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. Here, we investigate what information can be gleaned from these general databases 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. We use this framework to integrate measurements from the protein data bank (PDB) and UniProt (FASTA) and to quantify 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 improve parameter estimation of the cell signaling models. The impact of the database-derived measurements on model predictions depends on the complex relationship between model prediction and parameter values. Nonetheless, this study demonstrates that measurements from databases have the potential to be generalized to better inform parameters in models 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 such 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 models are useful, having been applied to optimize CAR-T cell therapies or elucidate cellular 'decision making' [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 limited to relative concentrations and are subject to significant noise. When these measurements are used for parameter estimation, these limitations can result in significant uncertainty in the model predictions, limiting the applicability of the model [5]. To overcome this limitation, in this work we investigate how much information is gained from integrating a diverse set of multiscale data into the parameter estimation of cell signaling models. We hypothesize that augmenting the data used to calibrate signaling models will better inform model parameters and predictions.

While experimental observations are still limited for the parameters and variables of cell signaling models, there have been significant breakthroughs in our ability to measure other biological variables. For example, researchers have been able to collect measurements of protein structure and amino acid sequence with great success. This has resulted in rich databases such as the Protein Data Bank (PDB) and UniProt [6], [7]. We propose augmenting timeseries measurements of protein concentrations with sequence and structure data from UniProt and PDB to improve parameter inference of cell signaling models.

However, integrating data on sequence and structure into the parameter estimation of cell signaling models is non-trivial. This is 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. This raises the question of how one can leverage experimental measurements of sequence and structure to estimate parameters for models of cell signaling. Here, we propose a novel, user-friendly machine learning (ML) pipeline that uses sequence or structure to predict the binding affinity parameters of cell signaling models. We may 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 [8], [9] [10]. We note that the

latter study on steady state signaling was tested on only one protein in the system, limiting the generalizability of the results. Excitingly, ours is the first such approach tested on dynamic signaling models. Furthermore, we aim to increase the generalizability of our results by presenting an expanded sample size.

Ultimately, we want to evaluate how much information is gained from adding sequence and structure measurements to our signaling model parameter inference. Indeed, while the idea of expanding the utility large datasets using ML models holds promise, it is still an area of active research. There are only a limited number of studies on how much information is being gained from augmenting measurements traditionally used to calibrate systems biology models with large datasets generalized via ML models. For example, only recently has another study revealed the limited applicability of the ML model used to predict parameters of cellular metabolism models outside of the ML model's training data set [11]. To fill this important gap, we use Bayesian inference to thoroughly quantify how much information is being gained from augmenting our training data with sequence and structure.

Bayesian inference offers an established means to quantify the information gained from data. For example, one particularly relevant study used it to quantify the information gained from absolute measurements in the context of cell signaling models, which typically only leverage relative measurements for parameter inference [12]. Importantly, Bayesian inference characterizes the posterior distribution—that is, the distribution of plausible model parameters given a particular dataset. Inference begins by defining a prior distribution—that is, the distribution of plausible model parameters given prior information. For example, the prior might formalize the fact that physical parameters can only be positive. It then proceeds by incorporating experimental datasets. To quantify the information gained from a particular dataset, one can compare the distribution of plausible parameter values before and after integrating data using statistical distance metrics. One common distance metric used to quantify the difference between distributions is Kullback Leibler (KL) divergence [13]. The more the posterior distribution changes, the larger the KL divergence becomes, indicating that more information is gained from the data. KL divergence has been applied in experimental design and hypothesis generation [14], [15]. Here,

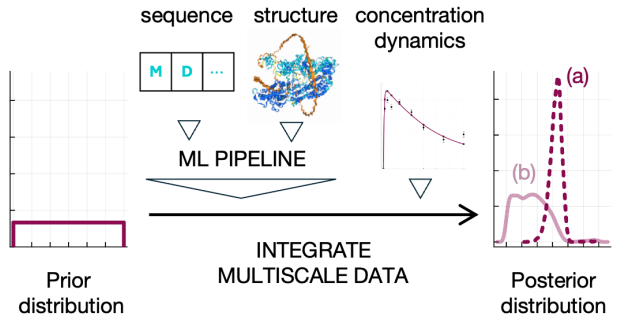

Figure 1. Multiscale Probabilistic Modeling Framework. *Starting with the same prior distribution, we integrate data from different scales into our Bayesian parameter inference using a ML pipeline. We perform inference using two data sets. One dataset comprises only concentration dynamics. The other comprises concentration dynamics as well as amino acid sequence and protein structure. We may then compare the difference between the posterior distribution with, (a), and without, (b), measurements of sequence and structure. Solid, dark pink probability density, prior; dotted, dark pink probability density, posterior given sequence, structure, and concentration dynamics; solid, light pink probability density, posterior given concentration dynamics.*

we use it to quantify the information gained from incorporating experimental observations of sequence or structure into the parameter inference of cell signaling models.

Figure 1 provides a graphical abstract of our investigation. Overall, we are proposing a multiscale, probabilistic modeling framework that enables data integration across scales and quantification of the benefit of this integration. We propose a novel ML 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 out approach on two, well characterized ODE models—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 [16]. In the case of GPCR signaling, 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 [17].

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 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 novel 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 an information theoretic distance to quantify the information gained from augmenting the baseline data. Use of this distance—called KL divergence—is enabled by our probabilistic framework. 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 ML pipeline to calibrate the bimolecular, reversible binding reactions of cell signaling models using amino acid sequence and structure

