Integrated Mechanistic-Empirical Modeling of Cellular Response Based on Intracellular Signaling Dynamics

Proceedings of the ASME 2013 Dynamic Systems and Control Conference
DSCC2013
October 21-23, 2013, Palo Alto, California, USA

DSCC2013-3806

INTEGRATED MECHANISTIC-EMPIRICAL MODELING OF CELLULAR RESPONSE BASED ON INTRACELLULAR SIGNALING DYNAMICS

Michaëlle N. Mayalu
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139 USA
mmayalu@mit.edu

H. Harry Asada
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139 USA
asada@mit.edu

ABSTRACT

A hybrid modeling framework integrating a highly specific mechanistic model with highly abstract empirical model is presented. With the growing interest in the scientific and medical community for identification of therapeutic targets in treatment of disease, it is necessary to develop predictive models that can describe cellular behavior in response to environmental cues. Intracellular signaling pathways form complex networks that regulate cellular response in both health and disease. Mechanistic (or white-box) models of biochemical networks are often unable to explain comprehensive cellular response due to lack of knowledge and/or intractable complexity (especially in events distal from the cell membrane). Empirical (or black-box) models may provide a less than accurate representation of cellular response due to data deficiency and/or loss of mechanistic detail. In the proposed framework, we use a mechanistic model to capture early signaling events and apply the resulting generated internal signals (along with external inputs) to a downstream empirical sub-model. The key construct in the approach is the treatment of a cell's biochemical network as an encoder that creates a functional internal representation of external environmental cues. The signals derived from this representation are then used to inform downstream behaviors. Using this idea, we are able to create a comprehensive framework that describes important mechanisms with sufficient detail, while representing complex or unknown mechanisms in a more abstract form. The model is verified using published biological data describing T-Cells in immune response.

INTRODUCTION

Building accurate dynamic models of intracellular signaling events in response to extracellular cues is the key step towards the development of predictive models for cells or whole organisms. These models will ultimately provide scientific explanations of behavior of biological systems in health and disease as well as the potential for control of these systems. In particular, there is growing interest in the pharmaceutical industry for control of cellular response in the development of therapeutic targets for treatment of diseases, including cancer metastasis and autoimmunity [1] . In order to understand the effects of extracellular cues on cellular response, it is crucial to examine and identify the internal mechanisms involved in normal functioning of cells and also the defects associated with disease [1].

Cellular response is regulated by the transfer of information from the environment to within the cell. This transfer is realized through a complex biochemical network in response to different extracellular cues. With recent advancements in measurement and sensing, useful mechanistic models (white-box models) of biochemical networks have been developed. Some examples of relatively detailed mechanistic models include Markov chains [2], and differential equations [3]. However, these white-box models require an enormous amount of mechanistic detail due to the inherent complexity of the biological processes. Furthermore, the details behind some mechanisms, especially complex events downstream of the cell membrane, remain unclear. Therefore, due to lack of knowledge and/or intractable complexity, white-box models are often unable to explain a comprehensive cellular response from input cue to observable output or phenotypic change.

To decrease complexity, another common modeling approach is to use experimental data to empirically model phenotypic responses to external stimuli. Purely empirical (or black-box models) such as Clustering [4] and Partial Least Squares [5] correlate experimentally measured variables to elucidate model components and potential relationships. However, in order to accurately represent a given data set, empirical models require a large amount of data. Accumulating appropriate data may be time-consuming and impractical due to limited high throughput technologies for certain biological processes. In addition, empirical models contain little to no mechanistic detail and consequently

understanding of the true mechanisms is lost. Therefore due to data deficiency and loss of mechanistic detail, empirical models may provide a less than accurate representation of the cellular process being studied.

Our goal is to create a comprehensive model with moderate complexity and marginal loss of mechanistic detail that will accurately represent a biological process and its dynamics. It is therefore necessary to consider approaches outside of standard engineering methodologies for control, modeling, simulation and analysis. Our approach attempts to create a hybrid (grey-box) modeling framework that methodically connects a highly specific mechanistic (white-box) model with highly abstract empirical (black-box) model. Previous integrated multi-scale models have made connections between adjacent scales of specificity ([4, 6]), but few have attempted to merge models at such contrasting levels of abstraction.

