interval are displayed for each protein degradation rate and for (A) the number of T cells at the peak, (B) the time of the peak of the response, (C) the number of memory cells at the end of the response, (D) the ratio between the number of memory cells at the end of the response and the number of T cells at the peak of the response. Color code is the same as in Figure 4.

on $d_{1,2}$, followed by $d_{1,3}$, and the time to peak mostly depends on $d_{1,3}$, followed by $d_{1,2}$. Memory cell numbers at the end of the response mostly depend on $d_{1,2}$, and to a lesser but significant extent on $d_{1,7}$ and $d_{1,6}$. The ratio of memory cells at the end of the response to T cell count at the peak mostly depends on $d_{1,7}$, followed by $d_{1,6}$ and $d_{1,3}$.

Overall cell population characteristics are consistent with results obtained with the local sensitivity analysis. In particular, variations of key features are similar (Supplementary Figure 6). The absolute variation of T cell count at the peak of the response was approximately 3-fold, while the time of the peak varied by around 2 days, with variations between days 5 and 8 (Figure 6.D), and the ratio between memory T cells at the end of the response and the number of T cells at the peak ranges in the interval $[0.05; 0.15]$.

Interestingly, although the number of memory cells at the end of the simulation depends strongly on $d_{1,2}$ (first-order index equals approximately 60%), no influence of $d_{1,2}$ was observed on the ratio between memory cells at the end of the response and the number of T cells at the peak (Figure 7.D), as already identified from the local sensitivity analysis (Supplementary Figure 5.B). Yet, this ratio depends on $d_{1,3}$ and $d_{1,6}$ (around 20% each), see Figure 7.D and Supplementary Figure 5.B. In addition, it strongly depends on $d_{1,7}$ (almost 40%). This latter degradation rate is associated with the gene responsible for differentiation in effector cells, and its impact on the ratio between memory cells at the end of the response and the number of T cells at the peak can only occur through indirect mechanisms.

Sensitivity analysis can therefore show the functional impact of a molecular parameter value of the molecular GRN at the cell population level: a higher degradation rate for the protein controlling cell death induces an earlier peak in the response. The accumulation of $P3$ up to its threshold value to induce death takes longer time with a more unstable protein, suggesting a higher degradation rate allows the expanding population to reach its maximum more rapidly (even though this earlier peak is associated with less T cells at the peak, see Supplementary Figure 5 and Table 7), although more complex explanations might be at stake. This is however a very interesting result which illus-

trates the benefit of this model, since it can only be obtained with the proper coupling of all scales and the inclusion of a mechanistic gene model. To a lesser extent, a partial global sensitivity analysis highlights the robustness of the cell population dynamics with respect to the protein degradation rates. In particular, it stresses the indirect influence of a molecular parameter (degradation rate of the protein controlling the differentiation in effector cells) on a measure of the efficiency of the response (ratio memory cell count at the end of the response over T cell count at the peak).

4 Discussion

In this work, we have presented a multiscale model of the CD8 T cell immune response, from naive to memory cells, in which cell behavior (proliferation, apoptosis or differentiation) is determined by molecular intracellular dynamics. We have demonstrated that this modeling approach can be used to correctly reproduce the expected population dynamics of CD8 T cells [2], using a 9-component principle-based GRN.

The use of a mechanistic model of gene expression, the two-state model in its bursty regime [30], allowed us to obtain at all time points a realistic distribution of the amount of mRNAs. It is important to point out that these types of expression distributions are the main output of scRNAseq experiment and therefore comparing distributions represent a natural way to compare a model output to experimentally observed values [61]. The model-generated distributions can be used for comparison with experimental distributions, for instance using the Kantorovitch distance [41]. Interestingly, the output of the model makes it also possible to extract single-cell proteomics, therefore while it is currently challenging to perform single-cell proteomics [62] we can make predictions using our modeling and simulation framework.

UMAP representation is also a natural way of comparing the output of a model and an experimental single-cell dataset (see e.g. [61]). In our case, the GRN we introduced was able to correctly generate a time-dependent evolution of single cells at the molecular level, and we could observe the expected time-dependent trajectory in the UMAP space,

as well as the branching resulting from the decision making process through the toggle switch as observed experimentally [2].

Cell dynamics illustrated in Figure 2 highlight a very limited run-to-run variability. Although we do not have a clear explanation for this observation, it may be noted that the model comprises both stochastic and deterministic parts, and sometimes stochastic process may be negligible before deterministic ones, the reasons being not straightforward. For instance, when the initial cell population is smaller then run-to-run variability is larger, yet this may not be a consequence of the initial amount of cells solely: the size of the computational domain, cells movement speed, activation times, dynamics of interacting genes, to name but a few processes, certainly play a role in that observation too. A consequence of this low variability is the possibility to theoretically develop a mean-field model for the cell population dynamics. Nevertheless, such a model would be close to several CD8 T cell population dynamics models available in the literature (in the form of an evolution equation with reaction terms accounting for cell proliferation and death) and would not account for the stochastic molecular dynamics driving cell behavior, while this is the purpose of developing multiscale models.

