Multiscale Probabilistic Modeling: A Bayesian Approach to Augment Mechanistic Models of Cell Signaling with Machine-Learning Predictions of Binding Affinity

Holly A. Huber[a] and Stacey D. Finley[a ,b,c]

[a]Department of Biomedical Engineering, University of Southern California, Los Angeles, CA 90089, USA

[b]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089, USA.

[c]Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, CA 90089, USA.

Corresponding Author: Stacey D. Finley, email - sfinley@usc.edu

Abstract

Computational models in systems biology are often underdetermined—that is, there is little data relative to the complexity and size of the model. The lack of data is primarily due to limits in our ability to observe specific biological systems and restricts the utility of computational models. However, there are a growing number of experimental databases in biology. While these databases provide more observations, they often do not have observations that match the system of interest exactly. For example, database measurements might be collected at different experimental conditions or on a different scale compared to the system of interest. Here, we investigate what information can be gleaned from generalizing databases across these differences in the context of modeling a specific system – cell signaling. Ultimately, our goal is to better determine models of specific systems, thereby increasing their utility. To do this, we propose a novel, multiscale, probabilistic framework. We use this framework to integrate measurements of protein structure from the Protein Data Bank and measurements of amino acid sequence from the Universal Protein Resource into the parameter inference of cell signaling models. Then, we quantify exactly what information is gained from these measurements when modeling cell signaling. We choose to investigate the utility of these databases in the context of dynamic cell signaling models because experimental measurements of the variables of interest, protein dynamics, are still quite limited. We find that we can successfully integrate measurements from these databases to significantly improve parameter estimation of signaling models. The impact of sequence and structure measurements on model predictions depends on the sensitivity of the prediction to perturbations in the parameter values. Overall, this study demonstrates that measurements of protein structure and amino acid sequence can be leveraged to better inform parameters in models of cell signaling.

Author Summary

Computational models of cell signaling have provided mechanistic insights into complex biological systems, including in physiological and disease settings. Accurate and predictive modeling critically depends on the precise estimation of model parameters, which is often hindered by the limited availability of experimental data. In this study, we present a novel multiscale probabilistic inference framework that broadens the scope of data types that can be leveraged for parameter estimation for models of cell signaling. The framework integrates a machine learning pipeline with a generalizable parameter inference approach, enabling the use of experimental data across scales. Specifically, we demonstrate that incorporating protein amino acid sequence and 3D structural data enhances parameter estimation compared to traditional measurements such as protein concentrations over time. Improving parameter estimation increases the robustness and applicability of cell signaling models. Ultimately, our framework facilitates use of a broader range of data and supports the development of predictive computational models that increase our understanding of cell signaling.

1 Introduction

The application of computational modeling to biological systems has generated a substantial surge in discoveries and insights into these systems; however, especially as systems become more complex, accurately determining computational models remains challenging (1). For example, it can be challenging to parameterize systems biology models, as we often do not have direct measurements of model parameters. Instead, parameters must be inferred using noisy, partial observations of the system of interest. For instance, a system of ordinary differential equations (ODEs) may be used to model intracellular signaling (2). Such mechanistic models are useful, having been applied to optimize CAR-T cell therapies or elucidate control mechanisms of cellular responses (3,4). However, we rarely have observations of all the model parameters—such as binding affinities or protein shuttling rates. Instead, these parameters must be inferred using observations of the model variables: protein concentrations over time. Timeseries measurements of protein concentrations are often sparse, limited to relative concentrations, and noisy. When such measurements are used for parameter estimation, these limitations can result in significant uncertainty in the model predictions, which limits the applicability of the model (5).

To overcome this limitation, in this work we propose a novel, multiscale, probabilistic modeling framework that expands the experimental measurements that may be integrated into the parameter estimation of cell signaling models. Figure 1 illustrates this framework. We hypothesize that despite differences in scale, measurements of protein structure and amino acid sequence can still facilitate the parameter inference of ODE models of cell signaling. To overcome this difference in scale, our framework includes a user-friendly machine learning (ML) pipeline that predicts binding affinity parameters of cell signaling models using sequence or structure. To

integrate these affinity parameter predictions into the estimation of signaling models, our framework also includes a generalizable Bayesian inference approach. This probabilistic approach to parameter estimation allows us to formally quantify the information gained from augmenting traditional experimental measurements for parameter estimation of cell signaling modes, such as protein concentration dynamics, with measurements of sequence and structure.

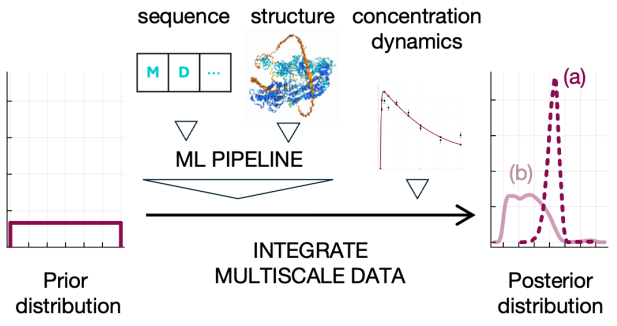

Figure 1. Multiscale Probabilistic Modeling Framework. Starting with the same range of possible parameter values—prior distribution—for the ODE model of cell signaling, we integrate data from different scales into our Bayesian parameter inference using a ML pipeline. We perform inference using two data sets. One dataset comprises only concentration dynamics. The other comprises concentration dynamics as well as amino acid sequence and protein structure. We may then compare the difference between the plausible parameter values—posterior distribution—with, (a), and without, (b), measurements of sequence and structure.

We use measurements of protein structure and amino acid sequence to augment our cell signaling model parameter inference. While experimental observations are still limited for the parameters and variables of signaling models, there have been significant breakthroughs in the ability to measure other biological variables. For example, researchers have been able to collect many more measurements of protein structure and amino acid sequence. This has resulted in rich databases such as the Protein Data Bank (PDB) and Universal Protein Resource (UniProt) (6,7). Thus, we propose augmenting timeseries measurements of protein concentrations with more plentiful data on sequence and structure.

Our proposed multiscale, probabilistic framework leverages a ML pipeline to integrate measurements of sequence and structure into the parameter inference of cell signaling models. Integrating such data into the parameter estimation is non-trivial because the data are not observations of variables or parameters of the biological system being modeled. Indeed, amino acid sequences and protein structures are on a different scale than the protein concentrations modeled in cell signaling. To overcome this challenge, we propose a ML pipeline that predicts the binding affinity parameters of cell signaling models using sequence and structure. We then compare these predicted parameters to the model parameters. ML models have been leveraged to similarly predict the parameters of models of cellular metabolism and steady-state signaling

91 (8,9) (10). We note that the latter study on steady state signaling was tested on only one protein
92 in the system, limiting the generalizability of the results. Excitingly, ours is the first such approach
93 tested on dynamic signaling models. Furthermore, we aim to increase the generalizability of our
94 results by presenting an expanded sample size.

95 Ultimately, we want to use our proposed framework to evaluate how much information is gained
96 from adding sequence and structure measurements to our signaling model parameter inference.
97 Despite interest in generalizing large datasets using ML models, there are a limited number of
98 studies quantifying how much information is being gained from augmenting measurements
99 traditionally used to calibrate systems biology models. For example, only recently has the limited
100 applicability of the ML model used to predict parameters of cellular metabolism models outside of
101 the ML model's training data set been revealed (11). To fill this important gap, our proposed
102 framework also comprises a generalizable Bayesian inference approach.

103 Bayesian inference is an established means to quantify the information gained from data. For
104 example, one particularly relevant study used it to quantify the information gained from absolute
105 measurements in the context of cell signaling models, which typically only leverage relative
106 measurements for parameter inference (12). Importantly, Bayesian inference characterizes the
107 posterior distribution, which is the distribution of plausible model parameters given a particular
108 dataset. Inference begins by defining a prior distribution, the distribution of plausible model
109 parameters given prior information. For example, the prior might formalize the fact that physical
110 parameters can only be positive. The inference approach then proceeds by incorporating
111 experimental datasets. To quantify the information gained from a particular dataset, one can
112 compare the distribution of plausible parameter values before and after integrating data using
113 statistical distance metrics. One common distance metric used to quantify the difference between
114 distributions is Kullback Leibler (KL) divergence (13). The more the posterior distribution changes,
115 the larger the KL divergence becomes, indicating that more information is gained from the data.
116 KL divergence has been applied in experimental design and hypothesis generation (14–16). Here,
117 we use it to quantify the information gained from incorporating experimental observations of
118 sequence or structure into the parameter inference of cell signaling models.

119 Overall, we are proposing a novel, multiscale, probabilistic modeling framework that enables data
120 integration across scales and quantifies the benefit of this integration (Figure 1). We use an ML
121 pipeline to integrate multiscale data (amino acid sequence and protein structure) into the

probabilistic, Bayesian parameter inference problem of ODE models of intracellular signaling. This probabilistic inference allows us to quantify the information gained from augmenting parameter inference with sequence and structure.

