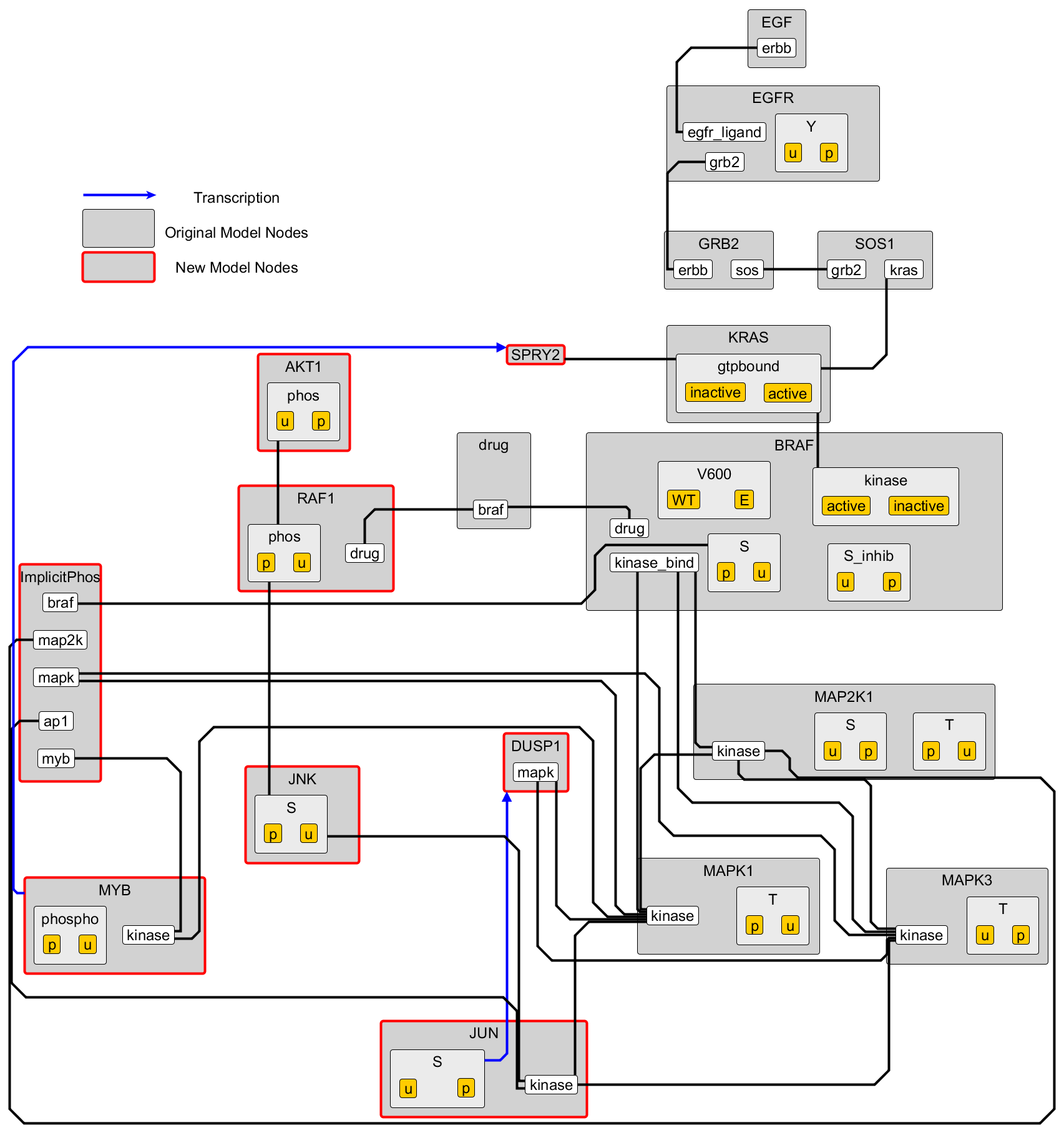
Previously we built a model aiming for the smallest possible network, with the plan to build up as needed. We started with a model built around the very basics of map kinase signaling, focusing on the proteins EGF, EGFR, SOS1, GRB2, KRAS, BRAF, MAPK1, MAPK3, and MAP2K1. Additionally in an attempt to fit the desired dynamics we added a placeholder generic phosphatase capable of dephosphorylating ERK. We fit the model to the data, allowing a parameters to vary across cell lines to give the model maximum flexibility in fitting the data, with the hopes of eventually holding these parameters constant across cell lines. Even with this additional flexibility, the model struggled to fit the data. Although certain parameter regimes showed promising qualitative behavior, the final fit was quantitatively poor, pointing to a structural weakness within the model.