In the proposed framework, we use a mechanistic model to capture early signaling events and apply the generated internal signals, along with external inputs, to a downstream empirical sub-model. The key construct in our approach is the treatment of a cell's biochemical network as an encoder that creates a functional internal representation of external environmental cues. The signals derived from this representation are then used to inform downstream behaviors. The integrative approach of the proposed framework allows flexibility in the design and construction of each sub-model. In addition, the proposed framework essentially decouples the rapid dynamics involved in the signal transduction process from the slower processes occurring inside the nucleus, such as gene regulation. To our knowledge the proposed modeling framework is a novel approach to analysis of dynamic signaling data.

The paper is organized as follows. In the next section, a brief review of the physiochemical and kinetic mechanisms involved in intracellular signaling is discussed. Next, we introduce the model framework and validate on published data regarding T-Cells in immune response We conclude with discussion of various results and potential implications.

CELLULAR RESPONSE: THREE STAGE PROCESS

We may consider cellular response as structured in three stages: input (receptor activation); intermediate (signal transduction); and output (gene regulation and observable response) [1].

Stage One: Receptor Activation

In the first stage, external environmental cues (e.g. to mechanical stresses, biochemical factors, or communication from adjacent cells) interact with a receptor on the cell membrane. Depending on the combination of external cues, the cell responds differently. The interaction between the external cue and membrane receptor causes activation (change in configuration or polarity) of the receptor. Upon activation,

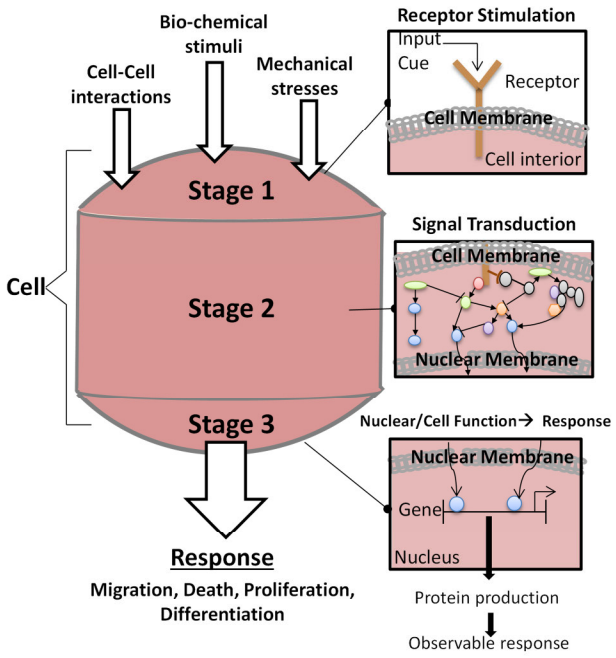

FIGURE 1: CELLULAR RESPONSE STRUCTURED INTO THREE STAGES.

the receptor may activate other molecules and proteins within the cell.

Stage Two: Signal Transduction

In the second stage the cell undergoes the signal transduction process, which consists of a cascade of protein activations originating from the cell membrane. In practical terms, this process serves to encode external signals into an internal representation of the outside environment. The encoded signals are then applied to later stage mechanisms downstream of the cell membrane.

Stage Three: Cellular Function and Observable Response

The third stage involves the transfer of encoded signals (from stage two) to the nucleus where a specific function (such as including gene expression and/or protein fabrication) is performed. After a set of intermediate mechanisms has occurred, the internal state of the cell is altered leading to phenotypic change.