One of the main difficulties we encountered deals with the parameterization of the model. Numerous parameters have to be defined both at the molecular and at the cellular levels (see Section 2). For instance, all parameter values used in the gene-gene interaction matrix were based on previous experiences with such models, since values of such parameters cannot be found in the literature (mainly because such parameters depend on modeling choices). Since the aim of this work was to establish a proof-of-concept that we could reproduce expected dynamics, parameters were tuned for such a purpose. In trying to improve the realism of our model, more values should be estimated from experimentally measured datasets. However, it is a notoriously difficult task to obtain realistic values for certain molecular parameters, such as degradation rates, from experiments [63]. The runtime in `Simuscale` being significantly faster and more competitive than previous software, it is an advantage for estimating more parameter values.

Although a full sensitivity analysis was beyond the scope of the present work, we nevertheless could analyze the impact of a few parameters through a local sensitivity analysis and a global sensitivity analysis on protein degradation rates. Quite interestingly, some characteristic population-level behavior of the model proved to be very robust to specific influences. Besides robust parameters, there are also some for which a more in-depth sensitivity analysis will be required, such as the interaction strength of the toggle switch between $G5$ and $G6$, or the time-dependent parameters for cell-cell interaction. Preliminary testing shows that varying those parameters may well change the population dynamics, as well as the differentiation states.

The partial global sensitivity analysis performed on the protein degradation rates showed the feasibility of such an analysis applied to population models driven by molecular interactions. It highlighted dependencies on parameters that had not emerged from the local sensitivity analysis. As expected, parameters $d_{1,2}$ and $d_{1,3}$, controlling cell proliferation and cell death, accounted for most of the variance in the response variables associated with the expansion phase. Yet, parameters $d_{1,7}$, $d_{1,2}$ and $d_{1,6}$, associated with the effector phenotype, the control of cell proliferation and the memory phenotype, were found important for the formation of memory cells and the balance between memory cells at the end of the response and number of cells at the peak, which is a measure of the quality of the CD8 T cell response. Such a global sensitivity analysis may provide hints on how complex interactions between molecular processes result in specific cell population dynamics.

The next step in our approach will consist in changing the principle-based GRN to a more realistic one. We are currently in the process of using CARDAMOM [61] to infer a GRN from the time-stamped scRNAseq data from [64]. In particular, CARDAMOM would infer gene-gene interaction parameters. Interestingly, the model described in the present work already gives us an idea of 1) what to expect, i.e. we need at least a toggle switch, which will lead to the differentiation state of the cell; and 2) where the signals should affect GRN behavior, i.e. APC signalling should influence the proliferation, and TCC signalling should impact the death. Moreover, our experience in choosing the

degradation rate parameter values will be useful too, as long as they have not been experimentally determined. Finally in the case where we will try to fit experimental data, the mRNAs distributions observed experimentally will be compared directly to the model output, and the model parameters will be adjusted to minimize any discrepancy.

There are a number of changes that can be envisioned to improve the modeling approach:

1. The 3D domain represented in `Simuscale` is currently a fixed-size cube, which can be reduced to speed up the contraction phase. It could also be enlarged or its shape be modified, to assess the influence on the population dynamics.

2. All cellular decisions are built upon the crossing of a given threshold level. We could assess the impact of changing this rule for a more realistic one that would make the *probability* of a decision to depend upon a given protein level.

3. As of now, signalling can only be modeled by cell-to-cell contact. However, we have previously demonstrated the impact of the local diffusion of IL2 on the CD8 T cell response [20]. We could therefore implement a diffusion process in the extracellular space, although it would have to be carefully drafted not to slow down the code execution too significantly.

4. Different CD8 T cell phenotypes have been considered in this work, based on Todorov et al. (2022), but noticeably not all phenotypes have been considered. This is the case for exhausted CD8 T cells, which would be highly relevant when studying chronic infections or cancer-induced immune response. Adding another phenotype would require developing a branch of the GRN associated to this differentiation, and specific to the considered phenotype.

Acknowledgments

We thank all members of the ARN MEMOIRE project (MEMOIRE ANR-18-CE45-0001) for helpful discussions. We thank Ulysse Herbach (Inria Nancy) who was very helpful

to implement the bursty regime model within `Simuscale` and the Pôle Scientifique de Modélisation Numérique (PSMN) of ENS de Lyon for providing the computational resources. We thank the BioSyL Federation and the LabEx Ecofect (ANR- 11-LABX-0048) of the University of Lyon for inspiring scientific events.

Funding

This work was supported by the ARN MEMOIRE project (MEMOIRE ANR-18-CE45-0001).

Author Contributions

Conceptualization, TNTN, CA, OG, FC; Funding acquisition, OG, FC.; Investigation, All; Methodology, TNTN, SB, OG, FC; Project administration, OG, FC; Software, TNTN, MM, SB; Supervision, OG, FC; Visualization, TNTN, OG, FC; Writing – original draft, TNTN, OG, FC; Writing – review & editing, All.

References

[1] M. L. Dustin, The immunological synapse, Cancer Immunol Res 2 (11) (2014) 1023–33.

[2] H. Todorov, M. Prieux, D. Laubreton, M. Bouvie, S. Wang, S. de Bernard, C. Arpin, R. Cannoodt, W. Saelens, A. Bonnaffoux, O. Gandrillon, F. Crauste, Y. Saeys, J. Marvel, CD8 memory precursor cells generation is a continuous process, iScience 25, 104927 (2022).

[3] S. M. Kaech, W. Cui, Transcriptional control of effector and memory $CD8^+$ T cell differentiation, Nat Rev Immunol 12 (11) (2012) 749–61.

