Flexible fitting of PROTAC concentration-response curves with Gaussian Processes

Elizaveta Semenova[1], Maria-Luisa Guerriero[1], Bairu Zhang[1], Andreas Hock[2], Philip Hopcroft[2], Ganesh Kadamur[2], Avid M. Afzal[1], and Stanley E. Lazic[3]

[1]Data Sciences and Quantitative Biology, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK
[2]Mechanistic Biology and Profiling, Discovery Sciences, R&D, AstraZeneca, Cambridge UK
[3]Prioris.ai Inc, Ottawa, Canada

Abstract

A proteolysis targeting chimera (PROTAC) is a new technology that marks proteins for degradation in a highly specific manner. During screening, PROTAC compounds are tested in concentration-response (CR) assays to determine their potency, and parameters such as the half-maximal degradation concentration (DC_{50}) are estimated from the fitted CR curves. These parameters are used to rank compounds, with lower DC_{50} values indicating greater potency. However, PROTAC data often exhibit bi-phasic and poly-phasic relationships, making standard sigmoidal CR models inappropriate. We therefore used a Bayesian Gaussian Processes (GP) model that can flexibly fit both non-sigmoidal and sigmoidal CR curves without user input. Parameters, such as the DC_{50}, the maximum effect $D_{\max}$, and the point of departure (PoD) are estimated from the fitted curves. We then rank compounds based on one or more parameters, and propagate the parameter uncertainty into the rankings, enabling us to confidently state if one compound is better than another. Hence, we used a flexible and automated procedure for PROTAC screening experiments. The code and data are provided on Github (https://github.com/elizavetasemenova/gp_concentration_response).

Keywords: concentration-response curve, PROTAC, Gaussian Process, changepoint kernel, compound ranking, Bayesian

Introduction

A proteolysis targeting chimera (PROTAC) is a new drug modality that induces degradation of disease-causing proteins. It is a molecule composed of two active domains and a linker and forms complex consisting of the target,

an E3 ligase, and the PROTAC molecule. PROTAC molecules ubiquitinate their target proteins, thereby tagging them for degradation by the proteasome. Factors to consider when designing PROTACs include: which E3 ligase to target, the required binding affinities between the target protein and E3 ligase, the target protein basal turnover rate, and the E3 ligase expression level in the relevant tissue. Small molecules typically inhibit protein activity, but PROTACs completely remove the protein from cells. As the PROTAC molecule itself is not consumed during proteasomal degradation, it can be recycled and used many times. Catalytic knockdown of proteins in vivo has been shown in some studies [1]. Precisely estimating parameters of CR curves, such as potency and efficacy is essential for robust drug development.

Traditional small molecule screening for inhibitors typically yield sigmoidal monotonic concentration response curves, characterised by a plateau at high drug concentrations. In contrast, PROTACs CR curves can have a distinct profile, marked by a loss of efficacy at higher doses. Accounting for such a "hook effect" in data analysis is essential to understand the underlying mechanism. A mathematical framework has been developed to understand the ternary complex equilibria [2]: it explains that the concentration-response behaviour of such a complex should form a hook. The theoretical model predicts a symmetric curve with respect to the point of the maximal effect. However, experimental data frequently displays non-symmetric patterns. There is therefore an unmet need for an automated method to fit models to experimental data allowing for non-sigmoidal shapes.

The challenge can be addressed in several ways - either by adjusting parametric models, or by non-parametric modelling. Parametric models are widely used in drug discovery as a tool to fit CR curves. They have a pre-defined functional form, such as a linear, exponential decay [3], or Hill model [4]. The parameters governing these models often have a straightforward interpretation, such as the slope and intercept for a linear model, or the concentration that gives the half-maximal degradation (DC_{50}) and the maximal effect ($D_{\max}$) for a sigmoidal model. Parametric models reduce an unknown and potentially complicated true function, to a simple form with few parameters. Such models can produce consistent results when observations follow a pre-defined class of shapes but may not fit the data well when a pattern exhibited by data does not fall into that class. Then the simplicity of a parametric model might lead to erroneous estimates. Another price to pay for the simplicity of parametric models is their global behaviour, i.e. a change in one parameter can change the shape of the whole curve.

Non-parametric models such as LOESS (LOcally Estimated Scatterplot Smoothing), splines, and Gaussian processes (GPs) are more flexible and learn the shape of a curve from data. However, the parameters of these non-parametric models are often not directly interpretable. In addition, they are usually more computationally intensive and harder to implement.

GPs can incorporate interpretable parameters, such as points of departure (PoD), i.e. the concentration at which the curve crosses the detection band [5]. In addition, some parts of a curve can be fixed while other modelled more flexibility. Parametric framework offers two approaches to fitting when the shape of a curve is not known in advance: model averaging [6] and model selection. Both approaches start with a set of parametric models and then undergo two stages. At the first stage, model fits to the data are obtained. At the second stage, either models are averaged or one model is selected. Both methods would not work well if the initial set does not contain a realistic model for the given data. Non-parametric approaches can avoid the two-stage process, while the necessary decision-influencing parameters can still be derived from the fitted curves.

The aim of this work is to demonstrate the usefulness of a Bayesian GP model for fitting PROTAC CR curves. This approach can quantify uncertainty and infer PoD. The composite GP kernel captures the biology such as a flat (negligeable) effect at concentrations up to the PoD, and more flexibility at higher concentrations to capture the behaviour of interest. An advantage of a Bayesian framework is that the model can account for several sources of uncertainty and represents the parameter uncertainty in the form of a distribution, which is propagated to the uncertainty of the compound ranks.

Materials and methods

Data

The data was collected in a set of 2 experiments for 9 compounds as follows: 384 well plates (PE cell carrier ultra, PerkinElmer, 6057308) containing 12,000 cells per well in RPMI1640 medium (Sigma, R7638) supplemented with glutamine (Gibco, 25030) and 10% FCS were dosed with an Echo 555 (labcyte) at the indicated doses. Maximum effect and neutral controls DMSO were also used. After one day of incubation, cells were fixed with PFA (4% final concentration) for approx. 20 minutes, washed (3X PBS by using a BioTek plate washer) and subsequently stained with primary followed by secondary antibody for target detection and DRAQ5 (Abcam, Ab108410). Each staining step was followed by a washing protocol described above. The signal for the target antibody stain per well was measured by automated microscopy (Cellinsight, Thermo). Controls distributed at several positions across the plate were normalised to 0% (DMSO, neutral control) and -100% (maximal effect compound). Data were normalised by Genedata Screener software (Genedata Screener®).

Model

Bayesian inference

The Bayesian model formulation consists of two parts: preliminary information, expressed as prior distributions, and a likelihood. Inference is made by updating the prior distributions of the parameters according to the likelihood. The likelihood reflects the assumptions about the data generating process and allows to evaluate a model against observed data. In the process of inference, samples of parameters are drawn from posterior distributions. These samples can be consequently used to quantify uncertainty of parameter estimates in the form of distributions, as well as to derive statistics.

A summary of the model is provided in the Supplement and here we lay out the details.

