Heuristic Search Algorithms for the Determination of Rate Constants and Reaction Mechanisms from Limited Concentration Data

David B. Terry and Michael Messina*

Department of Chemistry, University of North Carolina at Wilmington, Wilmington, North Carolina 28405

We present a chemical kinetics search algorithm that can be used to predict rate constants and reaction orders for chemical reactions from a limited set of concentration versus time data. The algorithm is based on the functional optimization of a well-defined error function. This error function is defined as the squared difference between the known set of limited concentration vs time data and the concentrations that are computed from a postulated set of rate constants and reaction orders. The best sets of rate constants and reaction orders are found by minimizing this error function. The error function contains many local minima, and therefore a heuristic search algorithm is used to find a useful local minimum in the error function. In this work we critically compare the performance of the chemical kinetics search algorithm when two different heuristic optimization schemes are used: the Genetic Algorithm and Simulated Annealing. We test the search algorithm on three types of chemical reactions: a consecutive reaction with a reversible step, a parallel reaction, and a reaction that represents the chemical fate of a pollutant in the environment.

I. INTRODUCTION

Many important chemical reactions have complex reaction mechanisms that involve multiple elementary steps. The complexity of these reactions often makes it difficult to experimentally measure complete concentration vs time data for all the species involved in the reaction. The lack of such data means that there are three important pieces of chemical information missing about the reaction: (1) the reaction mechanism, (2) the complete set of rate constants that govern each step in the elementary reaction, and (3) the concentration vs time profiles for each of the species involved in the reaction for all times. Experimentalists often do have some set of limited concentration vs time data for reactants and products for some of the elementary reaction steps. In this work we develop a chemical kinetics search algorithm that uses this limited concentration vs time data to obtain the reaction mechanisms, the complete set of rate constants, and the concentration profiles for all species.

We formulate a problem here that is rather different from the conventional formulation of chemical kinetics. Conventionally in chemical kinetics one is handed a well-defined reaction mechanism and a set of known rate constants that governs the elementary steps in the reaction mechanism and uses this information to determine an overall concentration vs time profile for the reactants and products. Here we will present a solution to what can be considered the inverse problem in chemical kinetics: "Given a set of limited concentration vs. time data, what are the values of the complete set of rate constants that give rise to these reaction profiles, and what is the overall reaction mechanism?" Although this problem is rather easy to state it can be rather difficult, and in fact impossible to solve unless it is confronted carefully.

In a previous work we presented a computational strategy that can be used to determine rate constants from limited concentration data for a chemical reaction *if supplied with a well-defined reaction mechanism.*[1] The general idea behind the strategy is to define an error function that is the difference between experimentally measured and computed concentration vs time data. The computed concentration vs time data are found by solving the kinetics equations for the reaction based on a set of guessed rate constants. The best set of rate constants is then found by minimizing this error function. Since the error function can contain many local minima, we have found it is extremely hard to minimize by conventional minimization methods such as gradient descent or the downhill Simplex Algorithm. In this previous work we developed a computational strategy that utilized the Genetic Algorithm (GA) to find a useful local minimum in this error function,[1] i.e., one that leads to accurate values of the unknown rate constants. The GA[2−5] distinguishes itself as one of several heuristic optimization methods, such as Simulated Annealing (SA)[6−9] and Tabu search,[10−13] that has the ability to locate a minimum of a function near enough to the global minimum to yield accurate results without getting stuck in local minima.

The purpose of this work is 2-fold: (1) to develop a related computational strategy based on a heuristic search algorithm that can predict the set of rate constants *and the reaction mechanism* from a set of limited concentration vs time data for a chemical reaction and (2) critically compare the relative efficiencies between the two most ubiquitous and most widely available heuristic search algorithms, these being SA and the GA.

We want to define very carefully here what we mean by determining the unknown reaction mechanism. In this work we assume we know the overall reaction scheme. We define the reaction scheme as the set of elementary chemical equations that governs the overall chemical reaction mech-

* Corresponding author e-mail: messinam@uncwil.edu; fax (910) 962-3013.

Search Algorithms from Limited Concentration Data

J. Chem. Inf. Comput. Sci., Vol. 38, No. 6, 1998 **1233**

anism. A well-defined reaction mechanism is obtained from the reaction scheme *if all the stoichiometric coefficients of all the reactants and products are known.* In this work we assume that the reaction scheme is known, but not the overall reaction mechanism because we treat some of these stoichiometric coefficients as unknowns. The stoichiometric coefficients of the products and reactants in the elementary reactions that comprise the overall reaction mechanism translate into the reaction orders of each of these species in the kinetic equations.

In this work we will study the relative efficiencies between chemical kinetics search algorithms based on the SA and GA for several of the most common types of chemical reactions. These include a consecutive reaction with a reversible step, a parallel reaction, and a reaction representing the chemical fate of a pollutant in the environment. For each of these reactions we use a synthesized set of limited concentration vs time data for each of the reacting species and then compare the relative efficiencies of the SA and GA based chemical kinetics search algorithms in finding the complete set of rate constants that govern the overall reaction and the orders of some of the reactants and products in the elementary reaction steps.