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

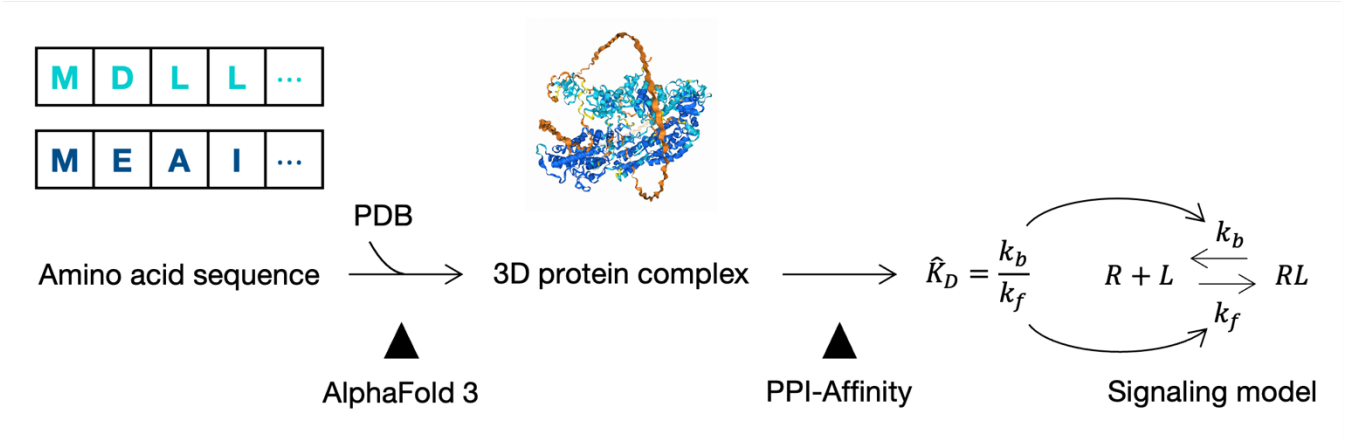

Figure 2. ML pipeline to predict K_D. *K_D is predicted using an SVM regression model, PPI-Affinity. PPI-Affinity needs the complex structure as an input. If the structure exists, it may be extracted from PDB. If it does not exist, the structure may instead be predicted using AlphaFold 3. AlphaFold 3 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.*

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. It is displayed in Figure 2.

Upon reviewing the literature, we found that there exist several ML models that translate the information in an amino acid sequence and/or structure to a scale that is meaningful for cell signaling[18], [19], [20], [21]. Specifically, there are several ML models that predict protein binding affinity, K_D, in units of concentration, from structure or sequence. 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. These ML models are generally limited to bimolecular, reversible binding reactions. However, this is not a limitation in the context of cell signaling models, which are often comprised of bimolecular, reversible binding reactions. This is because even when the product of the reaction contains more than two proteins, the process usually involves a series of bimolecular reactions [22].

We found two ML models with convenient webservers for making K_D predictions: Prodigy and PPI-Affinity [18], [23]. We moved forward with PPI-Affinity, a support vector machine (SVM) regression model, as it shows modest improvements on Prodigy's performance. Further, it performs comparably to more complicated K_D predictors, like those based on deep learning models [18]. PPI-Affinity requires the bound protein complex as an input; however, only a fraction of experimentally available protein structures are complexes—instead, most are monomers. Thus, we sought an alternative input to PPI-Affinity.

Excitingly, the latest iteration of the Nobel-prize winning, deep learning model AlphaFold, AlphaFold 3 (AF3), is designed to predict the structures of biomolecular interactions—including protein complexes [24]. 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. Together, 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. The full pipeline is shown in Figure 2.

2.2 ML pipeline outperforms uninformative prior distribution when predicting K_D

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.

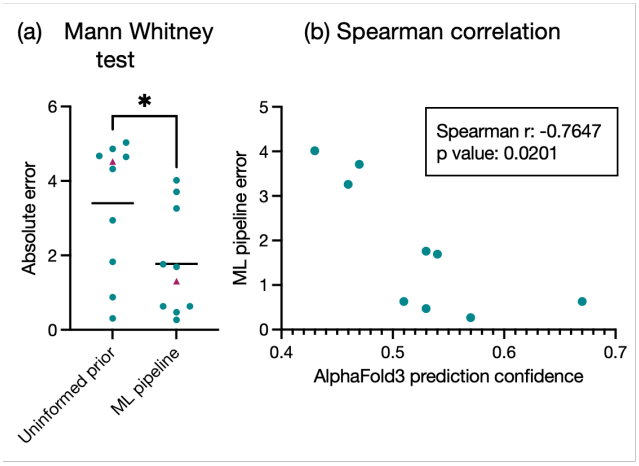

Figure 3. ML Pipeline Performance. *Cyan, EGFR reaction; pink, GPCR reaction. Circles, predicted structure; triangles, experimental structure. (a) Mann-Whitney test comparing absolute error of structure-informed K_D prediction with a random sample from a log-uniform distribution. n=10 binding reactions. *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 for which we used predicted structure, given there was no experimental structure.*

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 [25], [26]. 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 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 5th 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 performance (Supplementary Figure 1a). However, we did find that AlphaFold 3's ranking score was significantly correlated with the pipeline's performance (Figure 3b).

