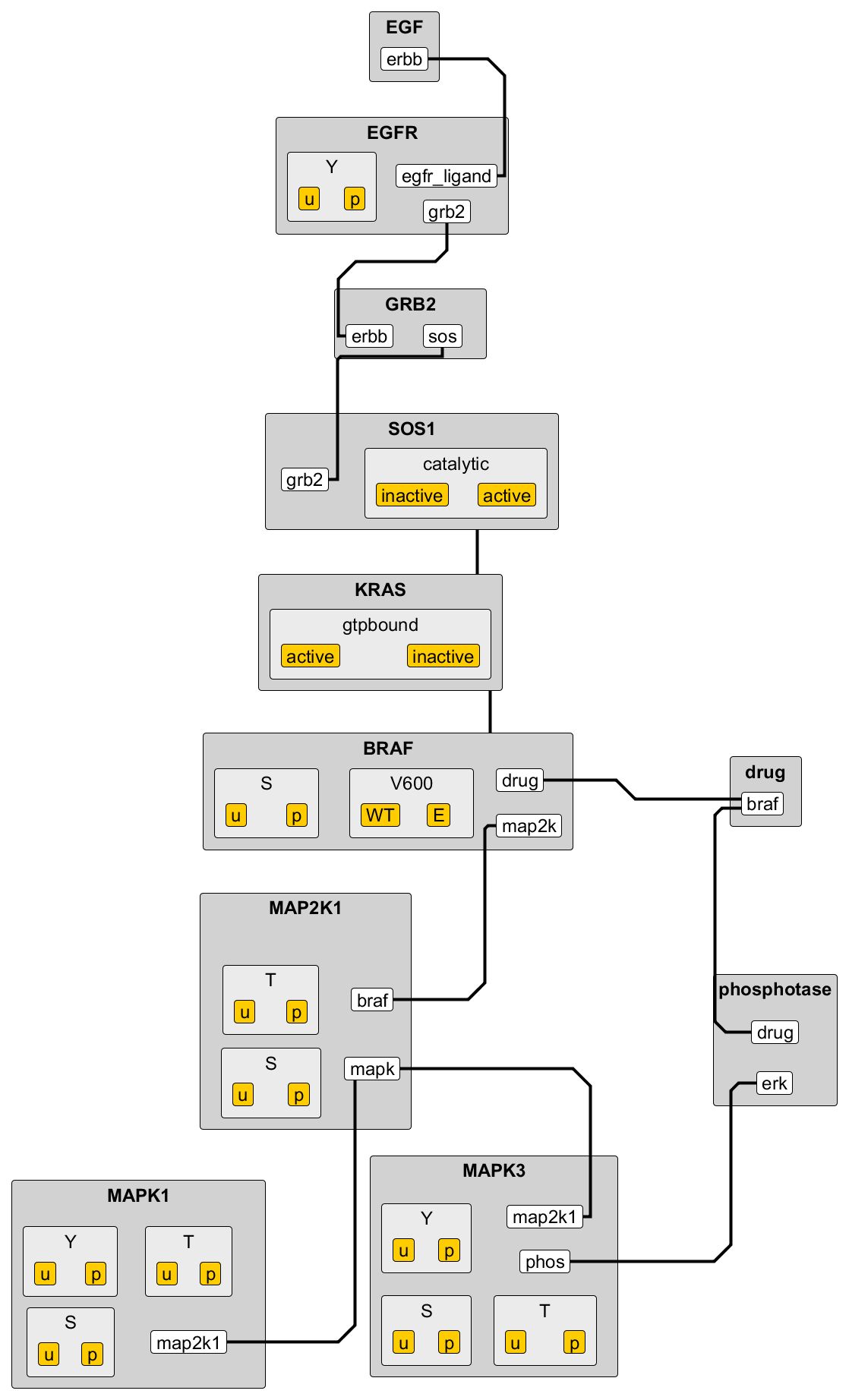
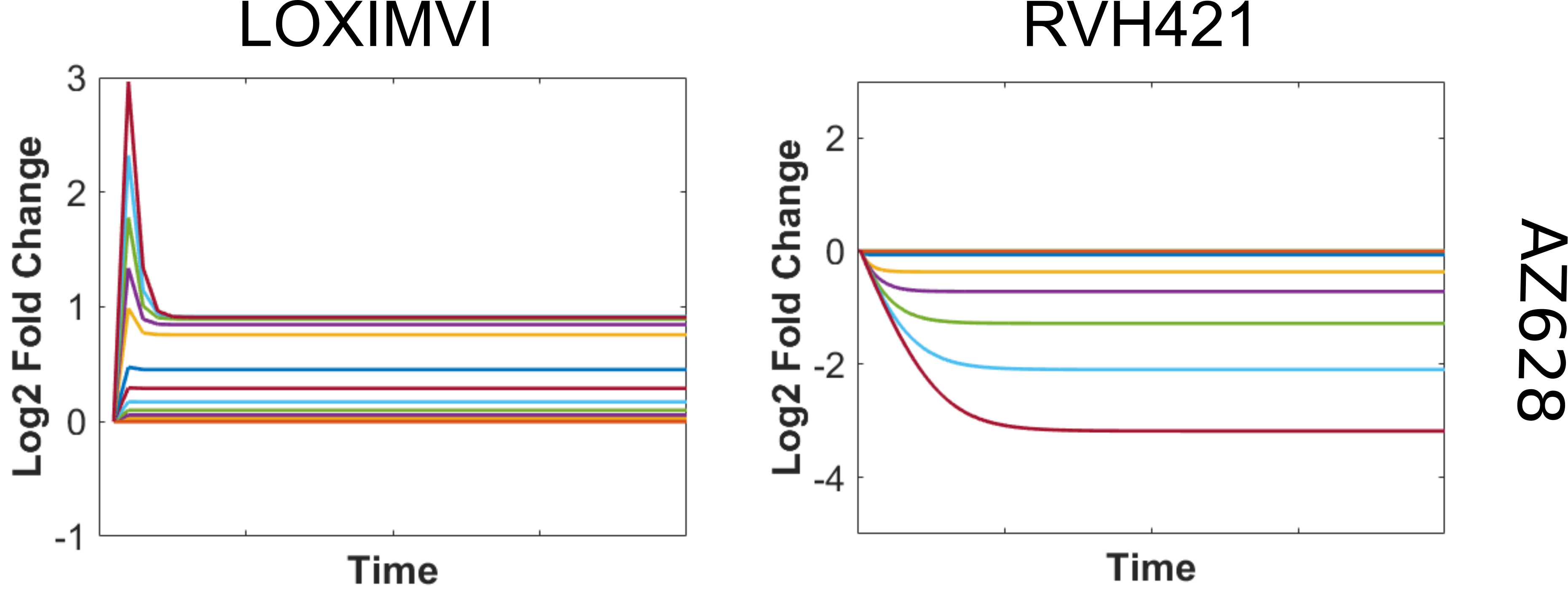
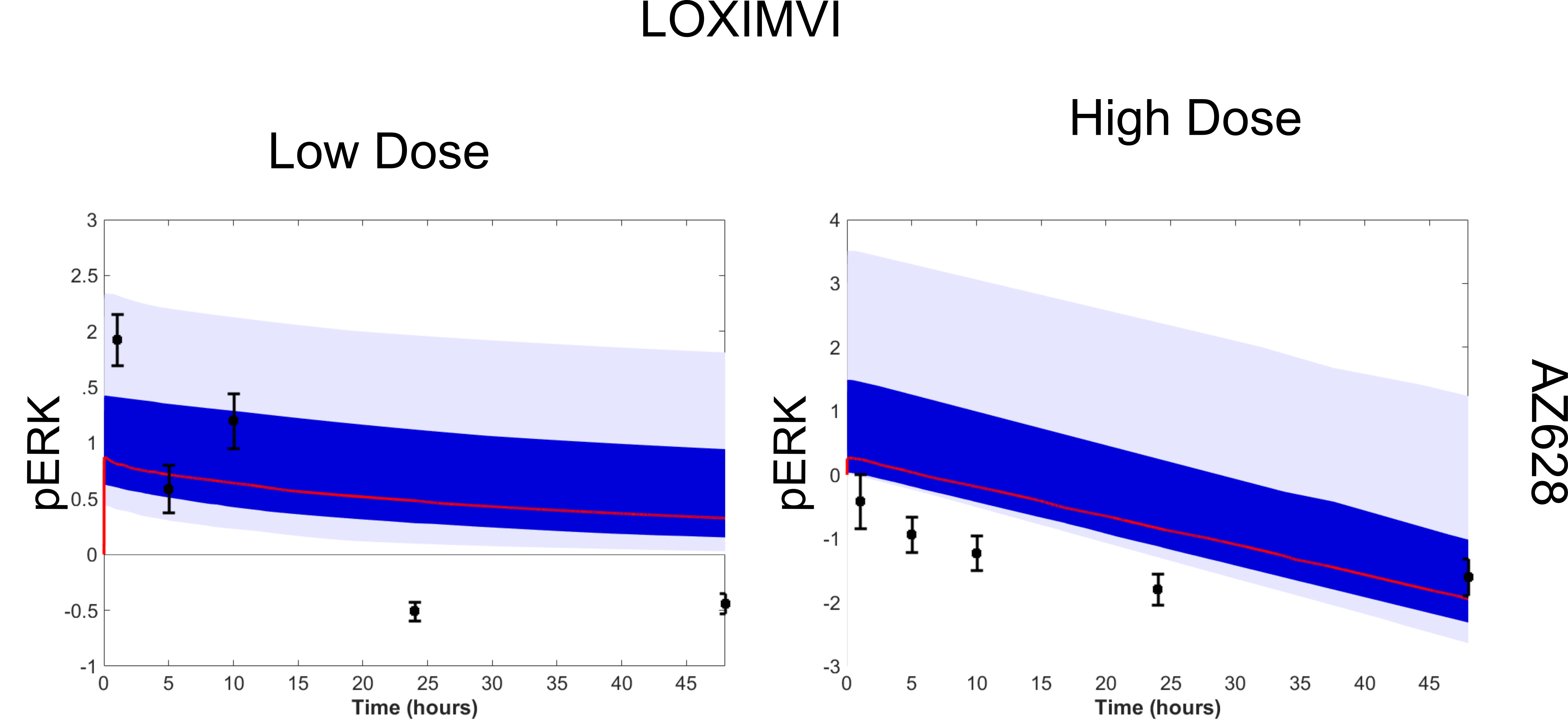
To assemble a smaller, more flexible model we took the assembled INDRA statements and performed additional filtering with the goal of making a minimal model for the desired biological use case, Task 2. We filtered statements based on a small, manually selected set of nodes to encapsulate the EGFR signaling cascade upstream of ERK phosphorylation, the desired readout. This required ignoring the mTOR/AKT signaling machinery which would add substantial complexity and computational burden with minimal benefit to the model for explaining this task. Additionally we manually added a minimal mechanism that could help explain the experimental results, a phosphotase capable of dephosphorylating ERK, which could also be targeted by drugs of interest. This maintained the goal of keeping the model as small as possible to start, allowing the possibility to build more detailed explanatory mechanisms as needed.

 To increase simulation speed we then employed automated methods to coarse grain the model and extrapolate rule context. For simplicity we collapse all phosphorylated residues of the same type to a single site. Additionally we combined any sites modified by the same kinase to a single site. Additional coarse graining included removing all dimerization reactions and mutations that were not directly relevant to model simulations. Finally we used automated methods to add context required for all binding events, enforcing a strictly ordered signaling cascade. 