This paper is organized as follows. In section II we sketch the general theory and equations that govern the general chemical kinetics search algorithm and the reaction schemes for the three types of reactions to be studied. In section III we present numerical results that include the computational details of the algorithm and numerical results for each of these reactions. In section IV we discuss the results and conclude.

II. THEORY

(A) General Theory. Given a general reaction scheme we define three vectors: $\vec{c}(t)$, the concentration of all the species at time t, $\vec{k}$, the set of rate constants governing the reaction, and $\vec{v}$, the set of unknown reaction orders in the reaction mechanism. The *i*th vector component of $\vec{c}$ (t), $c_i(t)$, represents the concentration of species i at time, t. The kinetic equation that governs the time evolution of the concentration of species i is

$$\dot{c}_i(t) = f_i[\vec{c}(t), \vec{k}, \vec{v}; t] \tag{1}$$

where f_i is a function dependent upon the instantaneous concentrations $\vec{c}(t)$, the rate constants, $\vec{k}$, the unknown reaction orders $\vec{v}$, and the time. The solution of the set, $\vec{c}(t)$, then yields the complete set of concentration vs time profiles $\vec{c}(t)$ for a given set of rate constants, $\vec{k}$, and reaction orders, $\vec{v}$.

Consider an incomplete set of concentration vs time data, which we term the target concentrations. For species i, there may exist target concentration data, d_{i_j}, at several discrete times, t_{i_j}. Here, d_{i_j} is the target concentration of species i at discrete time, t_{i_j}. The error function for species i is defined as

$$\epsilon_i(\vec{k}, \vec{v}) = \sum_j \left(\frac{c_i(t_{i_j}) - d_{i_j}}{d_{i_j}} \right)^2 \tag{2}$$

The error function for species i is the squared difference between the computed concentration, $c_i(t_j)$, for a guessed set of rate constants, $\vec{k}$, the guessed set of reaction orders $\vec{v}$, and the target concentrations, d_{i_j} at the same times, t_{i_j}. The error function is normalized by the target concentrations, thus preventing species that have higher concentrations during the course of the reaction from being weighted more heavily than species with lower concentrations. The total error function for the complete reaction then becomes

$$\epsilon(\vec{k}, \vec{v}) = \sum_i \epsilon_i(\vec{k}, \vec{v}) \tag{3}$$

This error function depends on $\vec{k}$ and $\vec{v}$ implicitly since the time-dependence of the concentrations $\vec{c}(t)$ are governed by the kinetic equations in eq 1.

The chemical kinetics search algorithm is comprised of the following steps: (1) make an initial guess of the set of rate constants $\vec{k}$ and reaction orders $\vec{v}$, (2) solve the set of differential equations for the set of chemical species $\vec{c}(t)$ based on these guesses according to eq 1, (3) construct the error function from eqs 2 and 3, (4) minimize this error function to obtain updated values of set of rate constants $\vec{k}$ and reaction orders $\vec{v}$, (5) use these values in step 1, and continue this iterative process until the error function is zero within some predefined tolerance.

The minimization of the error function is the critical step in the algorithm, and a mechanism is needed to perform this minimization. As mentioned above, the gradient descent algorithm is prohibitive, since one must compute the gradient of the error function with respect to each rate constant, k_j, and reaction order v_i, from the sets $\vec{k}$ and $\vec{v}$, respectively. For example consider the gradient of the error function with respect to the rate constant, k_j

$$\frac{d\epsilon(\vec{k}, \vec{v})}{dk_j} = \sum_i \sum_j 2[c_i(t) - d_{i_j}] \frac{dc_i(t)}{dk_j} \tag{4}$$

In order to implement a gradient descent algorithm, $dc_i(t)/dk_j$ is needed. If eq 1 cannot be solved analytically, which is the usual case in chemical kinetics, then $dc_i(t)/dk_j$ must be computed numerically, which is both numerically unstable and computationally exhaustive. Further, any conventional functional minimizer such as the gradient descent or downhill Simplex algorithm will converge toward the local minimum nearest the initial parameter guesses which can lead to inaccurate values for the rate constants and reaction orders. Instead the error function is minimized via a heuristic search algorithm. In this work we explore both the GA and SA as tools for the error function minimization. To implement the chemical kinetics search algorithm using the GA or SA one supplies the function to be optimized, in this case the error function defined in eq 3 to the SA or GA, and the SA or GA finds the set of parameters, in this case the sets $\vec{k}$ and $\vec{v}$, that minimize the value of $\epsilon(\vec{k}, \vec{v})$.