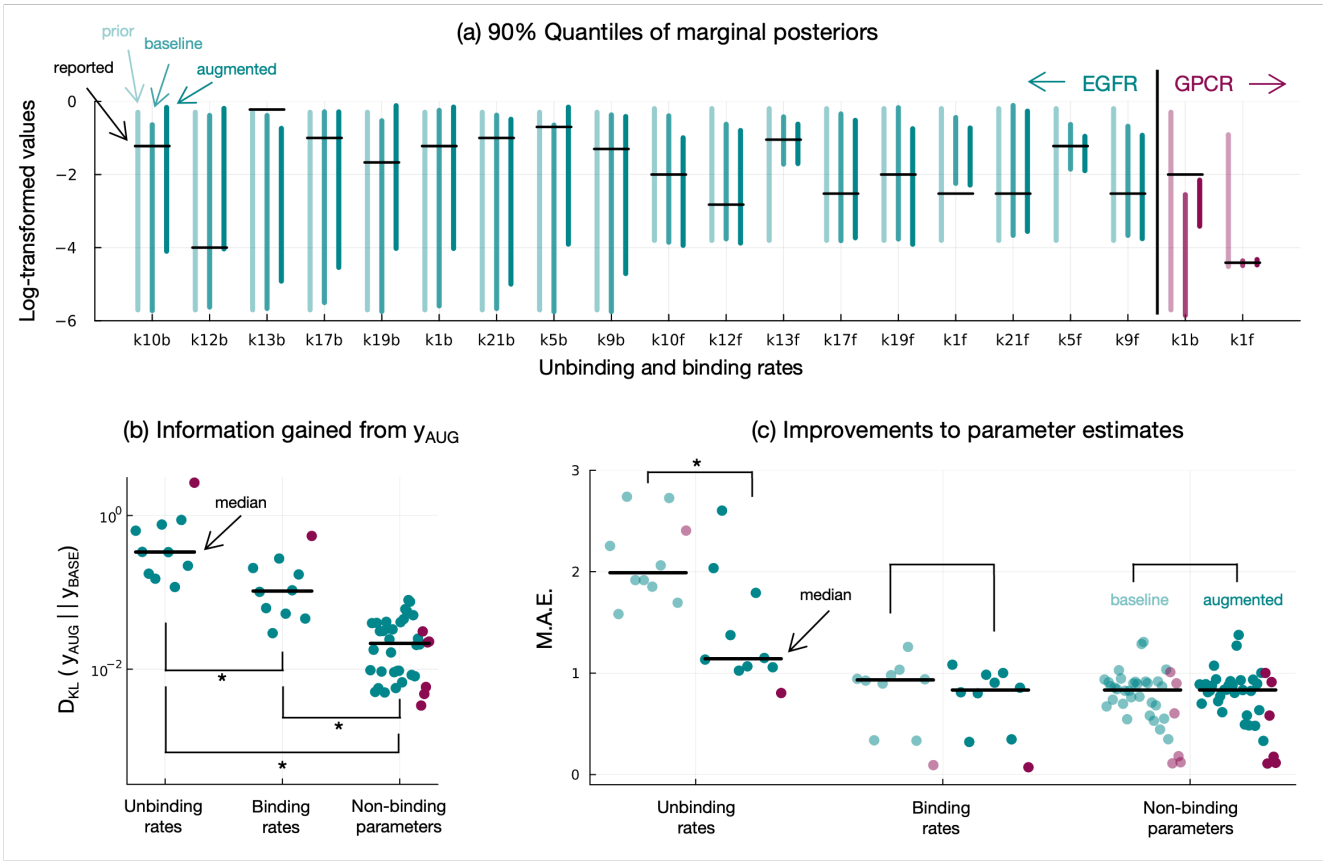

Figure 4. Parameter inference with and without data augmentation. *Cyan, EGFR results; pink, GPCR results. *p-value < 0.05 (a) 90% quantiles of marginal posterior distributions of binding parameters. All samples on log10 scale. Light cyan line, prior; medium cyan line, baseline posterior; dark cyan line, 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.*

2.3 Amino acid sequence and protein structure add information on the rate of protein unbinding

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 Figure 3), 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 y_{BASE} to 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.

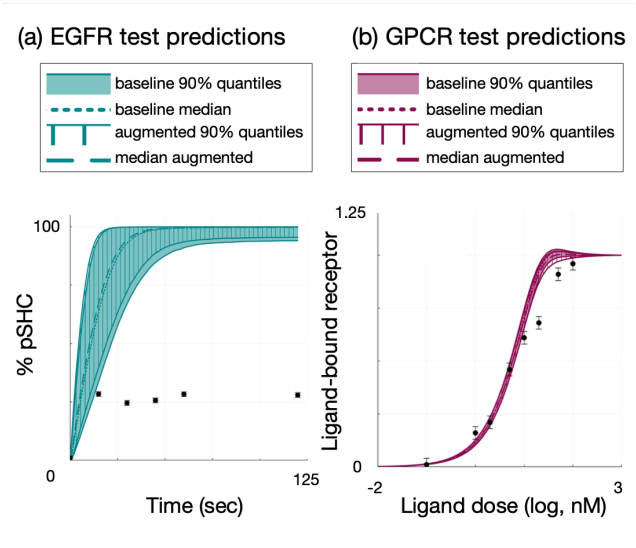

***Figure 5. Performance on test set**. (a) Predictions for EGFR test set, the percent of phosphorylated signaling protein SHC from 0-120 seconds. (b) Predictions for GPCR test set, the amount of ligand bound receptor 60 seconds post-stimulation, at different ligand doses, relative to 1000 nM of ligand. 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 Amino acid sequence and protein structure do not have a significant impact on predictions for test data

Figure 5 shows the difference in the medians and 90% quantiles of the test predictions generated by sampling the augmented posterior compared to the baseline posterior. The test/train split is detailed in the Materials and Methods. While we see some differences at middle timepoints of the EGFR test set (Figure 5a) and at certain doses of the GPCR test set (Figure 5b), the overall predictions are not qualitatively different. Further, in the case of GPCR, at no point does the difference between the medians of the two approaches become greater than the reported experimental error. In other words, the difference in the predictions between the two approaches falls within the error of the test set. In the case of EGFR, the difference between the medians does approach the level of the reported error. However, we note the unusually small scale of this reported error, which is +/- 1% change in phosphorylated SHC protein.