MODEL CONSTRUCTION AND VALIDATION ON PUBLISHED DATA

We may validate our approach by applying our grey-box model framework to published T-Cell receptor (TCR) signaling data. T-Cell lymphocytes play a key role in the immune system homeostasis because they have the ability to recognize foreign agents and initiate immune response [1]. In stage one of T-Cell response, the cell detects foreign agents by means of the TCR which interacts with bacterial or viral derived molecules called MHC (Major Histocompatibility

Complex) [1]. During stage two, upon MHC binding, the TCR initiates multiple signal transductions that cascade through a complex signaling network which controls the activation of several important transcription factors. Finally, stage three of the response involves the travel of transcription factors to the nucleus, leading to production of key proteins, such as IL-2 protein. Kemp et al. explores the correlation between MHC peptide binding affinity and cellular response [5]. As shown in Fig. 2, results suggest that levels of IL-2 production from stimulation with various MHC peptides correlate with apoptotic response (cell death) [5]. For a detailed description on materials and methods used to collect data we refer to the publication [5].

Using our proposed grey-box framework, we would like to model the observed relationship between MHC binding affinity and apoptotic response. The mechanistic (white-box) portion of our framework models the signaling pathways triggered by MHC/TCR binding (stages one and two of T-Cell response). The internal signals (concentration levels of key signaling molecules) generated from the mechanistic model are used as input variables to the downstream empirical sub-model. The empirical portion of our framework uses Partial Least Squares Regression to relate the external inputs (MHC/TCR binding affinity) and internal inputs (levels of key signaling molecules) to IL-2 production. IL-2 production is chosen as a relevant cellular function correlated to apoptotic response.

Mechanistic Sub-model

We have chosen to mechanistically model the relationship between ligand avidity (stage one) and signal transduction (stage two) using a stochastic chemical kinetic simulation algorithm [2]. This method will allow us to simulate multiple observation experiments of activated states of key signaling molecules within the desired network. We model the cell as a well-mixed biochemical reactor with uniformly distributed molecules. Consider the network shown in Fig. 4 [5]. We have modeled the outlined signaling pathway with four linear cascaded reactions ultimately leading to the activation of Akt protein. The activation state of Akt is integral in defining the downstream cellular function of IL-2 production. Consequently, we have chosen to monitor the time-course of activated Akt and use it as the generated internal input signal. The time-dependent number of each molecule in the outlined pathway is given by the state vector: $X(t) = [X_1, X_2, ..., X_N]$ where N is the number of molecular species in both their activated and inactivated states. Each molecule has the ability to undergo activation or inactivation reactions ($R_\nu, \nu = 1,...,m$), that will occur with probability:

$$a_\nu(X(t)) \cdot dt = h_\nu c_\nu \cdot dt \tag{1}$$

where $a_\nu\left(X(t)\right)$ is the propensity function for the $\nu - th$ reaction, c_ν is a constant which depends only on the

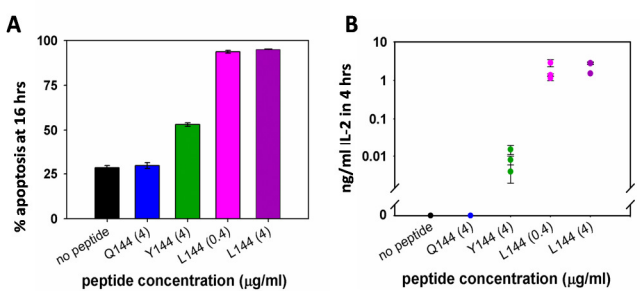

FIGURE 2: A. PRECENTAGE OF DEAD TCELLS AFTER STIMULATION WITH VARIOUS MHC PEPTIDES. B. IL-2 LEVELS IN TCELLS AFTER SIMULATION WITH VARIOUS MHC PEPTIDES. SOURCE:[5]

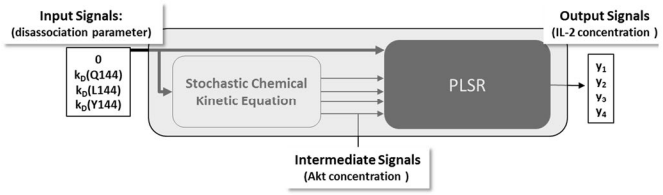

FIGURE 3: MODEL FRAMEWORK SPECIFIC TO PUBLISHED TCELL SIGNALING DATA.

temperature and physical properties of the system, and h_ν is the number of distinct molecular reactant combinations available in the current state $X(t)$. Each reaction defines a transition from the current state $X(t) = q$ to another state $X(t) = q + \zeta$.