(B) Chemical Reactions Under Study. (i) A Consecutive Reaction with a Reversible Step. Consecutive reactions with a reversible step occur widely in chemical kinetics and examples of these reactions include the Lindemann−

Hinshelwood mechanism for unimolecular decay and the Michaelis−Menten mechanism of enzyme catalysis.[14] We consider the reaction scheme for this reaction type shown below

$$v_1 A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} v_2 B$$

$$v_2 B \overset{k_2}{\longrightarrow} C \tag{5}$$
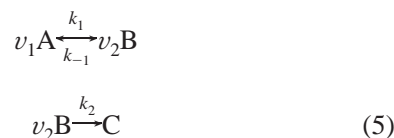
We will consider the set of rate constants $\vec{k} = (k_1, k_{-1}, k_2)$ and the set of orders of the reaction for each of the elementary steps $\vec{v} = \{v_1, v_2\}$ as unknowns to be determined from the heuristic search algorithm. The differential equations whose solutions yield the complete concentration vs time profile are

$$\frac{d[A]}{dt} = -k_1[A]^{v_1} + k_{-1}[B]^{v_2} \tag{6a}$$

$$\frac{d[B]}{dt} = -k_2[B]^{v_2} - k_{-1}[B]^{v_2} + k_1[A]^{v_1} \tag{6b}$$

$$\frac{d[C]}{dt} = k_2[B]^{v_2} \tag{6c}$$

(ii) A Parallel Reaction. Parallel reactions occur widely in chemical kinetics and can represent three types of general processes: (1) an initial reactant decomposes into different products, (2) a single substance reacts with several different reactants, or (3) a set of different reactants decompose into the same product.[14] We consider the parallel reaction of type 1 shown below

$$v_1 A \overset{k_1}{\longrightarrow} B$$

$$v_2 A \overset{k_2}{\longrightarrow} C \tag{7}$$

We treat the set of rate constants $\vec{k} = \{k_1, k_2\}$ and the set of orders of the reaction for each of the elementary steps $\vec{v} = \{v_1, v_2\}$ as unknowns to be determined from the heuristic search algorithm. The differential equations whose solutions yield the complete concentration vs time profile for the species are

$$\frac{d[A]}{dt} = -k_1[A]^{v_1} - k_2[A]^{v_1} \tag{8a}$$

$$\frac{d[B]}{dt} = k_1[A]^{v_1} \tag{8b}$$

$$\frac{d[C]}{dt} = k_2[A]^{v_2} \tag{8c}$$

(iii) The Chemical Fate of a Pollutant in the Environment. The last type of chemical reaction we will consider is the chemical fate of the pesticide chlorpyrifos in the environment. The removal of chlorpyrifos from the aqueous

phase has a known mechanism that involves six rate constants, and we show this mechanism below[15,16]

$$C_W \overset{k_1}{\longrightarrow} C \tag{9a}$$

$$C_W \overset{k_2}{\longrightarrow} B \tag{9b}$$

$$C_W \overset{k_3}{\longrightarrow} v_1 C_F \overset{k_4}{\longrightarrow} I \tag{9c}$$

$$v_2 C_W \underset{k_6}{\overset{k_5}{\rightleftharpoons}} C_S \tag{9d}$$

Here C_W, C_F, and C_S, are the chlorpyrifos concentrations in the aqueous phase, in fish, and in the soil, respectively. In eqs 9 C, B, and I, are the chlorpyrifos hydrolysis product, chlorpyrifos vapor, and chlorpyrifos metabolite from fish, respectively. What is typically measured are the concentrations C_W, C_F, and C_S. The set of differential equations that govern their time dependence are

$$\frac{dC_W}{dt} = -\frac{k_1 C_W}{d} - k_2 C_W - \frac{k_3 F C_W}{V} - \frac{k_5 S C_W^{v_2}}{V} + \frac{k_6 S C_S}{V} \tag{10a}$$

$$\frac{dC_S}{dt} = k_5 C_W^{v_2} - k_6 C_S \tag{10b}$$

$$\frac{dC_F}{dt} = k_3 C_W - k_4 C_F^{v_1} \tag{10c}$$

The set $\vec{k} = \{k_1, k_2, ..., k_6\}$ is the set of rate constants for chlorpyrifos' transfer between the phases, d is the depth of the water, V is the volume of the water, F is the mass of the fish, and S is the mass of the soil. The values of d, V, F, and S are 88.8 cm, 3.59×10^8 mL, 8580 g, and 1.49×10^7 g, respectively.[15,16] Equation 10 shows that the kinetic equations for the species concentration are independent of the rate constants k_1, and k_2 individually but depend on the sum $k_1/d + k_2$. With this in mind we redefine eq 10a above as

$$\frac{dC_W}{dt} = -k_1' C_W - \frac{k_3 F C_W}{V} - \frac{k_5 S C_W^{v_2}}{V} + \frac{k_6 S C_S}{V} \tag{10d}$$

where $k_1' = k_1/d + k_2$. The reaction orders, $\vec{v} = \{v_1, v_2\}$, and the set of rate constants, $\vec{k} = \{k_1', k_3, ..., k_6\}$, in eqs 9 and 10 are known and have the values shown in column two in Table 3.

III. COMPUTATIONAL RESULTS

(A) Computational Details. Concentration vs time data was synthesized for each of the reactions presented above by solving the differential equations in eqs 6, 8, and 10. These differential equations were solved numerically via a fourth-order Runge−Kutta adaptive, step-size integrator. The differential equation solver is also called as a subroutine by the SA or GA in the chemical kinetics search algorithm in order to evaluate the error function in eq 3.