Gaussian Process prior

The outcome variable (protein degradation) is modelled using a GP with a composite kernel proposed in [7]. A GP defines a probability distribution over functions, i.e. each sample from a GP is an entire function, indexed by a set of coordinates (concentrations) belonging to some set X, which it maps into a real-valued output. Any finite realisation of a GP represents a multivariate normal distribution, defined by a vector mean $\underline{\mu}$ and a covariance matrix K. A GP is specified by a mean and covariance function of one and two arguments, respectively: $y \sim \mathrm{GP}(\mu_f(\cdot), k(\cdot,\cdot))$. The mean function $\mu_f(\cdot)$ describes the baseline of the process. Its value at point x is an average of an infinite number of realisations of the GP evaluated at point x. By linearity, the mean function can be decoupled from a GP: $y = \mu_f + f, f \sim \mathrm{GP}(0, k)$, and that is why it is typically modelled separately, while GPs are being assigned zero mean. The covariance function or covariance kernel, controls variance-covariance structure: its value at a set of two points $k(x_1, x_2)$ quantifies the covariance of two random variables $f(x_1)$ and $f(x_2)$. Kernel design is crucial for GP modelling since it describes the class of possible functions which can be captured: both the global shape, as well as local properties, such as smoothness [8]. Most commonly used kernels, such as exponential $\eta^2 \exp(-\frac{d}{\rho})$, squared exponential $\eta^2 \exp(-\frac{d^2}{\rho^2})$ or Matern $C_{\nu,\rho}(d)$, depend on the distance $d = \|x_1 - x_2\|$ between two points and, hence, are stationary. They are inappropriate for modelling curves and surfaces which might display different behaviour across the input space X. Kernels derived from exponential and squared exponential kernels would have smooth shapes, while Matern kernels under certain values of ν result in curves with more local fluctuations and less smoothness. The length-scale ρ, common for many kernels, governs the flexibility of the curves: smaller length-scales induce correlations between points at smaller distances and hence more oscillatory behaviour - such behaviour, when exaggerated, can

lead to overfitting; too large length-scales, on the other hand, make curve fits too inflexible and may lead to underfitting the data. The amplitude η, frequently applied to kernels as a constant multiplier, controls the expected variation in the output. The amplitude controls how strong the functions fluctuate and the length-scale controls how rapidly they fluctuate. Choice of priors of the GP parameters plays a big role in their identifiability.

Covariance functions can be scaled, multiplied, summed with each other and further modified to achieve a desired effect. For instance, if it is known that a curve behaves differently in different areas of its input values, changepoint kernel would provide a reasonable model. The desired global GP can be then constructed as a weighted average of two or more local GPs,i.e. if the two local GPs are $f_1(x) \sim \mathrm{GP}(0, k_1)$ and $f_2(x) \sim \mathrm{GP}(0, k_2)$, we can define a global GP as $f(x) := (1 - w(x))f_1(x) + w(x)f_2(x) \sim \mathrm{GP}(0, k)$. Here $w(x)$ defines a weighting function and the kernel k can be computed as $k(x, x^{'}) = (1 - w(x))k_1(x, x')(1 - w(x')) + w(x)k_2(x, x')\sigma(x')$. Standard choice of the weighting function is a sigmoidal $w(x) = \sigma(x)$, providing a smooth transition from f_1 to f_2, centred around zero. However, the weighting function can be modified to shift the transition point from 0 to θ: $\sigma(x) \to \sigma(x - \theta)$, or make the transition more rapid $\sigma(x) \to \sigma(kx), k > 1$. If parameter k is a large number, the transition function can be substituted with a step function. In practice, due to the difficulties associated with differentiation of the step function, it is easier to formulate a changepoint kernel with a step-function as a weight, via an if-else statement.

In the context of concentration-response curve fitting, the data can be represented as a set of pairs (x_i, y_i), $i = 1, ...N$. Here $x = x_i$ is the value of the unique input coordinate (concentration), $y = y_i$ is the corresponding response, and N is the number of observations (see Figure 1). For PROTACs it is assumed that the expected response is constant at low concentrations up to a threshold concentration θ (the concentration threshold at which a chemical induces an effect, i.e. PoD). For concentrations above θ, the expected response is assumed to vary smoothly as a function of concentration and it can both grow or decline. Hence, exponential or squared exponential kernels would make a reasonable choice for concentrations above θ, but the overall kernel needs to be non-stationary as the behaviour of a curve is different before and after the threshold . The GP model is defined for the expected response in the additive form: $y = \underline{\mu} + f + \epsilon$. Here $\underline{\mu} = \mu * [1, ..., 1]^T$ is a constant vector of length N defining the baseline estimated both from control and treatment data, $f \sim \mathrm{MVN}(0, K)$ is the GP defining the shape of the mean curve, and $\epsilon \sim N(0, \sigma^2 I)$ is the noise term responsible for the curve uncertainty σ, unexplained by the general shape f. The covariance matrix K, determining the range of possible expected mean curves is defined using

the kernel

$$k(x_i, x_j) = \begin{cases} 0, & \text{if } x_i \text{ or } x_j \leq \theta, \\ \eta^2 (x_i - \theta)^2 (x_j - \theta)^2 \exp(-\frac{(x_i - x_j)^2}{\rho^2}), & \text{if } x_i \text{ and } x_j > \theta. \end{cases}$$

This is a changepoint kernel with transition point at θ and step-function as a weight $w(x)$. To the left side of the threshold ($k_1 = 0$), it ensures linear behaviour with no slope. The kernel $k_2 = \eta^2 (x_i - \theta)^2 (x_j - \theta)^2 \exp(-\frac{(x_i - x_j)^2}{\rho^2})$, defining the function to the right side of the threshold, is a product of the Gaussian and squared linear kernels. The Gaussian kernel $\eta^2 \exp(-\frac{(x_i - x_j)^2}{\rho^2})$ provides smoothness, and squared liner kernel $(x_i - \theta)^2 (x_j - \theta)^2$ allows the range of plausible values to increase together with the distance from the point of departure θ. The same class of curve shapes can be obtained via a steep sigmoidal weight function $w(x)$ (see Supplement).

Measurement model and likelihood

The model takes two sources of variation into account: curve uncertainty and replicate-to-replicate variation. Variability in replicates for the same concentration $y^{i,r}_{\text{rep}}, r = 1, ..., N_{\text{rep}}$ is modelled with a Student's t-distribution with three degrees of freedom, centred on the mean response and replicate-to-replicate scale σ_{rep}: $y^{i,r}_{\text{rep}} \sim t_3(y^i, \sigma_{\text{rep}}), r = 1, ..., N_{\text{rep}}$. Student's t-distribution has heavier tails than the normal distribution and, hence, allows for more extreme observations. This choice makes the model less sensitive to outliers. Data for controls $(x_1, y^c_{\text{control}}), c = 1, ...N_{\text{control}}$ has mean μ and two sources of variation: curve uncertainty and replicate-to-replicate uncertainty. That is why, heuristically, the likelihood for it is derived as $y^c_{\text{control}} \sim t_3(\mu, \sqrt{\sigma^2 + \sigma^2_{\text{rep}}})$, $c = 1, ...N_{\text{control}}$. Distributions for y_{rep} and y_{control} define the likelihood of the model.

