

Model checking for studying timing of events in T cell differentiation

Paolo Zuliani¹, Natasa Miskov-Zivanov², Penelope Morel³, James R. Faeder², and Edmund M. Clarke¹

Short Abstract — We use computational modeling and formal analysis techniques to study the temporal behavior of a logical model of the naïve T cell differentiation. The model is analyzed formally and automatically by performing temporal logic queries via statistical model checking.

I. INTRODUCTION AND MOTIVATION

The goal of this study is to identify key factors and pathways that contribute to the discrimination of the T-cell receptor (TCR) signal strength (*i.e.*, antigen dose/duration/affinity presented to TCR) by the differentiating T cell (Figure 1(a)). Different T cell phenotype ratios play an important role in T-cell mediated immunity, in both autoimmune diseases and in cancer. The two primary phenotypes we consider are: 1) regulatory (Treg) cells that express the transcription factor Foxp3 but do not express the cytokine IL-2; 2) and helper (Th) cells that do not express Foxp3 but do express and secrete IL-2. Control of the Treg vs. Th cell phenotype induction is a promising approach to either eliminate antigen-specific Treg cells and decrease (or even reverse) immune suppression in cancer, or enhance Treg induction to prevent autoimmune diseases. Previous studies have indicated that the timing of T cell stimulation, both antigen dose and the duration of antigen stimulation, strongly influence the T cell phenotype choice [1].

To study this system, we apply computational modeling approaches and formal methods from electronic design automation (EDA). The model used in this work (described in [2]) couples exogenous signaling inputs to T cell phenotype decisions. This model was developed using a discrete, logical modeling approach, and simulated using random asynchronous approach and *BooleanNet* tool [3]. Model simulations described in [2] allow for recapitulating a number of experimental observations and provide new insights into the system. However, to test new properties of the model, it is usually necessary to write new parts of the simulator code, or manually analyze a significant amount of simulation data. This approach quickly becomes tedious and error-prone.

In this work, we apply temporal logic model checking to automatically analyze the behavior of the model. Since the underlying semantic model of *BooleanNet* is essentially a discrete-time Markov chain, we need to verify probabilistic (stochastic) models. The verification problem for stochastic systems amounts to compute the probability that a given temporal logic formula is satisfied by the system. One approach to the verification problem uses precise numerical methods to compute exactly the probability that the formula is true (*e.g.* [4]). However, these methods suffer from the state explosion problem, and do not scale well to large-scale systems. Statistical model checking can be effectively used for verifying temporal logic specifications for systems

affected by the state explosion problem. The technique relies on system simulation, thereby avoiding a full state space search. This implies that the answer to the verification problem (*i.e.*, the probability that the property holds) is only approximate, but its accuracy can be arbitrarily bounded by the user. In return, statistical model checking is more scalable and hence more useful for large models.

II. METHODOLOGY

The steps of our methodology are presented in Figure 1(b) and described below. We encode relevant properties of the model as temporal logic formulae, which are then verified via statistical model checking. We use Bounded Linear Temporal Logic (BLTL) as our specification language. BLTL restricts the well-known Linear Temporal Logic (LTL) with time bounds on the temporal operators. For example, a BLTL formula expressing the specification “it is not the case that in the Future 10 time steps CD25 is Globally activated (*i.e.*, it equals 1) for 17 time steps” is written as

$$\neg \mathbf{F}^{10} \mathbf{G}^{17} (\text{CD25} = 1)$$

where the \mathbf{F}^{10} operator encodes “future 10 time steps”, \mathbf{G}^{17} expresses “globally for 17 time steps”, and CD25 is a state variable of the model. The syntax of BLTL is given by:

$$\psi ::= y \sim v \mid \psi_1 \wedge \psi_2 \mid \psi_1 \vee \psi_2 \mid \neg \psi_1 \mid \psi_1 \mathbf{U}^t \psi_2$$

where $\sim \in \{\leq, \geq, =\}$, $y \in \text{SV}$ (the finite set of state variables), $v \in \mathcal{R}$, $t \in \mathcal{R}_{>0}$, and \neg, \vee, \wedge are the usual Boolean connectives. Formula of the type $y \sim v$ are also called atomic propositions. The formula $\psi_1 \mathbf{U}^t \psi_2$ holds true if and only if, within time t , ψ_2 will be true and ψ_1 will hold until then. Note that the operators \mathbf{F}^t and \mathbf{G}^t referenced above are easily defined in terms of the until \mathbf{U}^t operator: $\mathbf{F}^t \psi = \text{true} \mathbf{U}^t \psi$ requires ψ to hold true within time t (*true* is the atomic proposition identically true); $\mathbf{G}^t \psi = \neg \mathbf{F}^t \neg \psi$ requires ψ to hold true up to time t .