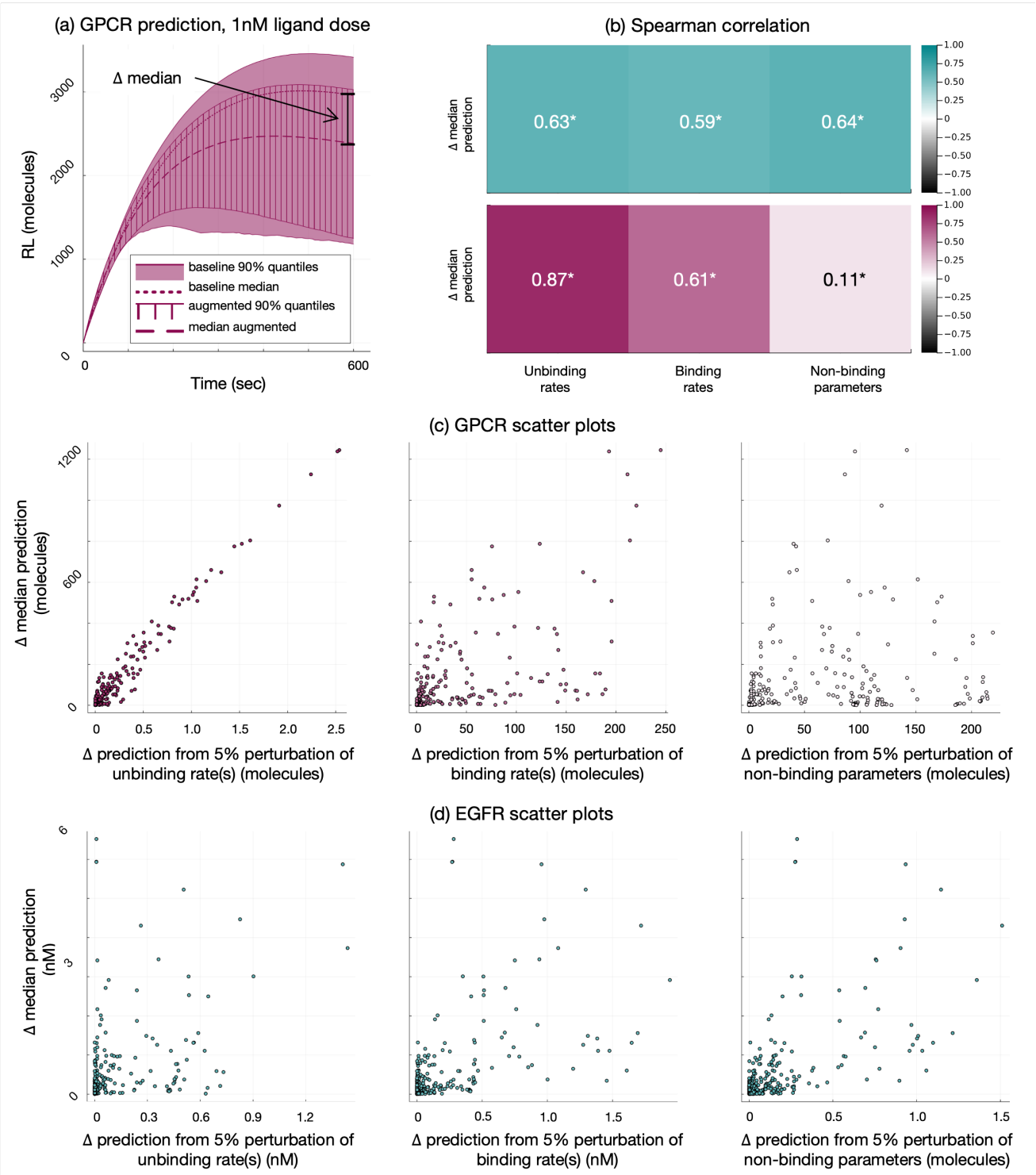

***Figure 6* Local Sensitivity Analysis of GPCR.** *(a) GPCR timeseries prediction of ligand-bound receptor, (RL), after 1 nM ligand stimulation. Units, molecules. Shaded region, 90% quantiles of baseline approach; patterned region, 90% quantiles of augmented approach; dotted line, median prediction 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.*

231

232 ***2.5 Impact of amino acid sequence and protein structure on cell signaling model predictions is a***
233 ***function of the model's local sensitivity to changes in rates of protein unbinding***

234 When we investigate the medians and 90% quantiles for other outputs besides the test set, we observe
235 larger differences in the predictions. For example, if we consider the timeseries of receptor-bound ligand,
236 RL, of GPCR at times other than 60 seconds post-stimulation, we see greater differences between both
237 the quantiles and medians (Figure 6a). Similarly, several species comprising EGFR signaling show larger
238 differences in the predicted median and 90 percent quantiles. (Supplementary Figure 2).

239 To investigate what leads to changes in some, but not all, ODE model predictions, we conducted a local
240 perturbation analysis of both the GPCR and EGFR outputs to model parameters. We perturbed the
241 maximum likelihood parameter values of the baseline case by 5%, and used the local sensitivity at the
242 maximum likelihood to calculate the projected change in an output from this perturbation. We analyzed
243 360 outputs for GPCR and 345 outputs for EGFR. These outputs correspond to five model species at
244 eight timepoints and nine doses and 23 model species at five timepoints and three doses, respectively.
245 Doses and timepoints match experimental conditions, though we exclude timepoint zero. This is because
246 the sensitivity of the initial concentrations to any parameter at time zero is zero. More details about the
247 local sensitivity analysis may be found in section the Materials and Methods.

248 Next, we calculated the Spearman correlation between the change in the median prediction of an output
249 and the sensitivity of that output to perturbations of either unbinding rates, binding rates, or non-binding
250 parameters. The resulting Spearman coefficients for both EGFR and GPCR are displayed in Figure 6b.
251 Both test cases display a positive and significant correlation with the sensitivity to perturbation of
252 unbinding rates. In the case of GPCR, the correlation decreases from unbinding to binding to non-binding
253 parameters. This decreasing correlation was clear visually (Figure 6c). Interestingly, this is the case
254 despite the sensitivity to perturbations being around an order of magnitude lower for unbinding rates than
255 the other parameter types. In the case of EGFR, there is a significant positive correlation for all parameter
256 types.

257 ***2.6 Results are consistent if confidence of protein structure prediction is considered when***
258 ***integrating data on amino acid sequence***

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

However, considering the confidence of the protein structure prediction did not significantly change the results of our parameter inference with and without data augmentation (Supplementary Figure 4). 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 (Supplementary Figure 5). As in Figure 5, the predictions with and without data augmentation were qualitatively 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 Results are consistent, independent of whether the prior is more informative or less informative

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 Figures 6,7,8, and 9). 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 Figure 7a). In the case of the less informative prior, the information gained from data augmentation was no longer significantly different between the unbinding and binding rates (Supplementary Figure 8b). This was because more information was gained for binding rates. Interestingly, this did not result in a corresponding significant change in the mean absolute error for binding rates.

3. Discussion

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

294 We use amino acid sequence and protein structure measurements from UniProt and PDB, respectively,
295 to augment the published datasets of two test cases—EGFR and GPCR signaling. The published
296 datasets include partial, relative measurements of protein concentration and protein co-localization over
297 time. We chose EGFR model and GPCR model as our test cases because both are well-established and
298 biologically significant. Taken together, these models offer a comprehensive test of our approach. For
299 example, EGFR provides a test of a moderately sized model, while GPCR provides a test of a minimal
300 model. In addition, our ML pipeline can make K_D predictions using sequence or structure. EGFR provides
301 a test of predictions using sequence, while GPCR provides a test of predictions using structure.

302 The GPCR and EGFR models' predictions occur on the scale of protein concentrations. To make
303 measurements of sequence and structure applicable to the parameter estimation of these models, we
304 needed to convert them to the same scale as the GPCR and EGFR models'. Excitingly, we were able to
305 integrate two, user-friendly ML models—PPI-Affinity and AlphaFold 3—into one pipeline to do this [18],
306 [24]. This pipeline uses sequence or structure to predict the binding affinity, K_D, of a given bimolecular,
307 reversible reaction in a cell signaling model. We find that this ML pipeline outperforms an uninformative
308 prior when predicting the K_D. Thus, it is appropriate to use when no experimental measurements exist
309 for K_D. Importantly, this pipeline is generalizable to other dynamic signaling models. It is limited only by
310 access to amino acid sequence and comprises two, user-friendly web servers.

