# 10

# MODELING, OPTIMIZATION, AND APPLICATIONS OF KINETIC MECHANISMS WITH OPENCHEM

JOHN E. TOLSMA, BRIAN SIMPSON, AND TAESHIN PARK
*RES Group, Inc., Cambridge, MA, USA*

JASON MUSTAKIS
*Pfizer, Inc., Groton, CT, USA*

## 10.1 INTRODUCTION

Pharmaceutical manufacturing involves large-scale transformation of raw materials into drugs through various processes such as chemical synthesis, separation, and purification. Central to pharmaceutical manufacturing is the chemical synthesis of API (active pharmaceutical ingredient). The API is the active ingredient of the drug and its synthesis often involves complex chemical transformations of raw materials and strongly depends on operating conditions such as temperature, pressure, and agitation. As the mechanisms of API synthesis (e.g., relationship between operating conditions and yields of API and by-products) are better understood, the pharmaceutical manufacturing process can be optimized.

Current practice in the design and optimization of pharmaceutical manufacturing processes depends on the empirical knowledge obtained from experimental observations. Without a detailed mechanistic understanding of how APIs are synthesized, there is limited knowledge of the complex trade-offs that exist when developing the pharmaceutical manufacturing process. This is consistent with an observation made by the FDA in their Critical Pathway Initiative [1]. In pharmaceutical manufacturing, the FDA identified opportunities in adopting systematic methodology and modern science and technology into the manufacturing process to increase product yield, reduce waste, and improve process monitoring and control. Mechanistic understanding of API synthesis is one of the key components to realize these opportunities.

The API synthesis mechanism describes the chemical steps involved in the transformation of initial reagents into desired products, intermediates, and undesired by-products. Developing the API synthesis mechanism consistent with experimental observation is a complex and challenging process involving several activities. Some of these activities are described in Table 10.1 and Figure 10.1.

The goal of this chapter is to convey some of the challenges in developing API synthesis mechanisms and to describe several approaches for addressing these challenges. OpenChem software [2] will be used to illustrate some of these approaches for developing API synthesis mechanisms. OpenChem is a modeling and optimization software platform designed to help chemists to build, visualize, analyze, calibrate, and apply chemical reaction mechanisms.

The remainder of this chapter is organized as follows: first, we describe an example mechanism, involving a Buchwald–Hartwig amination reaction, which will be used throughout the chapter. Next, several approaches for building the mechanism and describing it on a computer are considered with their advantages and disadvantages. Mechanism building is followed by mechanism analysis with emphasis on extracting the most information from the model. In particular, multiple ways of visualizing the results of the simulation are described. Calibrating the mechanism with experimental data is discussed next with emphasis on a calibration workflow for

*Chemical Engineering in the Pharmaceutical Industry: R&D to Manufacturing* Edited by David J. am Ende

**TABLE 10.1 Activities, Purpose, and Requirements in API Synthesis Mechanism Development**

| Activities | Purpose | Requirements |
|---|---|---|
| Building | Create one or more candidate mechanisms describing the API synthesis | Mechanism must capture correct reaction steps and reaction rates, including intermediate and unmeasured species |
| Analysis | Identify important reactions and parameters in mechanism influencing the predictions of interest | Analysis results should be valid even when values of parameters are highly uncertain |
| Calibration | Calibrate parameters such that model predictions best fit all available data | Calibration must be able to handle multiple data types (e.g., concentration, spectra, heat flow, hydrogen uptake, etc). Most methods will often converge to poor fits, but best fit should be ensured |
| Discrimination | Select, or discriminate, most likely mechanism from collection of candidate mechanisms | Correct mechanism should be selected with confidence from the set of proposed mechanisms. All available data should be used for this selection |

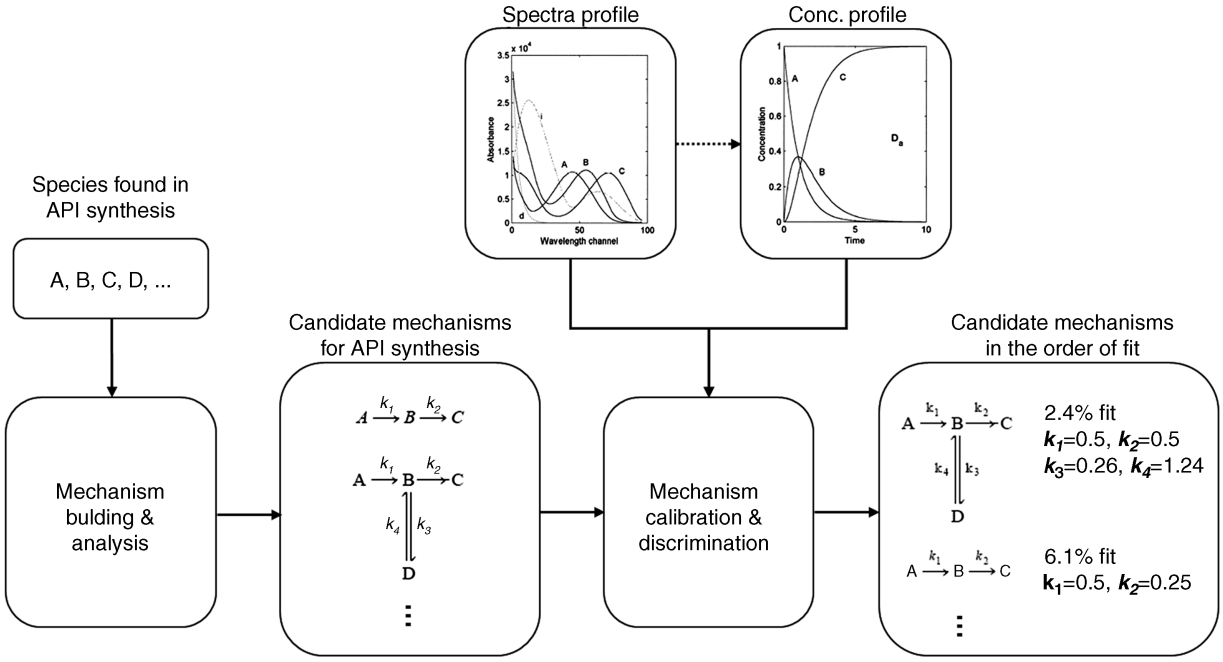



**FIGURE 10.1** Activities in API synthesis mechanism development including mechanism building, analysis, calibration, and discrimination. Both concentration and spectra profile data are used for mechanism calibration and discrimination.

systematizing the complex task of model calibration. Finally, several applications of the calibrated model are considered.

## 10.2 DESCRIPTION OF THE EXAMPLE MECHANISM

The chemical synthesis mechanism describes the reaction steps and rates for producing the product, intermediates, and by-products from the initial reagents. Figure 10.2 shows a mechanism for a coupling reaction between an aryl halide (ArX) and an amine (Amine) in the presence of a base (Base) and a catalyst (Cat). This catalytic cycle offers a convenient example of API synthesis where a relatively complex mechanism is present but only a limited amount of observations are available. An additional complexity is present in this case as the base (potassium hydroxide, KOH) is heterogeneous requiring in addition to the standard chemical steps a mass transfer term. As can be seen in the upper right corner of Figure 10.2 the base appears to have an effect on both catalyst

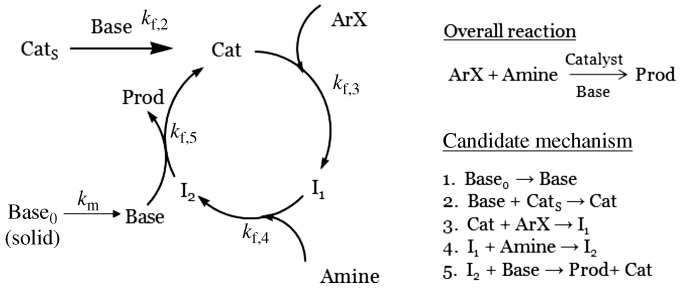



**FIGURE 10.2** Example API synthesis mechanism. Overall, an aryl halide reacts with an amine in the presence of a catalyst and strong base to form the desired aryl amine product. The schematic on the left shows the proposed mechanism steps, including formation of intermediate species. The corresponding candidate synthesis mechanism is shown on the right.

regeneration and initial catalyst activation. The proposed scheme and synthesis steps of the candidate mechanism are shown in Figure 10.2.

Overall, the aryl halide reacts with the amine in the presence of the catalyst and base to form the desired product, Prod. The solubility of the solid base is very low (e.g., KOH in toluene), so only a small concentration exists in solution to participate in the reaction. The equation used to describe the mass transfer term takes into account the effect of agitation for given reactor configurations. Please note that the formulation used for the base will still be valid if base reactions are happening on the surface. The first two steps in the proposed mechanism involve the solid base dissolving and the catalyst being reduced to the required oxidation state. This is followed in step 3 by the oxidative addition of the aryl halide to the modified catalyst, forming an intermediate $I_1$. In step 4, the halide (X) in the intermediate is replaced by the nitrogen in the amine to form an intermediate $I_2$, which then undergoes reductive elimination in the presence of the base to form the desired aryl amine (Prod) in step 5.