Priors for model parameters

Full Bayesian model formulation requires specification of prior distributions for parameters $\mu, \sigma, \sigma_{\text{rep}}, \eta, \rho$ and θ. Baseline response μ is assigned the weakly informative prior $\mu \sim N(0, 0.1)$. Scale σ, governing the concentration-to-concentration variation is assigned the inverse gamma prior $\sigma \sim \text{InverseGamma}(1, 2)$, implying a mode of 1. Scale σ_{rep}, governing the replicate-to-replicate variation, is assigned the inverse gamma prior $\sigma_{\text{rep}} \sim \text{InverseGamma}(1, 0.1)$ implying a mode of 0.05. In this way we express our belief that replicate-to-replicate variation for the same concentration is smaller than the deviation of measurements from the mean curve $\mu + f$. The length-scale, ρ, should not be smaller than the minimal distance between concentratiions (about 0.5 on the log10-concentration scale), and not larger than the range of concentrations (about 6 on the log10-concentration scale). We use a gamma prior

$\rho \sim \text{Gamma}(50, 20)$ which does respect the constraints and has a mode of approximately 2.5. This implies that for log10-concentrations between -10.5 and -4.5, correlations are non-negligeable at distances of about 40% of the total length of the concentration interval. For the amplitude η, we use a normal prior $\eta \sim N(1, 1)$, and for the threshold θ, a uniform distribution between the lowest and highest concentrations $\theta \sim U(x_{\min}, x_{\max})$.

GP predictions

Only evaluations of the GP at measured concentrations $x_i = x_i^{\text{obs}}$ are used for inference, but we are interested in finding continuous curves. To calculate curve trajectories, we define a grid of points (concentrations) at which we want to make predictions and calculate the values of f at the selected locations x_i^{pred}, $i = 1, ..., N^{\text{pred}}$. Jointly, observed and unobserved values are distributed as a multivariate normal

$$\begin{pmatrix} f^{\text{obs}} \\ f^{\text{pred}} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu^{\text{obs}} \\ \mu^{\text{pred}} \end{pmatrix}, \begin{pmatrix} K^{\text{obs}} & K^{\text{obs, pred}} \\ (K^{\text{obs, pred}})^T & K^{\text{pred}} \end{pmatrix} \right).$$

Due to the properties of the multivariate normal distribution, the distribution for the unobserved values given the observed values is given by $N(\underline{m}, \Sigma)$, with $\underline{m} = \mu^{\text{pred}} + (K^{\text{obs, pred}})^T * (K^{\text{obs}})^{-1} * (f^{\text{obs}} - \mu^{\text{obs}})$ and $\Sigma = K^{\text{pred}} - (K^{\text{obs, pred}})^T * (K^{\text{obs}})^{-1} * K^{\text{obs, pred}}$. Here $\mu^{\text{pred}} = \mu * [1, ..., 1]^T$ is a constant vector, K^{obs} is the covariance matrix evaluated at observed locations, K^{pred} is the covariance matrix evaluated at locations for predictions, and $K^{\text{obs, pred}}$ is the covariance matrix evaluated at the pairs of observed and prediction locations.

Calculating biologically relevant parameters and ranks

In the process of Bayesian inference we collect samples of the model parameters and the mean GP curve, which form the posterior distribution. If N^{iter} draws have been saved, each of the curves can be used to derive further statistics, such as biologically meaningful quantities - DC_{50}, $\text{D}_{\max}$ or PoD. We perform the computations based on each drawn GP to create a sampling distribution of the quantities. These distributions reflect uncertainty of the estimates. A PoD is directly represented in the model by its parameter θ; $\text{D}_{\max}$ is calculated as the maximal amount of degradation, i.e. the minimal value on the y-axis; and DC_{50} is calculated as the abscissa of the half-maximal degradation point. To find DC_{50} numerically, we first compute its ordinate y_{DC50} and then search for a point with the lowest $x-$value on the predictive grid with the minimal difference between the GP and $y_{\text{DC}_{50}}$.

For each iteration in this process we create a ranking of compounds based on the magnitude of DC_{50}. If $(\text{DC}_{50})_j^i$ is an estimate obtained from iteration $i, i \in (1, N^{\text{iter}})$ for compound $j, j \in (1, N^{\text{cpds}})$ (here N^{cpds} is the

total number of compounds), then for a fixed i, the N^{cpds} estimates can be sorted. The order of the sorting yields the ranking of compounds, and the final rank of a compound is its most frequent rank of this compound across all iterations.

Software

The model was fit in Stan [9] which uses a Hamiltonian Monte Carlo sampler. We used Rstudio [10] to process the data and interface with Stan. The Bayesplot [11] library was used to visualise convergence statistics. We used 4 chains of 10000 steps each to produce draws from the posterior. Convergence has been evaluated numerically using the $\hat{R}$ statistic (measures the ratio of within chain variation to between chain variation), effective sample size statistic (an estimate of the number of independent draws from the posterior distribution) and visually by using posterior distributions of model parameters (Figure 5 in the Supplement shows results for one compound). Posterior distributions from different chains agree well and demonstrate convergence. Mean elapsed CPU time (on a laptop with 2.3 GHz Intel Core i5 and 8 GB Memory) for 4 chains was 9.23 seconds with standard deviation of 1.22 seconds. Estimates of model parameters for one compound are presented in Table 2 in the Supplement. Predictions were made on a regular grid of $N^{\text{pred}} = 1000$ points spaced between the minimal and maximal concentrations. The code for fitting the data to one compound using a steep inverse logit function as a weight function $w(x) = \text{logit}^{-1}(kx), k = 10$ is available on Github (https://github.com/elizavetasemenova/gp_concentration_response).

Results

First, we fit the model to one compound with 12 concentrations (the dosage currently used in practice) and two replicates, and discuss the results in detail. Next, we analyse 9 compounds with 18 concentrations. Data of the two experiments were available, and each experiment contained two replicates. In this way, the combined dataset provides four replicates per concentration for each of the nine compounds. We demonstrate uncertainty in the produced ranking based on DC_{50} and suggest extending this metric to also include D_{max}.

Example of one compound

This dataset contains 12 concentrations with 2 replicates per compound. Figure 1 shows the mean prediction, which closely follows the main trend in the data, and the qunatile-based (95%) Bayesian credible Interval [12] (BCI; shaded region around the curve). Despite the non-sigmoidal shape, the model accurately detects the lowest part of the curve, which indicates

the maximal effect (D_{max}). In addition, the model automatically estimates a sensible value for the DC_{50} – similar to where a person might draw a vertical line. The PoD is estimated with less precision, but covers the range of concentrations where the curve is largely flat.

