STOCHASTIC SIMULATION AND SYSTEMS ANALYSIS OF INSULIN-STIMULATED GLUT4 TRANSLOCATION

Eric C. Kwei, Kevin R. Sanft, Jason E. Shoemaker, Linda R. Petzold, and Francis J. Doyle III
University of California, Santa Barbara, CA 93106

Introduction

The estimated number of people in the US with diagnosed diabetes has more than doubled in the last 15 years to 14.6 million, with associated annual medical costs of $174 billion; type 2 diabetes mellitus (T2DM) represents 90 to 95% of these cases [1].

T2DM is primarily linked to obesity and insulin resistance—a decreased sensitivity of glucose response to normal insulin levels—which have been linked to defects in the insulin signaling pathway. Analysis of detailed mathematical models of insulin signaling should yield a better understanding of the underlying mechanisms of insulin resistance and its subsequent progression to T2DM. This may be very relevant for maximizing treatment efficacy and minimizing side effects, which can ultimately improve the quality of life for those who suffer from T2DM.

We have brought a variety of engineering tools, including stochastic modeling, parameter sensitivity analysis, and robust performance analysis, to bear on an existing insulin signaling model, published by Sedaghat et al. [2]. This differential equation (DE) model, based primarily on mass action kinetics, largely reproduces the mechanisms in adipocytes (fat cells) that transduce an insulin input signal to movement of GLUT4, a glucose transporter responsible for glucose uptake, to the cell surface.

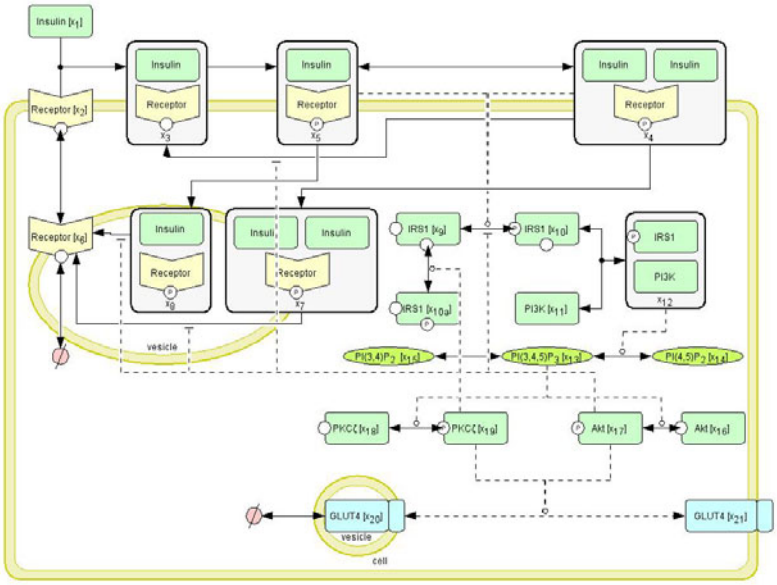

Figure 1. Sedaghat insulin-stimulated GLUT4 translocation model with feedback mechanisms

The Sedaghat model of insulin-stimulated GLUT4 translocation can be decomposed into 3 sub-models. The first sub-model describes insulin receptor dynamics—insulin binds to insulin receptor, causing subsequent receptor autophosphorylation; receptor recycling dynamics are also included. The second sub-model describes the phosphorylation cascade downstream from the insulin receptor: phosphorylated insulin receptor has tyrosine kinase activity, leading to the activation of a signaling cascadeconsisting of insulin receptor substrate 1 (IRS1), phosphatidylinositol 3-kinase (PI3K), phosphatidylinositol triphosphate (PIP_3), followed by protein kinases B (Akt) and C (PKC). The final sub-model describes the movement and fusion of specialized GLUT4 storage vesicles with the plasma membrane by Akt and PKC —the extent of GLUT4 translocation is quantified by the percentage of the total amount of GLUT4 that exists in the plasma membrane (surface GLUT4). Two versions of the model were proposed by Sedaghat et al., one incorporating both positive and negative feedback mechanisms with 21 states and one without these mechanisms with 20 states (Figure 1).