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

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

325 Overall, augmenting our data with data on sequence or structure provides more information on rates of
326 protein unbinding compared to using only the original data. This was consistent across both the EGFR

327 test case and GPCR test case. Further, the information gained from sequence or structure tends to push
328 the unbinding rate towards the rates reported in the original publications. These reported rates were all
329 taken from experiments, rather than fit to data [25], [26]. Constraining unbinding rates with our approach
330 can be useful, as unbinding rates tend to vary over a larger range than binding rates [22]. These results
331 offer an understanding of when our approach is useful: in the case of underdetermined rates of protein
332 unbinding. We have confidence in this recommendation of utility, since the results were consistent even
333 with a more or less informative prior distribution.

334 When we investigate how predictions change due to the data augmentation, the results are less clear.
335 We first find that augmenting the baseline data set does not fundamentally change the predictions of our
336 test data for either test case. We do see more fundamental changes for other outputs besides the test
337 data. We thus investigated when predictions inferred with y_{BASE} will differ from predictions inferred with
338 y_{AUG}. Our approach changes unbinding parameter rates. Intuitively, one would then expect to see
339 differences in predictions for outputs influenced by unbinding rates. We thus investigated the relationship
340 between the local sensitivity of an output to unbinding rates, binding rates, and non-binding parameters.
341 When we analyzed the GPCR model, there emerged a striking, significant, positive and linear correlation
342 between changes in a prediction and the local sensitivity of that prediction to perturbations in the
343 unbinding rate. A significant positive correlation also emerged with the EGFR model. However, this
344 relationship was similar across unbinding rates, binding rates, and nonbinding parameters. These results
345 may be a function of the complexity of the models. While GPCR is a minimal model comprised of eight
346 parameters, EGFR is a more moderately sized model comprised of 50 parameters. It is thus plausible
347 that the relationship between sensitivity and predictions is impacted by higher-order interactions between
348 parameters in the case of EGFR [27]. These interactions would not be captured by the local sensitivity
349 analysis we conduct here.

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

358 We acknowledge some limitations of our study. We test our ML pipeline on only the 10 binding reactions
359 reported in our test cases. We could test our pipeline with more reactions from databases of paired

structures and affinities, like PDBbind [28]. We chose to focus only on binding affinities reported in the context of ODE models as these affinities are known to recapitulate intracellular protein dynamics. Future work could investigate how our ML pipeline could be even more specialized to binding reactions occurring inside the cell. For example, future work could take in to account the crowded intra-cellular environment [29]. With future improvements to the ML pipeline in mind, we developed a Bayesian framework for incorporating $\widehat{K}_D$ that can evolve with our pipeline. For example, we use the error reported by PPI-Affinity to weigh the predicted K_D likelihood against the likelihood of the conventional experimental measurements. This weight can be adjusted as that error improves.

There are also ML models that use amino acid sequence to predict binding affinity. In theory, we could use one of these models instead of our pipeline. However, in practice, we found that there were no such models hosted on convenient webservers. Even more limiting was the prohibitively large RAM these models needed to make a prediction for the relatively large signaling protein complexes we are interested in here [19]. We instead focus on more user-friendly models—both in terms of RAM and implementation—for our pipeline.

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

4. Materials and Methods

4.1 ODE models of intracellular signaling

We base our investigation on two ODE models: a 50 parameter, 23 species model of EGFR signaling and an eight parameter, seven species model of GPCR signaling [25], [26]. In the case of EGFR signaling, a mixture of mass action and Michaelis Menten kinetics are used to describe cell signaling events such as protein binding and phosphorylation. In the case of GPCR signaling, mass action kinetics are used to describe cell signaling events such as protein binding, activation, and degradation. Binding reactions are of particular interest to our study: in the case of EGFR, 18 parameters characterize nine binding reactions, while in the case of GPCR, two 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* [30].

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 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 [31], [32]. 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 [31], [32]. 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 [31], [33]. Based on our requirements, we choose to use *DifferentialEquation.jl*'s QNDF solver, which is equivalent to Matlab's ode15s solver [33]. It is a stiff solver that can handle medium tolerances; thus, it meets our problem's requirements [31], [33], [34].

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 intracellular signaling dynamics (y_{BASE})

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 **7**a) and in

the context of the GPCR model, y_{BASE} are measurements of GPCR signaling dynamics (Figure **7**b). We extracted the data and errors with Plot Digitizer [35] .