### 10.2.1 Reaction Formula

The reaction formulas of the candidate mechanism shown in Figure 10.2 specify the reaction stoichiometry. For example, in reaction step 4, one mole of intermediate $I_1$ reacts with one mole of amine to form one mole of intermediate $I_2$.

### 10.2.2 Rate Laws

To fully define the API synthesis mechanism, the rate of each reaction as a function of species concentrations, temperature, and other variables must also be specified. The reaction rates for the candidate mechanism shown in Figure 10.2 are the following:

$$r_1 = k_{m,1}\left(\frac{\text{RPM}}{\text{RPM}_{\text{ref}}}\right)^n ([\text{Base}_0]-[\text{Base}])$$

$$r_2 = k_{f,2}[\text{Base}][\text{Cat}_s]$$

$$r_3 = k_{f,3}[\text{ArX}][\text{Cat}]$$

$$r_4 = k_{f,4}[\text{I}_1][\text{Amine}]$$

$$r_5 = k_{f,5}[\text{I}_2][\text{Base}]$$

where $r_i$ denotes the rate of reaction $i$ in moles per unit time per unit volume, $k_{m,1}$ and $k_{f,i}$ ($i = 2, \ldots, 5$) denote the rate constants, and $[\cdot]$ denotes the species concentration in moles per unit volume. Parameters RPM and $\text{RPM}_{\text{ref}}$ are the agitator speed and reference speed, respectively. Parameter $n$ indicates how the solid base dissolution rate scales with agitator speed. The reaction rate constant is usually a function of temperature, but may also depend on other quantities such as pressure and other species concentrations. Typically, temperature dependence of the reaction rate constant is given by the Arrhenius expression:

$$k(T) = AT^{\beta}\, \mathrm{e}^{-E_A/RT}$$

where $k(T)$ is the rate constant, $A$ is the preexponential factor, $\beta$ is the temperature exponent, $E_A$ is the reaction activation energy, $R$ is the gas constant, and $T$ is the temperature. Parameters $A$, $\beta$, and $E_A$ are typically adjusted in the synthesis mechanism to fit experimental data. This process is described in detail in Section 10.4. The expression defining the reaction rate is often referred to as the reaction rate law.

The mechanism is fully defined by the reaction formulas and rate laws. The following section describes how the mechanism can be specified and analyzed on a computer.

## 10.3 MECHANISM BUILDING AND ANALYSIS

### 10.3.1 Mechanism Building

When constructing a mechanism for the synthesis of an API, we usually know the reactants, main products, and some byproducts. From this we can postulate various reaction steps and develop a candidate mechanism. These steps can also be performed automatically with a computer, where standard

chemical reactions or steps can be applied to create a graph connecting the reactants to the products. Automatic generation techniques have been applied for the creation of detailed homogeneous gas-phase reactions for many years [3] and have also been applied to more general reactions including API synthesis [4]. Automatic mechanism generation is outside the scope of this chapter and will not be covered here.

In practice, the experimentalist may only be able to measure a subset of the species actually involved in the reaction network. Also, the exact reaction steps that occur when transforming initial reagents into products, intermediates, and by-products are typically not known exactly. This uncertainty in reaction steps and limited observables creates a situation where several mechanisms may potentially explain the same API synthesis. For these reasons, mechanism building is an iterative process where several candidate mechanisms may be developed and tested. To facilitate this, the software tool used to describe the mechanism should be easy to use and allow rapid mechanism building and modifying. OpenChem supports two main approaches for specifying the synthesis mechanism: tabular and graphical. These are described below.

***10.3.1.1 Tabular Input*** The API synthesis mechanism consists of a collection of one or more reaction formulas and associated rate laws. A natural input for this type of information is a table. Figure 10.3 shows the OpenChem tabular interface for specifying a reaction mechanism.

The user enters into the "New Reaction" input at the top of the form the reaction formula for the reaction to be added to the mechanism, for example, "Cat + ArX → $I_1$." OpenChem will parse this reaction formula to identify products and reactants. If new species are encountered, they are added to the mechanism. The new reaction is added to the mechanism and the user can add additional information, including a reaction identifier, the rate law type and expression, and annotation. The advantage of the tabular form is that it provides a compact representation of the mechanism. The table can also be sorted and queried. For example, only reactions involving a particular species can be listed. One disadvantage of the tabular form is that it can be difficult to visualize the relationships between various species in the mechanism. This drawback is addressed with the graphical view of the mechanism, described below.

***10.3.1.2 Graphical Input*** The OpenChem Pathway Diagram provides a graphical interface for the mechanism. Figure 10.4 shows the Pathway Diagram for the mechanism shown in Figure 10.2. The round-cornered rectangles represent the species in the mechanism. The circles represent the reactions in the mechanism.

An arrow from a species node to a reaction node indicates the species is a reactant, or consumed in the reaction. An arrow from the reaction node to a species node indicates the species is a product, or produced in the reaction. Located on the top left of the OpenChem Pathway Diagram panel is a list of the species in the mechanism. Selecting one or more species in this list will highlight the corresponding species in the graph. Similarly, selecting one or more reactions in the "Reaction List," located below the "Species List," will highlight the corresponding reaction nodes in the graph. This feature is particularly useful when the mechanism is very large. An advantage of the graphical interface of the mechanism is that it provides an intuitive view of the system. Species interactions and reaction channels are more easily identified, providing greater insight into the mechanism. The graphical view also offers additional ways of viewing the simulation results, as described in Section 10.3.3.


**FIGURE 10.3** OpenChem form for specifying a mechanism in tabular format. The reaction formula, rate law type and expression, reaction identifier, and annotation are specified in this tab.

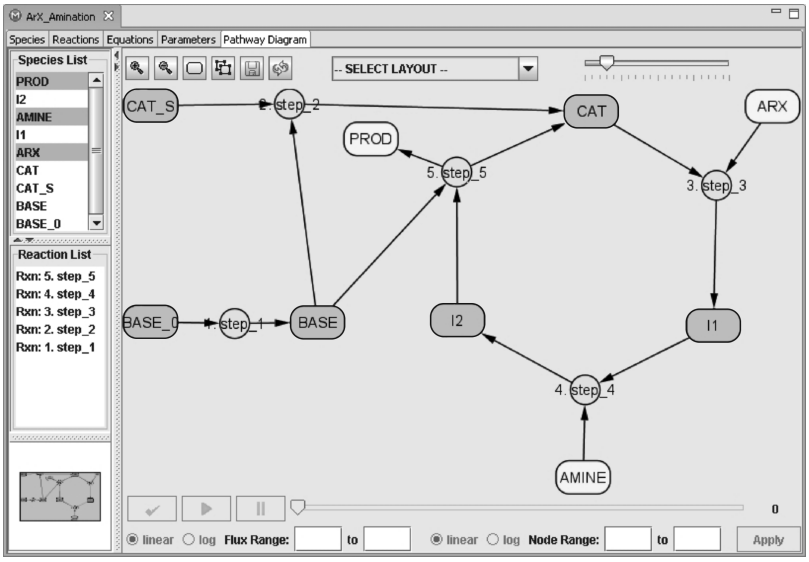



**FIGURE 10.4** OpenChem form for specifying a mechanism in graphical format. Species are denoted by round-cornered rectangles and reactions are denoted by circles. Arrows point from the reactant nodes to the reaction nodes and from the reaction nodes to the product nodes. A list of species and reactions is shown on the left of the form. Selecting one or more species and/or reactions in these lists will highlight the corresponding nodes in the graph.

### 10.3.2 Connecting the API Synthesis Mechanism to the Reactor Model

The previous section describes how to specify the reaction stoichiometry and rates in the API synthesis mechanism. In this section, the mechanism will be coupled with a reactor model to simulate the behavior of the API synthesis. There are many reactor models one may choose to represent the laboratory experiment, pilot or commercial plant. OpenChem provides a library of reactor models, including

- isothermal, isobaric batch reactor for gas;
- isobaric batch reactor for gas;
- isothermal batch reactor for liquid;
- batch reactor for liquid;
- steady-state isothermal plug flow reactor;
- steady-state plug flow reactor;
- transient, isothermal plug flow reactor for gas with heterogeneous catalyst; and
- transient plug flow reactor for gas with heterogeneous catalyst.

OpenChem also provides an input language that can be used to customize the reactor models listed above or create entirely new reactor models. In the example that follows, the liquid-phase isothermal batch reactor will be used to simulate the laboratory experiments.