[4] F. Crauste, E. Terry, I. Mercier, J. Mafille, S. Djebali, T. Andrieu, B. Mercier,

G. Kaneko, C. Arpin, J. Marvel, O. Gandrillon, Predicting pathogen-specific CD8 T cell immune responses from a modeling approach, J Theor Biol 374 (2015) 66–82.

[5] K. Murali-Krishna, L. L. Lau, S. Sambhara, F. Lemonnier, R. Ahmed, Persistence of memory CD8 T cells in MHC class I-deficient mice, Science 286 (5443) (1999) 1377–81.

[6] S. Baral, R. Raja, P. Sen, N. M. Dixit, Towards multiscale modeling of the CD8(+) T cell response to viral infections, Wiley Interdiscip Rev Syst Biol Med 11 (4) (2019) e1446.

[7] M. H. Swat, G. L. Thomas, J. M. Belmonte, A. Shirinifard, D. Hmeljak, J. A. Glazier, Multi-scale modeling of tissues using CompuCell3D, Methods Cell Biol 110 (2012) 325–66.

[8] A. Ghaffarizadeh, R. Heiland, S. H. Friedman, S. M. Mumenthaler, P. Macklin, Physicell: An open source physics-based cell simulator for 3-d multicellular systems, PLOS Computational Biology 14 (2) (2018) 1–31.

[9] S. Hoehme, D. Drasdo, A cell-based simulation software for multi-cellular systems, Bioinformatics 26 (20) (2010) 2641–2.

[10] A. E. Cowan, I. Moraru, J. C. Schaff, B. M. Slepchenko, L. M. Loew, Spatial modeling of cell signaling networks, Methods Cell Biol 110 (2012) 195–221.

[11] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, U. Kummer, Copasi–a complex pathway simulator, Bioinformatics 22 (24) (2006) 3067–74.

[12] S. Andrews, Smoldyn: particle-based simulation with rule-based modeling, improved molecular interaction and a library interface, Bioinformatics 33:, (2017) 710.

[13] E. Agmon, R. K. Spangler, C. J. Skalnik, W. Poole, S. M. Peirce, J. H. Morrison, M. W. Covert, Vivarium: an interface and engine for integrative multiscale modeling in computational biology, Bioinformatics 38 (7) (2022) 1972–1979.

[14] M. Verma, J. Bassaganya-Riera, A. Leber, N. Tubau-Juni, S. Hoops, V. Abedi, X. Chen, R. Hontecillas, High-resolution computational modeling of immune responses in the gut, Gigascience 8 (6) (2019) giz062.

[15] P. Varela, C. Ramos, P. Monteiro, C. Chaouiya, EpiLog: A software for the logical modelling of epithelial dynamics [version 2; peer review: 3 approved], F1000Research 7 (1145) (2019).

[16] M. E. Tejero, D. Lashgari, R. García-Valiente, X. Gao, F. Crauste, P. A. Robert, M. Meyer-Hermann, M. R. Martínez, S. M. van Ham, J. E. J. Guikema, H. Hoefsloot, A. H. C. van Kampen, 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, Frontiers in Immunology 11 (2021).

[17] T. J. Sego, J. P. Sluka, H. M. Sauro, J. A. Glazier, Tissue forge: Interactive biological and biophysics simulation environment, PLOS Computational Biology 19 (10) (2023) e1010768.

[18] G. Letort, A. Montagud, G. Stoll, R. Heiland, E. Barillot, P. Macklin, A. Zinovyev, L. Calzone, Physiboss: a multi-scale agent-based modelling framework integrating physical dimension and cell signalling, Bioinformatics 35 (7) (2019) 1188–1196.

[19] L. Calzone, V. Noël, E. Barillot, G. Kroemer, G. Stoll, Modeling signaling pathways in biology with maboss: From one single cell to a dynamic population of heterogeneous interacting cells, Comput Struct Biotechnol J 20 (2022) 5661–5671.

[20] X. Gao, C. Arpin, J. Marvel, S. Prokopiou, O. Gandrillon, F. Crauste, IL-2 sensitivity and exogenous IL-2 concentration gradient tune the productive contact duration of CD8+ T cell-APC : a multiscale modeling study, BMC Sys Biol 10 (2016) 77.

[21] S. Prokopiou, L. Barbarroux, S. Bernard, J. Mafille, Y. Leverrier, C. Arpin, J. Marvel, O. Gandrillon, F. Crauste, Multiscale modeling of the early CD8 T cell immune response in lymph nodes : an integrative study, Computation 2 159-181 (2014).

[22] A. Becskei, B. B. Kaufmann, A. van Oudenaarden, Contributions of low molecule number and chromosomal positioning to stochastic gene expression, Nat Genet 37 (9) (2005) 937–944.

[23] A. Coulon, O. Gandrillon, G. Beslon, On the spontaneous stochastic dynamics of a single gene: complexity of the molecular interplay at the promoter, BMC systems biology 4 (4) (2010) 2.

[24] M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, Stochastic gene expression in a single cell, Science 297 (5584) (2002) 1183–1186.

[25] A. Raj, A. van Oudenaarden, Nature, nurture, or chance: stochastic gene expression and its consequences, Cell 135 (2) (2008) 216–226.

[26] J. M. Raser, E. K. O'Shea, Noise in gene expression: origins, consequences, and control, Science 309 (5743) (2005) 2010–3.