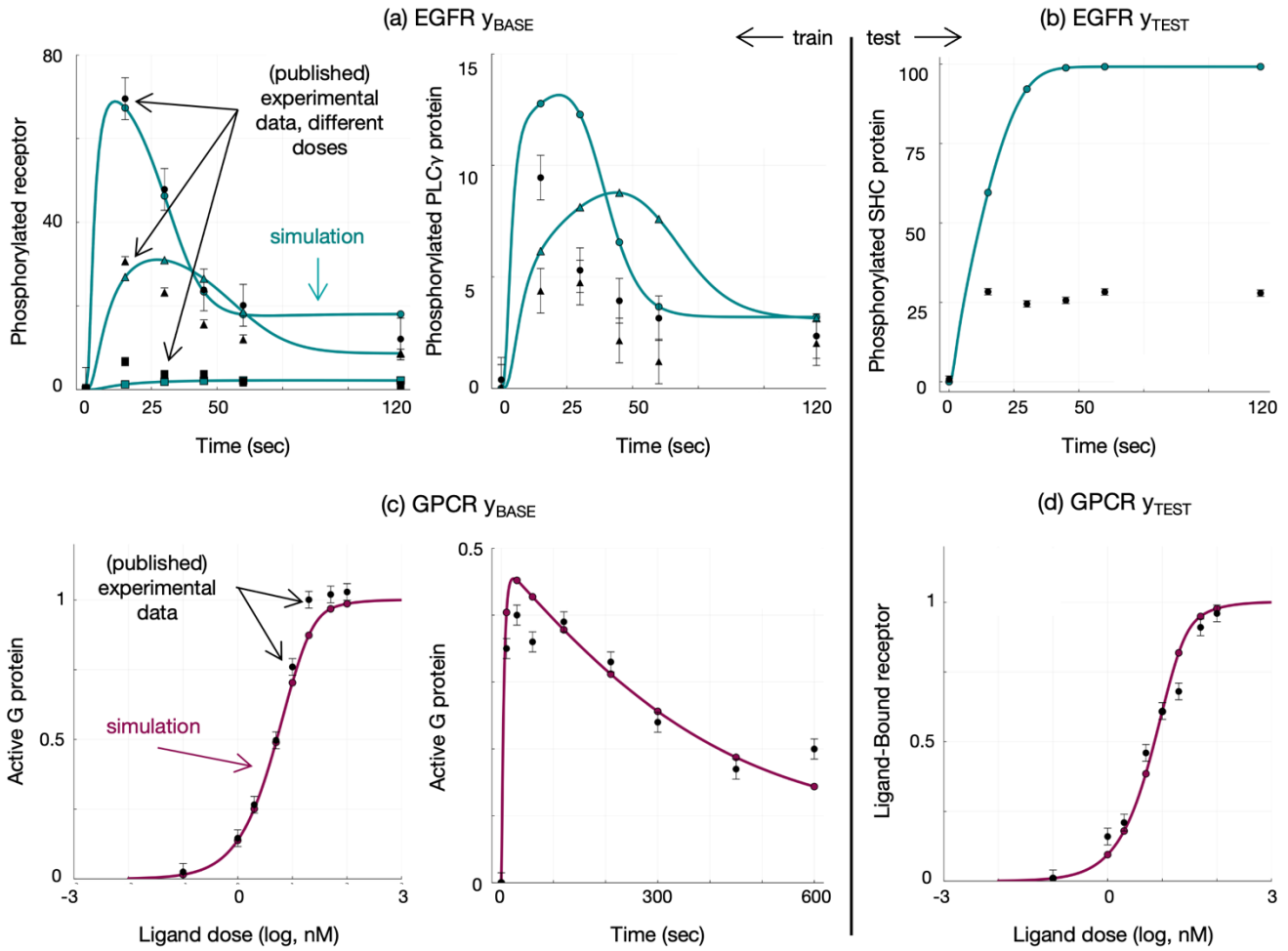

Figure 7. Baseline data set (y_{BASE}). *(a) Training data for EGFR parameter inference. (b) Test data for EGFR parameter inference. (c) Training data for GPCR parameter inference. (d) Test data for GPCR parameter inference. Cyan or pink lines, simulations using reported parameter values for EGFR and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

Eight outputs of EGFR are observed, and we show three of these outputs as a representative sample in Figure **7**a. 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 dosages. Further observations include the percent of an intracellular protein bound with another intracellular protein and the receptor. The original work does not fit parameters to data, so we must decide our test/train data split. We choose to reserve measurements of one of the intracellular proteins, SHC, as our test data. We choose SHC due to its clinical relevance, as this signaling species has been shown as an important marker in various cancers [36]. Furthermore, the published model predictions do not quantitatively match the experimental measurements of SHC [36]. Ideally, incorporating data on protein structure will improve

our predictions for this experimentally relevant target. Thus, we use the data for SHC to test our model predictions.

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

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

4.4 Experimental measurements from UniProt (y_{FASTA}) and PDB Databases (y_{PDB})

We augment y_{BASE} with amino acid sequences or structures of the proteins that are involved in binding for each system. Given two proteins that are binding, we first searched for any structures of the bound complex in PDB. One such structure exists, describing GPCR bound to alpha pheromone, PDB structure 7QBI [37], [38]. This structure corresponds to the single binding reaction in the GPCR model and is formatted as a PDB file. This measurement comprises y_{PDB}.

If no such structure was found, as was the case for the 9 EGFR binding reactions, we extracted the canonical amino acid sequences of the individual proteins, as FASTA files, from UniProt. They are extracted UniProt sequences P01133 (EGF), P29353 (SHC1), P19174 (PLCG1), Q07889 (SOS1), P62995 (GRB2), and P00533 (EGFR). These measurements comprise y_{FASTA}.

4.5 ML pipeline to predict K_D

Figure 2 provides an overview of the ML pipeline used to predict K_D using amino acid sequence (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 [18]. 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 [39], [40]. 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 [20]. 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 [24]. Both the ranking score and ipTM are significantly correlated with the performance of the ML pipeline (Figure 3 and Supplementary Figure 1b). 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 [18]. 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. MCMC approximations are particularly powerful, 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* [41]. 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 [42], [43]. Further

499 discussion of the convergence diagnostics of the samples to the posterior may be found in the
500 Supplementary Information.

501 **4.7 Prior**

502 The prior encodes our assumptions, ideally derived from empirical knowledge, about parameters. Here,
503 we choose to define independent, log-uniform priors. If the parameter is a binding rate, we limit the
504 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}].
505 This is in line with prior distributions for binding rates in the literature and database values[12], [44]. For
506 non-binding parameters, we limit the distributions to +/- two orders of magnitude above and below the
507 reported parameter value.

508 **4.8 Likelihood**

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

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

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

524 For parts of the investigation, we augment this likelihood term to account for AlphaFold3's confidence.
525 Specifically, we weigh the log-likelihood of a given $\widehat{K}_D^k$ by AlphaFold 3's interface predicted template
526 modeling (ipTM) score, which represents the confidence of $\widehat{K}_D^k$ scaled from zero to one. This likelihood
527 formulation is known as a power prior. Power priors consist of raising the likelihood, or a term in the
528 likelihood, to a discounting factor. This controls the amount of information taken from certain data [45],
529 [46]. When the log likelihood is evaluated, the product of likelihoods becomes a sum, and the discount

factor(s) become weight(s) on the terms of this sum. Further details on our likelihood distribution are included in the Supplementary Information.

4.9 Log Transformation of Parameters for Analysis

We note that analyses of the model parameters, and therefore posteriors, are done after a log10-transformation of the parameter samples. The reason for this is three fold—first, log-transformation tends to bring biochemical parameters like K_D towards a Normal distribution, which is desirable for most statistical analyses [42]. Second, by comparing order-of-magnitude changes in parameters, we can compare parameters of different scales. Finally, we are concerned with order-of-magnitude changes in K_D, as it has been shown that changes less than this do not significantly influence model predictions [47].

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 [48], [49], [50]. 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 [48]. 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 Analysis

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/hhuber2512/structure_informed_cell_signaling

6. Acknowledgements

The authors acknowledge members of the Finley research group for constructive feedback. This work was partially supported by the National Science Foundation (Award 2410753) to S.D.F.

6. References

[1] B. L. Puniya, M. Verma, C. Damiani, S. Bakr, and A. Dräger, "Perspectives on computational modeling of biological systems and the significance of the SysMod community," *Bioinforma. Adv.*, vol. 4, no. 1, p. vbae090, Jan. 2024, doi: 10.1093/bioadv/vbae090.