In general, the synthesis mechanism description is the same regardless of the reactor type. This important observation is exploited in OpenChem by separating mechanism creation from reactor selection. This allows the mechanism to be created once and then used with a variety of reactor types.

***10.3.2.1 Mathematical Formulation for the Batch Reactor Model*** The reaction stoichiometry and rates specified in Section determine the molar production rate of each species, or the time rate of change of species concentration per unit volume. As an example, consider species Base in Figure 10.2: one mole is produced in step 1, one mole is consumed in step 2, and one mole is consumed in step 5. The net rate of change in Base is given by $\dot{\omega}_{\text{Base}} = r_1 - r_2 - r_5$. OpenChem constructs automatically the molar production rates using the reaction formulas and rate laws. The molar production rates for the candidate mechanism are shown below:

$$\begin{array}{lll} \dot{\omega}_{\text{Base}_0} = -r_1 & \dot{\omega}_{\text{Cat}} = r_2 - r_3 + r_5 & \dot{\omega}_{\text{I}_1} = r_3 - r_4 \\ \dot{\omega}_{\text{Base}} = r_1 - r_2 - r_5 & \dot{\omega}_{\text{ArX}} = -r_3 & \dot{\omega}_{\text{I}_2} = r_4 - r_5 \\ \dot{\omega}_{\text{Cat}_s} = -r_2 & \dot{\omega}_{\text{Amine}} = -r_4 & \dot{\omega}_{\text{Prod}} = r_5 \end{array}$$

The mechanism is connected to the reactor model by the molar production rates. From the user-specified mechanism and the selected reactor type, OpenChem will automatically construct the mathematical model describing the reactor. The isothermal, liquid-phase batch reactor equations combined

with molar production rates are then

$$\frac{\mathrm{d[Base]}}{\mathrm{d}t} = \dot{\omega}_{\mathrm{Base}} \qquad \frac{\mathrm{d[Amine]}}{\mathrm{d}t} = \dot{\omega}_{\mathrm{Amine}}$$

$$\frac{\mathrm{d[Cat_s]}}{\mathrm{d}t} = \dot{\omega}_{\mathrm{Cat_s}} \qquad \frac{\mathrm{d[I_1]}}{\mathrm{d}t} = \dot{\omega}_{\mathrm{I_1}}$$

$$\frac{\mathrm{d[Cat]}}{\mathrm{d}t} = \dot{\omega}_{\mathrm{Cat}} \qquad \frac{\mathrm{d[I_2]}}{\mathrm{d}t} = \dot{\omega}_{\mathrm{I_2}}$$

$$\frac{\mathrm{d[ArX]}}{\mathrm{d}t} = \dot{\omega}_{\mathrm{ArX}} \qquad \frac{\mathrm{d[Prod]}}{\mathrm{d}t} = \dot{\omega}_{\mathrm{Prod}}$$

The above equations are a system of eight ordinary differential equations (ODEs) describing the time evolution of the species concentrations in a constant volume, constant temperature batch reactor. OpenChem also allows the user to define additional relationships to constrain variables and provide useful information. In this example, the following two algebraic equations are added to the model:

$$\mathrm{Base}_0 = 1$$

$$y_{\mathrm{ArX}} = \frac{[\mathrm{ArX}]}{[\mathrm{Prod}] + [\mathrm{ArX}]}$$

Algebraic equations do not explicitly contain time derivatives of the state variables. The first algebraic equation specifies that the solid base concentration is fixed at unity. The second algebraic equation defines a new variable, $y_{\mathrm{ArX}}$, equal to the fraction of reactant ArX that has reacted at any given time.

The eight differential equations and two algebraic equations form a system of differential algebraic equations (DAEs) [5]. DAEs offer more flexibility when formulating the problem and the cost of simulating (in terms of CPU time and memory requirements) is roughly the same as solving the stiff ODE systems typically found in chemical reaction simulations.

### 10.3.3 Mechanism Analysis

***10.3.3.1 Dynamic Simulation*** One of the first tasks after defining the mechanism and selecting the reactor type is to simulate the model. Before doing this, however, three pieces of information must be provided: (1) values of the time-invariant parameters (e.g., rate constants), (2) initial conditions for the state or differential variables, and (3) the simulation duration. In this example, the initial concentrations of all the species are zero except reactants ArX and Amine, which have initial concentrations of 1.0 and 1.2, respectively. The dynamic system is simulated for 50,000 s (14 h) from the specified initial conditions. Figure 10.5 shows the OpenChem Simulation Exploring environment.

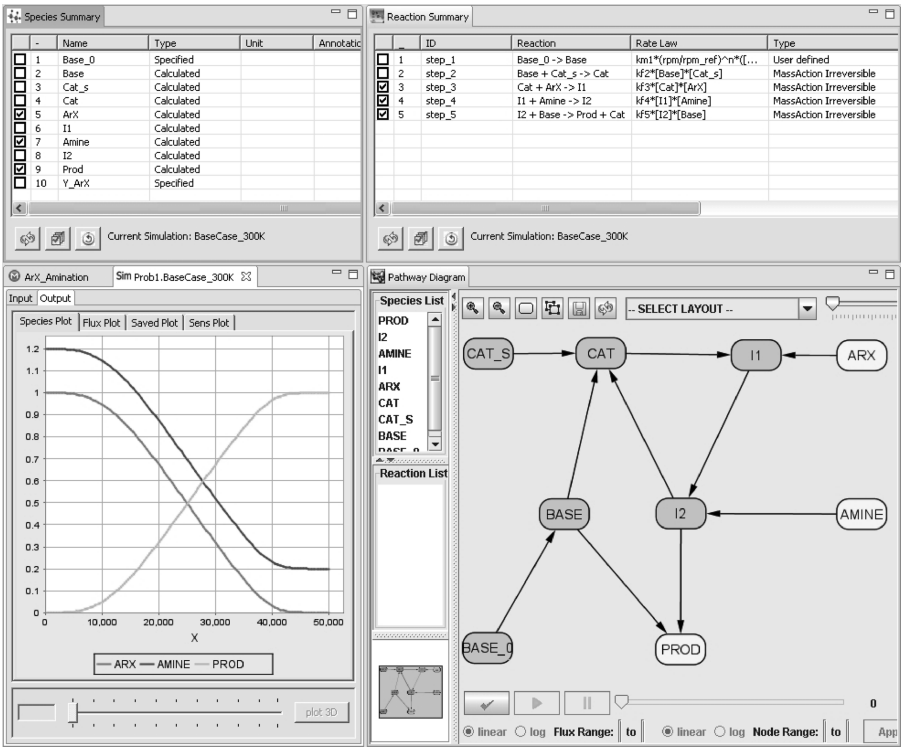



**FIGURE 10.5** OpenChem interactive four-panel Simulation Explorer. The upper left and upper right panels contain a summary of the species and reactions, respectively. The lower left and lower right panels contain the Plotting Environment and Pathway Diagram, respectively. These panels interact, providing a convenient overview of the mechanism behavior.

The Simulation Exploring environment provides an overview of the mechanism and simulation results in four panels. The Species Summary panel is shown in the upper left. The Reaction Summary panel is shown in the upper right. The bottom left and bottom right panels contain the Plotting Environment and Pathway Diagram, respectively. The four panels of the Simulation Explorer are interactive. For example, one or more species may be selected in the Species Summary panel by clicking on the checkbox to the left of the species name. The Reaction Summary panel will be updated and show only the reactions that involve the selected species. The plotting panel (lower left) will be updated, showing the concentration profiles of the selected species. Similarly, if one or more reactions are selected in the Reaction Summary, all species participating in the selected reactions will be displayed in the Species Summary. Further, the Flux Plot tab in the Plotting Environment panel will display the flux profiles for all selected reactions.

Figure 10.6 shows more closely (a) concentration profiles for species ArX, Amine, Prod, $I_1$, and $I_2$ and (b) flux profiles for reactions step_1, step_2, and step_5. The species profiles indicate how the concentration changes in the reactor as a function of time. Reactants ArX and Amine monotonically decrease, product Prod monotonically increases, and intermediates $I_1$ and $I_2$ exhibit a maximum value at an intermediate time. Further, $I_2$ was produced in higher concentration than $I_1$. From this plot (Figure 10.6), we can see that reactant ArX is the limiting reagent and the reaction goes to completion. The flux profiles indicate how the reaction flux changes as a function of time. Visualizing the reaction fluxes allows the reactions to be compared to determine which reactions are most important. The reaction flux analysis described next provides an alternative method for identifying important reaction channels.