Upon scaling the deterministic model for a typical human adipocyte cell volume of 0.93 nL [3], we found that molecule counts revealed that several model species were present in small numbers (O(1) in some instances). When such small populations are involved, fluctuations may have significant effects on system behavior. To investigate this possibility, we have developed and simulated a stochastic version of the Sedaghat model.

While the Sedaghat model is the most complete available in the literature, it does not incorporate a number of known signaling components and has, to date, not been validated. Thus, we wanted to design an optimal experiment to better identify the model. To do this, we estimated the kinds of fluctuations one might observe in a reasonably-scaled experiment using

our stochastic simulation of the Sedaghat model. Using these fluctuation estimations combined with parameter sensitivity analysis, we optimized experimental design for model identification.

We have also approached this model from a different but complementary perspective. It may not be terribly important to identify each model parameter precisely, as the signaling network structure may be robust to variations in that parameter. Conversely, the signaling network may be particularly fragile to perturbations in certain parameters, which may pinpoint potential causes for insulin resistance or even suggest potentially potent targets for drug therapy. Therefore, we performed a robustness analysis on the Sedaghat model to identify specific parameters that are crucial for robust performance of the GLUT4 response to insulin.

Stochastic Simulation of Sedaghat Model

The accuracy of a deterministic model fitted to data depends on measurement precision, which is limited by, among other factors, stochastic fluctuation magnitude. Because modern experimental methods can quantify the dynamic response of single target cells to insulin, we became interested in establishing whether a deterministic model can completely describe a system at this scale.

A discrete stochastic model can be derived from an ordinary differential equation (ODE) model by first representing the ODE's as individual chemical reactions. The rates of change in the state variables are then converted to *propensity functions* that describe the probability that a corresponding reaction will occur. When the ODE model is represented using only mass action kinetics and the chemical species are assumed to be well- mixed, the conversion to a stochastic model is straightforward and the resulting stochastic model has a time-dependent probability of being in any possible state that is described by the *chemical master equation* (CME). Due to its high dimensionality, the CME can rarely be solved exactly. However, using Gillespie's stochastic simulation algorithm (SSA) to run many realizations of the stochastic model, it is possible to get reasonable estimates of the solution of the CME [4].

A stochastic version of the Sedaghat model with feedback mechanisms was implemented using the StochKit software package. Custom code was added to modify the propensity functions when the deterministic model does not follow simple mass-action kinetics. Additional custom code was written to allow arbitrary insulin input functions, such as the pulse input used by Sedaghat et al. [5].

To estimate the effect of stochastic fluctuations, we used the SSA to generate 100 realizations of the stochastic model. A 15 minute pulse input of insulin was used, and state information was collected at every minute of simulation time up to 60 minutes.

Stochastic Simulation Results

For this model, appropriate metrics of the adipocyte response to insulin include the peak values of GLUT4 (glucose transporter) at the plasma membrane and the distribution of these values. There is qualitative agreement in the trajectory of the deterministic model and the mean trajectory of the stochastic models; however, the peak surface GLUT4 percentage is roughly 34% in the ODE model compared to 30% (1.51×10^6 molecules/cell) for the mean of the

stochastic simulations (Figure 2). We attribute this difference to a non-mechanistic upper bound on the effect of Akt and PKC on GLUT4 translocation in the deterministic model (see [5] for a more complete explanation).