[27] A. Sigal, R. Milo, A. Cohen, N. Geva-Zatorsky, Y. Klein, Y. Liron, N. Rosenfeld, T. Danon, N. Perzov, U. Alon, Variability and memory of protein levels in human cells, Nature 444 (7119) (2006) 643–6.

[28] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, F. Naef, Mammalian genes are transcribed with widely different bursting kinetics, Science 332 (6028) (2011) 472–4.

[29] S. Bernard, F. Crauste, O. Gandrillon, C. Knibbe, D. Parsons, Simuscale: A modular framework for multiscale single-cell modelling, Tech. Rep. RT-0520, Inria Lyon (2024).

[30] U. Herbach, A. Bonnaffoux, T. Espinasse, O. Gandrillon, Inferring gene regulatory networks from single-cell data: a mechanistic approach, BMC Systems Biology 11 (2017) 105.

[31] S. Girel, C. Arpin, J. Marvel, O. Gandrillon, F. Crauste, Model-based assessment of the role of uneven partitioning of molecular content on heterogeneity and regulation of differentiation in CD8 T-cell immune responses, Front. Immunol 10, article 230 (2019).

[32] M. Ko, A stochastic model for gene induction, J Theor Biol. 153 (1991) 181–194.

[33] J. Peccoud, B. Ycart, Markovian modelling of gene product synthesis, Theoretical population biology 48 (1995) 222–234.

[34] U. Herbach, Gene regulatory network inference from single-cell data using a self-consistent proteomic field, arXiv 2109.14888 (2022).

[35] E. Ventre, Reverse engineering of a mechanistic model of gene expression using metastability and temporal dynamics, In Silico Biology 14 (2021) 89–113.

[36] S. M. Kaech, R. Ahmed, Memory $CD8^+$ T cell differentiation: initial antigen encounter triggers a developmental program in naïve cells, Nat Immunol. 5 (2) (2001) 415–422.

[37] M. J. B. van Stipdonk, E. E. Lemmens, S. P. Schoenberger, Naïve CTLs require a single brief period of antigenic stimulation for clonal expansion and differentiation, Nat. Immunol. 2 (2019) 423–429.

[38] D. Kyburz, D. E. Speiser, T. Aebischer, H. Hentgartner, R. M. Zinkernagel, Virus-specific cytotoxic T cell-mediated lysis of lymphocytes in vitro and in vivo, J. Immunol. 150 (1) (1993) 5051–5058.

[39] M. W. Su, P. R. Walden, D. B. Golan, H. N. Eisen, Cognate peptide-induced destruction of $CD8^+$ cytotoxic T lymphocytes is due to fratricide, J. Immunol. 1051 (2) (1993) 658–667.

[40] J. C. Nolz, Molecular Mechanisms of CD8$^+$ T cell Trafficking and Localization, Cell. Mol. Life Sci. 72 (13) (2015) 2461–2473.

[41] A. Bonnaffoux, U. Herbach, A. Richard, A. Guillemin, S. Gonin-Giraud, P.-A. Gros, O. Gandrillon, Wasabi: a dynamic iterative framework for gene regulatory network inference, BMC Bioinformatics 20 (2019) 220.

[42] U. Herbach, Harissa: stochastic simulation and inference of gene regulatory networks based on transcriptional bursting, in: Pang, J., Niehren, J. (eds) Computational Methods in Systems Biology. CMSB 2023. Lecture Notes in Computer Science, Vol. 14137, 2023.

[43] S. Tarantola, D. Gatelli, T. A. Mara, Random balance designs for the estimation of first order global sensitivity indices, Reliability Engineering & System Safety 91 (6) (2006) 717–727.

[44] J.-Y. Tissot, C. Prieur, Bias correction for the estimation of sensitivity indices based on random balance designs, Reliability Engineering & System Safety 107 (2012) 205–213.

[45] E. Plischke, An effective algorithm for computing global sensitivity indices (easi), Reliability Engineering & System Safety 95 (4) (2010) 354–360.

[46] D. A. Lewis, T. Ly, Cell Cycle Entry Control in Naïve and Memory CD8$^+$ T Cells, Front. Cell Dev. Biol. 9 (2021) 727441.

[47] V. Kalia, S. Sarkar, Regulation of Effector and Memory CD8 T Cell Differentiation by IL-2—A Balancing Act, Front. Immunol. 9 (2018) 2987.

[48] J. M. Depper, W. J. Leonard, C. Drogula, M. Krönker, T. A. Waldmann, W. C. Greene, Interleukin 2 (IL-2) augments transcription of the IL-2 receptor gene, Proc Natl Acad Sci U.S.A. 82 (12) (1985) 4230–4234.

[49] P. H. Krammer, CD95's deadly mission in the immune system, Nature 407 (2000) 789–795.

[50] S. Nagata, P. Goldstein, The Fas Death Factor, Science 267 (5203) (1995) 1449–1456.

[51] N. S. Joshi, S. M. Kaech, Effector CD8 T cell development: a balancing act between memory cell potential and terminal differentiation, J. Immunol. 180 (3) (2008) 1309–15.

[52] A. M. Intlekofer, A. Banerjee, N. Takemoto, S. M. Gordon, C. S. Dejong, H. Shin, C. A. Hunter, W. E. J., T. Lindstein, S. L. Reiner, Anomalous type 17 response to viral infection by CD8+ T cells lacking T-bet and eomesodermi, Science 321 (5887) (2008) 408–11.