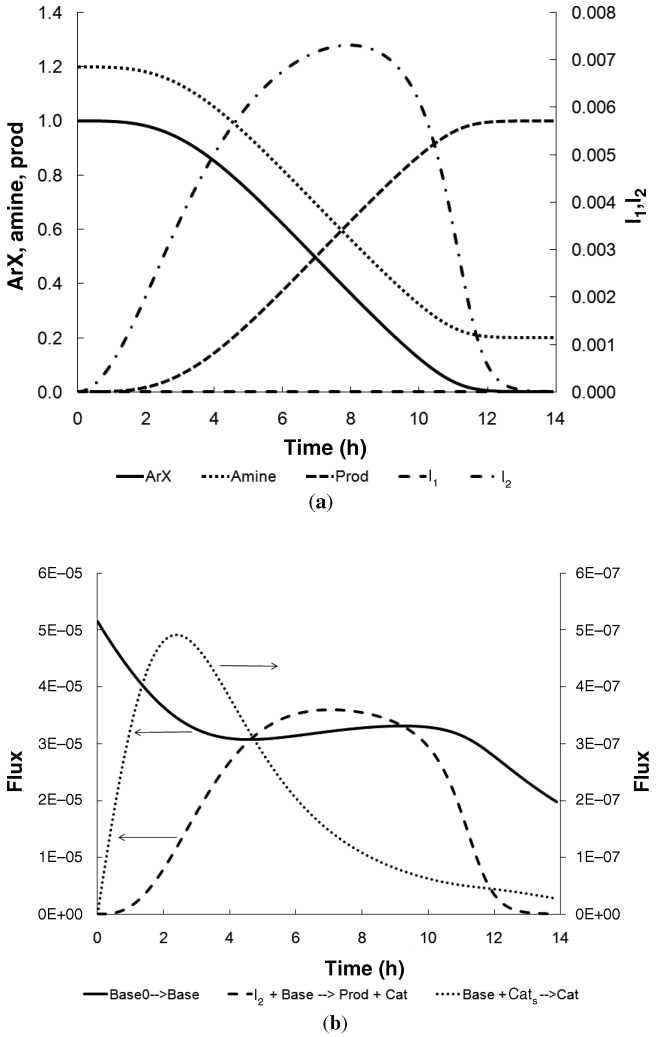
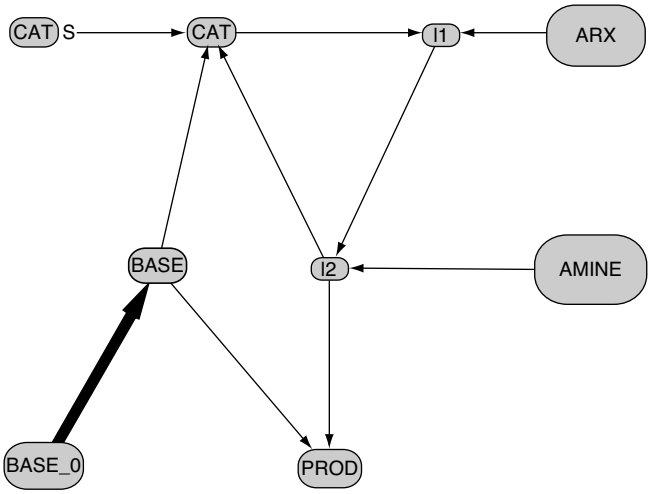



**FIGURE 10.6** Selected species and flux profiles for the synthesis mechanism shown in Figure 10.2. Part (a) shows the concentration profiles for species ArX, Amine, Prod, $I_1$, and $I_2$. Part (b) shows the flux profiles for reactions $Base_0 \rightarrow Base$, $I_2 + Base \rightarrow Prod + Cat$, and $Base + Cat_s \rightarrow Cat$.

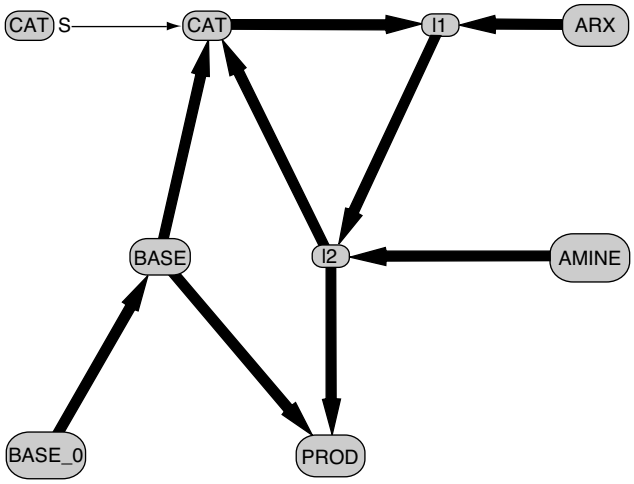
***10.3.3.2 Reaction Flux Analysis*** The Pathway Diagram panel of the Simulation Explorer provides another way to visualize the results of a simulation. Selecting the checkmark button in the lower left corner of the Pathway Diagram panel enables the flux animation. Rather than viewing time profiles of the species and/or fluxes as shown in the previous section, the flux animation displays the reaction progress by changing the size of the species nodes and thickness of the reaction arrows. The size of the species node is proportional to the species concentration and the thickness of the arrow is proportional to magnitude of the flux. Figure 10.7 shows a snapshot of the flux animation at 1, 7, and 14 h of simulation time.

At 1 h of simulation time, the primary reaction occurring is the dissolution of the solid base ($Base_0 \rightarrow Base$). The arrow in the flux diagram for this reaction is thick and the nodes corresponding to the reactants ArX and Amine are large, indicating high concentration. At 7 h, sufficient solid base has dissolved and all reactions are proceeding. Finally, at the end of the simulation, after 14 h, reactants ArX and Amine have reduced and product Prod is formed. The animation can be played by selecting the run button in the lower left corner and stopped at any time using the stop button. The sliding bar at the bottom of the Pathway Diagram can be positioned to view a snapshot of the reaction progress at any time in the simulation.

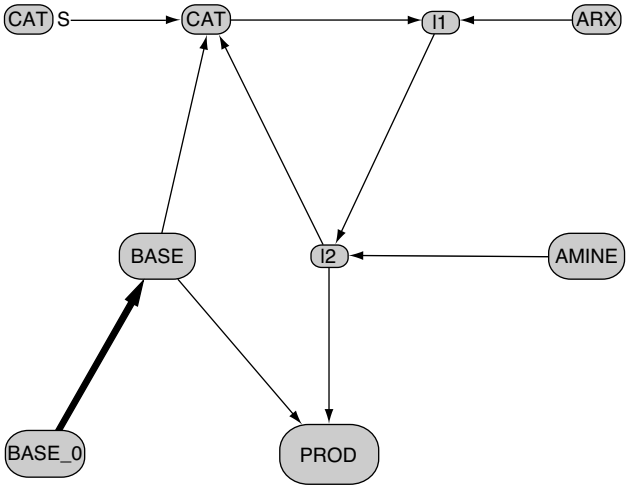
When viewing the flux profiles as described in the previous section, individual reactions can be compared and important reactions identified. The reaction flux analysis, however, can be used to identify important reaction channels, that is, groups of reactions connecting reactants to products via intermediates. This is difficult to grasp when viewing individual reaction fluxes. This analysis is helpful for identifying important intermediates and pathways as well as finding bottlenecks or rate-determining pathways.


CAT S
CAT
I1
ARX
BASE
I2
AMINE
BASE_0
PROD

(**a**) 1 h


CAT S
CAT
I1
ARX
BASE
I2
AMINE
BASE_0
PROD

(**b**) 7 h


CAT S
CAT
I1
ARX
BASE
I2
AMINE
BASE_0
PROD

(**c**) 14 h

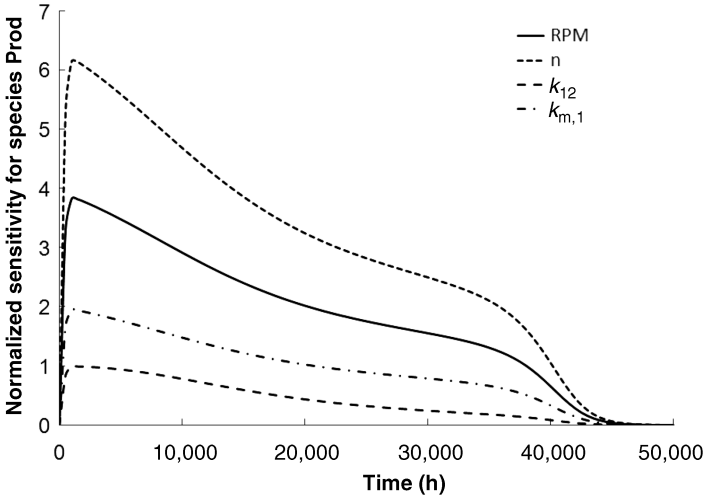



**FIGURE 10.8** Normalized parametric sensitivities for species Prod with respect to several parameters.

***10.3.3.3 Parametric Sensitivity Analysis*** In addition to computing species concentration and flux profiles, OpenChem can also compute the parametric sensitivities of the dynamic system. The parametric sensitivities are defined as

$$s_{ij}(t) = \frac{\partial y_i}{\partial p_j}$$

Here, $y_i$ is a model output and $p_j$ is a time-invariant parameter. This quantity is called the differential sensitivity. OpenChem can also compute the normalized sensitivities, defined as

$$\tilde{s}_{ij}(t) = \frac{\partial \ln y_i}{\partial \ln p_j} = \frac{p_j}{y_i}\frac{\partial y_i}{\partial p_j}$$

Sensitivity analysis has traditionally been used to determine how the parameters influence the model predictions [6]. This calculation involves solving a dynamic system (related to the original model) that computes the partial derivatives of the model outputs with respect to the time-invariant parameters as a function of time. The sensitivity trajectories can be compared to determine which parameters have the greatest influence on the outputs of interest and when during the simulation they are important. Figure 10.8 shows the normalized sensitivities for species Prod with respect to several parameters. Notice that the value of Prod is most sensitive to the parameter values early during the simulation. The sensitivity goes to zero as Prod reaches its final value.