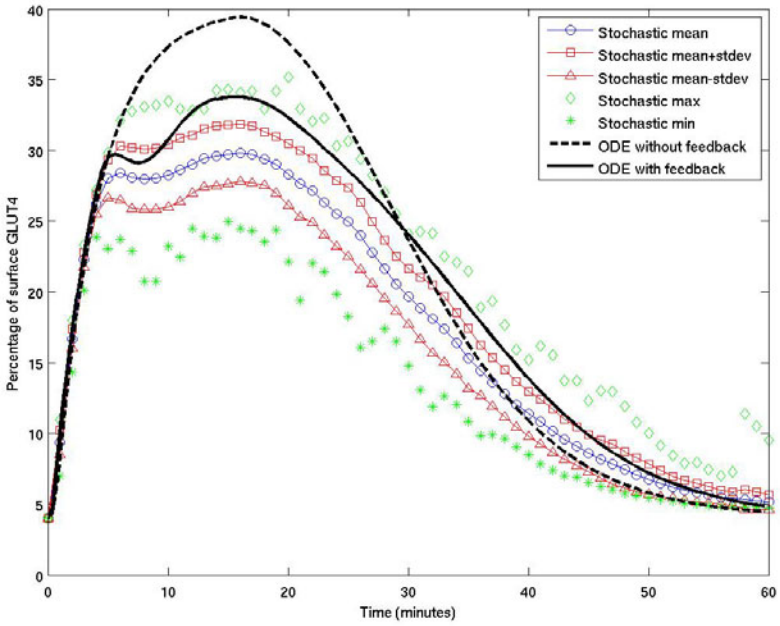

Figure 2. Means, standard deviations, and extreme values of percentage surface GLUT4 for 100 realizations of stochastic model versus Sedaghat model

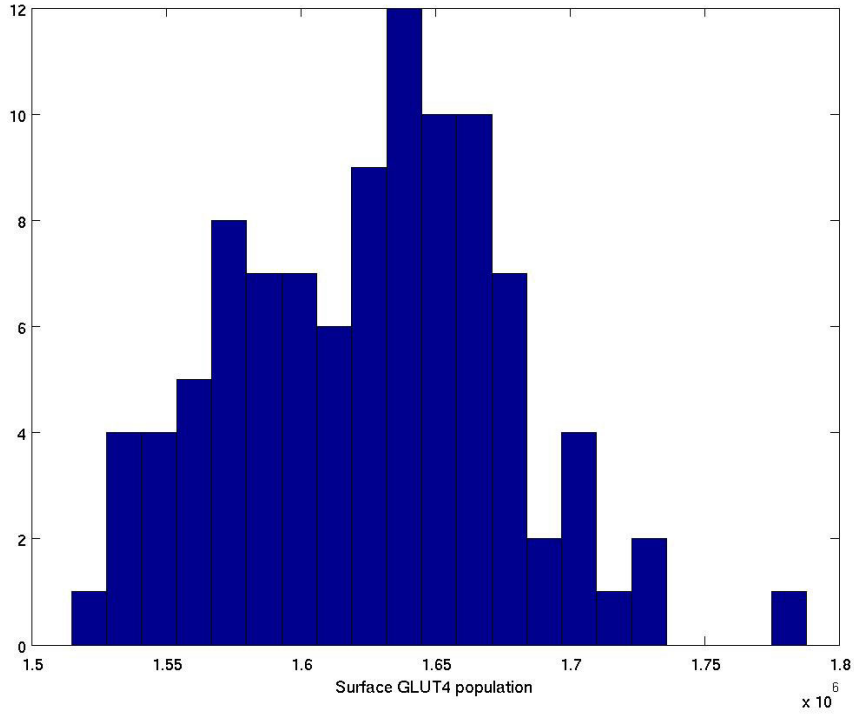

Figure 3. Histogram of the peak surface GLUT4 values for 100 realizations of the stochastic model.

The peak values of the individual stochastic realizations are summarized in Figure 3. It is of note that the peak of the mean values of surface GLUT4 percentage, 30%, does not equal the mean of the peak values, 32% (1.63 x10^6 molecules/cell); this is due to fluctuations in time to reach peak value (mean: 15.48 minutes).

The stochastic fluctuations around peak surface GLUT4 are approximately Gaussian; the standard deviation is 2.1% in surface GLUT4 percentage—or 7% of the mean (Figure 2). It is difficult to assess how physiologically relevant these fluctuations are. In addition, non-mass action kinetics in the original deterministic model may distort the computed stochastic fluctuations unrealistically.

Experimental Design Using Sensitivity Analysis

Because experimental measurements can be sparse in biological systems, it is vital to develop an experimental design that can maximize the amount of information about model parameter values from limited and noisy measurements. We believe that there is utility in optimizing experimental protocols in this particular case because the Sedaghat model lacks experimental validation. By taking into account the model structure and estimating measurement noise (using stochastic fluctuation as a surrogate), we have used parameter sensitivity analysis to create experimental designs that are efficient for model identification, given constraints on measurement error and cost.