Let: $$a_0(X(t)) = \sum_{\nu=1}^{m} a_\nu\ (X\ (t)) \tag{2}$$

be the total propensity over the pathway in the current state. Using the above definitions and equations, we may now write the Chemical Master Equation (or the forward Kolmogorov equation):

$$\frac{dP(q,t)}{dt} = \sum_{\zeta} a_0(\zeta; q-\zeta) P(\zeta; q-\zeta, t) - \sum_{\zeta} a_0(\zeta; q) P(q,t) \tag{3}$$

which describes the time rate of change of the probability ($P(q,t)$) that the system is in state q at time t. Statistically correct solutions to the Chemical Master Equation may be found through conducting Monte Carlo simulations such as the Gillespie stochastic algorithm [2]. It should be emphasized that all parameter values and initial conditions in the mechanistic sub-model were assumed to be known *apriori* and are obtained from literature [7, 8]. Therefore, no parameter fitting was conducted for this portion of the model.

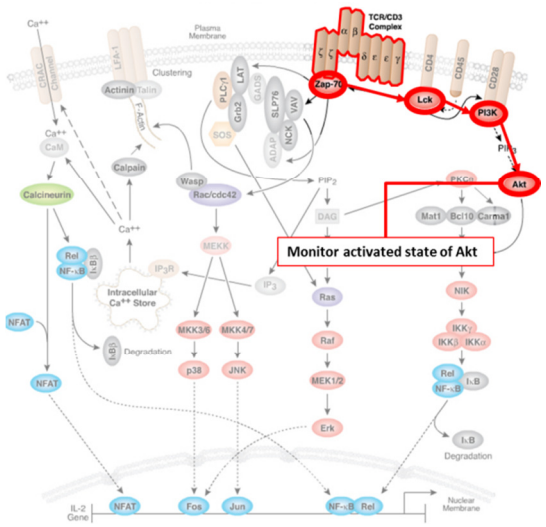

FIGURE 4: DIAGRAM OF MAJOR SIGNALING PATHWAYS INVOLVED IN MHC/TCR ACTIVATION. SOURCE: [5]

EMPIRICAL SUB-MODEL

Partial Least Square Regression was used to relate external and internal input values (MHC/TCR rate parameters for various peptides and corresponding generated Akt signals) to cell functional response (experimentally measured IL-2 concentration). The PLSR algorithm successfully finds the linear combinations of input variables that have maximum correlation with the output variables [5]. Specifically, PLSR fits data to a hyper-plane made up of the data's principal components. Projections of the input data onto the hyper-plane form new variables called scores $\left(t \in \mathbb{R}^{r \times K}\right)$ that are used to approximate the output variables through a linear regression:

$$U = t \cdot P^T + E \tag{4}$$

$$\Upsilon = t \cdot C^T + F \tag{5}$$

In Eq. (4), $U \in \mathbb{R}^{r \times n}$ denotes the input data matrix (containing both external and internal inputs signals); $P \in \mathbb{R}^{n \times K}$ denotes the cosine of the angles (or loadings) between the K number of principal components and original input space; and $E \in \mathbb{R}^{r \times n}$ denotes the error residuals between the projected and original data representations. In Eq. (5) $\Upsilon \in \mathbb{R}^{r \times 1}$ denotes the output data; $C \in \mathbb{R}^{1 \times K}$ denotes the regression coefficients used to relate the projected input data (t) to the output data (Υ); and $F \in \mathbb{R}^{r \times 1}$ denotes the error residuals between the measured calculated outputs. PLSR analysis was performed in Matlab. It should be noted that the data was log-transformed, mean-centered, and scaled by unit variance for ease of calculation.