[2] B. P. Ingalls, *Mathematical modeling in systems biology: an introduction*. Cambridge, Massachusetts: MIT Press, 2013.

[3] N. Nandagopal, L. A. Santat, L. LeBon, D. Sprinzak, M. E. Bronner, and M. B. Elowitz, "Dynamic Ligand Discrimination in the Notch Signaling Pathway," *Cell*, vol. 172, no. 4, pp. 869-880.e19, Feb. 2018, doi: 10.1016/j.cell.2018.01.002.

[4] J. A. Rohrs, E. L. Siegler, P. Wang, and S. D. Finley, "ERK Activation in CAR T Cells Is Amplified by CD28-Mediated Increase in CD3ζ Phosphorylation," *iScience*, vol. 23, no. 4, Apr. 2020, doi: 10.1016/j.isci.2020.101023.

[5] A. L. MacLean, Z. Rosen, H. M. Byrne, and H. A. Harrington, "Parameter-free methods distinguish Wnt pathway models and guide design of experiments," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 9, pp. 2652–2657, Mar. 2015, doi: 10.1073/pnas.1416655112.

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

[7] H. M. Berman *et al.*, "The Protein Data Bank," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, Jan. 2000, doi: 10.1093/nar/28.1.235.

[8] H. Yu, H. Deng, J. He, J. D. Keasling, and X. Luo, "UniKP: a unified framework for the prediction of enzyme kinetic parameters," *Nat. Commun.*, vol. 14, no. 1, p. 8211, Dec. 2023, doi: 10.1038/s41467-023-44113-1.

[9] F. Li *et al.*, "Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction," *Nat. Catal.*, vol. 5, no. 8, pp. 662–672, Aug. 2022, doi: 10.1038/s41929-022-00798-z.

[10] V. D. Zaballa and E. E. Hui, “Systems-Structure-Based Drug Design,” Oct. 14, 2024, *arXiv*: arXiv:2410.10108. doi: 10.48550/arXiv.2410.10108.

[11] A. Kroll and M. J. Lercher, “DLKcat cannot predict meaningful kcat values for mutants and unfamiliar enzymes,” *Biol. Methods Protoc.*, vol. 9, no. 1, p. bpae061, Jan. 2024, doi: 10.1093/biomethods/bpae061.

[12] B. Thijssen, T. M. H. Dijkstra, T. Heskes, and L. F. A. Wessels, “Bayesian data integration for quantifying the contribution of diverse measurements to parameter estimates,” *Bioinforma. Oxf. Engl.*, vol. 34, no. 5, pp. 803–811, Mar. 2018, doi: 10.1093/bioinformatics/btx666.

[13] S. Kullback and R. A. Leibler, “On Information and Sufficiency,” *Ann. Math. Stat.*, vol. 22, no. 1, pp. 79–86, Mar. 1951, doi: 10.1214/aoms/1177729694.

[14] L. M. Lomeli *et al.*, “Optimal experimental design for mathematical models of haematopoiesis,” *J. R. Soc. Interface*, vol. 18, no. 174, p. 20200729, Jan. 2021, doi: 10.1098/rsif.2020.0729.

[15] H. A. Huber, S. K. Georgia, and S. D. Finley, “Systematic Bayesian Posterior Analysis Guided by Kullback-Leibler Divergence Facilitates Hypothesis Formation,” *J. Theor. Biol.*, p. 111341, Nov. 2022, doi: 10.1016/j.jtbi.2022.111341.

[16] M. Glont, C. Arankalle, K. Tiwari, T. V. N. Nguyen, H. Hermjakob, and R. S. Malik-Sheriff, “BioModels Parameters: a treasure trove of parameter values from published systems biology models,” *Bioinformatics*, vol. 36, no. 17, pp. 4649–4654, Nov. 2020, doi: 10.1093/bioinformatics/btaa560.

[17] D. S. Eiger, U. Pham, J. Gardner, C. Hicks, and S. Rajagopal, “GPCR systems pharmacology: a different perspective on the development of biased therapeutics,” *Am. J. Physiol. - Cell Physiol.*, vol. 322, no. 5, pp. C887–C895, May 2022, doi: 10.1152/ajpcell.00449.2021.

[18] S. Romero-Molina *et al.*, “PPI-Affinity: A Web Tool for the Prediction and Optimization of Protein–Peptide and Protein–Protein Binding Affinity,” *J. Proteome Res.*, vol. 21, no. 8, pp. 1829–1841, Aug. 2022, doi: 10.1021/acs.jproteome.2c00020.

[19] F. Zheng, X. Jiang, Y. Wen, Y. Yang, and M. Li, “Systematic investigation of machine learning on limited data: A study on predicting protein-protein binding strength,” *Comput. Struct. Biotechnol. J.*, vol. 23, pp. 460–472, Dec. 2024, doi: 10.1016/j.csbj.2023.12.018.

[20] Z. Zhou *et al.*, “ProAffinity-GNN: A Novel Approach to Structure-Based Protein–Protein Binding Affinity Prediction via a Curated Data Set and Graph Neural Networks,” *J. Chem. Inf. Model.*, vol. 64, no. 23, pp. 8796–8808, Dec. 2024, doi: 10.1021/acs.jcim.4c01850.

[21] A. Vangone and A. M. Bonvin, “Contacts-based prediction of binding affinity in protein–protein complexes,” *eLife*, vol. 4, p. e07454, Jul. 2015, doi: 10.7554/eLife.07454.

[22] T. D. Pollard, “A Guide to Simple and Informative Binding Assays,” *Mol. Biol. Cell*, vol. 21, no. 23, pp. 4061–4067, Dec. 2010, doi: 10.1091/mbc.e10-08-0683.

[23] L. C. Xue, J. P. Rodrigues, P. L. Kastritis, A. M. Bonvin, and A. Vangone, “PRODIGY: a web server for predicting the binding affinity of protein–protein complexes,” *Bioinformatics*, vol. 32, no. 23, pp. 3676–3678, Dec. 2016, doi: 10.1093/bioinformatics/btw514.

[24] J. Abramson *et al.*, “Accurate structure prediction of biomolecular interactions with AlphaFold 3,” *Nature*, May 2024, doi: 10.1038/s41586-024-07487-w.

[25] B. N. Kholodenko, O. V. Demin, G. Moehren, and J. B. Hoek, “Quantification of Short Term Signaling by the Epidermal Growth Factor Receptor,” *J. Biol. Chem.*, vol. 274, no. 42, pp. 30169–30181, Oct. 1999, doi: 10.1074/jbc.274.42.30169.