Uncertainty in ranking of compounds

The data in Figure 2 are from 2 experiments at 18 concentrations and with 2 replicates per experiment. This is a very high resolution and would be rarely available in a screening process, and hence they yield tight uncertainty bounds for $\log_{10}(DC_{50})$. However, even low uncertainty in the estimates translates into the uncertainty of the rankings for many compounds. Figure 3 (a) displays posterior distributions for the $\log_{10}(DC_{50})$ values and Figure 3 (b) displays the resulting ranks and their uncertainty. Compounds ranked as the best (rank=1,2) and the worst (rank=9,8) are very distinct, while uncertainty in the ranks is higher for the other compounds. Decisions made solely on DC_{50} might overlook other important properties, such as the maximal degradation D_{max}. Figure 4 displays the estimates and BCIs for both parameters. Compounds in the bottom left quadrant are preferable. This agrees well with the ranking of compound nine - it has the lowest $\log_{10}(DC_{50})$ and $\log_{10}(D_{max})$; however, compound seven, confidently ranked second according to $\log_{10}(DC_{50})$, shows less maximal degradation, and potentially, would be de-prioritised if more information was taken into account. We suggest using a weighted combination of $\log_{10}(DC_{50})$ and $\log_{10}(D_{max})$: $D_{comb} = \alpha\widetilde{DC_{50}} + (1-\alpha)\widetilde{D_{max}}$. Here $\widetilde{DC_{50}}$ and $\widetilde{D_{max}}$ are $\log_{10}(DC_{50})$ and $\log_{10}(D_{max})$ re-scaled to the $[0,1]$ range, and $\alpha \in [0,1]$ is the weighting coefficient, which can be chosen by a scientist to give more preference to DC_{50} (α close to 1), weight DC_{50} and D_{max} equally ($\alpha = 1/2$) or to prioritise D_{max} (α close to 0). Using the weighted metric, we re-calculate ranks of the 9 compounds according to different values of α (Table 1). We have performed the scaling by subtracting the minimal value of all samples and dividing by the difference between minimal and maximal values of all samples.

α	Compound number								
	1	2	3	4	5	6	7	8	9
	Compound's rank (and its relative frequency)								
1	3 (0.97)	7 (0.94)	9 (1.00)	4 (0.95)	6 (0.88)	5 (0.91)	2 (1.00)	8 (1.00)	1 (1.00)
2/3	3 (0.83)	2 (0.84)	9 (1.00)	5 (0.78)	7 (0.48)	4 (0.92)	6 (0.50)	8 (0.70)	1 (1.00)
1/2	3 (0.89)	2 (1.00)	9 (1.00)	6 (0.63)	7 (0.77)	4 (0.81)	8 (0.99)	5 (0.68)	1 (1.00)
1/3	3 (0.47)	2 (1.00)	9 (0.99)	6 (0.59)	7 (0.59)	5 (0.42)	8 (0.99)	5 (0.37)	1 (1.00)
0	5 (0.63)	1 (0.62)	8 (0.70)	7 (0.52)	6 (0.60)	4 (0.61)	9 (1.00)	3 (0.94)	2 (0.62)

Table 1: Ranks of 9 compounds and uncertainty based on the combination metric $\alpha\widetilde{DC_{50}} + (1-\alpha)\widetilde{D_{max}}$ for different values of α.

Discussion

We have used a Bayesian GP model with a composite kernel to flexibly fit PROTAC CR curves. A Bayesian approach is advantageous for small amounts of data and can represent uncertainty of the parameters and derived estimates in the form of distributions. A GP model does not require strong assumptions on curve shapes, as parametric models do. The model is stable to outliers due to using a Student's distribution. Non-identifiability is a frequent problem for GPs: often only one of the length-scale or amplitude parameters can be identified, while the other parameter needs to be fixed for the analysis. We have used an informative prior for length-scale and have not encountered problems under the specified settings.

We have demonstrated how uncertainty in the estimates of DC_{50} influences uncertainty in ranks and have proposed a single metric to capture both DC_{50} and $\mathrm{D}_{\max}$. This metric involves a weighting parameter that allows a scientist to choose their decision-making approach and tackle ranking in a quantified manner. We have demonstrated that the ranks of the worst (rank=9) and the best (rank=1) compounds are fairly stable with respect to the varying weighting coefficient, while ranks of less clear leaders (e.g., compounds number 2 and 7) may change their ordering. Currently, decisions are mainly made based on DC_{50}, which corresponds to $\alpha = 1$. The weighting can be extended to more parameters, e.g. DC_{50}, $\mathrm{D}_{\max}$, and θ if they all represent interest for screening; the combination metric would weight all three estimates $\alpha_1\widetilde{\mathrm{DC}_{50}} + \alpha_2\widetilde{\mathrm{D}_{\max}} + \alpha_3\widetilde{\theta}$ with $\alpha_1 + \alpha_2 + \alpha_3 = 1, \alpha_k \in (0, 1)$. For parametric models, such as a sigmoidal or double-sigmoidal, the area under curve (AUC) can be used as a single measure of compound's potency. This approach, however, is less applicable to GP fits since such curves might experience local fluctuations. These fluctuations would have little influence on the estimates of parameters such as DC_{50}, but cumulatively might contribute non-negligeable amounts of AUC leading to erroneous conclusions. A potential solution could be computing AUC for GPs only in the domain between the minimal concentration and the first concentration where the maximal effect $\mathrm{D}_{\max}$ is achieved. In our example we have mapped DC_{50} and $\mathrm{D}_{\max}$ by scaling. Alterhative approaches can be used, such as desirability functions [13].

The proposed model has been used internally to answer a set of experimental design questions, such as "Is it better to have more concentrations, or more replicates?" A series of comparisons has been done between 4,3,2 and 1 replicates for 18 concentrations, and between different number of concentrations (9 and 18) for the same number of replicates. An example of 18 and 9 concentrations (2 replicates each), for the same compound as Figure 1 is shown in the Supplement (Figure 6). The plot shows slightly more uncertainty in the model fit produced using 9 concentrations, while the general shape of the fit and DC_{50} estimate remain close. Model fits based on 4,3,

and 2 replicates for the same compound are presented on Figure 7 in the Supplement. Both the curve shapes and amount of uncertainty are similar for all three fits. Our conclusion from this exercise, evaluating the trade-off between number of replicates and number of concentrations, was that more concentrations is preferred to more replicates and that 12 concentrations with 1 replicate is sufficient to produce a robust ranking.

The changepoint kernel, used in this work, fits a flat curve to the data to the left side of the PoD and does not use the mechanistic understanding of the model otherwise. As a result of that, some GP fits might display minor local fluctuations at higher concentrations, which a parametric model would not do. Nevertheless, these fluctuations would not have major impact on the estimates of the estimated parameters and ranking. If it is important to capture more mechanics in the curve-fitting step, it can be done by further expansion of the changepoint kernel. For instance, in case of PROTACs, one would expect the curve to flatten out after the hook. The changepoint kernel, in this case, would have three GP components (flat linear, smooth, and flat linear) and two inflexion point θ_1 and θ_2. Derivations of the kernel in this case, are provided in the Supplement. Another solution to the local fluctuations at high concentrations could be provided by a concentration-dependent length-scale, such as, for instance $\rho^2(1+(x_i-\theta)^2)(1+(x_j-\theta)^2)$ [14].

GP models are computationally more involved than parametric models. Running Markov Chain Monte Carlo samplers predictions might take longer than applying traditional (frequentist) parametric methods. As a result, incorporating Bayesian GP models into workflows might be not straightforward. However, as computational tools evolve, this limitation can be gradually elevated. Furthermore, frequentist optimisation techniques have their own difficulties, such as subjective choice of initial values, and might have convergence issues.

Unlike parametric models, it is more difficult to link GPs to a mechanistic understanding. And sometimes models might display minor local fluctuations, which a parametric model would not have. Nevertheless, these fluctuations would not have major impact on the parameters estimates and ranking. We believe that this approach has a lot of potential and can be used internally by pharmaceutical companies by taking steps towards its automation. If the assumption about possible curve shapes holds, important choices to make are the priors, and degrees of freedom in the Student's t-distribution.