We test our approach on two, well characterized ODE models of cell signaling: Epidermal Growth Factor Receptor (EGFR) and G-Protein Coupled Receptor (GPCR) signaling. Not only are these models well-established; they are also biologically significant. In the case of EGFR signaling, this model represents dynamics upstream of the most well-studied pathway in cell signaling: MAPK (17). In the second case, modeling GPCR signaling is important as GPCR's are considered a prime drug target due to their ubiquity and targetability. In fact, around a third of drugs approved by the U.S. Food and Drug Administration (FDA) target this protein (18).

To understand what information is gained from using sequence or structure in the parameter inference of GPCR and EGFR signaling, we compare the results of two parameter inference approaches. In what we will refer to as the "baseline approach", only the published data reported for the EGFR or GPCR model is used to infer model parameters. In the "augmented approach", the reported data is used together with data from PDB or UniProt. Data from PDB and UniProt is integrated using our multiscale probabilistic framework. Throughout the paper, we refer to these two datasets as y_{BASE} and y_{AUG}. We ultimately compare the predictions and posterior inferred using y_{AUG} to the predictions and posterior inferred using y_{BASE}. We perform this comparison using KL divergence, an information theoretic distance enabled by our probabilistic framework, to quantify the information gained from augmenting the baseline data. Ultimately, we answer the question of what information can be gained from protein structure and amino acid sequence in the context of dynamic models of cell signaling.

2 Approach and Results

2.1 Machine Learning $\mathbf{K_D}$ Predictor Pipeline

To leverage the large quantities of data on amino sequence and protein structure in the context of cell signaling models, we must translate the information in a sequence or structure to a scale that is meaningful for cell signaling—that is, on the scale of protein concentrations. Further, this translation approach should be easy to use to complement existing modeling approaches for cell signaling. We developed a ML pipeline for this task (Figure 2).

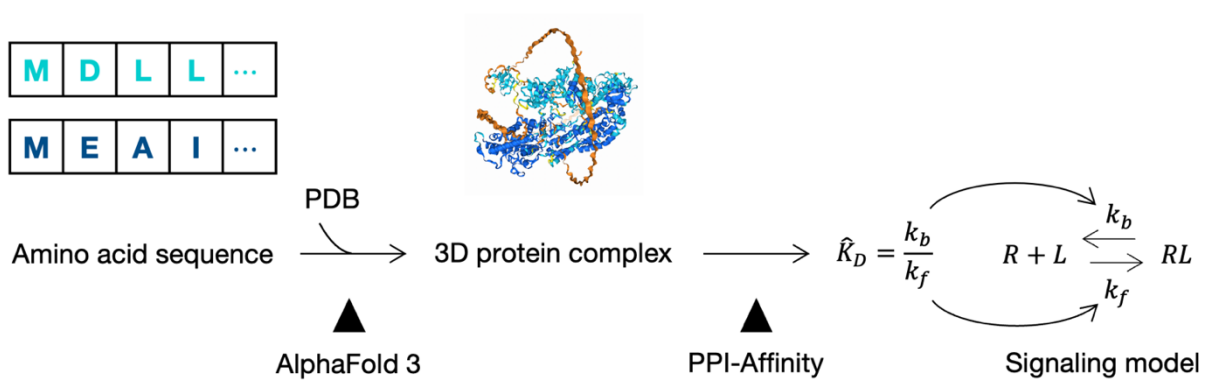

Figure 2. ML Pipeline to predict K_D. K_D is predicted using an SVM regression model, PPI-Affinity. PPI-Affinity needs the complex structure as an input. If the structure exists, it may be extracted from PDB. If it does not exist, the structure may instead be predicted using AlphaFold 3. AlphaFold 3 only needs amino acid sequences, which are typically available. The predicted K_D is the same K_D that characterizes reversible, bimolecular binding reactions comprising cell signaling models.

151 Upon reviewing the literature, we found that there exist several ML models that translate the
152 information in an amino acid sequence and/or structure to a scale that is meaningful for cell
153 signaling (19–22). Specifically, there are several ML models that predict protein binding affinity,
154 K_D, in units of concentration, from structure or sequence. We can then relate this predicted K_D to
155 the parameters of the ODE model, which are the unbinding (k_b) and binding (k_f) rates governing
156 the reaction, by using the definition of K_D: the ratio of the unbinding rate to the binding rate. These
157 ML models are generally limited to bimolecular, reversible binding reactions. However, this is not
158 a limitation in the context of cell signaling models, which are often comprised of such reactions.
159 This is because even when the product of the reaction contains more than two proteins, the
160 process usually involves a series of bimolecular reactions (23).

161 We found two ML models with convenient webservers for making K_D predictions: Prodigy and
162 PPI-Affinity (19,24). We moved forward with PPI-Affinity, a support vector machine (SVM)
163 regression model, as it shows modest improvements on Prodigy's performance. Further, it
164 performs comparably to more complicated K_D predictors, like those based on deep learning
165 models (19). PPI-Affinity requires the structure of the bound protein complex as an input; however,
166 only a fraction of experimentally available protein structures are complexes. Instead, most
167 structures available are for monomer proteins. Thus, we sought an alternative input to PPI-Affinity
168 to use in cases where the protein structure is not available.

169 Excitingly, AlphaFold (AF3), the latest iteration of the Nobel-prize winning, deep learning model
170 AlphaFold, is designed to predict the structures of biomolecular interactions, including protein

complexes (25). Although AF3 needs amino acid sequences of the binding partners to predict this structure, the amino acid sequence of a given protein of an intracellular signaling network is often known. AF3 is also hosted on a webserver, making predictions straightforward.

Altogether, the compatibility between signaling models and ML models, recent breakthroughs in predicting the structure of biomolecular interactions, and attention towards the useability of ML models enabled us to chain together existing ML models into one, user-friendly pipeline for incorporating amino acid sequence and protein structure into cell signaling models (Figure 2).

2.2 Test Case on Uninformative Prior

Before incorporating predictions from our ML pipeline into the parameter inference, we first tested its ability to accurately predict K_D for the ten binding reactions present in the ODE test cases. For each of the ten binding reactions, we predicted K_D using our proposed ML pipeline. We compared the accuracy of these predictions to an uninformed K_D prediction. To generate this uninformed prediction, we sampled our Bayesian prior ten times. This prior is log-uniform, and has been used as an uninformative distribution for parameter inference in systems biology (12). Thus, it is well suited for representing an uninformed K_D prediction. More details about this prior may be found in the Materials and Methods.

We calculate the absolute error of the ML pipeline predictions and prior samples with respect to the experimentally-derived K_D values reported in the GPCR and EGFR publications (26,27). Since we are concerned with order-of-magnitude changes in K_D, this error is calculated after a log10 transformation of the ML pipeline prediction and reported K_D values. Figure 3a shows how the ML pipeline's performance compares to the uninformed prediction's

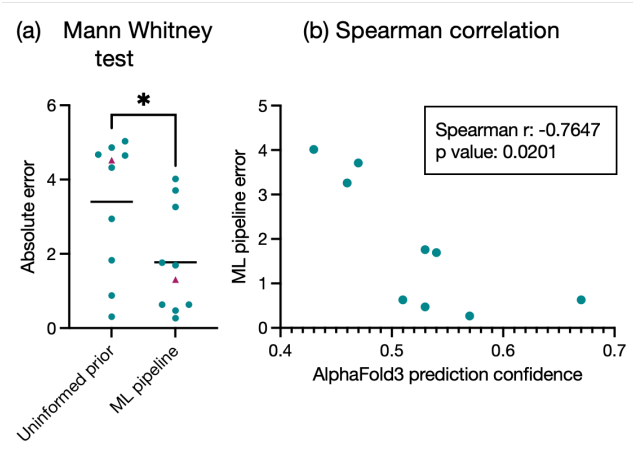

Figure 3. ML Pipeline performance. Cyan, EGFR reaction; pink, GPCR reaction. Circles, K_D predicted using predicted structure; triangles, K_D predicted using experimental structure. (a) Mann-Whitney test comparing absolute error of structure-informed K_D prediction with a random sample from a log-uniform distribution. n=10 binding reactions. *p-value< 0.05. Black horizontal line, sample mean. (b) Spearman's rank correlation between error of K_D prediction and AlphaFold 3 confidence metric, ranking score. n=9 binding reactions that were predicted using a predicted structure, given there was no experimental structure.

performance. The ML pipeline's predictions have a significantly lower absolute error than the uninformed prediction. One of these K_D predictions was made with an experimental structure. Interestingly, it resulted in only the fifth smallest error.