[53] M. R. Alderson, T. W. Tough, T. Davis-Smith, S. Braddy, B. Falk, K. A. Shooley, R. G. Goodwin, C. A. Smith, D. Lynch, Fas ligand mediates activation-induced cell death in human T lymphocytes, J. Exp. Med. 181 (1) (1995) 71–77.

[54] H. Huang, D. J. Tindall, Dynamic FoxO transcription factors, J cell Sci 120 (15) (2007) 2479–87.

[55] A. Delpoux, C.-Y. Lai, S. M. Hedrick, A. L. Doedens, FOXO1 opposition of $CD8^+$ T cell effector programming confers early memory properties and phenotypic diversity, Proc Natl Acad Sci U.S.A. 114 (42) (2017) E8865–E8874.

[56] J. M. Grayson, Z. M. J., S. D. Altman, R. Ahmed, Cutting Edge: Increased Expression of Bcl-2 in Antigen-Specific Memory CD8+ T Cells, J. Immunol 164 (8) (2000) 3950–3954.

[57] E. Becht, L. McInnes, J. Healy, C. A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, E. W. Newell, Dimensionality reduction for visualizing single-cell data using umap, Nat Biotechnol 37 (2018) 38–44.

[58] D. Pilling, A. Akbar, N. Shamsadeen, D. Scheel-Toellner, C. Buckley, M. Salmon, High cell density provides potent survival signals for resting T-cells, Cell. Mol. Biol. 46 (1) (2000) 163–174.

[59] J. Moore, H. Ahmed, D. McGuire, R. Akondy, R. Ahmed, R. Antia, Dependence of CD8 T Cell Response upon Antigen Load During Primary Infection, Bull. Math. Biol. 81 (2019) 2553–2568.

[60] V. P. Badovinac, B. B. Porter, J. T. Harty, Programmed contraction of $CD8^+$ T cells after infection, Nature Immunol. 3 (7) (2020) 619–626.

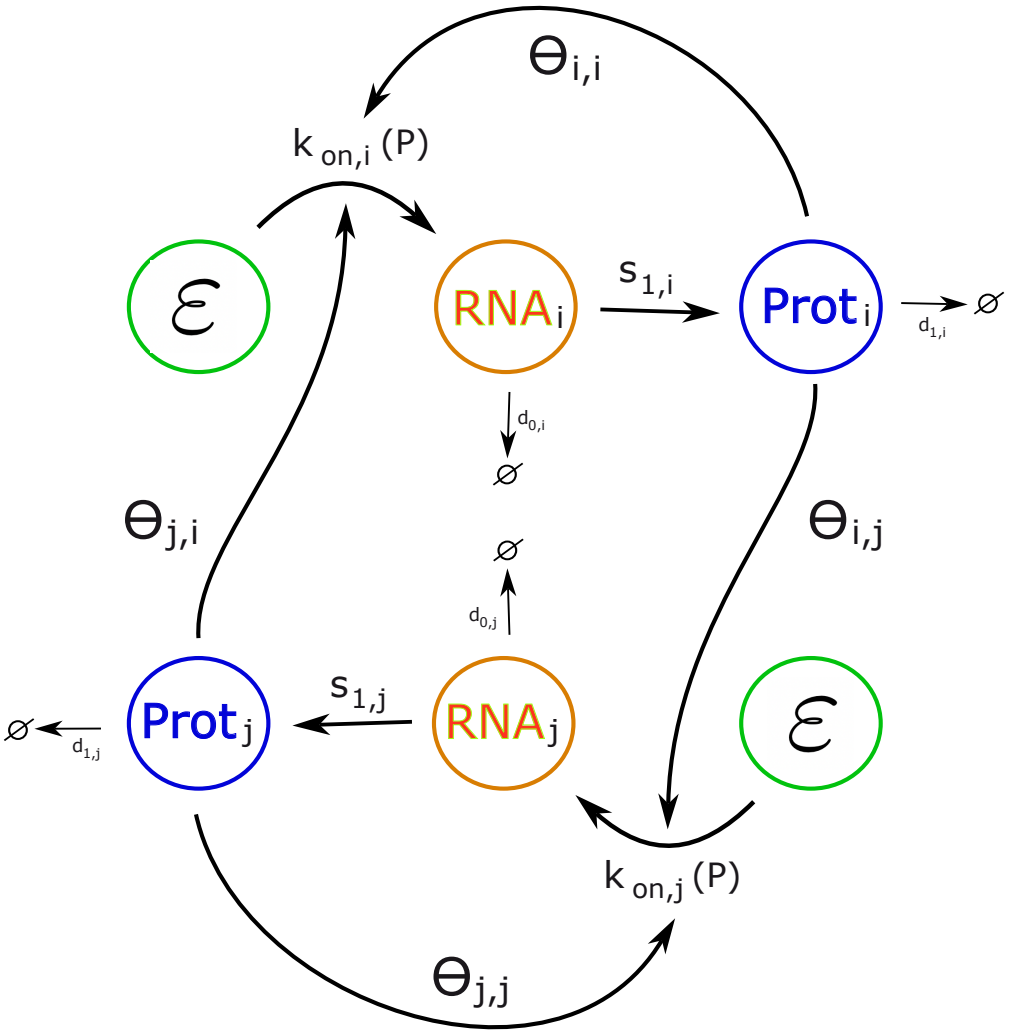
[61] E. Ventre, U. Herbach, T. Espinasse, G. Benoit, O. Gandrillon, One model fits all: combining inference and simulation of gene regulatory networks, PLoS Comput Biol 19 (3) (2023) e1010962.