We have combined *BooleanNet* with a parallel statistical model checker, so that verification of BLTL properties can be performed efficiently and automatically on a multi-core system. Statistical model checking treats the verification problem for stochastic systems as a statistical inference problem, using randomized sampling to generate traces (or simulations) from the system model, then using model checking methods and statistical analysis on those traces. Efficient Bayesian techniques were introduced and successfully applied to the verification of rule-based models of signaling pathways and other stochastic systems [5][6]. In particular, the approach is based on sequential estimation, and given a coverage probability and an interval width, it returns a Bayesian confidence interval for the probability that the BLTL formula is true.

III. RESULTS

Experimental observations from [1] that the induction/expansion of Foxp3+ Treg cells by low dose antigen is inversely correlated with the levels of signaling via the mTOR pathway suggest a complex interaction between cell surface receptors, signaling molecules and important transcription factors. The model in [2] captures critical

¹Computer Science Department, Carnegie Mellon University, E-mail: {pzuliani, emc}@cs.cmu.edu.

²Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, E-mail: {nam66,faeder}@pitt.edu

³Department of Immunology, School of Medicine, University of Pittsburgh, E-mail: morel@pitt.edu

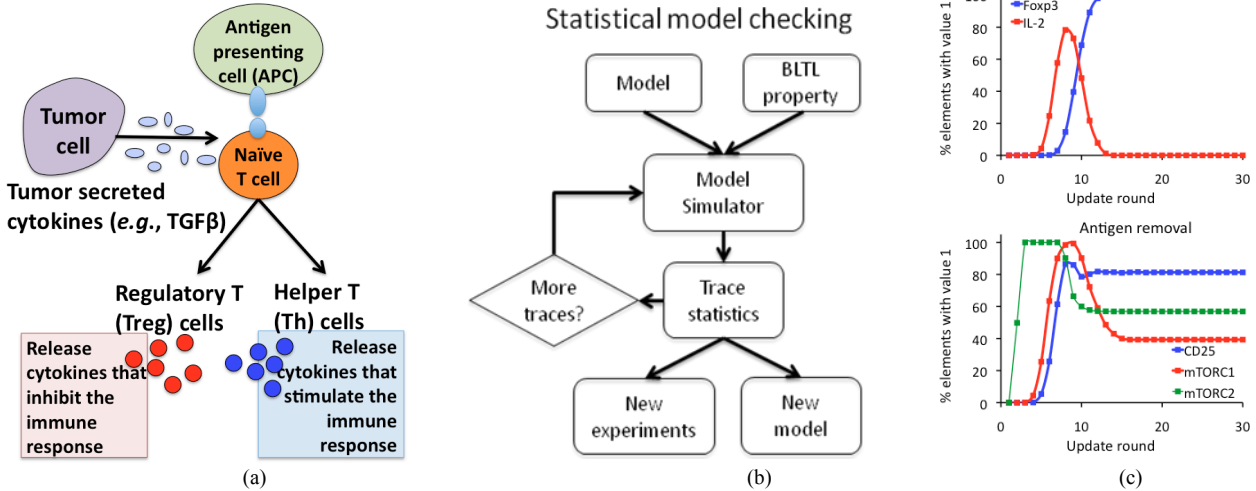


Figure 1. Modeling of immune systems cells: (a) Differentiation of naïve T cells into Treg or Th, induced antigen presented by APC, or cytokines secreted by tumor cells; (b) statistical model checking flow; (c) model simulation results for two scenarios.

signaling events, from stimulatory signals at receptors, through activation of transcription factors, to production of proteins representing different phenotypes.