We also tested whether the confidence metrics provided by the ML models that comprise the pipeline were indicative of the pipeline's overall performance. To do this, we calculated the Spearman correlation between the error of the ML pipeline's prediction and each confidence metric—the applicability domain (AD) for PPI-Affinity and the ranking score for AlphaFold 3. More details on these metrics can be found in the Materials and Methods. In the case of AF3, we test on only nine of the binding reactions. This is because for one of the reactions, there is an experimental structure. Thus, there was no need to use AlphaFold to predict the structure. We found that PPI-Affinity's AD metric was not significantly associated with the overall pipeline's

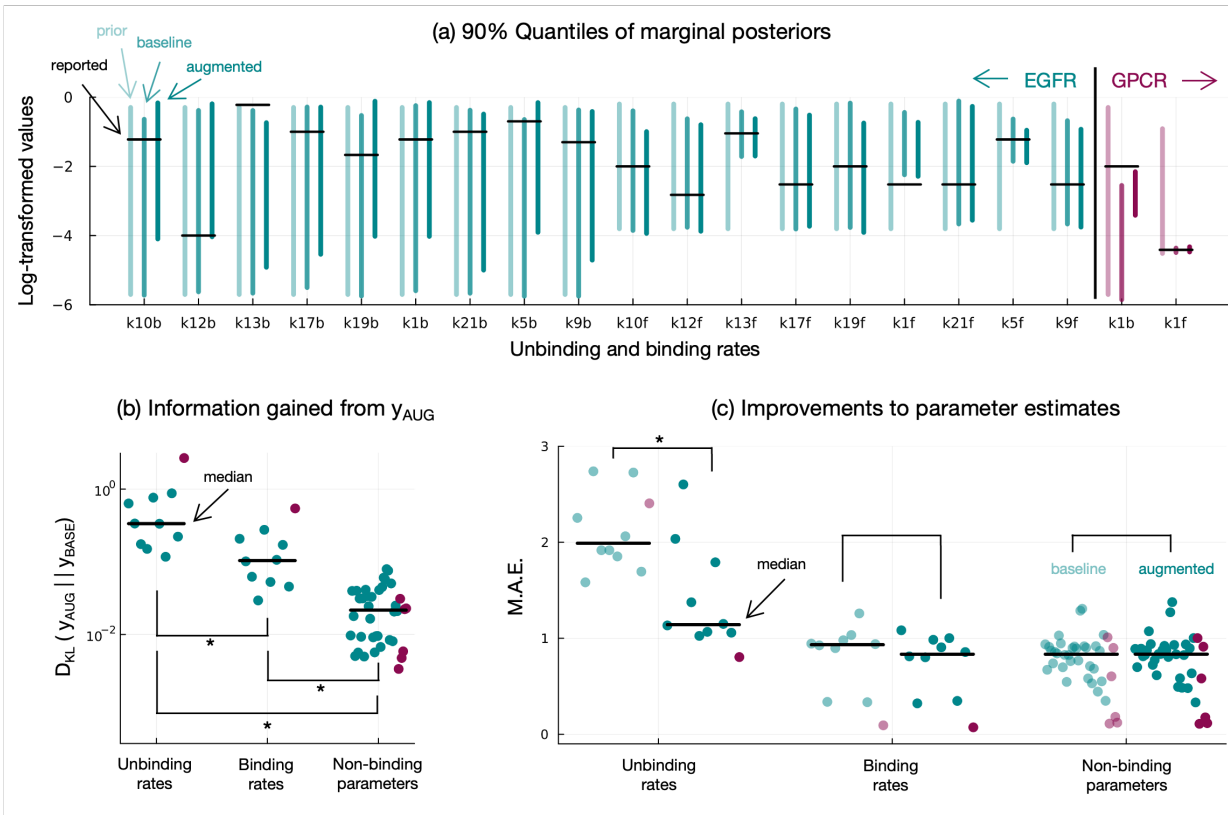

Figure 4. Parameter inference with and without data augmentation. Cyan, EGFR results; pink, GPCR results. All parameter samples were log-transformed prior to analysis. *p-value < 0.05 (a) 90% quantiles of marginal posterior distributions of binding parameters. Light cyan line, prior; medium cyan line, baseline posterior; dark cyan line, augmented posterior; black horizontal line, reported parameter value. (b) KL divergence, in bits, from baseline posterior to augmented posterior. Values are grouped by parameter function. (c) Mean absolute error (M.A.E.) of parameter samples. Mean taken with respect to each posterior distribution. Error calculated with respect to the values reported in the literature. Light cyan points, baseline; dark cyan points, augmented; black horizontal line, median M.A.E.

performance (Supplementary Information Figure S1a). However, we did find that AlphaFold 3's ranking score was significantly correlated with the pipeline's performance (Figure 3b).

2.3 Test Case on EGFR and GPCR Signaling: Parameters

Next, we infer the Bayesian posterior distributions of the ODE model parameters using y_{BASE} or y_{AUG}. Details on this inference can be found in the Materials and Methods. We analyze the posterior samples after a log10 transformation, as we want to compare order-of-magnitude changes in the parameters. The quantiles of the resulting marginal posterior distributions of the unbinding and binding rates are shown in Figure 4a. After checking that both inferences recapitulate the training data (Supplementary Information Figure S3), we analyzed the resulting posterior distributions. To measure the information gained from using the augmented dataset compared to the baseline dataset, we compute the KL divergence from the posterior estimated with y_{BASE} to the posterior estimated with y_{AUG}. The resulting KL divergences are shown in Figure 4b. Interestingly, the most information is gained on the rate of protein unbinding. Figure 4c plots the mean absolute log error the parameter estimates with and without data augmentation. Error is calculated with respect to the experimentally derived reported values from the published papers. Excitingly, we see that the information in y_{AUG} improves the parameter estimates, especially in the case of the unbinding rates, where the improvement is statistically significant.

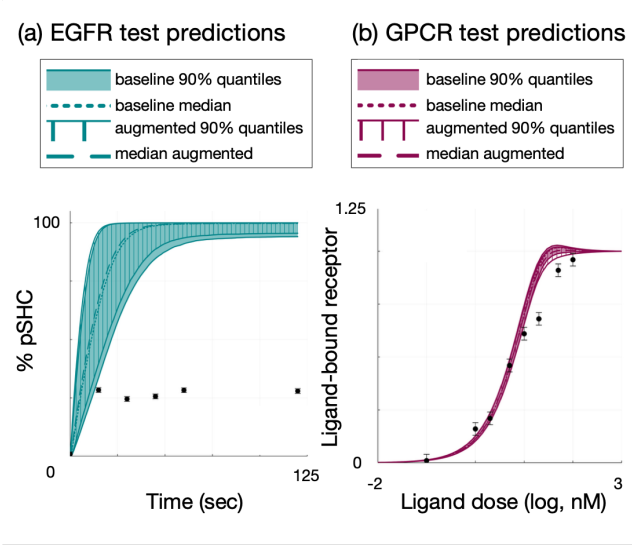

Figure 5. Performance on test set. (a) Predictions for EGFR test set, the percent of phosphorylated signaling protein SHC from 0-120 seconds. (b) Predictions for GPCR test set, the amount of ligand bound receptor 60 seconds post-stimulation, at different ligand doses, relative to 1000 nM of ligand. Cyan, EGFR; pink, GPCR. Shaded region, 90% quantiles of baseline approach; patterned region, 90% quantiles of augmented approach; dotted line, median prediction of augmented approach; dashed line, median prediction of augmented approach; black dots, experimental data with reported error.

2.4 Test Case on EGFR and GPCR Signaling: Predictions

Figure 5 shows the difference in the medians and 90% quantiles of the test predictions generated by sampling the augmented posterior compared to the

242 baseline posterior. The test/train split is detailed in the Materials and Methods. While we see some
243 differences at middle timepoints of the EGFR test set (Figure 5a) and at certain doses of the
244 GPCR test set (Figure 5b), overall the predictions are not qualitatively different. Further, in the
245 case of GPCR, at no point does the difference between the medians of the two approaches
246 become greater than the reported experimental error. In other words, the difference in the
247 predictions between the two approaches falls within the error of the test set. In the case of EGFR,
248 the difference between the medians does approach the level of the reported error. However, we
249 note the unusually small scale of this reported error, which is +/- 1% change in phosphorylated
250 SHC protein.

251 When we investigate the medians and 90% quantiles for other outputs besides the test set, we
252 observe larger differences in the predictions. For example, if we consider the timeseries of
253 receptor-bound ligand, RL, of GPCR at times other than 60 seconds post-stimulation, we see
254 greater differences between both the quantiles and medians (Figure 6a). Similarly, several
255 species comprising EGFR signaling show larger differences in the predicted median and 90
256 percent quantiles. (Supplementary Information Figure S2).

257 **2.5 Local Sensitivity Analysis**

258 To investigate what leads to changes in some, but not all, ODE model predictions, we conducted
259 a local perturbation analysis of both the GPCR and EGFR outputs to model parameters. We
260 perturbed the maximum likelihood parameter values of the baseline case by 5% and used the
261 local sensitivity at the maximum likelihood to calculate the projected change in an output from this
262 perturbation. We analyzed 360 outputs for GPCR and 345 outputs for EGFR. These outputs
263 correspond to five model species at eight timepoints and nine doses and 23 model species at five
264 timepoints and three doses, respectively for the two test cases. Doses and timepoints match
265 experimental conditions. However, we exclude timepoint zero because the sensitivity of the initial
266 concentrations to any parameter is zero. More details about the local sensitivity analysis may be
267 found in the Materials and Methods section.