[62] R. T. Kelly, Single-cell proteomics: Progress and prospects, Mol Cell Proteomics 19 (11) (2020) 1739–1748.

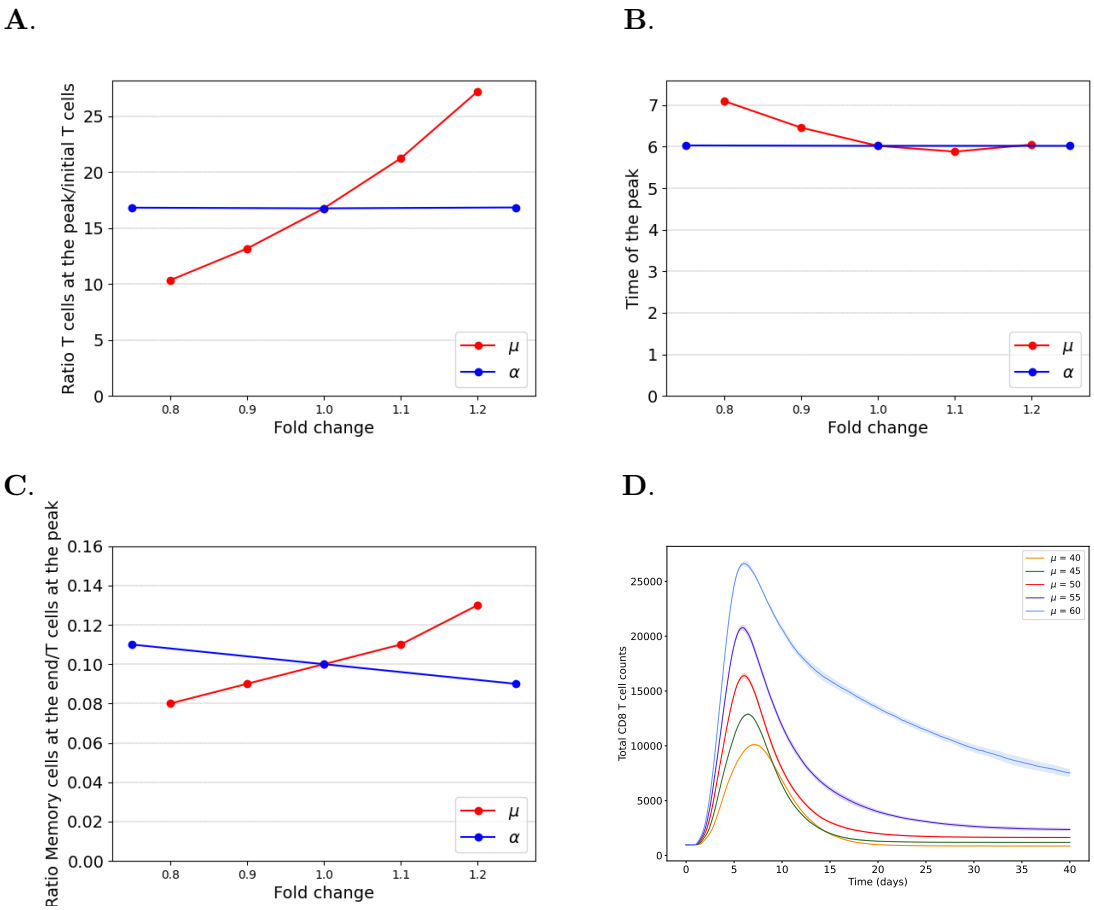
[63] Z. Rolfs, B. L. Frey, X. Shi, Y. Kawai, L. M. Smith, N. V. Welham, An atlas of protein turnover rates in mouse tissues, Nat Commun 12 (1) (2021) 6778.

[64] N. S. Kurd, Z. He, T. L. Louis, J. J. Milner, K. D. Omilusik, W. Jin, M. S. Tsai, C. E. Widjaja, J. N. Kanbar, J. G. Olvera, T. Tysl, L. K. Quezada, B. S. Boland, W. J. Huang, C. Murre, A. W. Goldrath, G. W. Yeo, J. T. Chang, Early precursors and molecular determinants of tissue-resident memory $CD8^+$ T lymphocytes revealed by single-cell RNA sequencing, Sci Immunol 5 (47) (2020) eaaz6894.