For an ODE model, a differential change in a parameter p_j from its nominal value may cause a change in state x_i, which is captured by a sensitivity coefficient, S_{ij} (where i ranges from 1 to the number of states N_x that will be measured and j ranges from 1 to the number of parameters N_p in the model that are of interest). For each measurement time point t_k (where k ranges from 1 to the number of timepoints N_t), a sensitivity coefficient matrix for each time point is generated (Equation 1).

$$S_{ij}(t_k) = \frac{\partial x_i(t_k)}{\partial p_j} \tag{1}$$

The sensitivity coefficient matrices are combined with estimations of measurement error (which are assumed to be Gaussian) for each state and timepoint to calculate the Fisher information matrix (FIM), where $_i(t_k)$ is the standard deviation for state x_i (Equation 2).

$$FIM = \sum_{k=1}^{N_t} S^T(t_k) V^{-1}(t_k) S(t_k)$$
$$V(t_k) = \begin{pmatrix} \sigma_1^2(t_k) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{N_x}^2(t_k) \end{pmatrix} \tag{2}$$

In the computations described below, we have selected $_i(t_k)$ to be the corresponding standard deviation from the stochastic simulations described above. The assumption that the stochastic trajectories have a normal distribution is not generally valid, particularly for states with O(1) copy count per cell volume.

A minimum measurement error of 1 molecule per cell volume is assumed, which allows $V(t_k)$ to be inverted in Equation 2. This minimum value is typically larger than the observed deviation from a normal distribution for states with O(1) copy count per cell volume; thus we can account somewhat for non-Gaussian behavior.

Finally, with the above assumptions, the FIM can be used to generate a lower bound the error that an unbiased estimation of parameter p_j, $_j$, would have if generated using measurements of a given choice of measured states, $\{x_i\}$, and measured timepoints, $\{t_k\}$ (Equation 3).

$$\sigma_j^2 \geq (FIM^{-1})_{jj} \tag{3}$$

For a more complete treatment, see [6].

To select between each choice of state and timepoint sets, we set a maximum allowable estimation error, the 95% confidence interval around the nominal parameter value, $1.96p_j$, for each “identified” parameter. For state and timepoint sets with identical numbers of identified parameters, the one with the smallest average normalized value of the 95% confidence interval for identified parameters (i.e., the most accurate estimation) is chosen.

Table 1. State measurement selection

# of states	Optimal states	# of identified p_j	Average $_j / p_j$
21	All	10	0.156
10	$x_2,x_5,x_6,x_8,x_9,x_{10},x_{13},x_{16},x_{20},x_{21}$	10	0.179
5	$x_2,x_5,x_9,x_{10},x_{20}$	8	0.236
3	x_2,x_9,x_{20}	4	0.288

Table 2. Timepoint measurement selection

# of timepoints	Optimal timepoints (min)	# of identified p_j	Average $_j / p_j$
60	All	10	0.156
15	1-5,10,19-24,46,56,57,60	9	0.179
10	1-5,10,24,46,56,57	8	0.150
5	1,3,4,46,56	8	0.200

Experimental Design Results

The sensitivity coefficient matrices were calculated for 31 parameters (after combining those that are *a priori* unidentifiable) in the Sedaghat model with feedback mechanisms over a 60 minute "experimental" duration using BioSens, a sensitivity analysis program [7]. As for the stochastic simulation, a 15-minute insulin pulse was used. Because the stochastic error estimations are in 1 minute intervals, the maximum number of timepoints that can be collected is 60. For any desired number of states and/or timepoints, a MATLAB routine was written that, first, maximizes the number of identified parameters and, second, minimizes the error in parameter estimation of identified parameters.

Selecting measured states that allow for maximum parameter estimation was the first experimental design undertaken, allowing all 60 timepoints to be measured. These results are summarized in Table 1 (Refer to Figure 1 for state numbers and signaling components). Overall, we observe that the model is not highly identifiable. Only 10 of 31 parameters can be identified with all state and timepoint measurements. More state measurements increases identified parameter count and accuracy. One would expect that directly measuring the states at the end of the signaling cascade would yield the highest amount of information, since later states would carry information about parameters that appear upstream. However, because different states have different measurement errors associated with them, this is not the case.

Selecting measured timepoints that allow for maximum parameter estimation was the next experimental design; these results are summarized in Table 2. Once again, taking more timepoints results in increased parameter identification and accuracy. The timepoints that contain the most parameter information seem to be clustered near the onset of insulin stimulation (1-5 minutes), with the remaining timepoints coming as the system relaxes back to a resting condition after the insulin stimulus is removed.