268 Next, we calculated the Spearman correlation between the change in the median prediction of an
269 output and the sensitivity of that output to perturbations of either unbinding rates, binding rates,
270 or non-binding parameters. The resulting Spearman coefficients for both EGFR and GPCR are
271 shown in Figure 6b. Both test cases display a positive and significant correlation with the

sensitivity to perturbation of unbinding rates. In the case of GPCR, the correlation decreases from

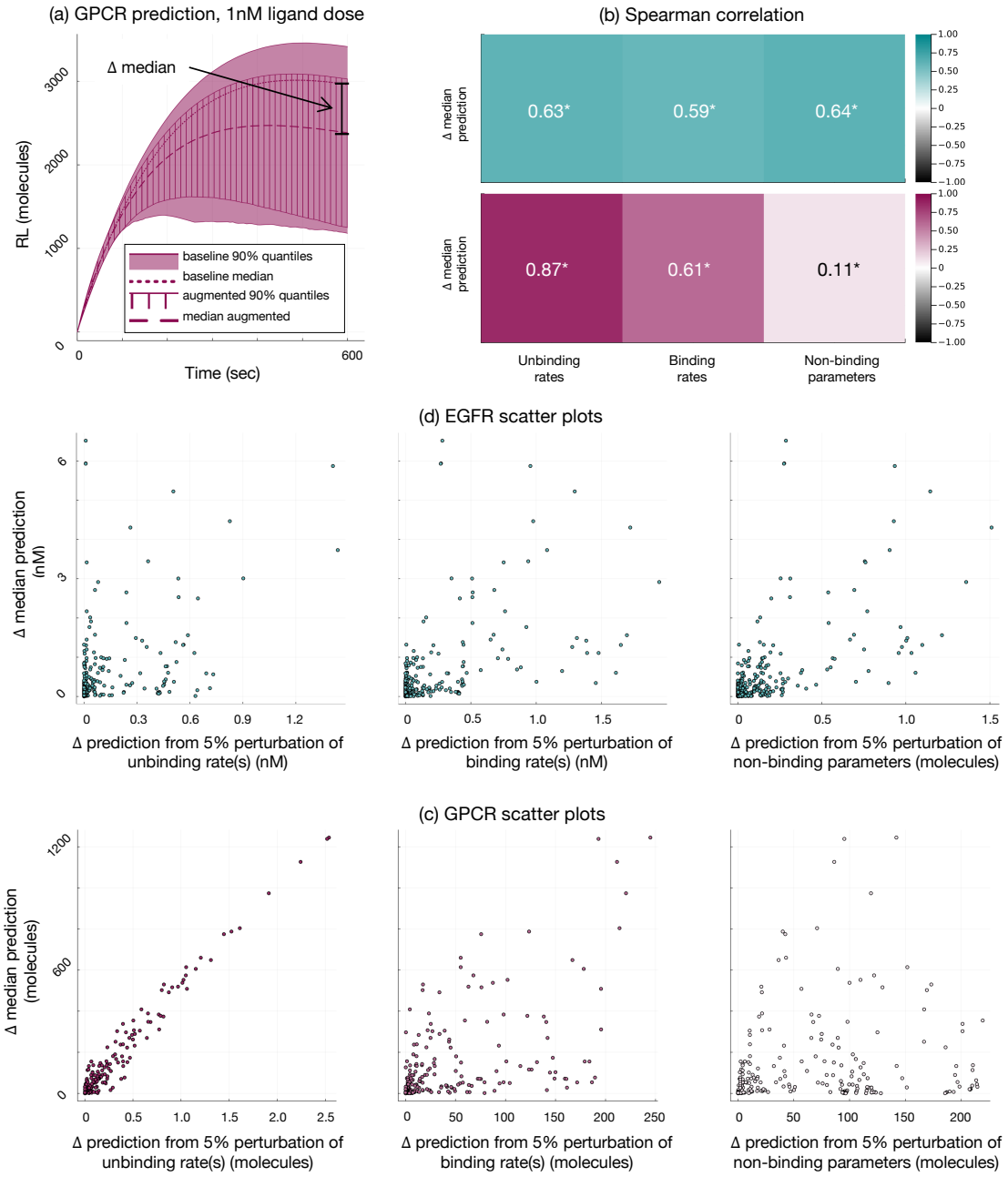

Figure 6 Local sensitivity analysis. (a) GPCR timeseries prediction of ligand-bound receptor, (RL), after 1 nM ligand stimulation. Units, molecules. Shaded region, 90% quantiles of baseline approach; patterned region, 90% quantiles of augmented approach; dotted line, median prediction of augmented approach; dashed line, median prediction of augmented approach. Difference between median prediction of baseline approach and median prediction of augmented approach, *Δ* median, indicated with black arrow. (b) Spearman correlation between *Δ* median and change in output due to 5% perturbation of either unbinding, binding, or non-binding parameters. Cyan, EGFR; pink, GPCR. *p-value< 0.05. n = 345 for EGFR; corresponding to 345 prediction outputs from 3 doses, 23 species, and 5 timepoints. n = 360 for GPCR; corresponding to 360 prediction outputs from 9 doses, 5 species, and 8 timepoints. Doses and timepoints match experimental observations. (c) Scatter plots illustrating relationship between perturbation and the *Δ* median prediction for GPCR. (d) Scatter plots illustrating relationship between perturbation to a type of parameter and the *Δ* median prediction for EGFR.

unbinding to binding to non-binding parameters. This decreasing correlation was clear visually (Figure 6c). Interestingly, this is the case despite the sensitivity to perturbations being around an order of magnitude lower for unbinding rates than the other parameter types. In the case of EGFR, there is a significant positive correlation for all parameter types.

2.6 Investigation of Machine Learning Prediction Confidence

Given the significant correlation between the confidence of the protein structure prediction and absolute error of the predicted K_D, (Figure 3b) we investigated whether weighing our data integration by the confidence of the predicted structure could improve our results. More details on how we implemented this confidence weight can be found in the Materials and Methods.

However, considering the confidence of the protein structure prediction did not significantly change most of the results of our parameter inference with and without data augmentation (Supplementary Information Figure S4). As in Figure 4, data augmentation provided the most information on unbinding rates. y_{AUG} also consistently improved parameter estimates of unbinding rates compared to y_{BASE}.

Predictions of the test data were similarly unchanged qualitatively (Supplementary Information Figure S5). As in Figure 5, the predictions with and without data augmentation were similar. One quantitative difference did emerge. When structure confidence was considered, the difference between median predictions for the EGFR test data was no longer greater than the reported error.

2.7 Robustness of Results to Changes in Prior Distribution

The prior distribution for each binding parameter is representative of a non-informative characterization of our knowledge of binding and unbinding rates. However, we also wanted to explore how the results might change given a more informative or less informative prior. To do so, we increased or decreased the upper and lower bounds of the prior distributions of all parameters by one order of magnitude. The results were qualitatively consistent across these changes in the prior (Supplementary Information Figures S6, S7, S8, and S9). Further, they were mostly quantitatively consistent as well; only a few quantitative differences emerged. In the case of the more informative prior, the difference between median predictions for the EGFR test data was no longer greater than the reported error (Supplementary Information Figure S7a). In the case of the less informative prior, the information gained from data augmentation was no longer

302 significantly different between the unbinding and binding rates (Supplementary Information Figure
303 S8b). This was because more information was gained for binding rates. Interestingly, this did not
304 result in a corresponding significant change in the mean absolute error for binding rates.

305 **3 Discussion**

306 Computational models of biological systems generate insights beyond what experiments alone
307 can offer. However, they also present their own challenges. Our work here explores the potential
308 of augmenting the conventional measurements used for parameter inference in models of
309 intracellular signaling. Ultimately, our goal was to reduce uncertainty in parameters for better
310 model determination. To do this, we proposed a multiscale, probabilistic modeling framework. We
311 first develop a ML pipeline for incorporating the new, multiscale measurements into parameter
312 inference. Then, we conduct a thorough quantification of the information gained from these new
313 measurements using a Bayesian framework.

314 We use amino acid sequence and protein structure measurements from UniProt and PDB,
315 respectively, to augment the published datasets of two test cases: EGFR and GPCR signaling.
316 The published datasets include partial, relative measurements of protein concentration and
317 protein co-localization over time. We chose the EGFR and GPCR models as our test cases
318 because both are well-established and biologically significant. Taken together, these models offer
319 a comprehensive test of our approach. For example, EGFR provides a test of a moderately sized
320 model, while GPCR provides a test of a minimal model. In addition, our ML pipeline can make K_D
321 predictions using sequence or structure. EGFR provides a test of predictions using sequence,
322 while GPCR provides a test of predictions using structure.

323 Predictions from the GPCR and EGFR models occur on the scale of protein concentrations. To
324 make measurements of sequence and structure applicable to the parameter estimation of these
325 models, we needed to convert the measurements to the same scale as the GPCR and EGFR
326 models. Excitingly, we were able to integrate two, user-friendly ML models, PPI-Affinity and
327 AlphaFold 3, into one pipeline to do this (19,25). This pipeline uses sequence or structure to
328 predict the binding affinity, K_D, of a given bimolecular, reversible reaction in a cell signaling model.
329 We find that this ML pipeline outperforms an uninformative prior when predicting the K_D. Thus, it
330 is appropriate to use when no experimental measurements exist for K_D. Importantly, this pipeline

331 is generalizable to other dynamic signaling models. It is limited only by access to amino acid
332 sequence and comprises two, user-friendly web servers.

