Application of Appropriate Kinetic Models in Developing Pharmaceutical Drug Substance Manufacturing Processes

Maitraye Sen\* and Alonso J. Arguelles

Synthetic Molecule Design and Development, Eli Lilly and Company, IN, USA

\*Corresponding author: sen\_maitraye@lilly.com

Abstract

The active pharmaceutical ingredient (API) is usually synthesized through a series of chemical reactions. Depending on the complexity of the reaction network, it is common to form multiple impurities. The API synthesis process should be designed to minimize the impurity formation. A kinetic model can provide better process insights for doing so. However, developing a kinetic model in the early stages of process development is challenging due to scarcity of data. This work presents a framework for developing fit-for-purpose kinetic models from limited information or data. Step-by-step guidance on model structure determination and parameter estimation have been discussed.

**Keywords**: Active pharmaceutical active ingredient, Kinetic model, Design space.

* 1. Introduction

Meeting the quality specification (i.e., purity) of the API is paramount. The API synthesis process must be designed to meet the quality specifications consistently. Hence, a detailed process understanding should be built from the outset. Kinetic models combined with experimentation can provide important insight into the process. Since the API synthesis often involves complex reaction network, developing the kinetic model is not trivial. Moreover, the chemistry is not fully known, and the analytical method for profiling the different chemical species is not matured during the early phases of process development. Hence, a comprehensive data and information is not available at this stage. This often hinders identifying appropriate kinetic models, since most practitioners aim for an optimal model structure and parameter estimates. An optimal kinetic model is not necessary to guide process development (Sen et al., 2021). This work presents a framework for developing fit-for-purpose kinetic models. A fit-for-purpose model is relevant only within a desired operating space, where the synthesis process is designed.

As such, a simple practice-based kinetic model development workflow has been demonstrated with the help of a complex telescoped reaction network described below.

* 1. Process chemistry

The simplified reaction scheme is depicted in Figure 1. It is a telescoped reaction with three sub-steps: Step 1a, 1b, 1c. In step 1a, the deprotection of hydrazide **1** is carried out in a system of aqueous HCl and m-xylene to produce the hydrazine hydrochloride **2.HCl**. The end of reaction (EOR) mixture undergoes a wash and TEA addition to neutralize the HCl. In step 1b, enol **3** is methylated using trimethyl orthoformate (TMOF) to produce methyl enol ether **4**. The EOR mixture from step 1b is concentrated via distillation. The EOR mixtures of steps 1a and 1b are combined in step 1c, where the enol ether **4** reacts with hydrazine **2** to form pyrazole **5** that is isolated via a crystallization step.



**Figure 1: Simplified step 1a reaction scheme**

**Step1a**: It is an acid-catalyzed multi-phasic reaction (slurry of compound **1** particles in water+m-xylene). Two clear liquid layers are produced at the end of reaction. The reaction completes in ~24 h at the target conditions. Traces of other hydrazide or di-hydrazide impurities in **1** forms hydrazine that may react in step 1c to form impurity **5-IM1** (Figure 3C).

**Step 1b (**Figure 2**)**: Enol **3** is O-methylated by treatment with neat TMOF to produce methyl enol ether **4**. The reaction starts as a suspension and turns clear in short order. Three compounds related to enol **3** are formed during step 1b: i) the desired product **4** (~90%), ii) N-methylated impurity **4-IM1** (~7%), and iii) methyl ester **4-IM2** (~1%). The amount of each compound formed at reaction completion at target conditions is provided in brackets (as %area from HPLC). This reaction takes ~12 h to complete at the target conditions. The methyl formate is removed by distillation following the 1b reaction.



**Figure 2: Mechanism of formation of 4, 4-IM1, 4-IM2 in step 1b**

**Step 1c** (Figure 3): Enol ether **4** and alkyl hydrazine **2** react to form pyrazole **5**. This reaction takes ~4 h to complete at target conditions. Many impurities are formed in step 1c (Figure 3B). Figure 3C depicts the major impurities and their amount (%area from HPLC) formed at target condition at the end of the overall reaction. The crystallization step must reject the impurities to the desired levels during the isolation of **5**.