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. This resulted in a very minimal model composed of 25 rules which expand to 53 unique species. This can be easily simulated as a set of differential equations in less than a second. This minimal model allows substantial flexibility, as it allows us to quickly explore large swaths of parameter space to test if the model will fit the data. The model can the be easily expanded or corrected based on original, more detailed reading output. We can add back removed reactions, expand coarse grained sites, and add or remove context as necessary.

To evaluate our model, first we tested its qualitative dynamics to see if it could recreate the types of results we see in the experimental data. After manual parameter tuning the model did appear to have to potential to mimic experimental outcomes. Examining the LOXIMVI and RVH421 cell lines and the actions of drug AZ628, we were able to use a single model structure to meet the two different behaviors. All kinetic parameters were kept constant, but two models were roughly contextualized to data, allowing for difference in initial conditions based on the provided expression data. Additionally, the initial concentration of the manually added generic phosphatase was allowed to vary. Results showed that LOXIMVI cell line could show a brief increase in ERK phosphorylation, followed by a decrease, similar to the dynamical patterns in the experimental data. The same model can show similar results to the RVH421 cell line, where there was a dose dependent but mostly time independent reduction in perk.



The next step is to fit the model to data to obtain a more quantitatively accurate model. Using Bayesian parallel tempering we attempted to fit our model to the lowest and highest dose of AZ628 stimulation in the LOXIMVI cell line. Once again, model structure was identical in both simulation routines Kinetic parameters were fit separately as a first pass, but will be held constant across doses in future fitting runs. Initial conditions were set using expression data for the cell line, which will allow us to fit multiple cell lines in a controlled way. The model was again able reproduce some of the qualities of the experimental data, but the quantitative fit was poor. This shows us that our simplified model is likely incomplete. We can examine the parameter space of the fit model to identify weak spots as likely targets for improvement. In particular, binding affinities for our generic phosphstase were predicted to be substantially stronger than is realistic. Going forward we will need to replace this mechanism with a more biologically accurate and detailed mechanism, using the additional information obtained from our automated reading already done.



Additionally, the framework is now set up for us to scale up our parameter estimation efforts for the final evaluation submission. We can simultaneously fit all drug doses across multiple cell lines. We can do this while keeping model structure and kinetic parameters static across these conditions, allowing cell line context to directly explain the differences in experimental results, and giving us strong predictions on what causes these differences. It is also trivial to add additional experimental observables beyond the single read out we are using in this task. Computational load will increase with the number of conditions being fit simultaneously, but with the current minimal model scale this is easily manageable, but could become a concern as we add additional detail to our model..