## 10.4 MECHANISM CALIBRATION

When constructing a candidate API synthesis mechanism, there will be a number of time-invariant parameters with unknown values. For example, the values of rate constants in the mechanism shown in Figure 10.2, $k_{m,1}$, $k_{f,2}$, $k_{f,3}$, $k_{f,4}$, and $k_{f,5}$, will likely be unknown at the beginning of mechanism development. These parameters must be inferred from experimental data through the process of mechanism calibration. Mechanism calibration involves the following tasks:

- identifying the parameters in the model that strongly influence the model outputs of interest;
- collecting necessary experimental data through a series of experiments; and
- adjusting systematically the important parameters so that the model outputs best fit the experimental data.

**FIGURE 10.7** OpenChem flux animation. This feature provides an additional view of the simulation results, showing how the reaction progresses in time by changing the relative sizes of the species nodes (to represent concentration) and thickness of the arrows (to represent flux magnitude) as an animation. Part (a) shows the reaction progress after 1 h of simulation time. Part (b) shows the reaction progress after 7 h of simulation time. Part (c) shows the reaction progress after 14 h of simulation time. For more complex mechanisms, this analysis is useful for identifying important intermediate species and reaction pathways.

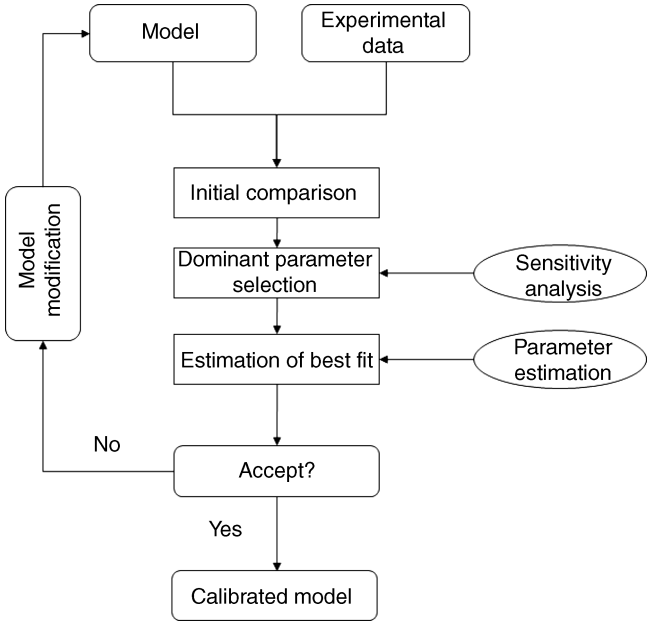



**FIGURE 10.9** Calibration workflow. The major steps involved in model calibration are illustrated in this figure. Techniques that can be applied are shown in the ellipses. An important part of this workflow is model modification; deficiencies in the model and/or data are often uncovered during model calibration.

There are several challenges that must be addressed when calibrating a mechanism. The process of identifying the important parameters is complicated by the fact that the parameter values are unknown. An uncalibrated model can play an important role in determining what experiments should be performed and under what conditions, provided rough estimates for the parameter values and reasonable upper and lower bounds can be supplied. Mechanism calibration is further complicated by the fact that experimental data are often in a form that is not directly comparable to the model. For example, the model may involve species concentrations whereas the data may be spectral measurements (e.g., ultraviolet and infrared). The ability to effectively handle mixed data types is an important part of model calibration. The process of systematically adjusting the parameter values to best fit the data is referred to as parameter estimation. A well-known issue with parameter estimation is that the parameter estimates obtained often depend strongly on the initial values provided for the parameters being adjusted.

### 10.4.1 Calibration Workflow

An effective workflow for model calibration involves identification of the important parameters, parameter estimation with special provisions for handling mixed data and ensuring that the best possible fit is obtained, and unbiased discrimination between multiple calibrated candidate mechanisms. Figure 10.9 shows a schematic of one calibration workflow.

At the top of Figure 10.9 are the candidate mechanism and experimental data. The first step is to match the model outputs with the experimental data and perform an initial comparison. The next step is to identify the dominant parameters in the model. Only the parameters that influence the model outputs corresponding to the data can be effectively calibrated, so correctly identifying these parameters is an important step of model calibration. Computations such as sensitivity analysis assist the user with this selection. The next step is to adjust the parameters to best fit the data. If successful, the model and estimated parameters can be analyzed to determine the quality of the fit. In many cases, the process of model calibration can identify weaknesses in the model and/or experimental data. An important part of model calibration is to modify the model and/or collect new experimental data to ensure that the best possible model is obtained. The remainder of this section describes this workflow as implemented in OpenChem.

### 10.4.2 Parameter Identification

The time-invariant parameters in a model influence the model predictions in various ways. Some parameters strongly influence the model outputs of interest whereas others have a limited effect. This becomes particularly true as the size of the mechanism and the number of parameters increase. As described earlier, parametric sensitivity analysis can be used

to quantitatively compare the influence of several parameters. Unfortunately, the approach described, referred to as a local sensitivity analysis, requires that the values of the parameters be specified prior to performing the calculation. Since these parameters are unknown or uncertain prior to calibration, the results of the sensitivity analysis depend on the initial estimates for the parameters.

To address the uncertainty in parameter values, OpenChem implements global sensitivity analysis (GSA) approach. This algorithm begins with the user selecting the parameters of interest and providing an appropriate range for the parameter values (e.g., upper and lower bound on each parameter). Even when the actual parameter values are not known, it is often possible to select reasonable values for the bounds. Next, values for the parameters are sampled from the parameter space defined by the bounds. This may be as simple as a uniform random sampling between the upper and lower bounds or a more complex sampling based on an *a priori* knowledge of the likelihood of the parameter values. A local sensitivity analysis is performed for each sample of parameter values. Each sensitivity profile is then converted to a scalar sensitivity metric, for example, maximum absolute value on the profile or integral of the absolute value of the profile:

$$\max_{0\le t\le T}|s(t)| \quad \text{or} \quad \int_0^T |s(t)|\mathrm{d}t$$

Here $s(t)$ denotes the parametric sensitivity profile. This sampling is repeated many times and the scalar metric of each sensitivity trajectory is averaged over all samples. Figure 10.10 shows the results of GSA applied to the synthesis mechanism described in Figure 10.2.

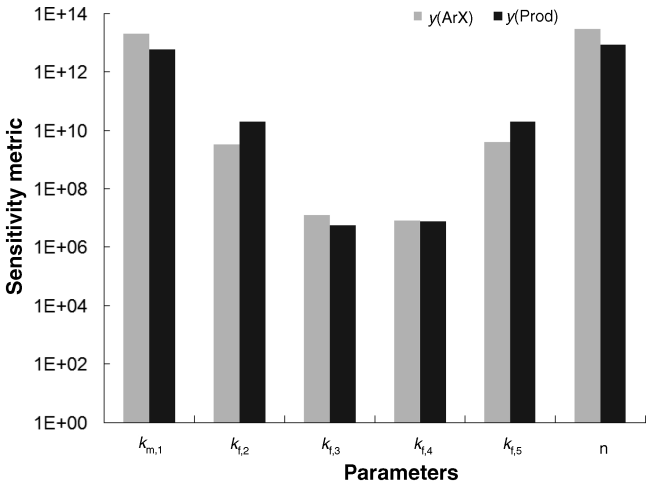



**FIGURE 10.10** Global sensitivity analysis results for the synthesis mechanism shown in Figure 10.2 for species ArX and Prod. The importance of the parameters, as measured by the averaged sensitivity metric, is valid not for a single set of parameter values but rather a four-order magnitude range of parameter values.