To address this, we’ve since expanded the model scope, adding mechanisms found via automated reading as well as manual reading and investigation. Model expansion was focused on identifying additional negative regulators of the pathway. We identified DUSP1 as a map kinase phosphatase from our assembled corpus of mechanisms, and searched for additional regulators of DUSP1. This lead to the additions of JUN, the transcription factor for DUSP1, and JNK, responsible for activation of JUN from our corpus. Regulation of JNK was not described, so we then manually searched the literature for any additional upstream regulators, and added AKT1 and RAF1 to our model, manually writing natural language sentences to account for their relevant mechanisms. Additionally, we added another negative feedback element to the pathway, SPRY2, which targets KRAS to inhibit the pathway. Manual reading again provided us with our final regulatory mechanisms, SPRY2’s transcription by MYB, and MYB’s activation by MAPK1. Adding these additional negative regulators to the pathway seemed likely to allow for the model to fit the non-intuitive dynamics of ERK phosphorylation in Task 2.

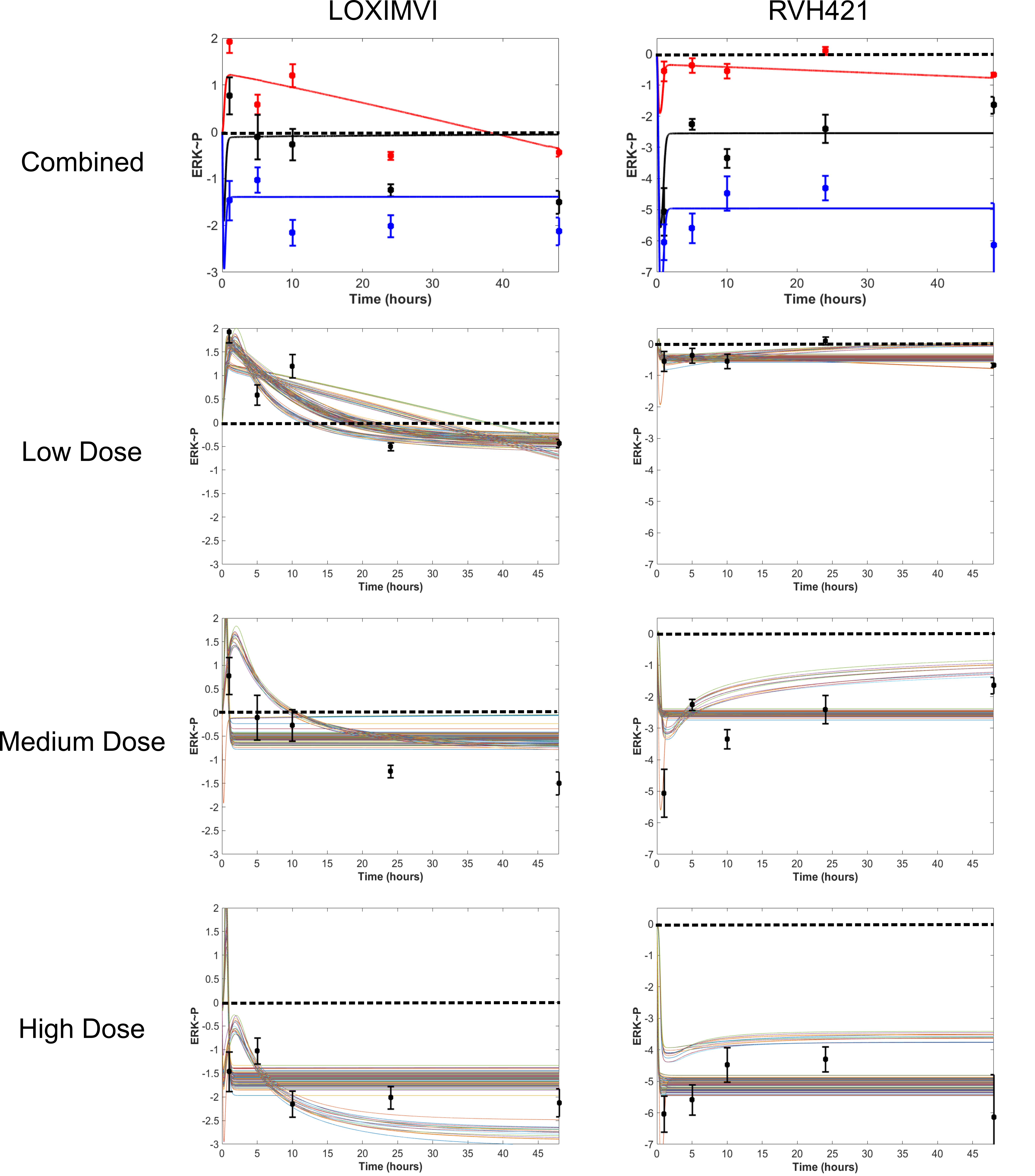
Our new model now featured seven new entities, involved in a variety of mechanisms. This type of expansion lead to an even greater increase in the combinatorial complexity of our model. Once again, we looked to simplify this complexity as much as possible with minimal sacrifice in accuracy. We again used automated methods to collapse phosphorylated residues of the same type where possible, with an additional check to separate inhibitory from activating sites to improve resulting model accuracy. We again used automated methods to add a context requirement for all reactions, enforcing an ordered cascade originating from the receptor, utilizing activating phosphorylations when present, and previous binding events otherwise. This greatly reduces the number of possible reactions by reducing circumstances in any which any binding rule can fire, and reducing the number of unique complexes formed. We additionally added dephosphorylation reactions using an generic phosphatase to correspond to every phosphorylation for which there is no corresponding phosphatase declared. This prevents a single phosphorylation leading to infinite activation for a protein, improving model accuracy. Next, we took the contact map generated for our model in the gml graph format and edited it so that it could be read by the networkx package in python. This was done manually but will be automated in the future. Networkx allows us to check our model for cycles, which could be indicative of a potential infinite oligomer in the model which will make simulation impossible. Networkx identified ['MAP2K1', 'MAPK3', 'BRAF'] as a cycle, so we had to add additional binding context to these reactions so that they could not bind in a repeating chain.