333 Interestingly, we found that our pipeline's performance did not significantly benefit from using an
334 experimental protein complex structure, rather than predicted complex structure, when predicting
335 K_D. The binding affinity between GPCR and its ligand was predicted using an experimental
336 structure. However, the error of this prediction was not smaller than all the other K_D prediction
337 errors. PPI-Affinity was trained using complex structure and K_D pairs from the PDBbind database.
338 Although the GPCR structure was within the applicability domain of PPI-Affinity, it seems that the
339 ligand-bound GPCR complex was not included in this training set. This could explain the lack of
340 improvement we see. We also note the small sample size (n=1) of experimental structures we
341 tested, which limits the certainty of this result.

342 We incorporate the K_D's that we predict using the ML pipeline into the likelihood that we use in
343 our signaling model's parameter inference. We combined this likelihood of the predicted K_D and
344 the likelihood of the data reported in the original publications to form an augmented dataset, y_{AUG}.
345 This augmented dataset is ultimately informed by the experimental data on sequence or structure
346 that we use to predict K_D. We then compared the posteriors and predictions inferred with y_{AUG} to
347 the posteriors and predictions inferred using only the original data, y_{BASE}, our baseline dataset.

348 Overall, augmenting our data with data on sequence or structure provides more information on
349 rates of protein unbinding compared to using only the original data. This was consistent across
350 both the EGFR test case and GPCR test case. Further, the information gained from sequence or
351 structure tends to push the unbinding rate towards the rates reported in the original publications.
352 These reported rates were all taken from experiments, rather than fit to data (26,27). Our results
353 offer an understanding of when our approach is useful: in the case of underdetermined rates of
354 protein unbinding. We have confidence in this recommendation of utility, since the results were
355 consistent even with a more informative or less informative prior distribution. In practice, it is
356 helpful to use our approach to constrain unbinding rates, as these rates tend to vary over a larger
357 range than binding rates (23).

358 When we investigate how predictions change due to the data augmentation, the results are less
359 clear. We first find that augmenting the baseline data set does not fundamentally change the
360 predictions of our test data for either test case. However, we do see more substantial changes for

361 other outputs besides the test data. We thus investigated when predictions inferred with y_{BASE} will
362 differ from predictions inferred with y_{AUG}. As our approach changes unbinding parameter rates,
363 intuitively, one would then expect to see differences in predictions for outputs influenced by
364 unbinding rates. Thus, we investigated the relationship between the local sensitivity of an output
365 to unbinding rates, binding rates, and non-binding parameters. When we analyzed the GPCR
366 model, there emerged a striking, significant, positive and linear correlation between changes in a
367 prediction and the local sensitivity of that prediction to perturbations in the unbinding rate. A
368 significant positive correlation also emerged with the EGFR model. However, this relationship was
369 similar across unbinding rates, binding rates, and nonbinding parameters. The different results
370 from the two test cases may be a function of the complexity of the models. While GPCR is a
371 minimal model comprised of eight parameters, EGFR is a more moderately sized model
372 comprised of 50 parameters. It is thus plausible that the relationship between sensitivity and
373 predictions is impacted by higher-order interactions between parameters in the case of EGFR
374 (28). These interactions would not be captured by the local sensitivity analysis we conduct here.

375 There was a significant negative correlation between AlphaFold 3's confidence metric and the
376 error of the K_D prediction. Based on this, we investigated weighing the likelihood of K_D by
377 AlphaFold's confidence. Interestingly, incorporating confidence in this way did not significantly
378 improve or change our results. This could be explained by the fact that we were conservative
379 when balancing the influence of data augmentation against the influence of the baseline data. For
380 example, weighing by confidence, which was scaled from [0,1] decreased the influence of
381 predicted K_D's. An interesting future direction could be investigating other probabilistic
382 frameworks for incorporating AlphaFold's confidence into parameter inference.

383 We acknowledge some limitations of our study. We test our ML pipeline on only the 10 binding
384 reactions reported in our test cases. We could test our pipeline with more reactions from
385 databases of paired structures and affinities, like PDBbind (29). We chose to focus only on binding
386 affinities reported in the context of ODE models as these affinities are known to recapitulate
387 intracellular protein dynamics. Future work could investigate how our ML pipeline could be even
388 more specialized to binding reactions occurring inside the cell. For example, future work could
389 take in to account the crowded intra-cellular environment (30). With future improvements to the
390 ML pipeline in mind, we developed a Bayesian framework for incorporating $\widehat{K}_D$ that can evolve
391 with our pipeline. For example, we use the error reported by PPI-Affinity to weigh the predicted

392 K_D likelihood against the likelihood of the conventional experimental measurements. This weight
393 can be adjusted as that error improves.

394 There are also ML models that use amino acid sequence to predict binding affinity. In theory, we
395 could use one of these models instead of our pipeline. However, in practice, we found that there
396 were no such models hosted on convenient webservers. Even more limiting was the prohibitively
397 large RAM these models needed to make a prediction for the relatively large signaling protein
398 complexes we are interested in here (20). We instead focus on more user-friendly models, both
399 in terms of RAM and implementation.

400 In conclusion, we have developed a multiscale, probabilistic framework to show that it is viable to
401 use measurements of amino acid sequence and protein structure to augment parameter inference
402 in the context of dynamic signaling models. Experimental data on protein structure and amino
403 acid sequence consistently adds information on the unbinding rate parameters of these signaling
404 models. This improves parameter estimation and thus model determination. In the future, our
405 framework has the potential to synergize with breakthroughs in ML and data collection. As ML
406 models better predict binding affinities, our framework can better constrain signaling models. As
407 databases of protein structure and amino acid sequence grow, our framework will generalize to
408 more binding reactions.

409 **4 Materials and Methods**

410 **4.1 ODE models of intracellular signaling**

411 We base our investigation on two ODE models: a 50 parameter, 23 species model of EGFR
412 signaling and an eight parameter, seven species model of GPCR signaling (26,27). In the case
413 of EGFR signaling, a combination of mass action and Michaelis Menten kinetics is used to
414 describe cell signaling events such as protein binding and phosphorylation. In the case of GPCR
415 signaling, mass action kinetics are used to describe cell signaling events such as protein binding,
416 activation, and degradation. Binding reactions are of particular interest to our study: in the case
417 of EGFR, 18 parameters characterize nine binding reactions, while in the case of GPCR, two
418 parameters characterize one binding reaction.

The EGFR model simulates signaling up to 120 seconds, while the GPCR model simulates signaling up to 600 seconds. We use the reported initial species pools to initialize the model simulations. We simulate both models with the Julia package *DifferentialEquations.jl* (31).

Because our model parameter values can vary across a wide range, the solution may be stiff. Thus, we want to choose an ODE solver that can handle stiffness. In addition, we need a solver that performs well given a higher-than-default accuracy requirement. This is due to the nature of our state variable: protein concentration. The accuracy of a particular problem is specified by the absolute and relative tolerances. The relative tolerance (rtol) is set based on the number of significant digits required for the model output, plus 1 (32,33). We note that the upper bound on the number of significant digits is between 15 and 17, as this is the maximum number of significant digits stored by a Float64 type variable, which we use here. In our case, we deem 5 significant digits sufficient for our problem and set the rtol to $10^{-(5+1)}$. The absolute tolerance (atol) is set based on the absolute value of the model output that may be deemed insignificant (32,33). In our case, we deem a protein concentration of 10^{-5} nM, or less than 1 protein/cell, as insignificant. A rtol of 10^{-6} and atol of 10^{-5} are more as stringent as the default tolerances for most scientific computation, 10^{-3} and 10^{-6}, respectively (32,34). Based on our requirements, we choose to use *DifferentialEquation.jl*'s QNDF solver, which is equivalent to Matlab's ode15s solver (34). It is a stiff solver that can handle medium tolerances; thus, it meets our problem's requirements (32,34,35).

4.2 Data Augmentation

Given an ODE model, $M(\cdot)$, and experimental data, y, our objective is to infer the model parameters, θ. In this work, we propose augmenting the published data for each system, y_{BASE}, with predicted K_D values, $\widehat{K}_D$. Equation (1) shows how the augmented dataset, y_{AUG} incorporates additional data into y_{BASE}. We predict K_D using experimental observations of amino acid sequences from UniProt, y_{FASTA}, or protein structures from PDB, y_{PDB}. y_{BASE} , y_{FASTA} and y_{PDB} are defined in detail in the subsequent sections.

$$y_{AUG} = \left[y_{BASE}, \widehat{K}_D(y_{PDB} \vee y_{FASTA})\right] \tag{1}$$

Throughout the paper, we refer to $\widehat{K}_D(y_{PDB}, y_{FASTA})$ as simply predicted K_D, for conciseness and clarity.