The GA was run with the following tuning parameters; a population size of ten, probability for mutation = 0.1, probability for crossover = 0.5, the GA was run with uniform

SEARCH ALGORITHMS FROM LIMITED CONCENTRATION DATA

J. Chem. Inf. Comput. Sci., Vol. 38, No. 6, 1998 **1235**

Table 1. Rate Constants and Reaction Orders for Consecutive Reaction

algorithm	k_1 [s^{-1}]	k_{-1} [s^{-1} M^{-1}]	k_2[s^{-1}]	v_1	v_2	error
SA	9.9	0.95	0.01	1.0	2.0	4.6×10^{-8}
GA	8.5	0.8	0.01	1.2	2.2	6.3×10^{-5}
Simplex	10.0	0.4	0.01	1.4	2.5	6.3×10^{-4}

crossover, and with creep mutations enabled. The search boundaries for each of the parameters were initially set at a minimum of zero and a maximum of unity. As the program ran, every time a search boundary was exceeded for a particular parameter, the search boundary for this parameter was increased, by ten, until the GA yielded values of the parameter within the search boundary.

The SA algorithm was run with an annealing schedule given by the following equation

$$T_{k+1} = \lambda T_k \qquad (11)$$

where T_{k+1} is the temperature after N function evaluations. The initial temperature and temperature reduction factor, λ, were chosen as 10.0 and 0.666 for all examples, respectively. The temperature is reduced after $N = N*100$ moves, where N is the number of unknown parameters in the particular problem. The lower bounds on all parameters were set to zero, and the upper bounds on all parameters were set to 100 for examples presented in this paper. For all reactions studied in this paper we have found that 1.5×10^4 iterations took about 15 min on an SGI Indigo2 workstation.

(B) Numerical Results. (i) Consecutive Reaction. For the consecutive chemical reaction scheme that is outlined in eq 5 we synthesized a limited set of known concentration data for the reaction by choosing the rate constants to have the values of $\vec{k} = \{10.0 \text{ s}^{-1}, 1.0 \text{ M}^{-1} \text{ s}^{-1}, 0.01 \text{ s}^{-1}\}$ and the reaction orders to have the values $\vec{v} = \{1, 2\}$. Thus the rate constants for this reaction span over four orders of magnitude. We synthesized the set of known concentration data by solving the differential equations in eq 6 with initial concentration values of $\{[A]_0 = 1.0 \text{ M}, [B]_0 = 0, [C]_0 = 0\}$. The concentration vs time profile for this reaction is shown in Figure 1a. The arrows indicate the data points which we use as the known concentration vs time data. It should be pointed out that we only choose two time data points each for reactant A and final product C. Further, we assume we have no knowledge about the concentration of the intermediate species B. For many reactions experimental concentration vs time profiles of intermediates are hard to measure so we mimic this experimental situation by assuming we have no knowledge of the concentration vs time behavior of the intermediate B.

We now treat the rate constants, $\vec{k} = \{k_1, k_{-1}, k_2\}$, and the set of orders of the reaction for each of the elementary steps, $\vec{v} = \{v_1, v_2\}$, as unknowns and use the chemical kinetics search algorithm to find them from the limited concentration data. We assume no knowledge of the rate constants or reaction orders and thus choose their initial values as zero. We show results for the SA-based chemical kinetics search algorithm in the second row in Table 1. Table 1 shows that all rate constants and reaction orders are reproduced by the search algorithm, and the value of the
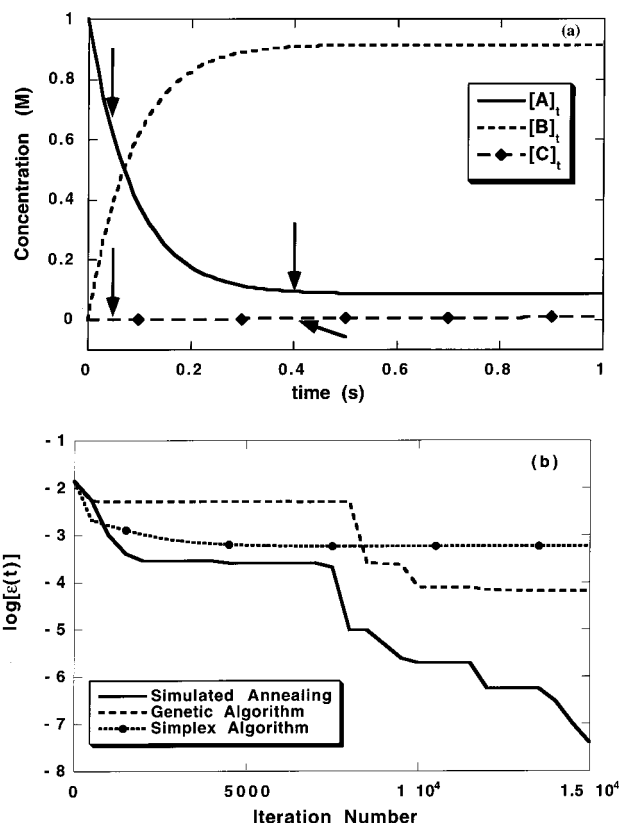