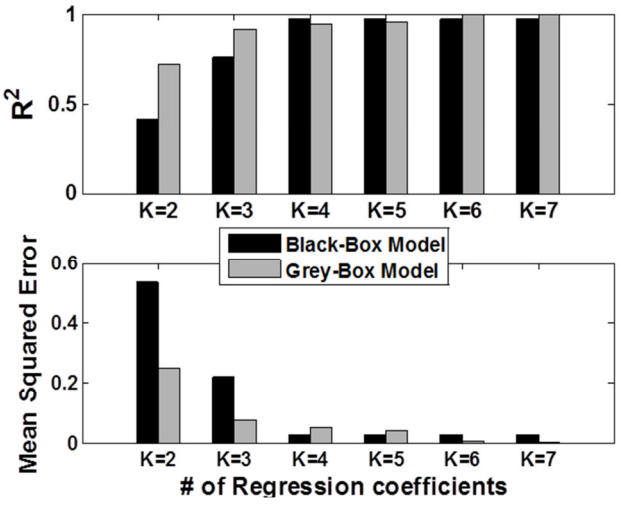

FIGURE 5: R^2 AND MSE FOR VARIOUS PARAMETER #'S IN BLACK BOX AND GREY-BOX MODELS

RESULTS AND DISCUSSION

Purely Empirical (Black-Box) Model Comparison

We would like to verify if internally generated input signals improved model representation over using a purely empirical (or black-box) approach. For comparison we constructed an empirical model consisting of a linear polynomial regression between input data (MHC/TCR affinity) and output data (experimentally measured IL-2):

$$\tilde{y} = \sum_{i=0}^{K-1} \tilde{c}_i \cdot \tilde{u}^i + \tilde{f} \tag{6}$$

Here $\tilde{u}$ denotes the input data (containing only external inputs: MHC binding affinity), $\tilde{c}$ denotes the regression coefficients used to relate the input data to the output data($\tilde{y}$), and $\tilde{f}$ denotes the error residuals between the measured and calculated outputs. Equation (6) may be thought of as a simplified form of the Volterra series expansion which contains products of increasing order of the input signal with itself.

Correlation between Measured and Calculated outputs

Fig. 5 compares the coefficients of determination (r^2 values) and mean squared errors for both the purely empirical model (black) and the proposed framework (grey). MSE and r^2 are generally accepted as a good theoretical basis for model selection. As can be seen, r^2 values are significantly higher for the grey-box model at low parameter numbers. This result signifies that calculated outputs from the grey-box framework have a higher correlation and therefore are a better representation for the presented data. Furthermore, this strengthens our claim that addition of internally generated

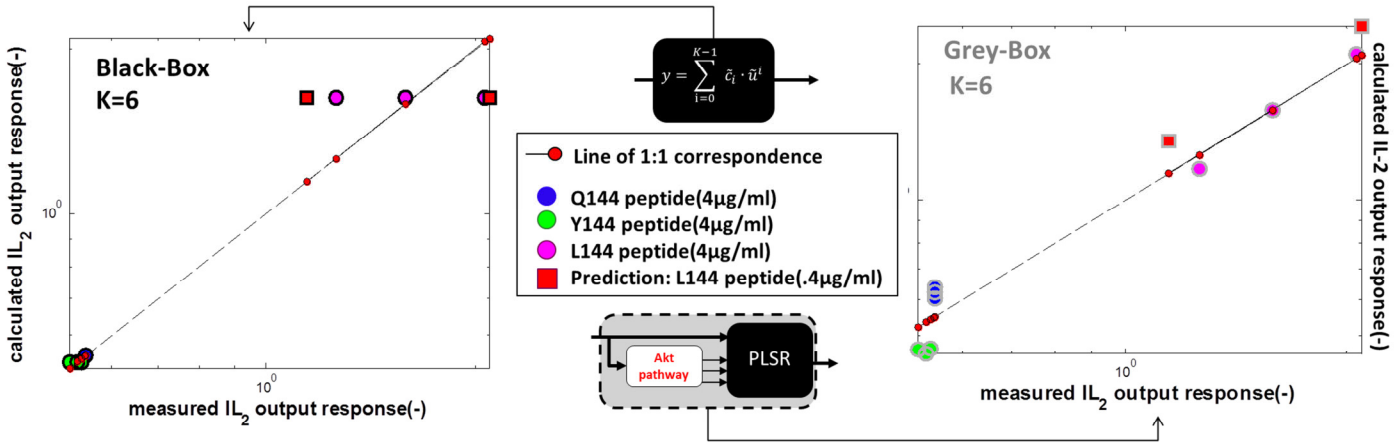

FIGURE 6: CORRELATION BETWEEN MEASURED AND CALCULATED VALUES FOR BLACK-BOX AND GREY-BOX MODELS

signals in our grey-box framework adds information to the downstream empirical sub-model. Figure 6 plots the calculated vs. measured IL-2 production (colored circles) for the black-box and grey-box models using K=6 regression coefficients.