4.3 Experimental measurements of EGFR and GPCR signaling

Both ODE models report experimental measurements of the system of interest. That is, measurements of in vitro protein dynamics, which correspond to the predictions of the ODE models (Figure 7). Note that we are not pooling data across systems, but, for simplicity, we will refer to each data set as y_{BASE}. In the context of the EGFR model, y_{BASE} are measurements of EGFR signaling dynamics (Figure 7a) and in the context of the GPCR model, y_{BASE} are measurements of GPCR signaling dynamics (Figure 7b). We extracted the data and errors with Plot Digitizer (36) .

Eight outputs of EGFR are observed, and we show three of these outputs as a representative sample in Figure 7a. These include the percent of phosphorylated receptor at various ligand dosages as well as the percent of phosphorylation of two intracellular proteins at various ligand

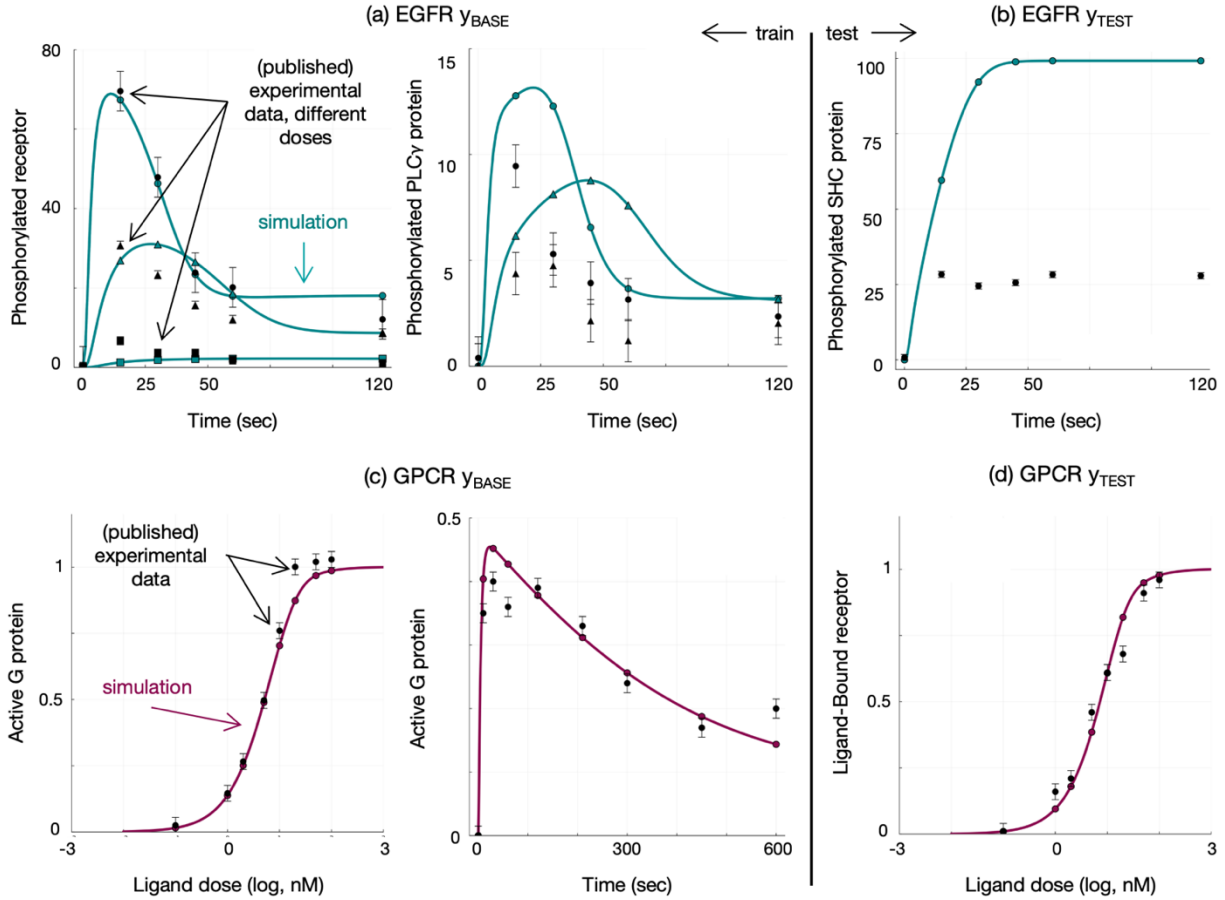

Figure 7. Baseline data set (y_{BASE}). (a) Training data for EGFR parameter inference. (b) Test data for EGFR parameter inference. (c) Training data for GPCR parameter inference. (d) Test data for GPCR parameter inference. Cyan or pink lines, simulations using reported parameter values for EGFR and GPCR, respectively; black markers, reported experimental data points for a particular dose; black error bars, reported standard error.

459 dosages. Further observations include the percent of an intracellular protein bound with another
460 intracellular protein and the receptor. The original work does not fit parameters to data, so we
461 must decide our test/train data split. We choose to reserve measurements of one of the
462 intracellular proteins, SHC, as our test data. We choose SHC due to its clinical relevance, as this
463 signaling species has been shown as an important marker in various cancers (37). Furthermore,
464 the published model predictions do not quantitatively match the experimental measurements of
465 SHC (37). Ideally, incorporating data on protein structure will improve our predictions for this
466 experimentally relevant target. Thus, we use the data for SHC to test our model predictions.

467 Three outputs of GPCR signaling are observed, all of which are shown in Figure 7b. These include
468 levels of active G-protein at various ligand dosages as well as levels of the ligand-bound receptor
469 at various ligand dosages. As in the original work, we use the two G-protein data sets as training
470 data. We reserve data on the bound receptor for testing.

471 Both models also report binding parameter values derived from experiments or physical bounds,
472 rather than fit to the data. We use these reported values as a test of the binding parameter values
473 themselves.

474 **4.4 Experimental measurements from UniProt and PDB Databases**

475 We augment y_{BASE} with amino acid sequences or structures of the proteins that are involved in
476 binding for each system. Given two proteins that are binding, we first searched for any structures
477 of the bound complex in PDB. One such structure exists, describing GPCR bound to alpha
478 pheromone, PDB structure 7QBI (38,39). This structure corresponds to the single binding reaction
479 in the GPCR model and is formatted as a PDB file. This measurement comprises y_{PDB}.

480 If no such structure was found, as was the case for the nine EGFR binding reactions, we extracted
481 the canonical amino acid sequences of the individual proteins, as FASTA files, from UniProt. They
482 are extracted UniProt sequences P01133 (EGF), P29353 (SHC1), P19174 (PLCG1), Q07889
483 (SOS1), P62995 (GRB2), and P00533 (EGFR). These measurements comprise y_{FASTA}.

484 **4.5 ML pipeline to predict K_D**

485 Figure 2 provides an overview of the ML pipeline used to predict K_D using amino acid sequence
486 (y_{FASTA}) or protein structure (y_{PDB}). Given either the sequences of the binding proteins or the

structure of the bound complex, our objective is to predict the binding affinity, K_D, in units of concentration, for a specific bimolecular binding reaction. We can then relate this predicted K_D to the parameters of the ODE model, which are the unbinding (k_b) and binding (k_f) rates governing this reaction, by using the definition of K_D: the ratio of the unbinding rate to the binding rate.

If the bound complex structure is available, we can use a support vector machine (SVM) regression model, PPI-Affinity, to predict ΔG, the free energy change in kcal/mol (19). PPI-Affinity uses a feature vector derived from the complex structure to make its prediction of ΔG. We then convert ΔG to K_D, given a gas constant and absolute temperature (40,41). If the bound complex is not available, we use a deep-learning model, AlphaFold 3, to predict the complex structure given the amino acid sequences of the individual proteins. If the complex is comprised of identical subunits, we extract the smallest subunit (21). We then use this predicted structure as in input to PPI-Affinity to predict K_D.

Both PPI-Affinity and AlphaFold 3 provide confidence metrics for their predictions. AlphaFold provides a 'ranking score' for each predicted structure. This score ranges from [-100, 1.5] with -100 representing low confidence. The ranking score is a function of various features of the protein structure, such as the fraction of the predicted structure that is disordered and the interface predicted template modeling (ipTM) score (25). Both the ranking score and ipTM are significantly correlated with the performance of the ML pipeline (Figure 3b and Supplementary Information Figure S1b). Conveniently, ipTM is scaled between [0,1], with zero representing low confidence. We take advantage of this property when considering AlphaFold 3 confidence in our Bayesian framework, detailed next. In the case of PPI-Affinity, it is noted whether each K_D prediction is within the applicability domain of the training dataset (19). We found that the accuracy of predictions was more dependent on ranking score than applicability domain.