References

[1] D. P. Bondeson, A. Mares, I. E. Smith, E. Ko, S. Campos, A. H. Miah, K. E. Mulholland, N. Routly, D. L. Buckley, J. L. Gustafson, *et al.*,

"Catalytic in vivo protein knockdown by small-molecule protacs," *Nature chemical biology*, vol. 11, no. 8, pp. 611–617, 2015.

[2] E. F. Douglass Jr, C. J. Miller, G. Sparer, H. Shapiro, and D. A. Spiegel, "A comprehensive mathematical model for three-body binding equilibria," *Journal of the American Chemical Society*, vol. 135, no. 16, pp. 6092–6099, 2013.

[3] C. N. Haas, J. B. Rose, and C. P. Gerba, *Quantitative microbial risk assessment*. John Wiley & Sons, 1999.

[4] A. V. Hill, "The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves," *j. physiol.*, vol. 40, pp. 4–7, 1910.

[5] K. R. Shockley, "Estimating potency in high-throughput screening experiments by maximizing the rate of change in weighted shannon entropy," *Scientific reports*, vol. 6, p. 27897, 2016.

[6] A. L. Gould, "Bma-mod: A bayesian model averaging strategy for determining dose-response relationships in the presence of model uncertainty," *Biometrical Journal*, vol. 61, no. 5, pp. 1141–1159, 2019.

[7] S. Hatherell, M. T. Baltazar, J. Reynolds, P. L. Carmichael, M. Dent, H. Li, S. Ryder, A. White, P. Walker, and A. M. Middleton, "Identifying and characterizing stress pathways of concern for consumer safety in next-generation risk assessment," *Toxicological Sciences*, 2020.

[8] C. Steinruecken, E. Smith, D. Janz, J. Lloyd, and Z. Ghahramani, "The automatic statistician," in *Automated Machine Learning*, pp. 161–173, Springer, Cham, 2019.

[9] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, "Stan: A probabilistic programming language," *Journal of statistical software*, vol. 76, no. 1, 2017.

[10] RStudio Team, *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA, 2020.

[11] J. Gabry and T. Mahr, "bayesplot: Plotting for bayesian models. r package version 1.6. 0," 2018.

[12] L. Hespanhol, C. S. Vallio, L. M. Costa, and B. T. Saragiotto, "Understanding and interpreting confidence and credible intervals around effect estimates," *Brazilian journal of physical therapy*, vol. 23, no. 4, pp. 290–301, 2019.

[13] S. Lazic, "Ranking, selecting, and prioritising genes with desirability functions. peerj 3: e1444," 2015.

[14] J. Reynolds, S. Malcomber, and A. White, "A bayesian approach for inferring global points of departure from transcriptomics data," *Computational Toxicology*, p. 100138, 2020.

Figures

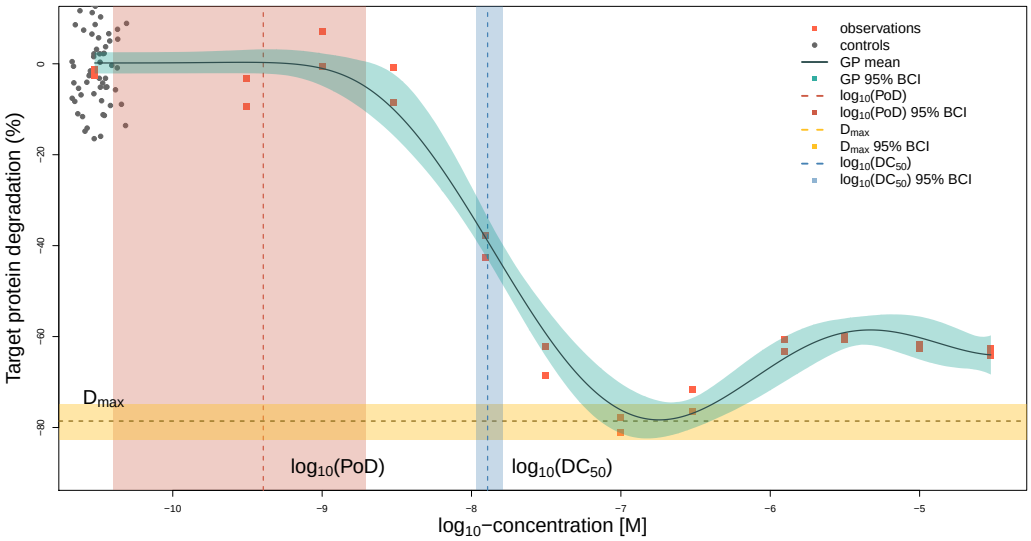


Figure 1: Prediction for one compound: The predicted curve, calculated estimates and uncertainty intervals are shown. The PoD is estimated less precisely compared to the DC_{50} and D_{max}.

