Anal. Chem. **2003,** *75,* 6701−6707

Kinetic Studies of Cascade Reactions in High-Throughput Systems

David Iron,[†] Hans F. M. Boelens,[‡] Johan A. Westerhuis,[‡] and Gadi Rothenberg*,[‡]

Chemical Engineering Department and KdV Institute, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

The application of robotic systems to the study of complex reaction kinetics is considered, using the cascade reaction A → B → C as a working example. Practical problems in calculating the rate constants k_1 and k_2 for the reactions A → B and B → C from concentration measurements of C_A, C_B, or C_C are discussed in the light of the symmetry and invertability of the rate equations. A D-optimal analysis is used to determine the points in time and the species that will give the best (i.e., most accurate) results. When exact data are used, the most robust solution results from measuring the pair of concentrations (C_A, C_C). The system's *information function* is computed using numeric methods. This function is then used to estimate the amount of information obtainable from a given cascade reaction at any given time. The theoretical findings are compared with experimental results from a set of two-stage cascade experiments monitored using UV−visible spectroscopy. Finally, the pros and cons of using a single reaction sample to estimate both k_1 and k_2 are discussed.

The last two decades have provided chemists with a variety of new experimental tools. Among these, the realization of the concepts of combinatorial synthesis and parallel experimentation are perhaps the most significant. These methods have revolutionized the pharmaceutical industry but, despite a promising start,[1−4] have not yet overwhelmed researchers in other fields.[5,6] This is due more to a psychological barrier than to equipment and infrastructure costs: Working with high-throughput setups necessitates a change in one's mode of thinking, because the value of the basic research unit, the scientific experiment, is changed. In high-throughput systems, "cheap experiments" can be performed to obtain quickly large amounts of rough-quality data that are then analyzed to point to the next generation of experiments. This differs from the conventional "scientific" mode of thinking, where every experiment must be thoroughly evaluated (reflecting perhaps the time and labor costs of the work).

The determination of reaction rate constants is a primary tool in mechanistic studies. Kinetic analysis is more time-consuming and labor-intensive than simple yes/no activity tests, because many samples have to be taken from every reaction to establish each kinetic profile. When working with high-throughput systems, such as reactor arrays, this multiple sampling and analysis often results in a bottleneck.[7] One way to prevent this bottleneck is to optimize the sampling times. Previously, we showed, using a mathematical model of the reaction rate law, that one can estimate the amount of future information that can be gained from sampling each reactor in an array of first-order reactions.[8,9] Here, we extend the application of the information function to the more complex system of cascade reactions of the type A $\rightarrow^{k_1}$ B $\rightarrow^{k_2}$ C and discuss the implications of simultaneous determination of two reaction rate constants in reaction arrays. A theoretical D-optimal analysis is compared with the results from a set of cascade experiments.

RESULTS AND DISCUSSION

General Considerations. To find any reaction rate constants requires measurements of concentrations that will necessarily contain errors. The effect these errors have on the calculation of a rate constant depends strongly on the times the concentrations are sampled and on the value of the reaction rate constants themselves. The sampling of first-order reactions was considered elsewhere.[10−12] However, most reactions are not governed by first-order kinetics. In this paper, we will discuss the complications that arise as the result of a two-stage cascade reaction given by eq 1,

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C \qquad (1)$$

The introduction of the intermediate B results in considerably more complex reaction dynamics.[13] If the initial concentration of

* Corresponding author. Fax: +31 20 525 5604. E-mail: gadi@science.uva.nl.
[†] KdV Institute.
[‡] Chemical Engineering Department.

(1) Jandeleit, B.; Schaefer, D. J.; Powers, T. S.; Turner, H. W.; Weinberg, W. H. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 2495−2532.
(2) Lavastre, O. *Actual. Chim.* **2000**, 42−45.
(3) Senkan, S. *Angew. Chem., Int. Ed.* **2001**, *40*, 312−329.
(4) Mirodatos, C. *Actual. Chim.* **2000**, 35−39.
(5) Hoffmann, R. *Angew. Chem., Int. Ed.* **2001**, *40*, 3337.
(6) Schlögl, R. *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 2333−2336.
(7) Lutz, M. W.; Menius, J. A.; Choi, T. D.; Laskody, R. G.; Domanico, P. L.; Goetz, A. S.; Saussy, D. L. *Drug Discovery Today* **1996**, *1*, 277−286.
(8) Boelens, H. F. M.; Iron, D.; Westerhuis, J. A.; Rothenberg, G. *Chem. Eur. J.* **2003**, *9*, 3876−3881.
(9) This function basically describes the relationship between the errors in the measured concentrations $\delta(\mathbf{a})$ (where **a** is a vector that contains the measured concentrations at the various time points) and the time vector **t**. See: Rothenberg, G.; Boelens, H. F. M.; Iron, D.; Westerhuis, J. A. *Chim. Oggi* **2003**, *21*, 80−83.
(10) Carr, P. W. *Anal. Chem.* **1978**, *50*, 1602−1607.
(11) Holler, F. J.; Calhoun, R. K.; McClanahan, S. F. *Anal. Chem.* **1982**, *54*, 755−761.
(12) Rothenberg, G.; Boelens, H. F. M.; Iron, D.; Westerhuis, J. A. *Catal. Today* **2003**, *81*, 359−367.