4.6 Bayesian inference

The objective of Bayesian inference is to find the posterior distribution, $p(\theta|y)$, which is defined using Bayes rule:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

While we often know the likelihood of the data given model parameters, $p(y|\theta)$, and the prior probability of parameters, $p(\theta)$ we usually do not know the probability of the data $p(y)$. Thus, we must approximate this distribution knowing $p(\theta|y)$ up to a proportionality constant. While there are various distribution approximation methods for this problem, the most powerful are Markov-Chain Monte Carlo (MCMC) approximations, because our initial guess of $p(\theta|y)$ does not need to be particularly close to the reality. These algorithms generate correlated samples of the posterior, which we may use to simulate our model.

To generate these samples for our ODE models, we use a Julia implementation of the affine invariant ensemble MCMC sampler from the package *AffineInvariantMCMC.jl* (42). Like all MCMC algorithms, the affine-invariant ensemble sampler algorithm constructs a Markov chain that is used to sample the posterior distribution. To better navigate the parameter space, this algorithm initializes an ensemble of samplers that propose the next sample based on the location of another sampler in the ensemble. Because all our parameter values are restricted to positive values and are often described on a logscale rather than linear scale, we conduct sampling in a log-transformed parameter space (43,44). Further discussion of the convergence diagnostics of the samples to the posterior may be found in the Supplementary Information.

4.7 Prior

The prior encodes our assumptions, ideally derived from empirical knowledge, about parameters. Here, we choose to define independent, log-uniform priors. If the parameter is a binding rate, we limit the distribution to [10^{-4} $nM^{-1}s^{-1}$, 100 $nM^{-1}s^{-1}$]; for an unbinding rate, we limit the distribution to [10^{-6} s^{-1}, 100 s^{-1}]. This is in line with prior distributions for binding rates in the literature and database values (12,45). For non-binding parameters, we limit the distributions to +/- two orders of magnitude above and below the reported parameter value.

4.8 Likelihood

Recall that our augmented data, y, is comprised of two categories, y_{BASE} and predicted K_D, $\widehat{K}_D$. Both y_{BASE} and $\widehat{K}_D$ represent vectors of data points. For example, in the case of EGFR, y_{BASE} is a vector of 42 points, corresponding to seven different outputs observed at six timepoints each, while $\widehat{K}_D$ is a vector of nine predicted affinities for the nine binding reactions of the model.

542 We assume that each data point is independent from the rest. Next, we assume that a data point
543 in y_{BASE}, y_{BASE}^{j}, is normally-distributed, given the ODE model prediction and experimental error for
544 data point j. For example, if j indicates the amount of active G-protein 300 seconds after
545 stimulation, then we use the model's prediction for active G-protein 300 seconds after stimulation
546 as the mean of the likelihood distribution and the reported error of this measurement as the
547 deviation of the likelihood distribution.

548 Finally, we assume that a data point in $\widehat{K}_D$, $\widehat{K}_D^k$, is log-normally distributed, given the ratio of the
549 unbinding and binding reaction rates for that data point and standard error of the SVM regression
550 model, PPI-Affinity. For example, if $\widehat{K}_D^k$ is the predicted binding affinity of a receptor binding a
551 ligand, then we will use the rate of receptor ligand unbinding divided by rate of receptor ligand
552 binding as the mean of the likelihood distribution and the standard error as the deviation. This
553 standard error is on the scale of one order of magnitude change in $\widehat{K}_D$.

554 For parts of the investigation, we augment this likelihood term to account for AlphaFold3's
555 confidence. Specifically, we weigh the log-likelihood of a given $\widehat{K}_D^k$ by AlphaFold 3's interface
556 predicted template modeling (ipTM) score, which represents the confidence of $\widehat{K}_D^k$ scaled from
557 zero to one. This likelihood formulation is known as a power prior. Power priors consist of raising
558 the likelihood, or a term in the likelihood, to a discounting factor. This controls the amount of
559 information taken from certain data (46,47). When the log likelihood is evaluated, the product of
560 likelihoods becomes a sum, and the discount factor(s) become weight(s) on the terms of this sum.
561 Further details on our likelihood distribution are included in the Supplementary Information.

562 **4.9 Log Transformation of Parameters for Analysis**

563 We note that analyses of the model parameters, and therefore posteriors, are done after a log10-
564 transformation of the parameter samples. The reason for this is threefold. First, log-transformation
565 tends to bring biochemical parameters like K_D towards a Normal distribution, which is desirable
566 for most statistical analyses (43). Second, by comparing order-of-magnitude changes in
567 parameters, we can compare parameters of different scales. Finally, we are concerned with order-
568 of-magnitude changes in K_D, as it has been shown that changes less than this do not significantly
569 influence model predictions (48).

570 **4.10 Quantifying information gained using KL divergence**

We use KL divergence between marginal posterior distributions to quantify the information gained from augmenting our data with experimental measurements of amino acid sequence and protein structure. Specifically, we calculate the KL divergence from the marginal probability of parameter θ_i given only the baseline data, $q(\theta_i|y_{BASE})$, to the marginal posteriors estimated with the augmented data, $p(\theta_i|y_{AUG})$:

$$D_{KL}(P||Q) = \int_0^\infty p(\theta_i|y_{AUG}) \log \frac{p(\theta_i|y_{AUG}):}{q(\theta_i|y_{BASE})}$$

Given posterior samples, we use a univariate kernel density estimate and Monte Carlo integration to approximate a divergence for each marginal posterior (15).

4.11 Sensitivity Analysis and Perturbation Analysis

We compute the local sensitivity of a model prediction with respect to model parameters using Julia's *ForwardDiff.jl* packages (49–51). Given the local nature of this analysis, parameters are not log-transformed. *ForwardDiff.jl* approximates the derivative of an output with respect to a model parameter using forward mode automatic differentiation. We use forward mode automatic differentiation as it is recommended for ODE's with less than 100 parameters (49). We take the local sensitivity with respect to the posterior parameter set that maximizes the likelihood of the baseline training data.

To calculate the projected change in an output due to a 5% change in a particular parameter, we multiple the local sensitivity by 5% of the maximum likelihood parameters. We take the absolute value of the resulting perturbation. We then need to down select one perturbation value for each type of parameter: unbinding, binding, or non-binding. In the case of GPCR, this is trivial for unbinding and binding rates, as there is only one unbinding parameter and one binding parameter. For the other cases, GPCR non-binding parameters and all EGFR parameters, we take the maximum resulting perturbation.

4.12 Statistical Analyses

All statistical analyses, Mann Whitney U test and Spearman rank correlation, were carried out using Prism 10.4.0 or a combination of Julia's *HypothesisTests.jl* package and *StatsBase.jl* packages.

4.13 Code Availability

Code is available at: https://github.com/FinleyLabUSC/structure_informed_cell_signaling.

5 Acknowledgements

The authors thank members of the Finley research group for critical comments and suggestions. This work was partially supported by the National Cancer Institute of the National Institutes of Health grant 1U01CA27580 and National Science Foundation award 2410753.

6 References

1. Puniya BL, Verma M, Damiani C, Bakr S, Dräger A. Perspectives on computational modeling of biological systems and the significance of the SysMod community. Zhu S, editor. Bioinforma Adv. 2024 Jan 5;4(1):vbae090.

2. Ingalls BP. Mathematical modeling in systems biology: an introduction. Cambridge, Massachusetts: MIT Press; 2013. 408 p.

3. Nandagopal N, Santat LA, LeBon L, Sprinzak D, Bronner ME, Elowitz MB. Dynamic Ligand Discrimination in the Notch Signaling Pathway. Cell. 2018 Feb 8;172(4):869-880.e19.

4. Rohrs JA, Siegler EL, Wang P, Finley SD. ERK Activation in CAR T Cells Is Amplified by CD28-Mediated Increase in CD3ζ Phosphorylation. iScience [Internet]. 2020 Apr 24 [cited 2025 Mar 27];23(4). Available from: https://www.cell.com/iscience/abstract/S2589-0042(20)30207-8

5. MacLean AL, Rosen Z, Byrne HM, Harrington HA. Parameter-free methods distinguish Wnt pathway models and guide design of experiments. Proc Natl Acad Sci U S A. 2015 Mar 3;112(9):2652–7.

6. UniProt: the Universal Protein Knowledgebase in 2025 | Nucleic Acids Research | Oxford Academic [Internet]. [cited 2025 May 16]. Available from: https://academic.oup.com/nar/article/53/D1/D609/7902999?login=false

7. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000 Jan 1;28(1):235–42.

624 8. Yu H, Deng H, He J, Keasling JD, Luo X. UniKP: a unified framework for the prediction of
625 enzyme kinetic parameters. Nat Commun. 2023 Dec 11;14(1):8211.

626 9. Li F, Yuan L, Lu H, Li G, Chen Y, Engqvist MKM, et al. Deep learning-based kcat prediction
627 enables improved enzyme-constrained model reconstruction. Nat Catal. 2022
628 Aug;5(8):662–72.

629 10. Zaballa VD, Hui EE. Systems-Structure-Based Drug Design [Internet]. arXiv; 2024 [cited
630 2025 Mar 27]. Available from: http://arxiv.org/abs/2410.10108

631 11. Kroll A, Lercher MJ. DLKcat cannot predict meaningful kcat values for mutants and
632 unfamiliar enzymes. Biol Methods Protoc. 2024 Jan 1;9(1):bpae061.