Table 3. Mixed state and timepoint selection

# of states	# of timepoints	# of identified p_j	Average $_j / p_j$
20	15	9	0.179
10	30	10	0.201
5	60	8	0.236

With the effects of state and timepoint measurement selection characterized separately, we turned our investigation to optimizing both state and timepoint measurement together. Thus, for a fixed total number of measurement datapoints (300), optimal state and timepoint measurements were determined and their effects on parameter identification are summarized in Table 3. For this example, an experimental design that tracks an intermediate number of states (10) and timepoints (30) would best identify model parameters. This method is easily adapted for modern high throughput methods of collecting biological data, including microarrays or phosphoproteomic methods, to select optimal measurement sets.

Insights from Robustness Analysis

Ultimately, it is hoped that the deterministic and stochastic models developed with the aid of the above techniques are useful in guiding modeling (in)validation and for (possibly combinatorial) drug targeting in insulin resistant cells. With this in mind, structured singular value (SSV or μ) analysis is performed to identify possible points of robust therapeutic intervention which minimize hazardous side effects. Fluctuations are not strictly limited to signaling pathways, as the intracellular environment is also subject to external noise; any therapeutic strategy applied must be robust to the uncertainty of the local environment as well as to the uncertainty in the therapy itself.

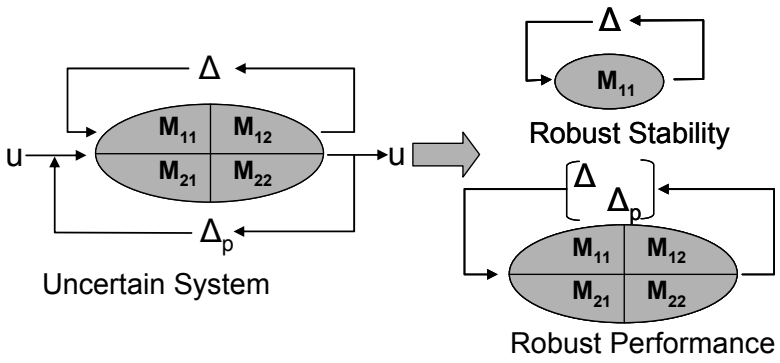

Figure 4. To test for robust stability of an uncertain system (left), one need only test the stability of the feedback of the uncertainty (top right). To extend the test to robust performance, the input/output channels are closed under feedback through a full uncertainty block, || p||•1. Reconfiguring the system (bottom left) and noticing the similarity, robust performance can be evaluated by evaluating the robust stability of the resulting block diagram.

We use SSV analysis is used to quantify the range of fluctuation a set of parameters can tolerate while maintaining robust insulin signaling performance. Briefly, SSV analysis identifies the smallest size perturbation, , which destabilizes the stable plant, **M**. The "size" of is measured in terms of its maximum singular value, $\bar{\sigma}(\ )$ (Figure 4). It is an extension of the Nyquist stability criterion, and μ is defined as

$$\mu(\mathbf{M}_{1,1}) = \min_{\Delta}\{\bar{\sigma}(\Delta) \mid \det(\mathrm{I} - \mathrm{M}_{1,1}\Delta) = 0 \text{ for structured } \Delta\}. \quad (4)$$

A full mathematical treatment of SSV analysis can found in [8].

The matrix is block-diagonal and is chosen to distribute parametric uncertainty to specific locations about the Jacobian, hence the uncertainty is structured. Furthermore, is generally weighted such that size of the system of uncertainty of interest is bounded, || ||•1; thus, when μ•1, the size of the perturbation to destabilize the system is greater than 1—outside of the uncertainty range of interest—and the system is robustly stable.

Extending robust stability to robust performance (RP) is achieved by applying a full uncertainty block, $_p$, to the input/output channels and closing the channels under negative feedback, as robust stability of this closed-loop structure ensures robust performance. The calculation of μ is an NP-hard problem; exact solutions can be found for small systems, but generally an upper and lower bound must be found for μ.