[26] T.-M. Yi, H. Kitano, and M. I. Simon, “A quantitative characterization of the yeast heterotrimeric G protein cycle,” *Proc. Natl. Acad. Sci.*, vol. 100, no. 19, pp. 10764–10769, Sep. 2003, doi: 10.1073/pnas.1834247100.

[27] G. M. Monsalve-Bravo *et al.*, “Analysis of sloppiness in model simulations: Unveiling parameter uncertainty when mathematical models are fitted to data,” *Sci. Adv.*, vol. 8, no. 38, p. eabm5952, Sep. 2022, doi: 10.1126/sciadv.abm5952.

[28] Z. Liu *et al.*, “PDB-wide collection of binding data: current status of the PDBbind database,” *Bioinformatics*, vol. 31, no. 3, pp. 405–412, Feb. 2015, doi: 10.1093/bioinformatics/btu626.

[29] G. Grassmann *et al.*, “Computational Approaches to Predict Protein–Protein Interactions in Crowded Cellular Environments,” *Chem. Rev.*, vol. 124, no. 7, pp. 3932–3977, Apr. 2024, doi: 10.1021/acs.chemrev.3c00550.

[30] C. Rackauckas and Q. Nie, “DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia,” *J. Open Res. Softw.*, vol. 5, no. 1, May 2017, doi: 10.5334/jors.151.

[31] L. F. Shampine, I. Gladwell, and S. Thompson, *Solving ODEs with MATLAB*. Cambridge ; New York: Cambridge University Press, 2003.

[32] K. E. Brenan, S. L. Campbell, and L. R. Petzold, *Numerical solution of initial-value problems in differential-algebraic equations*. in Classics in applied mathematics, no. 14. Philadelphia: Society for Industrial and Applied Mathematics, 1996.

[33] C. Rackauckas and Q. Nie, “DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia,” *J. Open Res. Softw.*, vol. 5, no. 1, p. 15, May 2017, doi: 10.5334/jors.151.

[34] The MathWorks, Inc, “Choose an ODE Solver.” The MathWorks Inc., 2023. [Online]. Available: https://www.mathworks.com/help/matlab/math/choose-an-ode-solver.html

[35] “PlotDigitizer: Version 3.1.6.” 2025. [Online]. Available: https://plotdigitizer.com

[36] Z. Guo *et al.*, “SHC1 serves as a prognostic and immunological biomarker in clear cell renal cell carcinoma: a comprehensive bioinformatics and experimental analysis,” *Sci. Rep.*, vol. 14, no. 1, p. 20150, Aug. 2024, doi: 10.1038/s41598-024-70897-3.

[37] V. Velazhahan, N. Ma, N. Vaidehi, and C. G. Tate, “Activation mechanism of the class D fungal GPCR dimer Ste2,” *Nature*, vol. 603, no. 7902, pp. 743–748, Mar. 2022, doi: 10.1038/s41586-022-04498-3.

[38] V. Velazhahan and C. G. Tate, “Structure of the GPCR dimer Ste2 in the active-like state bound to agonist.” Nov. 19, 2021.

[39] P. L. Kastritis *et al.*, “A structure-based benchmark for protein–protein binding affinity,” *Protein Sci.*, vol. 20, no. 3, pp. 482–491, Mar. 2011, doi: 10.1002/pro.580.

[40] Y. B. Ruiz-Blanco, W. Paz, J. Green, and Y. Marrero-Ponce, “ProtDCal: A program to compute general-purpose-numerical descriptors for sequences and 3D-structures of proteins,” *BMC Bioinformatics*, vol. 16, no. 1, p. 162, May 2015, doi: 10.1186/s12859-015-0586-0.

[41] J. Goodman and J. Weare, “Ensemble samplers with affine invariance,” *Commun. Appl. Math. Comput. Sci.*, vol. 5, no. 1, pp. 65–80, Jan. 2010, doi: 10.2140/camcos.2010.5.65.

[42] A. Bar-Even *et al.*, “The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters,” *Biochemistry*, vol. 50, no. 21, pp. 4402–4410, May 2011, doi: 10.1021/bi2002289.

[43] T. Lubitz and W. Liebermeister, “Parameter balancing: consistent parameter sets for kinetic metabolic models,” *Bioinformatics*, vol. 35, no. 19, pp. 3857–3858, Oct. 2019, doi: 10.1093/bioinformatics/btz129.

[44] N. Tötsch and D. Hoffmann, “Bayesian Data Integration Questions Classic Study on Protease Self-Digest Kinetics,” *ACS Omega*, vol. 5, no. 25, pp. 15162–15168, Jun. 2020, doi: 10.1021/acsomega.0c01109.

[45] L. M. Carvalho and J. G. Ibrahim, “On the normalized power prior,” *Stat. Med.*, vol. 40, no. 24, pp. 5251–5275, 2021, doi: 10.1002/sim.9124.

[46] J. G. Ibrahim, M.-H. Chen, Y. Gwon, and F. Chen, “The Power Prior: Theory and Applications,” *Stat. Med.*, vol. 34, no. 28, pp. 3724–3749, Dec. 2015, doi: 10.1002/sim.6728.

[47] A. R. Soltis and J. J. Saucerman, “Robustness portraits of diverse biological networks conserved despite order-of-magnitude parameter uncertainty,” *Bioinformatics*, vol. 27, no. 20, pp. 2888–2894, Oct. 2011, doi: 10.1093/bioinformatics/btr496.

[48] Y. Ma, V. Dixit, M. J. Innes, X. Guo, and C. Rackauckas, “A Comparison of Automatic Differentiation and Continuous Sensitivity Analysis for Derivatives of Differential Equation Solutions,” in *2021 IEEE High Performance Extreme Computing Conference (HPEC)*, Waltham, MA, USA: IEEE, Sep. 2021, pp. 1–9. doi: 10.1109/HPEC49654.2021.9622796.

[49] J. Revels, M. Lubin, and T. Papamarkou, “Forward-Mode Automatic Differentiation in Julia,” Jul. 26, 2016, *arXiv*: arXiv:1607.07892. doi: 10.48550/arXiv.1607.07892.

[50] V. K. Dixit and C. Rackauckas, “GlobalSensitivity.jl: Performant and Parallel Global Sensitivity Analysis with Julia,” *J. Open Source Softw.*, vol. 7, no. 76, p. 4561, Aug. 2022, doi: 10.21105/joss.04561.