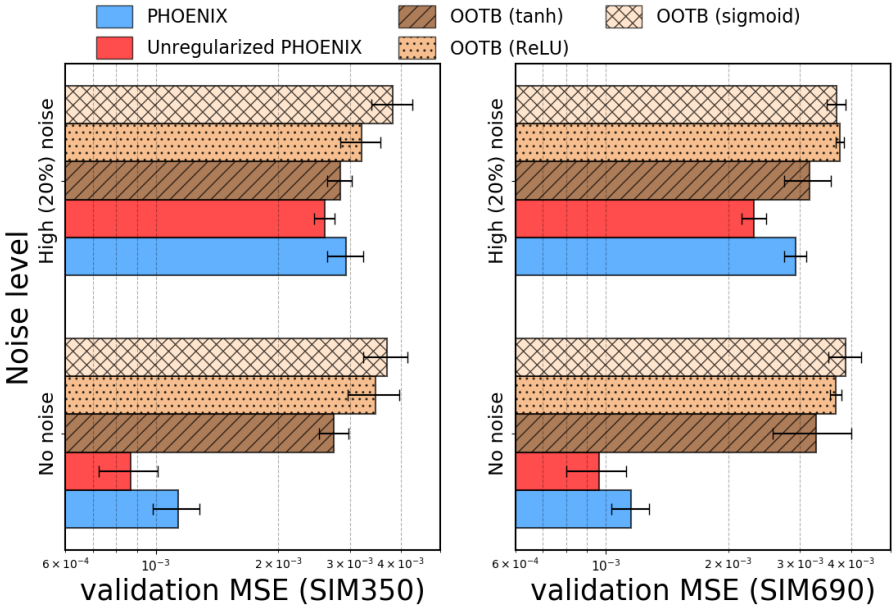


**Fig. SR1** We measured the marginal contribution of PHOENIX's architecture and incorporation of prior information by comparing against the baseline contributions of out-of-the-box NeuralODE models with three different activation functions, across both *in silico* dynamical systems SIM350 (**left**) and SIM690 (**right**). We display the performance of the out-of-the-box models, as well as PHOENIX ($\lambda_{\text{prior}}$ tuned using the validation set) and its unregularized version ($\lambda_{\text{prior}} = 0$), in terms of how well held out time points from pre-noise test set trajectories could be predicted after training on trajectories from different noise settings. Here high noise implies $\frac{\text{noise } \sigma}{\text{mean}} = 20\%$. The experiment was repeated five times to generate average mean-squared error (MSE) values and error bars.

that PHOENIX was outperformed in terms of temporal prediction by its unregularized ($\lambda_{\text{prior}} = 0$) version (**Figure SR1** and **Table SR2**). However, given that the prior can be interpreted as soft biological constraints on the estimated dynamical system [17], an important question is whether unregularized PHOENIX (as well as OOTB models) respect the underlying biology. In other words, we wanted to understand whether unregularized PHOENIX made accurate temporal predictions by correctly learning elements of the causal biology governing the dynamics, or whether the lack of prior information resulted in an alternate learned representation of the dynamics, which - despite predicting *these particular* held out trajectories accurately - was not reflective of the true biological regulatory process. Therefore, this "explainability" (or lack thereof) is what we investigated next.