To evaluate combinatorial therapies in the insulin signaling pathway, the Sedaghat model without feedback mechanisms is linearized (the model with feedback mechanisms is not amenable to linearization). Robust performance is then defined as a tracking error about the nominal surface GLUT4 response. Several parameters from the nonlinear model do not appear in the linearized model, and many of the remaining parameters appear in related pairs which may be lumped into single parameters for the purposes of RP analysis, leaving 17 model parameters to be considered as possible therapeutic targets. For this study, 3 parameters (drug targets) are considered at a time, resulting in 680 possible parameter combinations.

First we sought to identify which sets of parameters, when perturbed, most effectively removed an individual cell from healthy performance. For each set considered, μ is calculated when each parameter is allowed to fluctuate within 5.0% of its nominal value (Figure 5). .

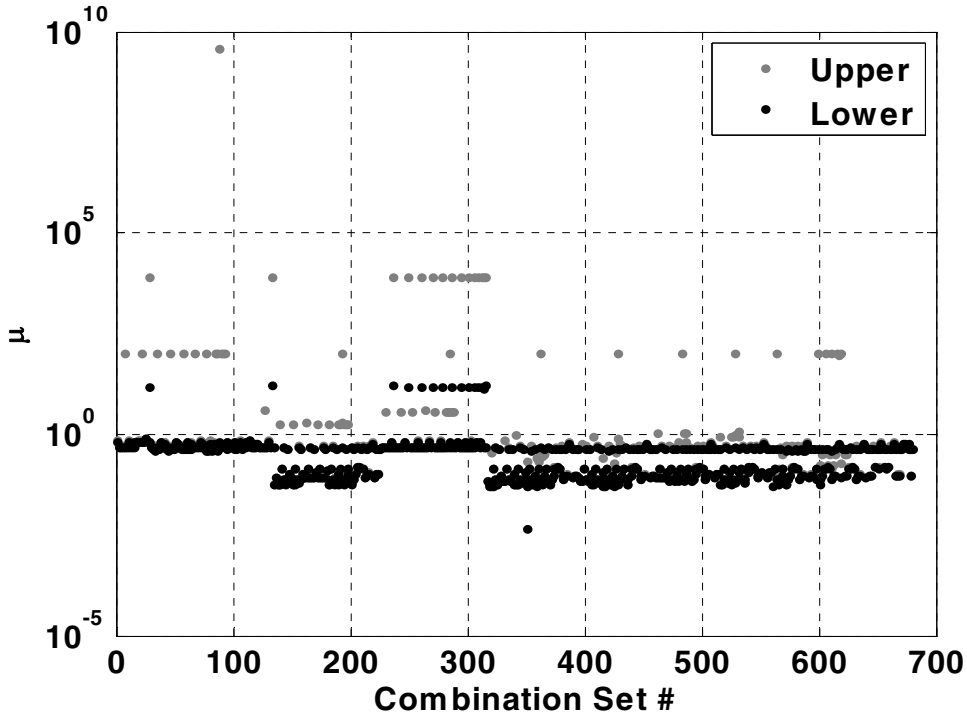

Figure 5. The upper and lower bound on μ for each of the possible 680 parameter combinations