Several model simulation results obtained using *BooleanNet* are shown in Figure 1(c). These results present the behavior of critical elements in the model averaged across 1000 simulation trajectories, for two different stimulation scenarios. When naïve T cells are stimulated with low antigen dose, they can differentiate into Treg cells expressing Fxp3. Similarly, model simulations that mimic the low antigen dose case result in steady state with Fxp3=1 (Figure 1(c) (top)). Model simulation results show that the behavior of IL-2 gene expression early after stimulation is similar for both low and high antigen dose. This is not so straightforward to measure in experiments as IL-2 is measured outside of cells, where it is consumed quickly after being expressed and secreted. What is not clear from averaged simulation trajectories (Figure 1(c) (top)) is whether IL-2 reaches value 1 on all trajectories, but at different update rounds, or whether it reaches value 1 on only 80% of trajectories. To test this, we consider the property $F^{20} (IL2 = 1)$. Statistical model checking shows that the probability that this property holds is close to 1. We have also computed the probability that IL-2 remains at level 0 until its inhibitor, Fxp3, becomes 1. This property:

$$(IL2 = 0) U^{15} (FOXP3 = 1)$$

is returned as a low-probability event. In other words, our model predicts initial increase in IL-2, irrespective of antigen dose scenario, and the criticality of variations in other element values for phenotype decision.

Another observation from experiments is that removal of antigen 18 hours after stimulation results in a mixed population of Treg and Th cells. Studies of the model have indicated that early events and relative timing of the Fxp3 activating and inhibiting pathways play crucial role in this differentiation. Figure 1(c)(bottom) shows transient behavior of CD25 (main element on Fxp3 activating pathway) and mTORC1/mTORC2 (inhibitors of Fxp3). With model checking, we were able to carry further and more efficient studies of early behavior of these elements. In Table I, we present a set of properties that we tested using statistical model checking and results obtained. We also include elapsed time that was necessary for checking those properties.

This scenario results in a mixed population of cells with

different phenotype. Model checking results outline that early events in CD25, mTORC1 and mTORC2 are good predictors of the mixed population, as most of the results show close to

Table I.

Tested properties and model checker runtime on a 48-core system. Coverage probability=0.999; half-interval=0.01, except for Property 1 (=0.001)

	Property	Probability estimate and sample size	Elapsed time [s]
1	$G^7 \sim (MTORC1 = 1 \ \& \ MTORC2 = 1)$	estimate = 0.0188048 samples = 200,160	1,946
2	$F^7 (MTORC1 = 1 \ \& \ MTORC2 = 1)$	estimate = 0.980884 samples = 2,352	23
3	$F^{10} (MTORC1 = 1 \ \& \ MTORC2 = 1 \ \& \ CD25 = 0 \ \& \ (F^{18} (CD25 = 1)))$	estimate = 0.60104 samples = 25,968	253
4	$F^{28} (MTORC1 = 1 \ \& \ MTORC2 = 1 \ \& \ CD25 = 0 \ \& \ (F^1 (CD25 = 1)))$	estimate = 0.592195 samples = 26,160	254
5	$F^{10} (MTORC1 = 1 \ \& \ MTORC2 = 1 \ \& \ CD25 = 0 \ \& \ (F^1 (G^{17} (CD25 = 1))))$	estimate = 0.39669 samples = 25,920	254

50% successes. In other words, although the tested properties would return more uniform behavior in other scenarios, in the case of antigen removal, we see more variability between possible trajectories. The next step now is to design further queries that could uncover exact relationship between early events and specific outcomes.

IV. CONCLUSION

Model checking is an efficient approach for studying cell signaling network models, as it allows for answering a variety of questions about the system. Instead of manually analyzing simulation trajectories and large output files, one creates properties that can be automatically verified. We uncovered several relationships between early behavior of elements in our T cell model. With the framework that we created, we will continue to study this model, focusing on several other key relationships, such as the one between Fxp3 and PTEN.

REFERENCES

- [1] M. S. Turner et al., "Dominant role of antigen dose in CD4+Foxp3+ regulatory T cell induction and expansion," in *J. of Immunology*, 183, pp. 4895-903, 2009.
- [2] N. Miskov-Zivanov et al., "Modeling and Analysis of Peripheral T Cell Differentiation," in preparation.
- [3] I. Albert et al. "Boolean network simulations for life scientists." *Source Code for Biology and Medicine* 2008, 3:16
- [4] M. Kwiatkowska et al. "PRISM 4.0: Verification of Probabilistic Real-time Systems." In *Proc. of CAV'11, LNCS 6806*, pages 585-591, 2011.
- [5] H. Gong et al. "Analysis and Verification of the HMGB1 Signaling Pathway." *BMC Bioinformatics* 2010, 11(Suppl 7):S10.
- [6] P. Zuliani et al. "Bayesian Statistical Model Checking with Application to Stateflow/Simulink Verification." In *HSCC 2010*, pages 243-252, 2010.