In small mechanisms such as the example shown above, almost all parameters have a strong influence on the few model outputs. Figure 10.11 shows the results of GSA applied to a larger mechanism. In this case, there are clearly groups of parameters that have a greater effect on the selected model outputs. For example, the parameters $k_1$ and $k_4$ have strong influence on concentration $C(1)$ while the parameters $k_2$ and $k_3$ have no influence at all.

### 10.4.3 Parameter Estimation

#### *10.4.3.1 Parameter Estimation with Concentration Data*

Calibration with concentration data is relatively well established and there are a number of techniques available [7]. The most widely used are those based on a descent direction of the calibration objective function, such as the Levenberg–Marquardt method. OpenChem implements a control parameterization approach for solving parameter estimation problems for dynamic models. In this case, the discrepancy between the time series experimental data and the model predictions (i.e., the calibration objective function) is computed by numerically integrating the dynamic system. The gradient of the objective function is determined by computing the local parametric sensitivities with respect to the parameters being estimated. The scalar objective function is minimized by applying a nonlinear programming algorithm, such as successive quadratic programming [8].

One well-known problem of descent direction methods like those described above is that they often converge to locally optimal solutions. That is, parameters are estimated such that they are the best fit in some neighborhood, but a better fit might be obtained with parameters outside this neighborhood. As a result, the parameter fits obtained often depend on the initial values provided for the parameters. OpenChem addresses this dependence on initial parameter values by applying a multistart parameter estimation approach. Like the GSA method described above, multistart parameter estimation begins with the user providing appropriate lower and upper bounds for the parameters. Values for the parameter to be estimated are then sampled between the lower and upper bounds to provide the starting guess for a parameter estimation calculation. The sampling procedures can be as simple as a Cartesian grid search or more complex, like starting from a set of seed values generated by examining the objective function during GSA. Upon successful completion of each run, the final objective function and optimal parameter estimates are recorded. Figure 10.12 shows the final objective function values versus sample iteration number for 46,656 runs, sorted by decreasing value of objective function.

The iteration number on the $x$-axis corresponds to a different starting value for the parameters being estimated. Figure 10.12 shows a number of flat, plateau regions. These

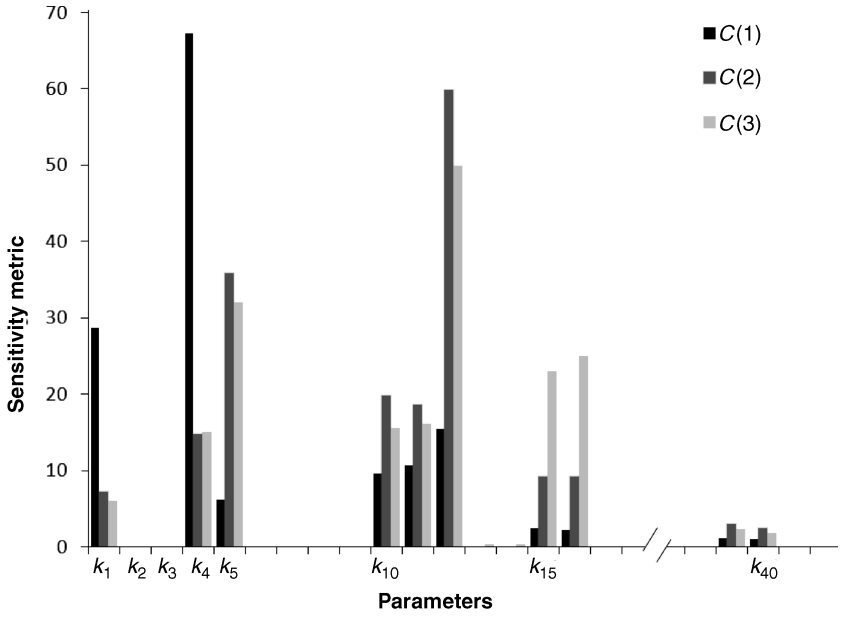



**FIGURE 10.11** Global sensitivity analysis results for a reaction mechanism involving 30 species and 42 parameters with results shown for three key model outputs, $C(1)$, $C(2)$, and $C(3)$. The importance of the parameters, as measured by the averaged sensitivity metric, is valid not for a single set of parameter values but rather a four-order magnitude range of parameter values.

correspond to locally optimal parameter estimates that are obtained from several starting values for the parameters. This phenomenon is typical of many parameter estimations encountered when calibrating kinetic mechanisms. Two regions on Figure 10.12 have been marked A and B, corresponding to two distinct estimates for the parameter values. Overlay plots for these two estimates are shown in Figure 10.13.

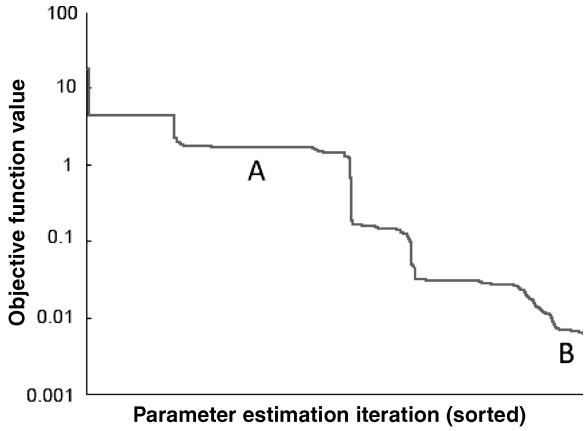



**FIGURE 10.12** Parameter estimation objective function versus multistart iteration number (sorted) for 46,656 samples. The flat, plateau regions correspond to locally optimal solutions that were found from several initial parameter values. Overlay plots corresponding to two parameter estimates, marked A and B, are shown in Figure 10.13.

In addition to multistart parameter estimation, other optimization methods such as genetic and simulated annealing algorithms are also available in OpenChem.

***10.4.3.2 Parameter Estimation with Mixed Concentration and Spectra Data*** The calibration method described above utilizes concentration data to compare model predictions directly to experimental values. However, most of the experimental data collected in practice are in a form other than concentration, such as spectra (e.g., ultraviolet and infrared) and HPLC (high-performance liquid chromatography). To use these data with traditional calibration methods, the measurements are first converted to concentration data. This process is time consuming, error prone, and sometimes not possible, for example, if the pure species in the system have similar or overlapping spectra. Calibration of the synthesis mechanism using concentration data is often referred to as a hard modeling approach. In contrast, soft modeling approaches utilize directly other forms of data, like spectra and HPLC, and can cope with data not applicable to hard modeling approaches [9, 10]. Curve resolution methods such as MCR-ALS (multivariate curve resolution–alternating least squares) are a category of soft modeling techniques.

The MCR-ALS algorithm represents a relatively recent advancement in calibrating the mechanism with spectra data. The basic idea is that the user provides experimentally measured spectra data ($D$) and initial guesses for the pure component spectra ($S_0$) and/or initial guesses for the species

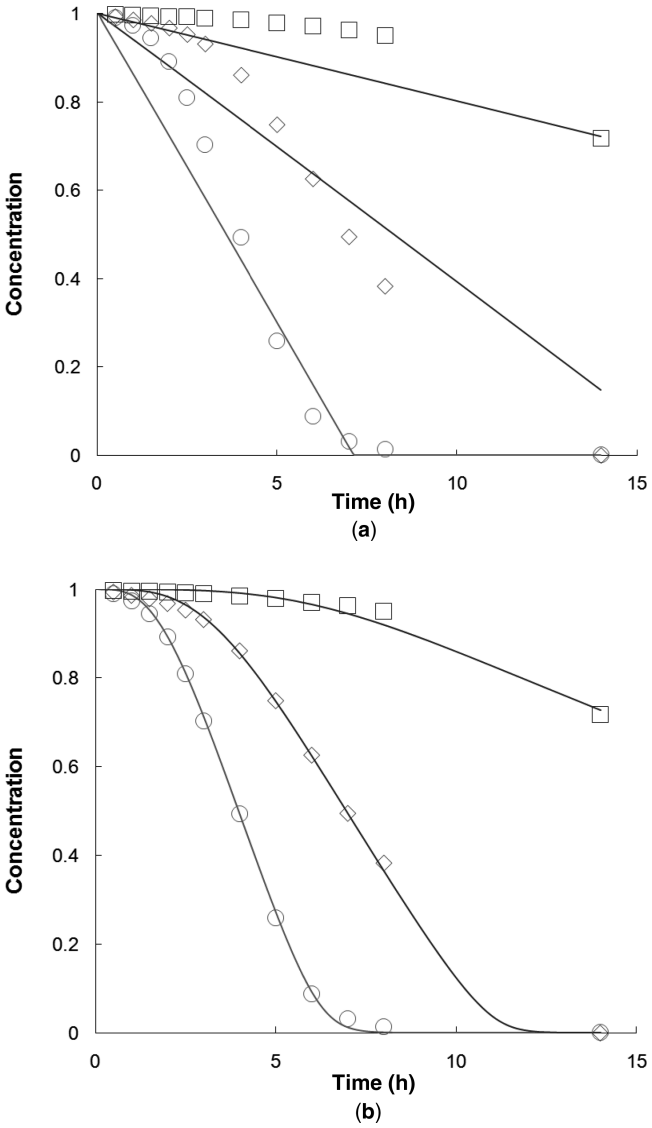



**FIGURE 10.13** Results of the OpenChem multistart parameter estimation for the mechanism shown in Figure 10.2. Part (a) shows the overlay plot for a locally optimal, but poor fit. Part (b) shows the overlay plot for the best parameter fit obtained and is clearly superior to the other fit.