Figure 1. (a) Concentration vs time profile for the consecutive reaction. The solid, short-dashed, and long-dashed (connecting diamonds) lines show species A, B, and C, respectively. The arrows indicate the target data. (b) Comparison of the convergence for the SA-, GA-, and Simplex-based chemical kinetics search algorithm for the consecutive reaction. The solid, long-dashed, and short-dashed (connecting circles) lines show the convergence of the SA-, GA-, and Simplex-based algorithms, respectively.

error function for the best set of parameters was $\epsilon = 4.6 \times 10^{-8}$ after 1.5×10^4 iterations.

We show results for the GA-based chemical kinetics search algorithm in the third row in Table 1. The GA-based algorithm is also able to reproduce the unknown rate constants and reaction orders quite well, although not as well as the SA-based kinetics search algorithm. The minimum value for the error found by the GA-based search algorithm was $\epsilon = 6.3 \times 10^{-5}$ after 1.5×10^4 iterations.

We now show results for the chemical kinetics search algorithm that is based on a conventional functional minimizer. This is the downhill Simplex algorithm which needs no derivative information but will become stuck in the local minimum that is closest to the initial guesses for the parameters. In the last row in Table 1 we show the results of the Simplex-based algorithm. As shown in Table 1 the Simplex algorithm gets stuck in a local minimum that is closest to the initial guesses yielding inaccurate values for the rate constants and the reaction orders. The best value of the error function for the Simplex algorithm based kinetics search algorithm is $\epsilon = 6.3 \times 10^{-4}$ after 1.5×10^4 iterations.

In Figure 1b we show the convergence characteristics of the SA-, GA-, and Simplex-based chemical kinetics search algorithms. In Figure 1b the log [ϵ] is plotted vs iteration number. We define the iteration number as the number of times the differential equation subroutine must be called by the kinetics search algorithm. Figure 1b shows that the SA-

Table 2. Rate Constants and Reaction Orders for Parallel Reaction

algorithm	k_1 [s^{-1}]	k_2 [s^{-1} M^{-1}]	ν_1	ν_2	error
SA	10.0	0.10	1.0	2.0	6.7×10^{-6}
GA	10.0	0.11	1.0	2.05	6.0×10^{-5}
Simplex	10.0	0.06	1.0	1.0	0.11

based algorithm is most efficient for this reaction, with the GA-based algorithm second-most efficient, while the Simplex-based algorithm gets stuck in a local minimum. We will see in the following examples that this general trend holds throughout.

(ii) A Parallel Reaction. For the parallel chemical reaction scheme that is outlined in eq 7 we synthesized a limited set of known concentration data for the reaction by choosing the two unknown rate constants to have the values of $\vec{k} = \{10.0 \text{ s}^{-1}, 0.10 \text{ M}^{-1} \text{ s}^{-1}\}$ and the reaction orders to have the values $\vec{v} = \{1, 2\}$. We synthesize the set of known concentration data by solving the differential equations shown in eq 8 with initial concentrations values of $\{[A]_0 = 1.0 \text{ M}, [B]_0 = 0, [C]_0 = 0\}$. The concentration vs time profile for this reaction is shown in Figure 2a. The arrows indicate the data points which we use as the known concentration vs time data.

We now treat the rate constants, $\vec{k} = \{k_1, k_2\}$, and the reaction orders $\vec{v} = \{v_1, v_2\}$, as unknowns and use the chemical kinetics search algorithm to find them from the limited concentration data. Again, we assume no knowledge of the rate constants or reaction orders, and we supply the search algorithm with initial values of zero for these parameters. We show results for the SA-based chemical kinetics search algorithm in the second row in Table 2. Table 2 shows that all rate constants and reaction orders are reproduced by the SA-based search algorithm, and the value of the error function for the best set of parameters was $\epsilon = 6.7 \times 10^{-6}$ after 1.5×10^4 iterations.

We show results for the GA-based chemical kinetics search in the third row in Table 2. The GA-based search algorithm is also able to reproduce the unknown rate constants and reaction orders quite well, although not as well as the SA-based kinetics search algorithm. The minimum value for the error found by the GA-based search algorithm was $\epsilon = 6.0 \times 10^{-5}$ after 1.5×10^4 iterations.

In the last row in Table 2 we show the results of the Simplex-based search algorithm. As demonstrated by Table 2, the Simplex-based search algorithm gets stuck in a local minimum that is closest to the initial guesses yielding inaccurate values for the rate constants and the reaction orders. The best value of the error function for the Simplex-based search algorithm is $\epsilon = 1.1 \times 10^{-1}$ after 1.5×10^4 iterations.

In Figure 2b we show the convergence characteristics of the SA-, GA-, and Simplex-based kinetics search algorithms. Figure 2b shows that, again, the SA-based algorithm is most efficient for this reaction, with the GA-based algorithm second-most efficient, while the Simplex-based algorithm gets stuck in a local minimum.