**Figure 3: Schematic of step 1c**

* 1. Kinetic model development

Step 1a has an aqueous and organic (m-xylene) phase. The solid particles of **1** is dispersed in the aqueous phase and assumed as a lumped solid phase. m-xylene is on the top of the aqueous phase. A schematic of the physical system is shown in Figure 4A. The reaction scheme was simplified for the model (Figure 4A). Sublimation of the benzoic acid (BzOH) is prevented by m-xylene.



**Figure 4: Schematic and model structure of step 1a**

The structure and estimated parameter values determine a models’ behavior. Three model structures incorporating the following physical phenomena were considered: i) reaction kinetics (RK) only, ii) liquid-liquid equilibrium (LLE) and RK and iii) dissolution rate (Disso), LLE and RK. Figure 4B shows a preliminary prediction from the three model structures when the kinetic parameters (k1, k2) are kept constant across them. The estimates of the kinetic parameters were obtained from historical data of similar reactions. The RK and LLE+RK structures predict reaction completion at less than 10 hours at target condition. The closest prediction to the experimental observation is the Disso+LLE+RK structure. RK model structure with different values of the kinetic parameters (i.e., with slower kinetics) could also fit the experimental data. However, the Disso+LLE+RK model structure is more appropriate and representative of the actual physical system since the reaction is multi-phasic with an observed dissolution rate and LLE. Hence, the model structure was selected based off the physical understanding of the system even though multiple candidates could potentially capture the experimental observation.

The final model presentation is detailed below. The model is a custom code written in gPROMS Formulated Products® (gFP) version 1.6 (Siemens, London, UK).

**Step 1a**: Compound **1** slowly dissolves from solid phase to the aqueous phase. An instantaneous partition of **1** between the aqueous and organic phase has been assumed. A constant volume of reaction has been assumed, which is equal to the summation of initial volumes of all chemical species present at time=0. Equations 1-5 present the model.

$K\_{i}=\frac{x\_{i}^{aq}}{x\_{i}^{org}}$ (1)

$\frac{dx\_{i}}{dt}=-r\_{i}+α\_{i}(x\_{i}^{s}-x\_{i}^{aq})$ (2)

$x\_{i}^{aq}=\frac{x\_{i}^{B}K\_{i}}{K\_{i}+1}$ (3)

$x\_{i}^{B}=x\_{i}^{org}+x\_{i}^{aq}$ (4)

$r\_{i}=f\left(x\_{i}^{aq}\right)$ (5)

Here, i= compound **1**, hydrazide impurity or di-hydrazide impurity (if present in **1**). ‘K’ is the partition coefficient. ‘x’ is the concentration (mol/l). Superscripts ‘aq’ stands for ‘aqueous’, ‘org’ is ‘organic’, ‘s’ is ‘solid’, and ‘B’ is ‘bulk phase’ (combined liquid phase of m-xylene+water). ‘α’ is a dissolution rate constant with units of (s-1) that lumps the term (DSw/Vh) in equation (1) of (Hattori et al. 2013). ‘r’ is the reaction rate ($\frac{mol}{l s}$). ‘K’ and ‘α’ have been expressed as a function of reaction temperature.

**Step 1b and 1c**: These two steps have been modeled as homogenous reaction (i.e., -rj = dxj/dt), where j is any chemical species. Step 1b also starts as a slurry that clears out quickly. Hence, model structure with RK only was used to represent step 1b.

Once the final model structure was decided upon, a formal parameter estimation exercise was done. The model parameters (reaction rate constants, α, and K) were estimated from the available experimental data. Batch isothermal experiments were conducted at different initial concentrations and temperature at lab-scale (~10 g). Compounds **1**, **3**, **4**, **5**, **4-IM1**, **4-IM2 and 5-IM1** were profiled over time (measured responses). Only three experiments with time-profiled concentration were available, hence limited data.

The sensitivity of each model parameter to the measured responses was calculated. Only the model parameters with reasonable sensitivities towards the measured responses were estimated. Including the less sensitive parameters in the parameter estimation deteriorates the overall performance of the algorithm (Sen et al., 2021). Hence, they were fixed at a value determined from previous experience with similar reactions. For example, Table 1 presents the sensitivity indices of the step 1a model parameters towards the concentration of **1** during a reaction at a given temperature and initial concentration. The sensitivity analysis was run by producing 80000 scenarios sampled using built-in Sobol sampling (Sobol, 1993) in gFP. The least sensitive parameter is k2, hence excluded from parameter estimation. It was fixed at a reasonable value determined from other similar reactions.