concentration profiles ($C_0$). The algorithm solves a sequence of linear least square problems, attempting to resolve the actual pure component spectra ($S$) and actual concentration profiles ($C$), which are related to the data according to the Beer–Lambert law [11]:

$$D = CS^{\mathrm{T}} + E$$

where $E$ is the experimental error. At each iteration, constraints are applied to $C_{(k)}$ and $S_{(k)}$ ($k$ is the current iteration number) to ensure that physically meaningful solutions are obtained. Also at each iteration, a standard parameter estimation is performed with the current $C_{(k)}$ to estimate the kinetic rate constants in the model.

When data from multiple sources are available, there are several advantages to using simultaneously all data sources when calibrating the mechanism, including

- more information is available leading to potentially better fits;
- estimated parameters will be consistent with all data; and
- combining all data sources in a single step simplifies the calibration and streamlines the calibration workflow.

OpenChem implements a modification of the MCR-ALS algorithm that simultaneously enables calibration with mixed spectra and concentration data [12]. The basic structure of this algorithm is similar to that of MCR-ALS described above, except that as part of one of the linear least squares problems, the concentration data are directly utilized in a hard modeling subproblem. In addition, ideas from the multistart parameter estimation approach described above are incorporated to ensure that the best possible fit is obtained.

The mixed data algorithm is illustrated with an example shown below involving three species and three reactions as shown in Figure 10.14.

Figure 10.15 shows the concentration and spectra data input used for calibration and optimal estimate output of the modified MCR-ALS algorithm. Figure 10.15a shows the concentration data for the three species in this mechanism. Figure 10.15b shows the spectra data for three measured wavelengths (this is the contents of matrix $D$ described above). The concentration and spectra data are used simultaneously in the modified MCR-ALS algorithm to produce the fit shown in Figure 10.15c, where the simulated species profiles using the optimal estimates for the parameter values (solid and dashed lines) are plotted with the concentration profiles (square, circle, and triangle markers).

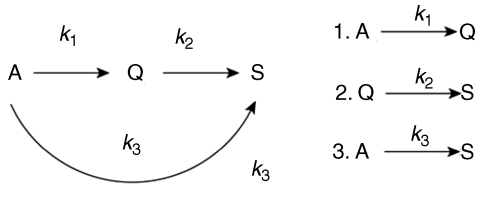



**FIGURE 10.14** Schematic of a synthesis mechanism involving three species and three reactions. Reactant A is converted to main product, S, directly and through an intermediate species Q.

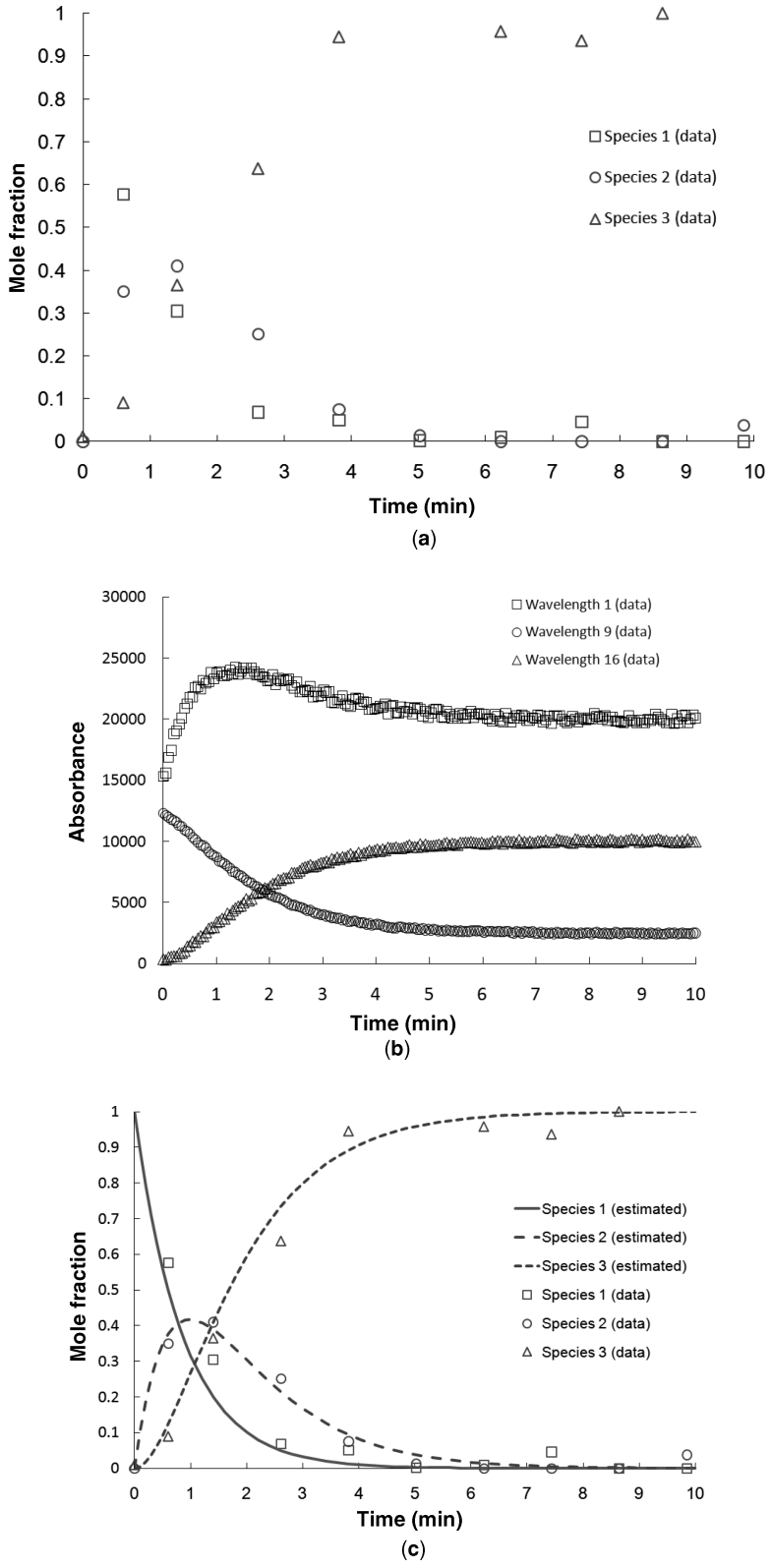



**FIGURE 10.15** Input data and results for the mixed data calibration method implemented in OpenChem. Part (a) shows the measured concentration data (algorithm input). Part (b) shows the measured spectra data (algorithm input). Part (c) shows the overlay plot for the optimal solution (simulated profiles and measured concentration data).

## 10.5 MECHANISM APPLICATION

The above sections described how to build, analyze, and calibrate an API synthesis mechanism. Once calibrated, there are several potential applications of the synthesis mechanism in pharmaceutical manufacturing, including

- scale-up of laboratory procedures;
- detailed design of manufacturing equipment;
- optimization of process equipment and operation;
- design of process control strategies for safety and quality assurance; and
- process improvements.

Scale-up involves taking a chemist's recipe for synthesizing an API in a test tube and creating a manufacturing process that is able to mass produce the API. Effects such as mixing, heat transfer, and by-product formation that are trivial or neglected in the laboratory can have a significant effect in large-scale production. The calibrated API synthesis mechanism can help an engineer determine early in the development process whether it is economically feasible to mass produce the API. Once scale-up procedures have been developed, the synthesis mechanism can be used for detailed design of manufacturing equipment and process operation. For example, accurately determining how operating conditions such as temperature, pressure, and residence time affect reaction conversion and selectivity is critical for designing reactors and separators in the process. After the individual process equipment has been designed, a model of the process, utilizing the API synthesis mechanism, can be used to optimize the process operation and design control strategies necessary for safety and product quality assurance.