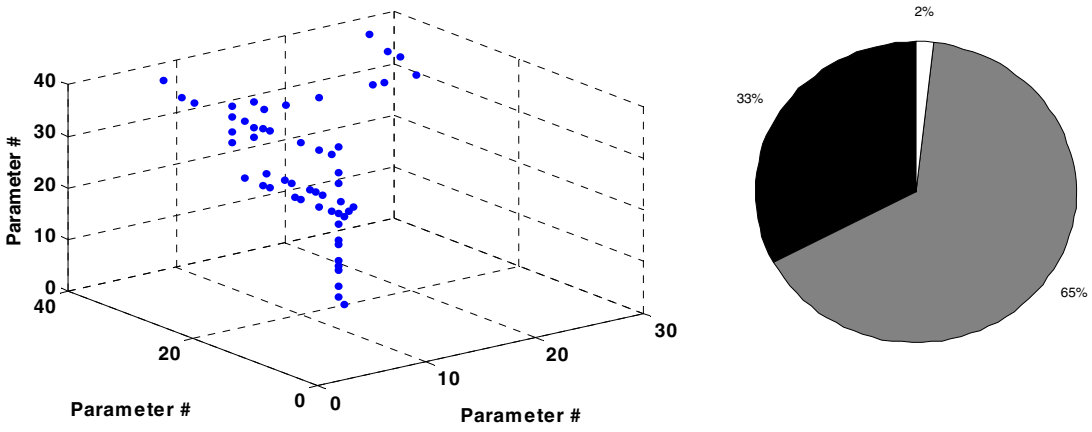

Figure 6. (Left) The set of all efficacious therapeutic combinations. (Right) The percentage of candidate therapeutic combinations that contain at least 1 (black), a combination of 2 (gray), or all three parameters PI3K, k_{-1}, and k_{9stim} (white).

Removing parameter sets for which the value of μ is poorly bounded (error difference> 0.1), 74.9% of the parameter sets remain to be analyzed for efficacy. Efficacy has been defined here as any remaining set with a μ value greater than 0.6. Applying this filter, only 8.1% of the original 680 possible combinations remain. Figure 6 shows how the remaining

parameter combinations cluster about key parameters. Three parameters from the Sedaghat model appear in every candidate set, PI3K, k_{-1}, and k_{9stim}, and at least 2 of these three parameters appear in 65% of the possible therapeutic vectors. The parameter PI3K accounts for the amount of PI3K in the cell; k_{-1} is the dissociation rate of insulin from its receptors; and k_{9stim} represents the conversion of PIP_2 to PIP_3 by phosphorylated IRS1/activated PI3K complex. PI3K is involved in several cellular functions, has some known tumorigenic effects, and, as such, does not represent a suitable target vector. Removing strategies directly targeting PI3K, future work will focus on exploring the remaining 21 possible therapeutic combinations. These therapies will be applied to various models to test for efficacy *in silico.*

Conclusion

Stochastic modeling and simulation of the Sedaghat insulin-stimulated GLUT4 translocation network reveals fluctuations in concentrations that cannot be entirely captured in a deterministic model; these inherent fluctuations represent limits on measurement precision and are important to understanding the robustness and performance of the underlying biological system. We have used this understanding, with sensitivity analysis, to create optimized selections of states and timepoints for this network.

The SSV analysis we performed illustrates an effective preliminary *in silico* procedure for the rapid identification and screening of potential drug targets in this network. Given the large amount of cross-talk between the insulin signaling network and other critical, auxiliary networks, multi-drug therapies were identified that best regulate the insulin response while minimizing side effects, identifying a number of potential drug therapy sets.

References

1. Centers for Disease Control and Prevention, *Diabetes: Disabling Disease to Double by 2050*. 2008.
2. Sedaghat, A.R., A. Sherman, and M.J. Quon, *A mathematical model of metabolic insulin signaling pathways.* American Journal of Physiology-Endocrinology and Metabolism, 2002. **283**(5): p. E1084-E1101.
3. Leonhardt, W., H. Schneide, H. Haller, and M. Hanefeld, *Human Adipocyte Volumes - Maximum Size, and Correlation to Weight Index in Maturity Onset Diabetes.* Diabetologia, 1972. **8**(4): p. 287-291.
4. Gillespie, D.T., *General Method for Numerically Simulating Stochastic Time Evolution of Coupled Chemical-Reactions.* Journal of Computational Physics, 1976. **22**(4): p. 403-434.
5. Kwei, E.C., K.R. Sanft, L.R. Petzold, and F.J. Doyle III, *Systems Analysis of the Insulin Signaling Pathway*, in *Proc. 17th IFAC World Congress*. 2008: Seoul, Korea.
6. Zak, D.E., G.E. Gonye, J.S. Schwaber, and F.J. Doyle III, *Importance of input perturbations and stochastic gene expression in the reverse engineering of genetic regulatory networks: Insights from an identifiability analysis of an in silico network.* Genome Research, 2003. **13**(11): p. 2396-2405.
7. Taylor, S.R., R. Gunawan, K. Gadkar, and F.J. Doyle III, *BioSens Sensitivity Analysis Toolkit (UCSB's contribution to Bio-SPICE) v2* . 2005 [Available from: http://www.chemengr.ucsb.edu/~ceweb/faculty/doyle/biosens/BioSens.htm].
8. Skogestad, S. and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design*. 1996: Wiley.