633 12. Thijssen B, Dijkstra TMH, Heskes T, Wessels LFA. Bayesian data integration for quantifying
634 the contribution of diverse measurements to parameter estimates. Bioinforma Oxf Engl.
635 2018 Mar 1;34(5):803–11.

636 13. Kullback S, Leibler RA. On Information and Sufficiency. Ann Math Stat. 1951 Mar;22(1):79–
637 86.

638 14. Lomeli LM, Iniguez A, Tata P, Jena N, Liu ZY, Van Etten R, et al. Optimal experimental
639 design for mathematical models of haematopoiesis. J R Soc Interface. 2021
640 Jan;18(174):20200729.

641 15. Huber HA, Georgia SK, Finley SD. Systematic Bayesian Posterior Analysis Guided by
642 Kullback-Leibler Divergence Facilitates Hypothesis Formation. J Theor Biol. 2022
643 Nov;111341.

644 16. George A, Zuckerman DM. From Average Transient Transporter Currents to Microscopic
645 Mechanism—A Bayesian Analysis. J Phys Chem B. 2024 Feb 29;128(8):1830–42.

646 17. Glont M, Arankalle C, Tiwari K, Nguyen TVN, Hermjakob H, Malik-Sheriff RS. BioModels
647 Parameters: a treasure trove of parameter values from published systems biology models.
648 Robinson P, editor. Bioinformatics. 2020 Nov 1;36(17):4649–54.

649 18. Eiger DS, Pham U, Gardner J, Hicks C, Rajagopal S. GPCR systems pharmacology: a
650 different perspective on the development of biased therapeutics. Am J Physiol - Cell
651 Physiol. 2022 May 1;322(5):C887–95.

652 19. Romero-Molina S, Ruiz-Blanco YB, Mieres-Perez J, Harms M, Münch J, Ehrmann M, et al.
653 PPI-Affinity: A Web Tool for the Prediction and Optimization of Protein–Peptide and Protein–
654 Protein Binding Affinity. J Proteome Res. 2022 Aug 5;21(8):1829–41.

655 20. Zheng F, Jiang X, Wen Y, Yang Y, Li M. Systematic investigation of machine learning on
656 limited data: A study on predicting protein-protein binding strength. Comput Struct
657 Biotechnol J. 2024 Dec;23:460–72.

658 21. Zhou Z, Yin Y, Han H, Jia Y, Koh JH, Kong AWK, et al. ProAffinity-GNN: A Novel Approach
659 to Structure-Based Protein–Protein Binding Affinity Prediction via a Curated Data Set and
660 Graph Neural Networks. J Chem Inf Model. 2024 Dec 9;64(23):8796–808.

661 22. Vangone A, Bonvin AM. Contacts-based prediction of binding affinity in protein–protein
662 complexes. Levitt M, editor. eLife. 2015 Jul 20;4:e07454.

663 23. Pollard TD. A Guide to Simple and Informative Binding Assays. Mol Biol Cell. 2010
664 Dec;21(23):4061–7.

665 24. Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A. PRODIGY: a web server for
666 predicting the binding affinity of protein–protein complexes. Bioinformatics. 2016 Dec
667 1;32(23):3676–8.

668 25. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure
669 prediction of biomolecular interactions with AlphaFold 3. Nature [Internet]. 2024 May 8 [cited
670 2024 May 15]; Available from: https://www.nature.com/articles/s41586-024-07487-w

671 26. Kholodenko BN, Demin OV, Moehren G, Hoek JB. Quantification of Short Term Signaling by
672 the Epidermal Growth Factor Receptor. J Biol Chem. 1999 Oct;274(42):30169–81.

673 27. Yi TM, Kitano H, Simon MI. A quantitative characterization of the yeast heterotrimeric G
674 protein cycle. Proc Natl Acad Sci. 2003 Sep 16;100(19):10764–9.

675 28. Monsalve-Bravo GM, Lawson BAJ, Drovandi C, Burrage K, Brown KS, Baker CM, et al.
676 Analysis of sloppiness in model simulations: Unveiling parameter uncertainty when
677 mathematical models are fitted to data. Sci Adv. 2022 Sep 23;8(38):eabm5952.

678 29. Liu Z, Li Y, Han L, Li J, Liu J, Zhao Z, et al. PDB-wide collection of binding data: current
679 status of the PDBbind database. Bioinformatics. 2015 Feb 1;31(3):405–12.

680 30. Grassmann G, Miotto M, Desantis F, Di Rienzo L, Tartaglia GG, Pastore A, et al.
681 Computational Approaches to Predict Protein–Protein Interactions in Crowded Cellular
682 Environments. Chem Rev. 2024 Apr 10;124(7):3932–77.

683 31. Rackauckas C, Nie Q. DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem
684 for Solving Differential Equations in Julia. J Open Res Softw [Internet]. 2017 May 25 [cited
685 2025 Mar 27];5(1). Available from:
686 https://openresearchsoftware.metajnl.com/articles/10.5334/jors.151

687 32. Shampine LF, Gladwell I, Thompson S. Solving ODEs with MATLAB. Cambridge ; New
688 York: Cambridge University Press; 2003. 263 p.

689 33. Brenan KE, Campbell SL, Petzold LR. Numerical solution of initial-value problems in
690 differential-algebraic equations. Philadelphia: Society for Industrial and Applied
691 Mathematics; 1996. 256 p. (Classics in applied mathematics).

692 34. Rackauckas C, Nie Q. DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem
693 for Solving Differential Equations in Julia. J Open Res Softw. 2017 May 25;5(1):15.

694 35. The MathWorks, Inc. Choose an ODE Solver [Internet]. The MathWorks Inc.; 2023.
695 Available from: https://www.mathworks.com/help/matlab/math/choose-an-ode-solver.html

696 36. PlotDigitizer: Version 3.1.6 [Internet]. 2025. Available from: https://plotdigitizer.com

697 37. Guo Z, Cai C, Zhou K, Song L, Wang X, Chen D, et al. SHC1 serves as a prognostic and
698 immunological biomarker in clear cell renal cell carcinoma: a comprehensive bioinformatics
699 and experimental analysis. Sci Rep. 2024 Aug 30;14(1):20150.

700 38. Velazhahan V, Ma N, Vaidehi N, Tate CG. Activation mechanism of the class D fungal
701 GPCR dimer Ste2. Nature. 2022 Mar;603(7902):743–8.

702 39. Velazhahan V, Tate CG. Structure of the GPCR dimer Ste2 in the active-like state bound to
703 agonist. 2021.

704 40. Kastritis PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AMJJ, et al. A structure-based
705 benchmark for protein–protein binding affinity. Protein Sci. 2011 Mar;20(3):482–91.

706 41. Ruiz-Blanco YB, Paz W, Green J, Marrero-Ponce Y. ProtDCal: A program to compute
707 general-purpose-numerical descriptors for sequences and 3D-structures of proteins. BMC
708 Bioinformatics. 2015 May 16;16(1):162.

709 42. Goodman J, Weare J. Ensemble samplers with affine invariance. Commun Appl Math
710 Comput Sci. 2010 Jan 31;5(1):65–80.

711 43. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, et al. The moderately
712 efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters.
713 Biochemistry. 2011 May 31;50(21):4402–10.

714 44. Lubitz T, Liebermeister W. Parameter balancing: consistent parameter sets for kinetic
715 metabolic models. Bioinformatics. 2019 Oct 1;35(19):3857–8.

716 45. Tötsch N, Hoffmann D. Bayesian Data Integration Questions Classic Study on Protease
717 Self-Digest Kinetics. ACS Omega. 2020 Jun 30;5(25):15162–8.

718 46. Carvalho LM, Ibrahim JG. On the normalized power prior. Stat Med. 2021;40(24):5251–75.

719 47. Ibrahim JG, Chen MH, Gwon Y, Chen F. The Power Prior: Theory and Applications. Stat
720 Med. 2015 Dec 10;34(28):3724–49.

721 48. Soltis AR, Saucerman JJ. Robustness portraits of diverse biological networks conserved
722 despite order-of-magnitude parameter uncertainty. Bioinformatics. 2011 Oct
723 15;27(20):2888–94.

724 49. Ma Y, Dixit V, Innes MJ, Guo X, Rackauckas C. A Comparison of Automatic Differentiation
725 and Continuous Sensitivity Analysis for Derivatives of Differential Equation Solutions. In:

2021 IEEE High Performance Extreme Computing Conference (HPEC) [Internet]. Waltham, MA, USA: IEEE; 2021 [cited 2025 Apr 3]. p. 1–9. Available from: https://ieeexplore.ieee.org/document/9622796/

50. Revels J, Lubin M, Papamarkou T. Forward-Mode Automatic Differentiation in Julia [Internet]. arXiv; 2016 [cited 2025 May 15]. Available from: http://arxiv.org/abs/1607.07892

51. Dixit VK, Rackauckas C. GlobalSensitivity.jl: Performant and Parallel Global Sensitivity Analysis with Julia. J Open Source Softw. 2022 Aug 17;7(76):4561.