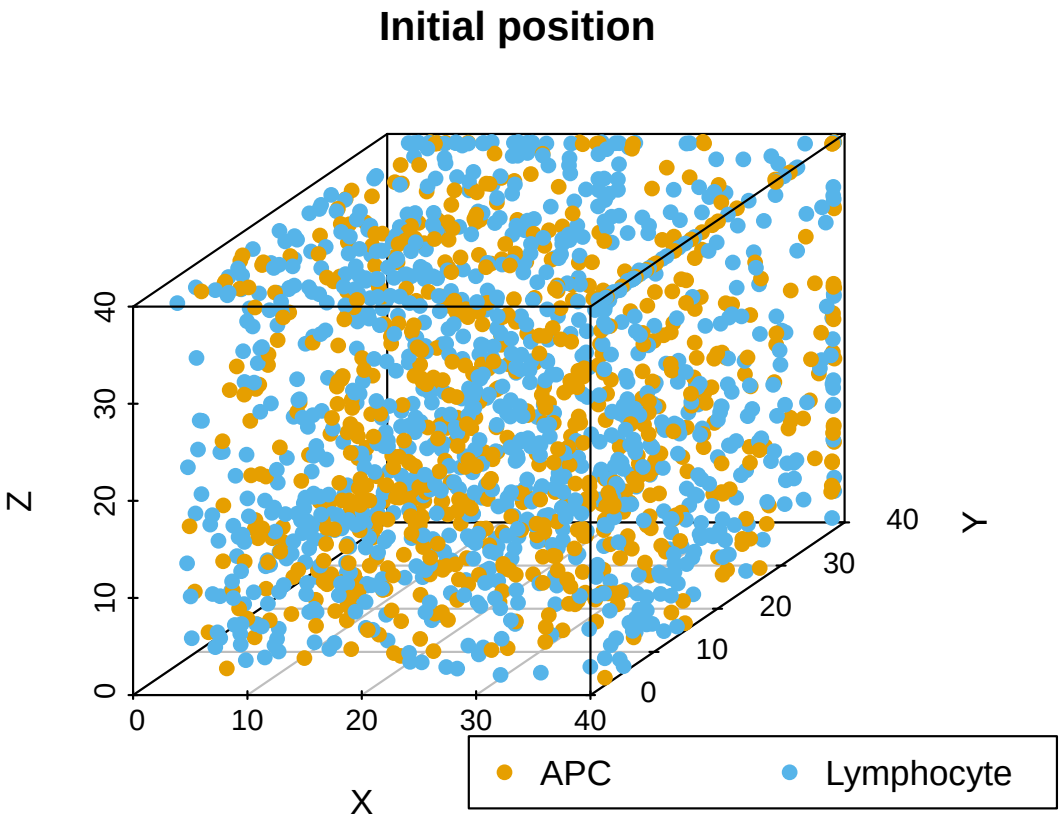
5 Supplementary Figures



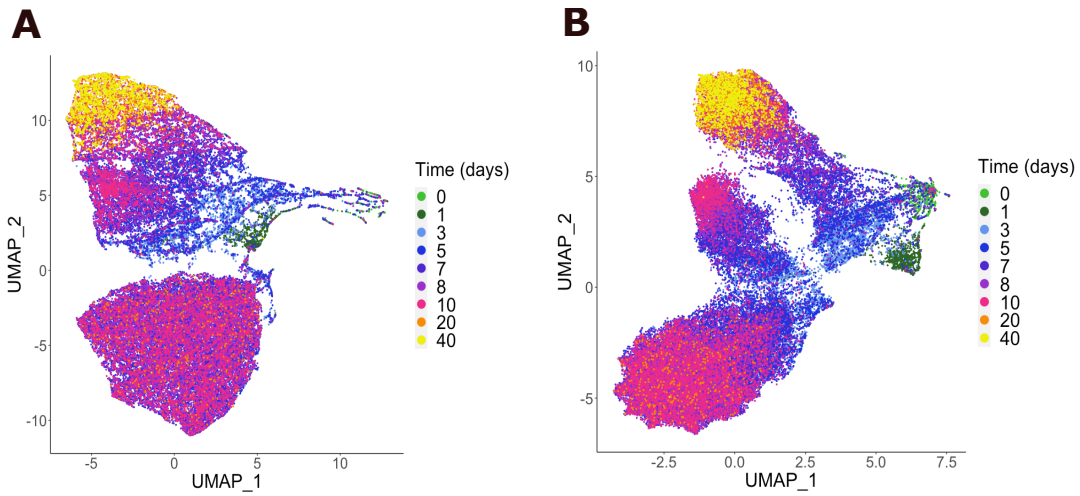
Supplementary Figure 1: A schematic representation of a GRN connecting genes in the bursty regime. Two genes G_i and G_j are represented.



Supplementary Figure 2: Effect of variations of parameters α (signalling parameter) and μ (mean of the exponential distribution of the bursts) on variables of interest. (A) The ratio between CD8 T cells at the peak of the response and the initial T cell number. (B) The time of the peak. (C) The ratio between memory T cells at the end of the simulation and the number of T cells at the peak of the response. Reference values are $\alpha = 20$ and $\mu = 50$, as indicated in the text. (D) Evolution of the total population of CD8 T cells for several values of μ, corresponding to $\pm 20\%$ of the reference value $\mu = 50$.

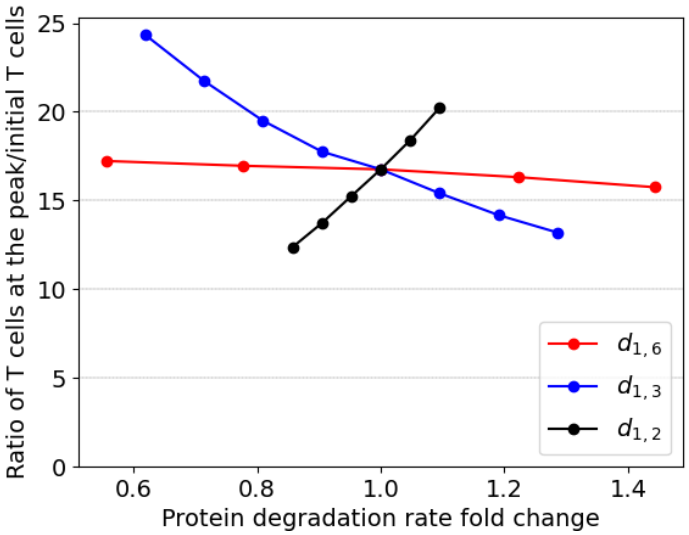


Supplementary Figure 3: Random positions of cells at the beginning of a simulation.