Model Validation

As the number of parameters increase, the r^2 values between black-box and grey-box models both approach 1. This would suggest that both models are good approximations of process being studied. However, upon validation using a new stimulation condition (not used in parameter estimation), the model prediction using the grey-box model had 17% prediction error, while the model prediction error of the black-box model was 32%. This suggests that grey-box framework has better predictive capabilities. This may be attributed to the model's ability to capture variance in the data more accurately since internal inputs are generated stochastically. In Fig. 6 the red boxes denote predicted data points.

CONCLUSION

A hybrid modeling framework integrating a highly specific mechanistic model with highly abstract empirical model is presented. Results suggest that mechanistically derived internal signals inform downstream behaviors since our grey-box framework better approximated the data. Future work will incorporate multiple downstream pathways involved in the network to examine effect on downstream cellular function and response. Furthermore, our current framework statically maps external and internal inputs to downstream behaviors using PLSR. This is with the assumption that when internal signaling dynamics reach steady state, if there are no added inputs, the cell will remain in that state for a period of time in which the change in phenotype or functional response will occur. Incorporation of dynamic (auto-regressive) models is a topic for future work. In addition, we would like to explore feedback and bi-directionality between sub-models.

ACKNOWLEDGMENTS

Special thanks to Dr. Melisa L. Kemp (Georgia institute of technology) for access to published TCR signaling data and valuable comments. This material is based on work supported in part by the Emergent Behaviors of Integrated Cellular Systems (EBICS) Science and Technology Center under grant number NSF STC-0902396 and the Singapore-MIT Alliance for Research and Technology, Bio-Systems and Micromechanics IRG.

REFERENCES

[1] J. Saez-Rodriguez, L. Simeoni, J.A. Lindquist, R. Hemenway, U. Bommhardt, B. Arndt, U. Haus, R. Weismantel, E.D. Gilles and S. Klamt, "A logical model provides insights into T cell receptor signaling," PLoS Computational Biology, vol. 3, pp. e163, 2007.

[2] D.T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," Journal of Computational Physics, vol. 22, pp. 403-434, 2005.

[3] B. Schoeberl, C. Eichler-Jonsson, E.D. Gilles and G. Muller, "Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors," Nat.Biotechnol., vol. 20, pp. 370-375, 2002.

[4] T. Rimchala, R.D. Kamm and D.A. Lauffenburger, "Endothelial cell phenotypic behaviors cluster into dynamic state transition programs modulated by angiogenic and angiostatic cytokines," Integrative Biology, 2013.

[5] M.L. Kemp, L. Wille, C.L. Lewis, L.B. Nicholson and D.A. Lauffenburger, "Quantitative network signal combinations downstream of TCR activation can predict IL-2 production response," The Journal of Immunology, vol. 178, pp. 4984-4992, 2007.

[6] H. Salis and Y. Kaznessis, "Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions," J.Chem.Phys., vol. 122, pp. 054103, 2005.

[7] M. Munder, E. Bettelli, L. Monney, J.M. Slavik, L.B. Nicholson and V.K. Kuchroo, "Reduced self-reactivity of an autoreactive T cell after activation with cross-reactive non–self-ligand," J.Exp.Med., vol. 196, pp. 1151-1162, 2002.

[8] R.N. Germain and I. Stefanová, "THE DYNAMICS OF T CELL RECEPTOR SIGNALING: Complex Orchestration and the Key Roles of Tempo and Cooperation 1," Annu.Rev.Immunol., vol. 17, pp. 467-522, 1999.