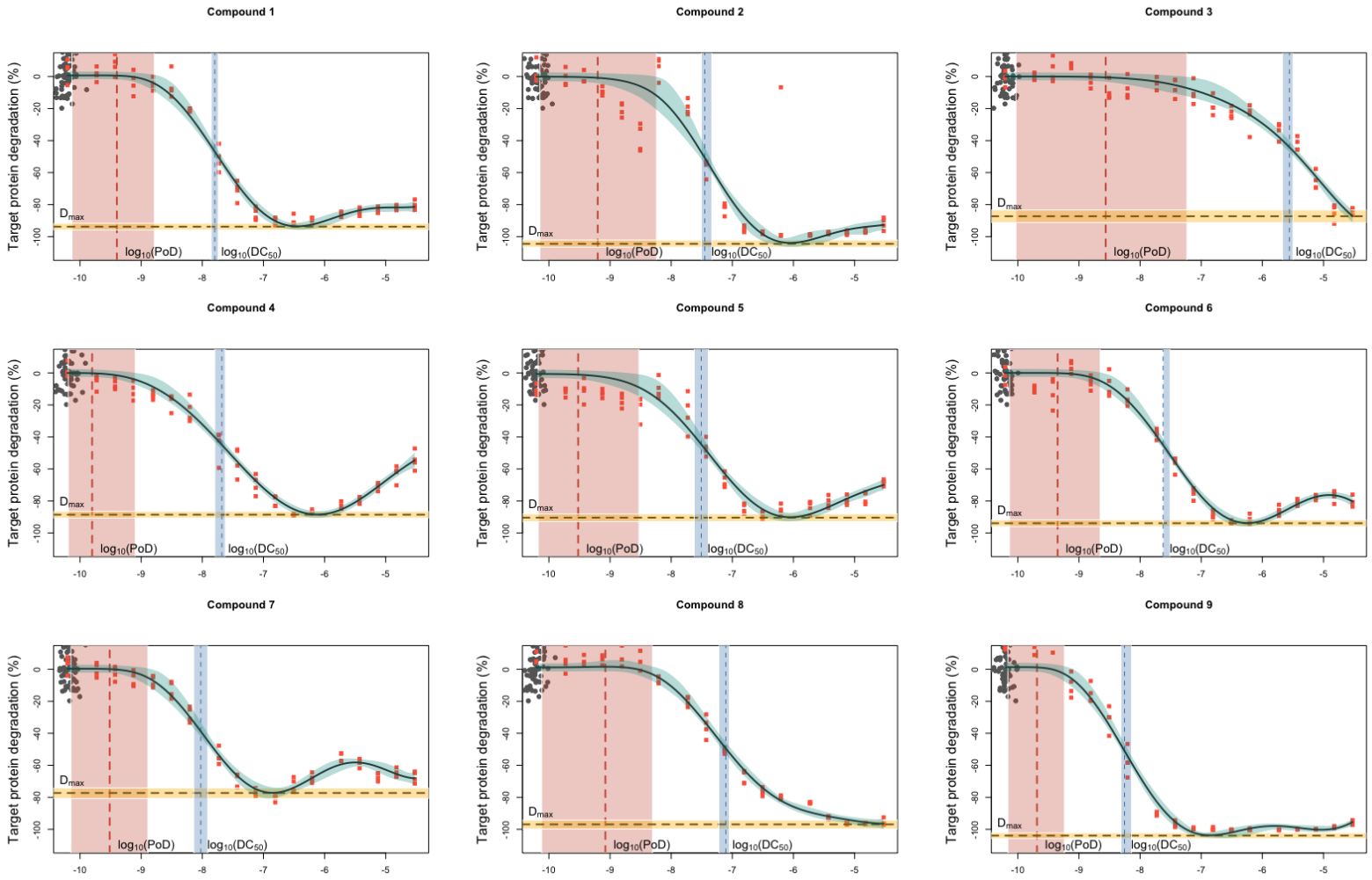
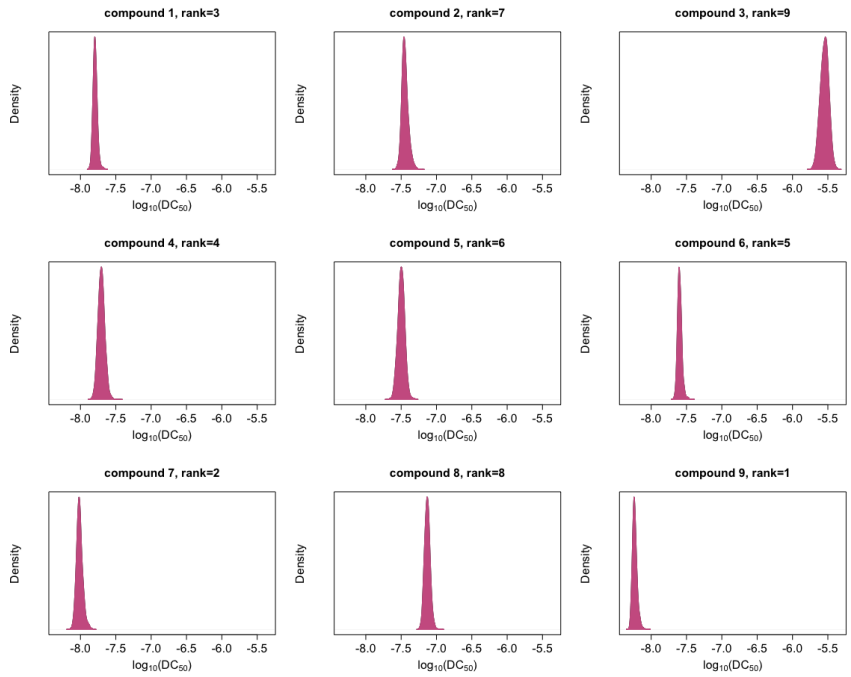
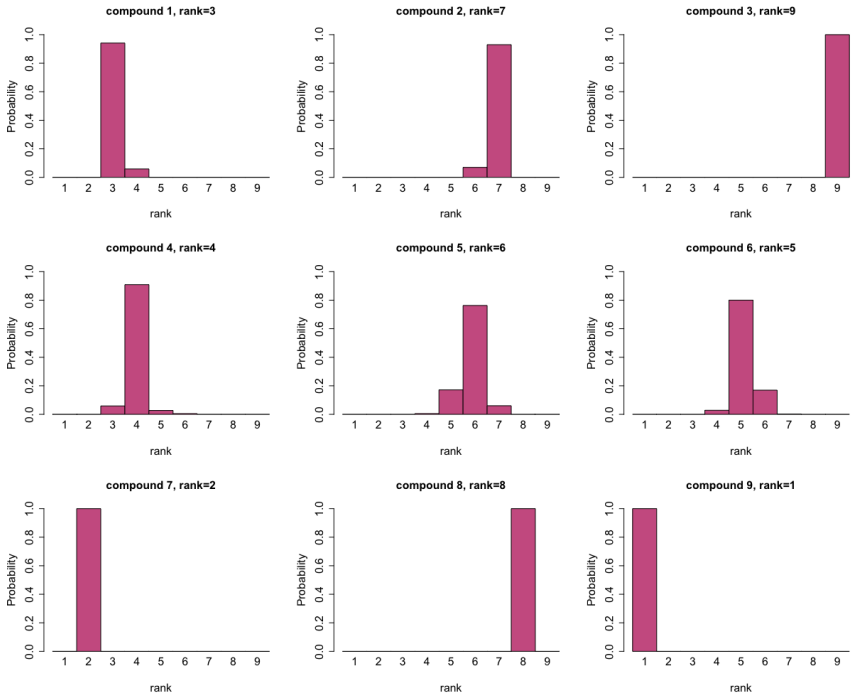


Figure 2: GP model fits of 9 compounds, with 4 replicates each. The flexibility of the model allows it to fit a variety of concentration-response relationships, and despite some non-sigmoidal shapes, the estimated DC_{50} and D_{max} values correspond to what a person might draw by hand.



(a) Estimated DC_{50} values



(b) Rank uncertainty

Figure 3: Posterior draws of $\log_{10}(DC_{50})$ values and resulting ranks of compounds together with their uncertainty.

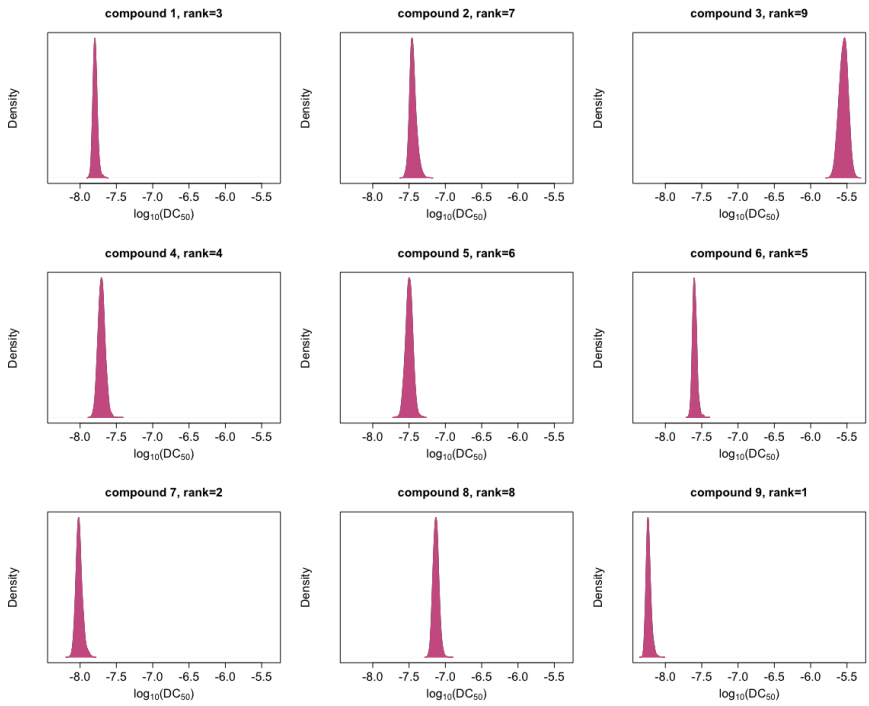


Figure 4: Two or more parameters can be used to rank compounds: Compounds with low values of $\widetilde{\mathrm{DC}_{50}}$ (scaled $\log_{10}(\mathrm{DC}_{50})$) and $\widetilde{\mathrm{D}_{\max}}$ (scaled $\mathrm{D}_{\max}$) are desirable as they indicate high potency and large effects. Compound 9 is distinguished as the best by these criteria. Error bars are 95% BCI.