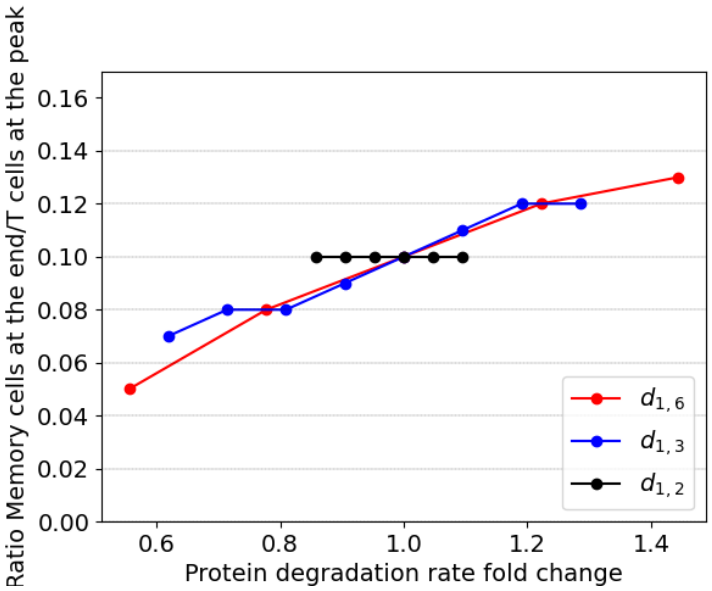


Supplementary Figure 4: UMAP representation of CD8 T cell population dynamics over time using either (A) only the 9-genes principle-based GRN, or (B) the augmented GRN, based on the 9 core genes plus the 27 "decorating" genes. The figure in (B) is repeated here from Figure 5A, to allow direct comparison.

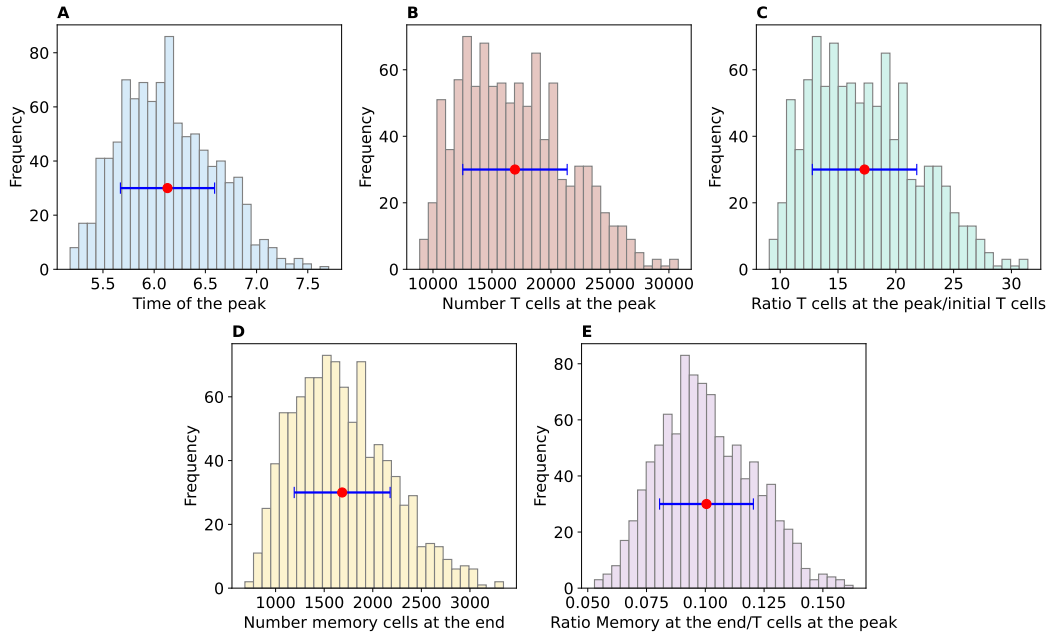
A.



B.



Supplementary Figure 5: Effect of proteins 2, 3 and 6 degradation rates on variables of interest: (A) the ratio between CD8 T cells at the peak of the response and the initial T cell number, (B) the ratio between memory T cells at the end of the simulation and the number of T cells at the peak of the response. In both cases, the ratio T cells/APC equals 1.4, and reference values for protein degradation rates are given in Table 3.



Supplementary Figure 6: Distribution of variables of interest obtained from 1,000 samples generated by LHS. (A) The time of the peak. (B) The number of T cells at the peak. (C) The ratio between T cells at the peak and the initial T cell number (initial T cell number equals 980, unchanged). (D) Memory cell counts at the end of the simulation. (E) The ratio between memory T cells at the end of the simulation and the number of T cells at the peak. Red points (blue lines) are the mean values (standard deviations).