All of these changes allowed us to simulate our model, however it still ran slowly, with network generation taking 2-3 minutes. This length of simulation makes parameter estimation difficult, so we looked for additional simplifications to improve simulation time. An avenue we identified for improvement was that all enzyme-substrate combinations bind to unique domains, allowing for the formation of large complexes involving multiple enzymes that are unlikely to naturally occur. To address this, we replaced multiple unique binding domains with a single domain that would be treated identically by potential binders, causing them to compete for binding rather than bind unrealistically large complex. This is an easy change that takes advantage of the rule-based modeling framework, and is a process we hope to automate as part of model assembly in the future.

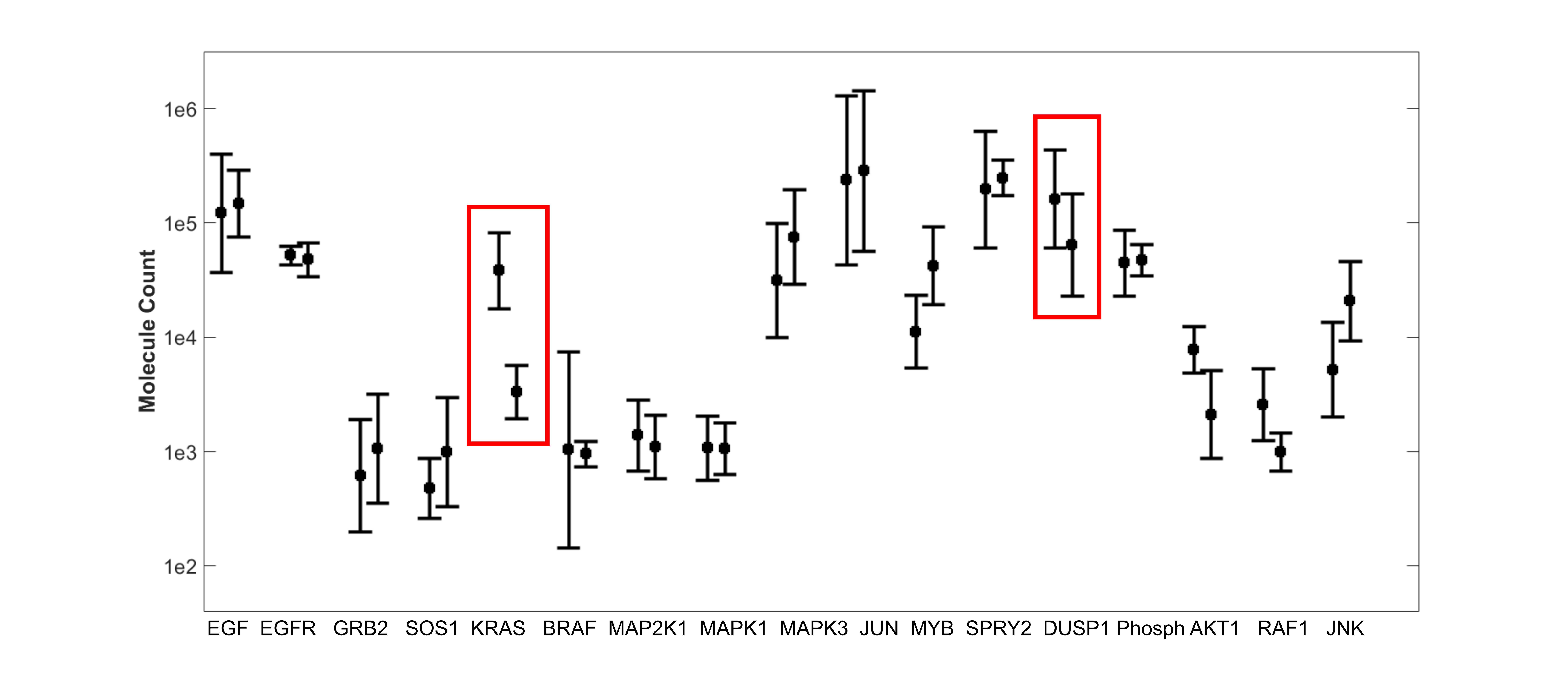
 This series of changes to model context results in a final model composed of 18 starting molecules that expand to a network of 81 unique species, and can be simulated in under 2 seconds (Figure 1). We used this model to perform a number of rounds of parameter estimation to attempt to fit the model to the experimental data provided. We focused on drug AZ628, using the low, medium, and high concentrations of the drug (0.0032 uM, 0.0316 uM, and 3.16 uM). We fit all three concentrations six conditions at once, keeping all kinetic parameters constant during each iteration, and allowing only initial conditions to vary across cell lines. This follows the assumption that networks and chemical kinetics are constant, and differential responses across cell lines are **Figure 1**. Contact map illustrating final rule-based model

controlled by differences in

expression. The model performed relatively well (Figure 2), looking at the average behavior of the estimated parameter ensemble, in the LOXIMVI cell line we were able to recover the 1 hour increase in ERK phosphorylation at low dose, and decrease in phosphorylation at later time points and higher doses. In the RVH421 cell line we recovered the dose-dependend decrease in ERK phosphoyrlation at all time points. The weakest result we obtained was in evaluating the middle dose in the LOXIMVI cell