Supplement

	mean	sd	10%	50%	90%	n_eff	Rhat
σ_{rep}	0.09	0.02	0.06	0.08	0.12	9082.12	1.00
σ	0.26	0.03	0.22	0.26	0.30	16686.96	1.00
μ	0.01	0.04	-0.04	0.01	0.05	16586.81	1.00
ρ	2.53	0.34	2.11	2.51	2.97	15696.80	1.00
η	0.66	0.45	0.22	0.54	1.26	10819.59	1.00

Table 2: Summary of the fitted model for one compound.

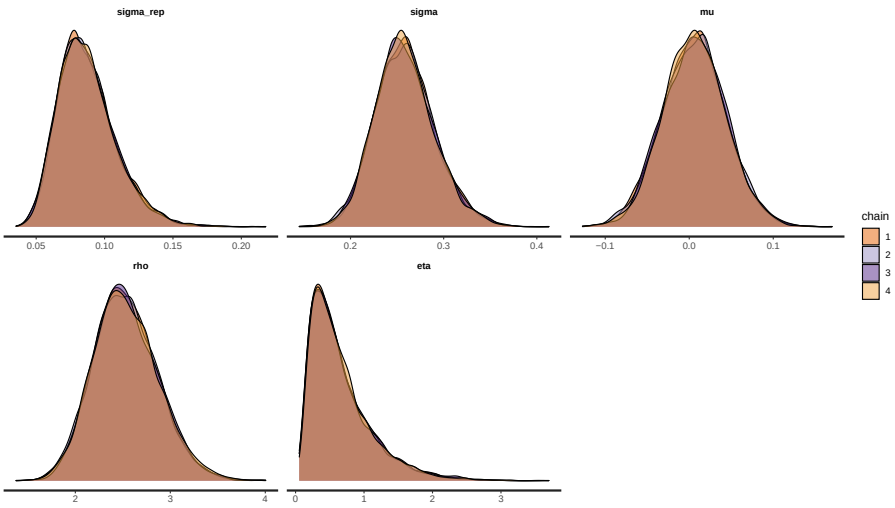


Figure 5: Posterior distributions of model parameters from different chains agree very well and show convergence.

Model Summary

Priors

$$
\begin{aligned}
y &= \underline{\mu} + f + \epsilon, \quad \underline{\mu} = \mu * [1, ..., 1]^T, \\
f &\sim \text{MVN}(0, K), \quad K = [k(x_i, x_j)]\text{- covariance matrix}, \\
\epsilon &\sim N(0, \sigma^2 I), \\
\mu &\sim N(0, 0.1), \\
\sigma &\sim \text{InverseGamma}(1, 2), \\
\sigma_{\text{rep}} &\sim \text{InverseGamma}(1, 0.1), \\
\rho &\sim \text{Gamma}(50, 20), \\
\eta &\sim N(1, 1), \\
\theta &\sim U(x_{\min}, x_{\max}),
\end{aligned}
$$

$$
k(x_i, x_j) = \begin{cases} 0, & \text{if } x_i \text{ or } x_j \leq \theta, \\ \eta^2 (x_i - \theta)^2 (x_j - \theta)^2 \exp(-\frac{(x_i - x_j)^2}{\rho^2}), & \text{if } x_i \text{ and } x_j > \theta. \end{cases}.
$$

Likelihood

$$y^{i,r}_{\text{rep}} \sim t_3(y^i, \sigma_{\text{rep}}), \quad i = 1, ..., N; r = 1, ..., N_{\text{rep}}$$
$$y^c_{\text{control}} \sim t_3(\mu, \sqrt{\sigma^2 + \sigma^2_{\text{rep}}}), \quad c = 1, ...N_{\text{control}}.$$

Changepoint models

Two functions

Changepoint between two GPs and one transition point:

$$\begin{aligned} f_1(x) \sim \text{GP}(0, k_1) \quad f_2(x) \sim \text{GP}(0, k_2), \\ f(x) := (1 - w(x))f_1(x) + w(x)f_2(x) \sim \text{GP}(0, k), \text{ where} \\ k(x, x^{'}) = (1 - w(x))k_1(x, x')(1 - w(x')) + w(x)k_2(x, x')w(x') \end{aligned}$$

These formulas are easy to understand if we chose the transition function as a step function $w(x) = s_\theta(d)$, where $d = x - \theta$ and

$$s_\theta(d) = \begin{cases} 0 & \text{if } d \leq 0, \\ 1 & \text{if } d > 0. \end{cases}$$

The most popular choice of $w(x)$ is the inverse logit (sigmoidal) function. A steep sigmoidal function $\sigma(k * x), k > 1$ is a good approximator of the step function.

Three functions

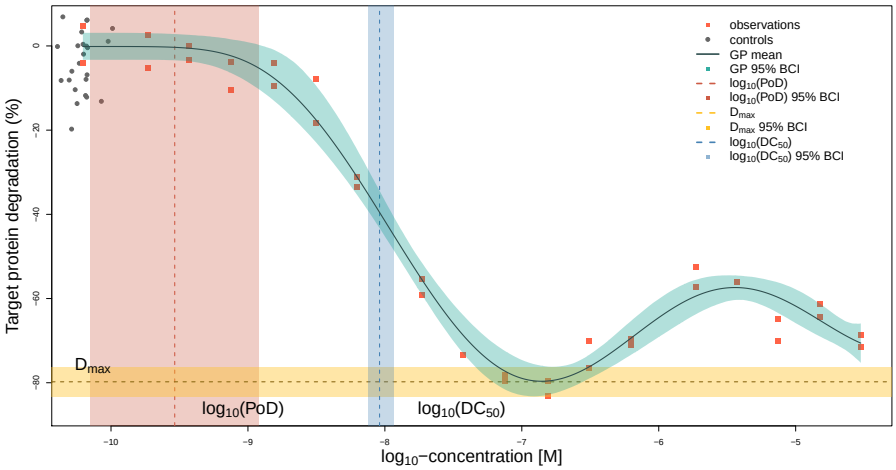
Changepoint between three GPs with two transition points θ_1, θ_2 can be constructed in a similar manner: we need to construct a function which behaves as $f_1(x)$ for $x \leq \theta_1$, as $f_2(x)$ for $\theta_1 < x \leq \theta_2$ and as $f_3(x)$ for $x > \theta_2$. This can be achieved by

$$\begin{aligned} f(x) &:= (1 - s_{\theta_1}(x))f_1(x) + s_{\theta_1}(x)f_{2,3}(x) \\ &= (1 - s_{\theta_1}(x))f_1(x) + s_{\theta_1}(x)\left[(1 - s_{\theta_2}(x))f_2(x) + s_{\theta_2}(x)f_3(x)\right] \\ &= (1 - s_{\theta_1}(x))f_1(x) + s_{\theta_1}(x)(1 - s_{\theta_2}(x))f_2(x) + s_{\theta_1}(x)s_{\theta_2}(x)f_3(x). \end{aligned}$$