(iii) The Chemical Fate of a Pollutant in the Environment. For the chemical fate of chlorpyrifos in the aqueous phase, the reaction scheme shown in eq 9, we synthesized a limited set of known concentration data for the reaction by choosing the rate constants to have the values that were
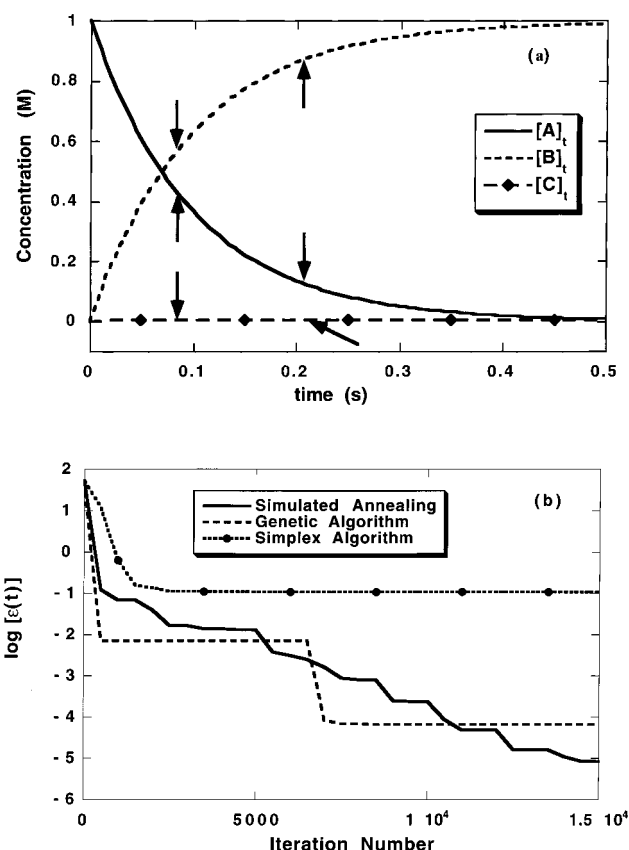

Figure 2. (a) Concentration vs time profile for the parallel reaction. The solid, short-dashed, and long-dashed (connecting diamonds) lines show species A, B, and C, respectively. The arrows indicate the target data. (b) Comparison of the convergence for the SA-, GA-, and Simplex-based chemical kinetics search algorithm for the parallel reaction. The solid, long-dashed, and short-dashed (connecting circles) lines show the convergence of the SA-, GA-, and Simplex-based algorithms, respectively.

proposed by Neely and Blau.[15] These proposed values are shown in the second column in Table 3. The two reaction orders in eq 9 have the known values of $\vec{v} = \{1, 1\}$. We synthesized the set of known concentration data by solving the differential equations in eq 10 with initial concentrations values of $\{[C_W]_0 = 5.75 \text{ ppb}, [C_s]_0 = 0, [C_F]_0 = 0\}$. Figure 3a−c show the concentration profiles for $\{C_W, C_S, C_F\}$, respectively for these values of the proposed rate constants and reaction orders. The arrows in Figure 3a−c indicate the target data chosen for this reaction.

We now treat the rate constants, $\vec{k} = \{k'_1, k_3, ..., k_6\}$ and the reaction orders, $\vec{v} = \{v_1, v_2\}$, as unknowns and use the chemical kinetics search algorithm to find them from the limited concentration data. We show results for the SA-based chemical kinetics search algorithm in the third column in Table 3. Table 3 shows that all rate constants and reaction orders are reproduced by the SA-based search algorithm and the value of the error function for the best set of parameters was $\epsilon = 3.0 \times 10^{-5}$ after 2.1×10^4 iterations.

We show results for the GA-based chemical kinetics search algorithm in the fourth column in Table 3. The GA-based algorithm is also able to reproduce the unknown rate constants and reaction orders quite well, although not as well as the SA-based kinetics search algorithm. The minimum value for the error found by the GA-based search algorithm was $\epsilon = 8.2 \times 10^{-4}$ after 2.1×10^4 iterations.