**Table 1: Sensitivity indices of step 1a model parameters**

|  |  |
| --- | --- |
| Model parameter | Sensitivity indices |
| k1 | 0.768 |
| k2 | 4.02E-8 |
| α | 0.236 |
| K | 0.034 |

* 1. Results and discussion

The kinetic model was used to obtain important process insights as detailed below.

**Step 1a**: Figure 5A presents the fractional yield of **2.HCl** (with respect to the initial concentration of **1**) at the end of 24 h, as a function of initial concentration of **1** (x-axis) and temperature (color coded). Total of 40000 simulations were run at various combinations of initial concentration of **1** and reaction temperature (sampled via Sobol, 1993 method). The reaction slows down significantly at lower temperatures. The desired conversion of **1** cannot be achieved at lower temperatures within the target completion time of 24 h. Since cycle time is important, maintaining a reasonable reaction temperature is critical. A temperature of 94 °C is affordable without slowing down the reaction significantly. It results in 98% conversion at the end of 24 h. Hence, reaction temperature should be always ≥ 94 °C. Moreover, at higher temperatures, slight variability in the initial concentration of **1** from target has no significant impact on the reaction rate.

**Step 1b**: Figure 5B shows **4-IM1**, a major impurity of step 1b. Total of 47000 simulations sampled via Sobol, 1993 method were run. The amount of **4-IM1** (y-axis) at the end of 12 h is plotted as a function of temperature (x-axis) and initial concentration of **3** (color coded). The spread of the colored dots representing the initial concentration of **3** is narrow compared to the change observed with variation in temperature. Clearly, the amount of **4-IM1** formed is more sensitive to the temperature than the initial concentration of **3**. A similar relationship of **4** (product of 1b) with temperature and initial concentration of **3** is seen. Since, both **4** and **4-IM1** form from **3** in two parallel and kinetically competitive reactions, there is no way to slow down one without slowing down the other. Hence, some amount of **4-IM1** will be formed that should be rejected during crystallization.

**Step 1c**: Total of 40000 simulations sampled via Sobol, 1993 method were run. Figure 5C presents the amount of **5** at the end of 4 h (y-axis) as a function of the ratio between initial concentration of **2** and **4** (x-axis) and reaction temperature (color coded). If the ratio is <1 the yield of **5** decreases significantly because the excess of **4** reacts to form impurities. **4** has a higher affinity to react with **2**, but in absence of **2**, it partakes in other side reaction. Figure 5C also shows that rate of formation of **5** is insensitive at temperature<30 °C (scattered dots). However, a distinct region appears at temperature>30 °C that shows that the rate decreases (amount of **5** formed decreases). This is because the impurity formation rate increases. Hence, it is important to maintain a ratio ≥1 and temperature<30 °C.



**Figure 5: Contour plots of step 1a, step 1b and step 1c**

* 1. Conclusion

Kinetic models enhance process understanding, bridging any knowledge gap resulting from limited data. A framework for developing fit-for-purpose kinetic model has been presented. The current framework is much simpler than the other alternatives of model structure determination and parametrization discussed by Sen et al., 2021. We have applied this framework to minimize impurity formation in a complex telescoped reaction.

**References**

* Y. Hattori, Y. Haruna, M. Otsuka, 2013. Dissolution process analysis using model-free Noyes-Whitney integral equation, Colloids and Surfaces B: Biointerfaces,102, 227-231.
* M. Sen, A.J. Arguelles, S.D. Stamatis, S. Garcia-Munoz, S. Kolis, 2021. An optimization-based model discrimination framework for selecting an appropriate reaction kinetic model structure during early phase pharmaceutical process development, React. Chem. Eng., 6 (11), 2092-2103.
* I.M. Sobol, 1993. Sensitivity Estimates for Nonlinear Mathematical Models, Mathematical Modelling and Computational Experiments, 4, 407-414.