The remainder of this section describes three features in OpenChem that facilitate the application of the API synthesis mechanism:

- Operating map generation
- OpenChem scripting
- Application programming interface to third-party software tools

### 10.5.1 Operating Map Generation

OpenChem automates the task of running multiple simulations for more common operations such as operating map generation. An operating map shows visually how key model outputs depend on multiple operating conditions and parameters. Operating conditions may include reactor residence time, operating temperature, agitator speed, and so on. Model outputs may include product yield, product selectivity, maximum temperature rise, and so on. The operating map is useful for a variety of tasks, including

- visualizing trade-offs between multiple operating conditions;
- providing a visual, more intuitive view of parameter sensitivity than that provided by the sensitivity analysis described above; and
- identifying optimal operating conditions.

In OpenChem, the user is able to select two operating conditions or parameters and one model output. With these user-specified selections, OpenChem will automatically execute a series of simulations by varying the specified operating conditions/parameters and plotting the selected model output in a contour plot.

Figure 10.16 shows the operating map for a model of a batch reactor and the synthesis mechanism described in Figure 10.2. The operating conditions for the operating map in Figure 10.16 are agitator speed ($x$-axis) and catalyst loading ($y$-axis). The model output is the processing, or residence, time necessary to achieve 98% product yield.

### 10.5.2 OpenChem Scripting

Application of the API synthesis mechanism typically involves running the reactor simulations and/or optimizations many times under a variety of conditions. OpenChem provides a scripting language enabling this task. The language allows the user to write high-level commands for sweeping through various operating conditions and parameter values, executing multiple simulations, plotting results, and writing files summarizing the analysis in any customized manner. For example, operating map generation is enabled by this OpenChem scripting language.

### 10.5.3 Interface to Third-Party Software

Finally, OpenChem implements an application programming interface that enables the synthesis mechanisms to be utilized within other third-party software. One example of this is developing a synthesis mechanism in OpenChem, and then utilizing this mechanism in a detailed reactor design implemented in a CFD (computational fluid dynamics) software package.

## 10.6 CONCLUSION

The process of building, analyzing, calibrating, and applying API synthesis mechanisms is described. Building API synthesis mechanisms is a challenging task for several reasons, including (1) the reaction steps involved are usually not

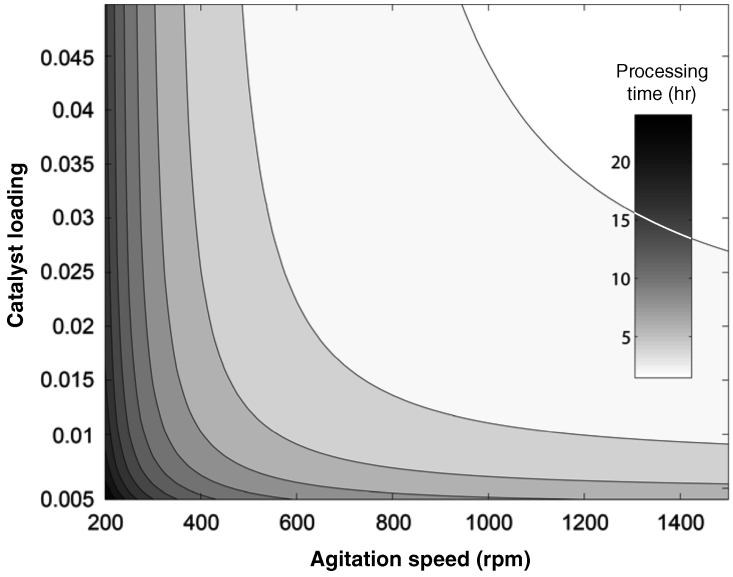



**FIGURE 10.16** Operating map for reactor operation for the synthesis described in Figure 10.2. The $x$-axis is the agitator speed and the $y$-axis is the catalyst loading. The contour plot shows the residence time in hours required to achieve the desired product yield. This figure indicates that operation is most sensitive to agitator speed at low rpm.

known with certainty and (2) all intermediates and by-products are not measured (or even known). Because of this uncertainty, it is often possible to postulate multiple candidate mechanisms to describe a synthesis. These mechanisms contain several parameters, for example, reaction rate constants, with highly uncertain values. Further, these parameters have different influence on outputs of interest. We must decide which parameters are important and which experiments must be performed in order to calibrate the mechanism. Mechanism calibration is complicated by the fact that the data are often not in a form suitable for traditional calibration methods and the quality of the fit obtained is often dependent on the initial estimates of the uncertain parameters.

The OpenChem software has been designed to address the issues described above and facilitate the mechanism development process. OpenChem provides several ways for specifying the mechanism, tabular and graphical. These two mechanism views offer various advantages, including compact representation of the reactions and easy visualization of the interactions between species in the mechanism. A collection of reactor models is provided and new reactors can be easily added. The results of simulations can be viewed as regular two- and three-dimensional plots as well as a graphical flux analysis animation. The problem of identifying important parameters in the presence of uncertainty is addressed through global sensitivity analysis. In OpenChem, multiple disparate data sources (e.g., concentration and spectra) can be used simultaneously to calibrate the mechanism and special provisions are applied to increase the likelihood that the best possible parameter fit is obtained. Finally, several features are provided for mechanism application, including operating map generation, a scripting language for automating tasks, and an interface to third-party software.

## NOTATION

| | |
|---|---|
| $A$ | Arrhenius expression pre-exponential factor |
| Amine | amine |
| ArX | aryl halide |
| Base | base |
| $C$ | actual or estimated concentration profiles |
| Cat | catalyst |
| $C_0$ | initial guesses for species concentration profiles |
| $D$ | experimentally measured spectra data |
| $E$ | experimental error matrix |
| $E_A$ | Arrhenius expression reaction activation energy |
| $I_1$, $I_2$ | intermediate species |
| $k_{m,1}$, $k_{f,i}$ | reaction rate constants |
| $k(T)$ | reaction rate constant |
| $n$ | parameter indicating how the solid base dissolution rate scales with agitator speed |
| $p_j$ | time-invariant parameter |
| Prod | product |
| $R$ | gas constant |
| $r_i$ | rate of reaction $i$ in moles per unit time per unit volume |
| RPM | agitator speed |
| $RPM_{ref}$ | reference agitator speed |
| $S$ | actual or estimated pure component spectra |

| | |
|---|---|
| $s(t)$ | parametric sensitivity profile |
| $s_{ij}(t)$ | parametric sensitivity of output variable $i$ with respect to parameter $j$ |
| $\tilde{s}_{ij}(t)$ | normalized parametric sensitivity of output variable $i$ with respect to parameter $j$ |
| $S_0$ | initial guesses for the pure component spectra |
| $T$ | temperature |
| X | halide |
| $y_{\mathrm{ArX}}$ | fraction of reactant ArX that has reacted at any given time |
| $y_i$ | model output |
| $[\cdot]$ | species concentration in moles per unit volume (e.g., [ArX] and [Cat]) |
| $\beta$ | Arrhenius expression temperature exponent |
| $\dot{\omega}_i$ | molar production rate of species $i$ in moles per unit volume per unit time |

## REFERENCES


1. Innovation or stagnation: challenge and opportunity on the critical path to new medical products, FDA White Paper, 2004.
2. RES Group, Inc. OpenChem: software for kinetic mechanism development and reactor design, RES Internal Report, June 2007.
3. Green WH, Barton PI, Matheu DM, Schwer DA, Song J, Sumathi R, Carstensen HH, Dean AM, Grenda JM. Computer construction of detailed chemical kinetic models for gas-phase reactors. *Ind. Eng. Chem. Res.*, 2001;40 (23):5362–5370.
4. Hopkinson GA.Computer-assisted organic synthesis design, Ph.D. thesis, University of Leeds, 1985.
5. Brenan KE, Campbell SL, Petzold LR. *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*, SIAM, New York, 1996.
6. Feehery WF, Tolsma JE, Barton PI. Efficient sensitivity analysis of large-scale differential-algebraic systems, *Appl. Numer. Math.*, 1997;25 (1):41–54.
7. Bard J. *Nonlinear Parameter Estimation*, Academic Press, New York, 2001.
8. Bazaraa MS, Sherali HD, Shetty CM. *Nonlinear Programming: Theory and Algorithms*, 2nd edition, Wiley, 1993.
9. de Juan A, Maeder M, Martinez M, Tauler R. Combining hard- and soft-modeling to solve kinetic problems. *Chemometr. Intell. Lab. Syst.*, 2000;54: 123–141.
10. Bezemer E, Rutan SC. Multivariate curve resolution with nonlinear fitting of kinetic profiles, *Chemometr. Intell. Lab. Syst.*, 2001;59: 19–31.
11. Ingle JDJ, Crouch SR. *Spectrochemical Analysis*, Prentice Hall, New Jersey, 1988.
12. RES Group, Inc. Calibration of kinetic models with combined concentration and spectra data, RES Internal Report, March 2008.