SEARCH ALGORITHMS FROM LIMITED CONCENTRATION DATA

J. Chem. Inf. Comput. Sci., Vol. 38, No. 6, 1998 **1237**
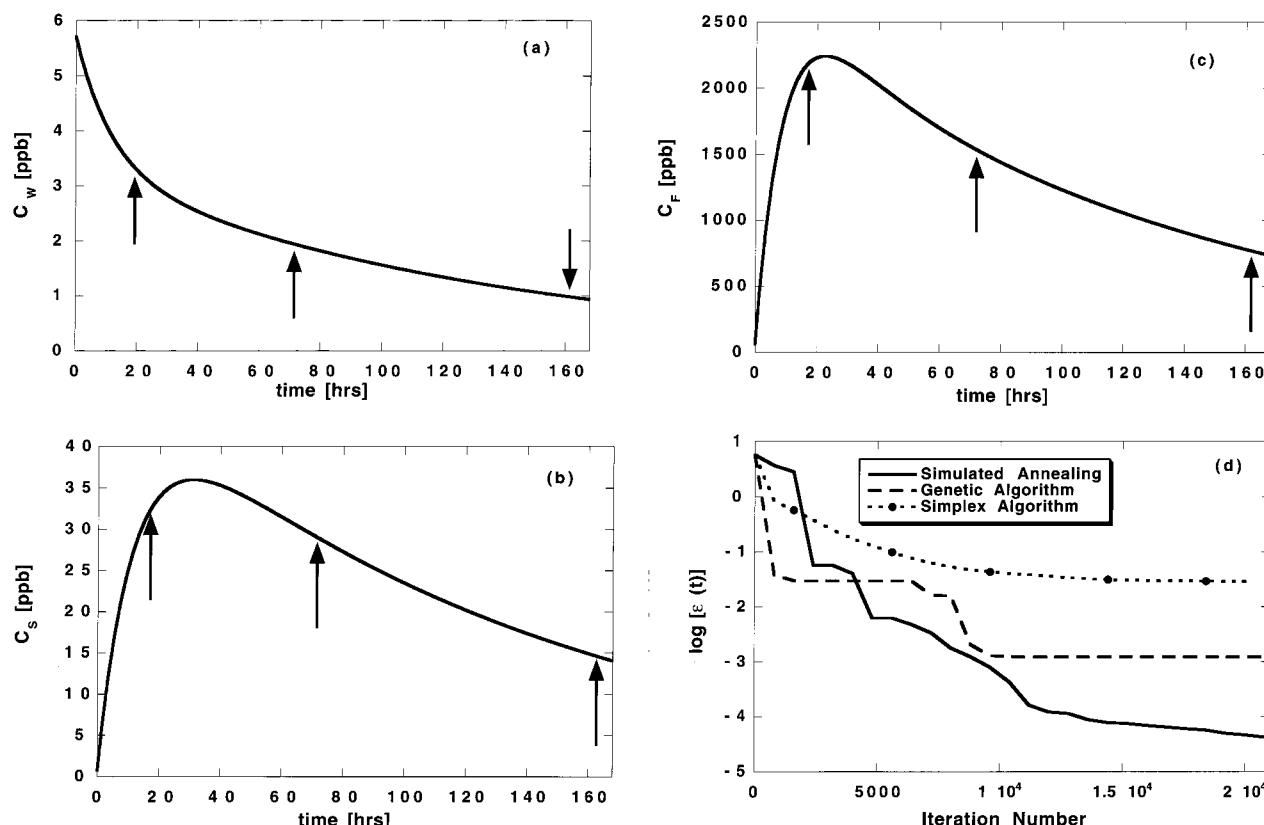

Figure 3. (a) Concentration vs time profile for chlorpyrifos in the aqueous phase. The arrows indicate the target data. (b) Concentration vs time profile for chlorpyrifos in the soil phase. The arrows indicate the target data. (c) Concentration vs time profile for chlorpyrifos in the fish phase. The arrows indicate the target data. (d) Comparison of the convergence for the SA-, GA-, and Simplex-based chemical kinetics search algorithm for the chlorpyrifos reaction. The solid, long-dashed, and short-dashed (connecting circles) lines show the convergence of the SA-, GA-, and Simplex-based algorithms, respectively.

Table 3. Rate Constants and Reaction Orders for the Fate of Chlorpyrifos

k	proposed	SA	GA	Simplex
k'_1	0.011 h^{-1}	0.011	0.011	0.013
k_3	55.5 mL g$_{fish}^{-1}$ h^{-1}	55.9	54.7	51.7
k_4	0.078 h^{-1}	0.072	0.06	0.068
k_5	0.67 mL g$_{soil}^{-1}$ h^{-1}	0.67	0.68	0.54
k_6	0.052 h^{-1}	0.052	0.052	0.042
v_1	1.00	1.00	1.00	0.8
v_2	1.00	1.00	1.03	1.4
	error	3.0×10^{-5}	8.2×10^{-4}	2.8×10^{-2}

In the last column in Table 3 we show the results of the Simplex-based chemical kinetics search algorithm. As shown in Table 3 the Simplex-based algorithm gets stuck in a local minimum that is closest to the initial guesses yielding inaccurate values for the rate constants and the reaction orders. The best value of the error function for the Simplex-based search algorithm is $\epsilon = 2.8 \times 10^{-2}$ after 2.1×10^4 iterations.

In Figure 3d we show the convergence characteristics of the SA-, GA-, and Simplex-based chemical kinetics search algorithms. Figure 3d shows that the SA-based algorithm is most efficient for this reaction, with the GA-based algorithm second-most efficient, while the Simplex-based algorithm gets stuck in a local minimum.

IV. DISCUSSION AND CONCLUSION

The results presented above for the chemical kinetics search algorithm lead to two major conclusions: (1) a chemical kinetics search algorithm that is based on a heuristic functional optimizer can be used to find unknown rate constants and reaction orders from a limited set of concentration data, and (2) the SA-based chemical kinetics search algorithm is more efficient than the GA-based search algorithm.