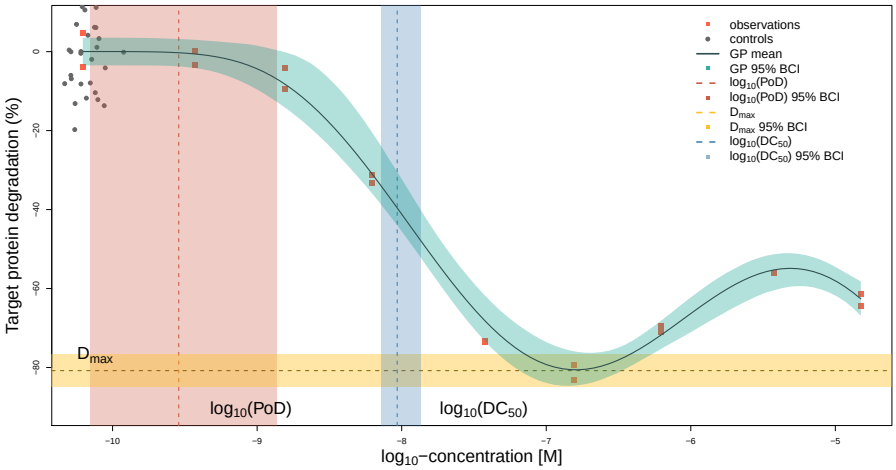
The kernel of this representation can be derived as

$$\begin{aligned} k(x, x') = (1 - s_{\theta_1}(x))k_1(x, x')(1 - s_{\theta_1}(x'))+ \\ s_{\theta_1}(x)(1 - s_{\theta_2}(x))k_2(x, x')s_{\theta_1}(x')(1 - s_{\theta_2}(x'))+ \\ s_{\theta_1}(x)s_{\theta_2}(x)k_3(x, x')s_{\theta_1}(x')s_{\theta_2}(x'). \end{aligned}$$

To enforce the hook behaviour with flat curve at low and high concentrations and flexible shape in-between, a reasonable choice of kernels would be linear (without slope) choices for k_1 and k_2 and a smooth kernel for k_2.

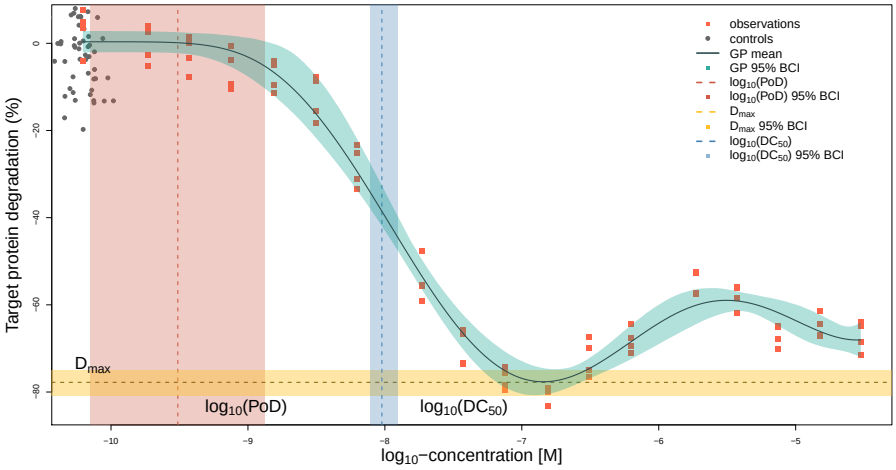


(a) 18 concentrations

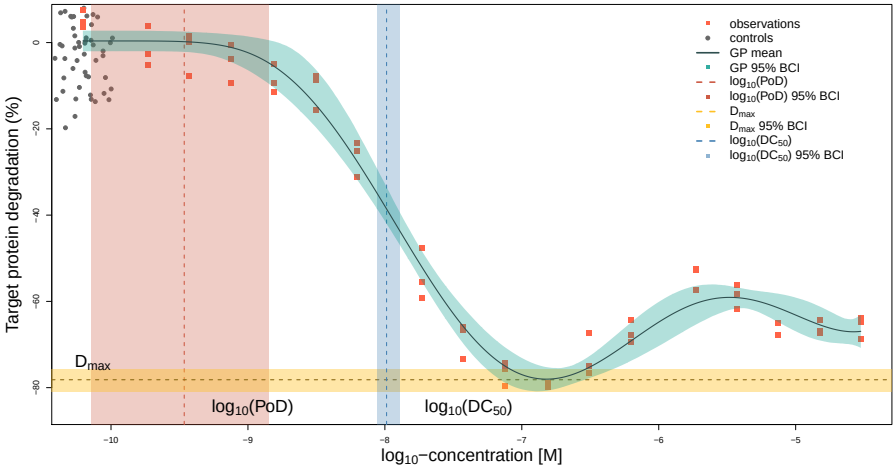


(b) 9 concentrations

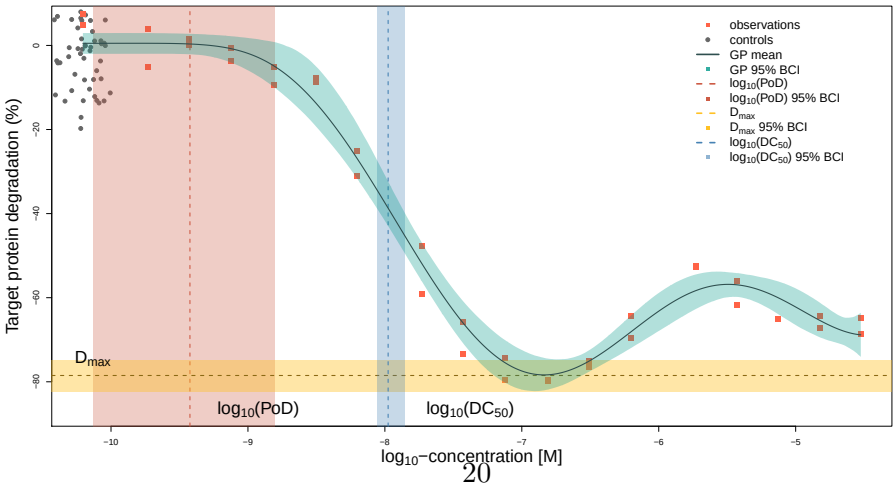
Figure 6: Example of one compound: fitted model and calculated estimates based on 18 and 9 measured concentrations in 2 replicates each.



(a) 4 replicates



(b) 3 replicates



(c) 2 replicates

Figure 7: Example of one compound: fitted model and calculated estimates based on 4,3 and 2 replicates, with 18 concentrations each.