line. Here, we saw a relatively static level of ERK phosphorylation near baseline, rather than an initial increase. To further examine this result, we broke down our ensemble of parameter sets to look at individual trajectories rather than average response. Here we see that the average behavior hides two separate potential behaviors, one of which much more closely aligns with the data. Each of our measure conditions shows the potential for multiple behaviors, showing the rich potential dynamics of this model structure, but also the difficulty of fitting six different conditions at once. It is likely that with additional optimization of the input priors for our Bayesian parameter estimation and additional computation time we could settle on the best set of behaviors and improve our model fit.

**Figure 2**. Model dynamics following calibration to experimental data. Average model behavior (top) and individual trajectories (bottom) are shown.

Finally we examined the initial conditions our model predicts for the two cell lines, leading to the difference in dynamics. The model outputs a distribution for each protein in the model (Figure 3). We can compare the predicted distribution for the LOXIMVI (left) and RVH421 cell line (right) side by side. Notably, the model predicts that LOXIMVI has an increased level of both KRAS and DUSP1. This agrees with the experimental data provided and points to the two negative feedbacks included in the new model, involving SPRY2 inhibition of KRAS and DUSP1 inhibition of ERK being critical mechanisms in achieving the unexpected dynamics seen in this cell line. Notably the model predictions do differ with the provided experimental data in some cases. The model predicts higher levels of RAF1 in LOXIMVI cells and higher levels of JNK in RVH421 cells when the opposite is true in both cases. This could potentially be explained by sloppiness in parameter sets across a closely linked regulatory pathway, but requires further examination. Additional fitting rounds using data points from more observables should help reduce sloppiness and identify the correct concentrations.

**Figure 3**. Predicted distributions of initial concentrations for all model proteins in LOXIMVI (left) and RVH421 (right) cells.