The chemical kinetics search algorithm that is based on a heuristic functional optimizer is not only efficient and accurate but also robust. In a previous work we tested the robustness of the GA-based search algorithm by performing two diagnostic tests.[1] In the first test we restarted the search algorithm with different initial values for the unknown parameters. The second test involved the stability of the search algorithm when there is an artificially induced error in the target concentration data. These two diagnostic tests were run for the SA- and GA-based kinetics search algorithms on the three reactions presented in this paper. For each test the search algorithms were shown to be stable and yield roughly the same values for rate constants and reaction orders as were presented in Tables 1−3.

A chemical kinetics search algorithm that is based on a conventional functional optimizer, such as the Simplex algorithm, was shown to become stuck in a local minimum that was closest to the initial guesses. The values of the rate constants and reaction orders found from the Simplex-based search algorithm were found to be inaccurate as compared to the values obtained from the SA- and GA-based search algorithms.

We now offer an explanation as to why the SA-based chemical kinetics search algorithm is more efficient than the GA-based chemical kinetics search algorithm. The SA algorithm that is used in this work is a continuous minimization method that is based on the Simplex algorithm.[17] At an annealing temperature of zero, i.e., the temperature in eq 11, the simulated annealing algorithm becomes the downhill Simplex algorithm. According to Figure 1b, 2b, and 3d, the GA- and SA-based search algorithms give comparable values for the error function at 1×10^4 iterations. After 1×10^4 iterations the annealing temperature is approximately zero, and the SA algorithm has become the downhill Simplex algorithm. Thus, for the first 1×10^4 iterations, the SA-based algorithm gets near a useful local minimum, and all subsequent iterations finds the precise location of this minimum with a conventional minimizer. The GA-based algorithm, on the other hand, stays a heuristic-type algorithm for all iterations. As is well-known the GA is efficient at getting near a useful local minimum but is not as efficient as a nonheuristic search algorithm for locating the precise location of this useful local minimum. Indeed in previous work[1] we have found that combining the GA-based chemical kinetics search algorithm with the downhill Simplex algorithm is more efficient than the GA alone. In previous work we used the GA to get near a useful local minimum and then use the parameter values found by the GA as initial guesses in a downhill Simplex algorithm to polish the parameter values, i.e., to locate the precise location of this useful local minimum.

REFERENCES AND NOTES

(1) Von Arx, K. B.; Huffman, S. W.; Manock, J. J.; Messina, M.. Using limited concentration data for the determination of rate constants with the genetic algorithm. *Environ. Sci. Technol.* **1998,** in press.

(2) Goldberg, D. *Genetic Algorithms in Search, Optimization, and Machine Learning;* Addison-Wesley: New York, 1989.

(3) Hasdorff, L. *Gradient Optimization and Nonlinear Control;* John Wiley & Sons: New York, 1976.

(4) Debock, G. J. *Trading on the Edge;* John Wiley & Sons: New York, 1994.

(5) Back, T. *Evolutionary Algorithms in Theory and Practice;* Oxford University Press: Oxford, 1996.

(6) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. Optimization by simulated annealing. *Science* **1983,** *220,* 671−680.

(7) Kirkpatrick, S. Optimization by simulated annealing: Quantitative studies. *J. Stat. Phys.* **1984,** *34,* 975−986.

(8) Otten, R. H. J. M.; van Ginneken, L. P. P. P. *The Annealing Algorithm,* Kluwer: Boston, 1989.

(9) Vanderbilt, D.; Louie, S. G. A Monte-Carlo simulated annealing approach to optimization over continuous variables. *J. Comput. Phys.* **1984,** *56,* 259−271.

(10) Glover, F. Heuristics for integer programming using surrogate constraints. *Decision Science* **1977,** *8,* 156−166.

(11) Glover, F.; Taillard, E.; de Werra, D. A users guide to Tabu Search. *Ann. Op. Res.* **1993,** *41,* 3−28.

(12) Glover, F.; Laguna, M. *Tabu Search;* Kluwer Academic Publishers: Boston, 1997.

(13) Bouju, A.; Boyce, J. F.; Dimitropoulus, C. H. D.; von Scheid, D.; Taylor, J. G.. In *Conference on Applied Decision Technologies: Modern Heuristic Methods*; Brunel University; 1995; pp 233−250.

(14) Steinfeld, J. I.; Francisco, J. S.; Hase, W. L. *Chemical Kinetics and Dynamics;* Prentice Hall: New Jersey, 1989.

(15) Neely, W. B.; Blau, G. E. In *Pesticides in Aquatic Environments*; Khan, M. A., Ed.; Plenum Press: New York, 1977; pp 145−163.

(16) Macek, K. J.; Waber, D. F.; Hogan, J. W.; Holz, D. D. *Trans. Amer. Fish Soc.* **1972,** *3,* 420−425.

(17) Press: W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in Fortran 77*; Cambridge University Press: U.S.A., 1992.

CI980407B