Incorporating model structure uncertainty in model-based drug discovery

Peter Gennemark, CVMD iMED DMPK, AstraZeneca R&D, Mölndal, Sweden

Background

Pharmacokinetic and pharmacodynamic (PKPD) modeling underpins drug discovery decision making by providing a quantitative system understanding and dynamical predictions with accompanying uncertainties. Ideally, a theoretical model can be derived from first principles. However, in practice, the model structure is defined from prior information and model parameters are inferred in two steps: first, initial guesses are derived, e.g., from graphical methods in combination with prior knowledge, and then the parameters are optimized by numerical methods, e.g., the least-squares method. In some cases there is high confidence in the choice of model structure. Then it is reasonable to calculate uncertainty in the parameter estimates (e.g., a point estimate and a 95 % confidence interval for each parameter).

However, in many cases there is uncertainty in the choice of model structure – a theoretical model structure cannot be derived and several model structures are plausible. In these cases, it is reasonable to define a model space of relevant model structures and then selecting a model from this space. Such decisions are based on a model selection criterion, and there are several fundamentally different ways of doing this, e.g., penalizing model complexity in the objective function (AIC, BIC, MDL), F-tests for a model space of nested models, and cross validation techniques.

In the drug discovery phase, several compounds are screened in an *in vivo* PD model. The turnover of pre-clinical PKPD data analysis and modeling is usually short and in depth structural model selected are seldom fully explored. In addition, a fundamental difficulty is that test compound data is sparse. This makes it hard to select the most appropriate model structure even without time constraints.

Objectives

The objective of this study is to improve standard model-based predictions from preclinical data sets by incorporating both the structure and parameter uncertainties, and not only parameter uncertainty, in an approach that is useful in practice.

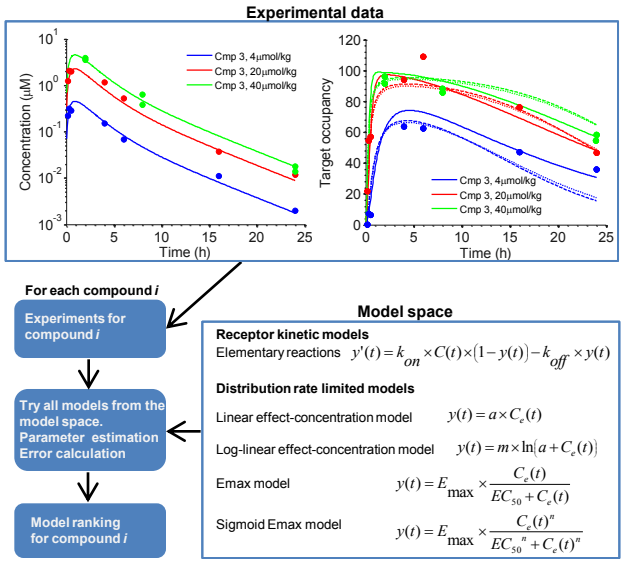

Figure 1. Typical experimental data (upper), workflow for model ranking (left) and model space composition (right). For compounds in the considered lead optimization series, both receptor kinetic models and distribution rate limited models are plausible.

Methods

Our approach was developed and tested using mice data for twelve compounds (data for one experiment is depicted in the upper part of Figure 1). Uncertainty in model structure was considered by evaluating several models from a model space using a model selection criterion. In this study we have used AIC. Model based predictions were generated from all models in the model space (Figure 1). Models were then weighted using the posterior model probabilities from the calculated Akaike weights (see Box 1 for details).

The time constraint of model selection was addressed by automatization of time consuming modeling steps.

Consider a model space of *N* models. For each model *i* from the model space, AIC is calculated as

$$AIC_i = -2 \times LL + 2 \times N_p$$

where *LL* is the log likelihood, and *Np* is the number of parameters in model *i*. Then, AIC differences with respect to the smallest (best) AIC is calculated as

$$\Delta AIC_i = AIC_i - \min_{i \in N}(AIC_i)$$

AIC differences represent the loss of information relative to the best model in the model space. Akaike weights are then calculated as

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta AIC_i\right)}{\sum_{i=1}^{N}\exp\left(-\frac{1}{2}\Delta AIC_i\right)}$$

In a Bayesian sense, w_i can be interpreted as the posterior probability of model *i*.

Box 1. Calculating posterior probability using Akaike weights (see, *e.g.*, Bonate (2006)).

Results

Two main obstacles for proper model selection in drug discovery are time constraints and sparse data. Using PKPD data from a drug discovery project, we addressed both issues.

To incorporate structure uncertainty we defined a model space including a space of PD structural models, and weighed the set of feasible models based on their posterior probability. In particular we analyzed twelve compounds from a lead optimization project where potency ranking traditionally was done using a receptor occupancy model with elementary reactions. For this data set, taking model structure uncertainty into account shifts the ranking of compounds with respect to potency (Figure 2).

Concerning the time constraint, we accelerated model selection by implementing a user-friendly computational process with input data in form of an Excel file and output in form of a PowerPoint presentation file. Taken together, we could rapidly obtain robust estimation with uncertainty.

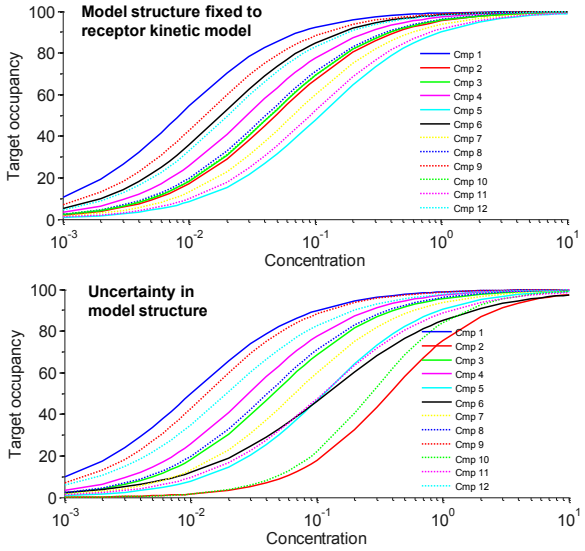

Figure 2. *Target occupancy versus steady state concentration for twelve compounds The upper plot is based on fitting a receptor kinetic model with elementary reactions to data. The lower plot is obtained by weighing (posterior probabilities based on AIC) all models from the model space. Note, for instance, how compound 6 (black solid line) falls from being ranked as the 3rd most potent to being ranked as the 8th –10th most potent compound.*

Conclusions

Model structure uncertainty, and not only parameter uncertainty for one single model structure, can be incorporated in drug discovery practice. This implies improved robustness in model selection, which is particularly important when data is sparse. A more realistic estimation of model prediction uncertainty can then be expected, which is pivotal in decision making such as compound selection.