10.1021/ac034719b CCC: $25.00
Published on Web 10/24/2003

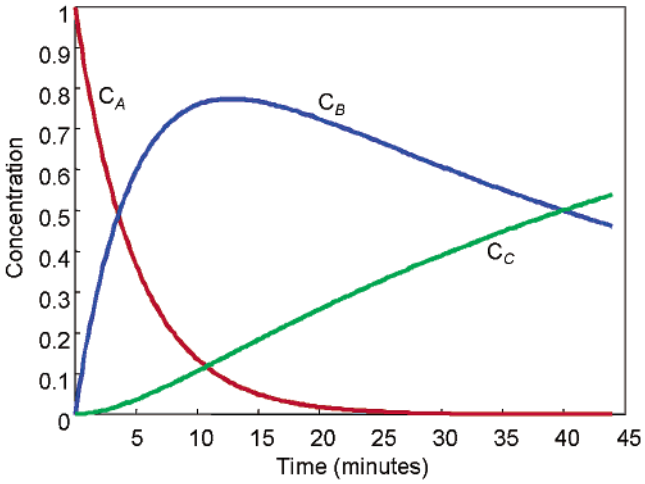


Figure 1. Concentration profiles for the cascade reaction A $\xrightarrow{k_1}$ B $\xrightarrow{k_2}$ C, with $k_1 = 0.2$ and $k_2 = 0.02$.

A is given by A_0 and the initial concentration of B and C are both 0, the reaction profiles are given by eqs 2a−2c.

$$C_A(t) = A_0 e^{-k_1 t} \tag{2a}$$

$$C_B(t) = A_0 \frac{k_1}{k_2 - k_1}(e^{-k_1 t} - e^{-k_2 t}) \tag{2b}$$

$$C_C(t) = \frac{A_0}{k_2 - k_1}(-k_2 e^{-k_1 t} + k_1 e^{-k_2 t}) + A_0 \tag{2c}$$

Here $C_A(t)$, $C_B(t)$, and $C_C(t)$ represent the concentrations, at time t, of A, B, and C, respectively (see Figure 1). The value of A_0 will have no importance in the remaining analysis, so from here on, we will scale the concentration values such that $A_0 = 1$.

The reaction is governed by two rate constants, so it will require a minimum of two measurements to determine k_1 and k_2. However, what should be measured? Is it possible to determine the rate constants by measuring the concentration of the same compound at two different times? Or should one measure the concentration of two different compounds at the same time or at different times? The following analysis assumes that all of the compounds can be measured with the same accuracy (in practice, the accuracy of the measurements may vary for each compound, and this must also be taken into account).

In this cascade system, even though there is sufficient information to determine the rate constants, no explicit expression for k_1 and k_2 exists. This is true even if we have exact data, because we cannot directly invert eq 2. Instead, we must use a numerical solver to find the rate constants. The timing of the measurements must also be carefully considered. If concentration measurements are taken at poorly chosen times, small errors in the measurements can be magnified when the rate constants are being estimated. All of these issues will be discussed in the next section.

Determining Rate Constants from Exact Data. We will now assume that we have exact concentration data and consider the problems associated with finding the rate constants, k_1 and k_2, that correspond to these data. For a single first-order reaction, there is a single rate constant k and the reaction profile is simply given by $C(t) = C_0 e^{-kt}$, where C_0 is the concentration at $t = 0$. For such a system, a single measurement is sufficient to determine k if C_0 is known. There is also an explicit expression for k (eq 3) in terms of the initial concentration C_0 and the concentration C_1 that is measured at time t_1

$$k = \ln(C_0/C_1)/t_1 \tag{3}$$

For a two-stage reaction, this is not the case. There are two reaction rate constants, and a minimum of two measurements is needed to have enough information to determine both of them. Even when we have sufficient information to determine k_1 and k_2, we cannot directly invert eq 2. This means that we have no explicit expression, such as eq 3, for the rate constants in terms of known constants and measured data. Instead we must use a numerical method, such as a generalized Newton root finder, to find approximate values of k_1 and k_2 that fit the experimental data.[14] We must also choose which concentrations to sample. In the case of a single first-order reaction, there is only one reaction and we just need to find an optimal time to sample it. Now we have three concentrations; this leads to many possible choices. It does not make sense to solely use data from C_A, since this gives us no information about k_2. However, we may attempt to find k_1 and k_2 solely from measuring C_B or C_C. We may also measure any two of the three concentrations at either the same or at different times. However, as eq 2 is highly nonlinear, the choice of which concentrations to measure can make it very difficult or even impossible to make an accurate approximation of the rate constants.

Let us now consider the difficulties associated with some of the choices of concentration measurements. The clearest problem results when we choose to measure the concentration C_C at two different times. Equation 2c is invariant under the transform $k_1 \leftrightarrow k_2$. This symmetry means that a cascade reaction with $k_1 = a$ and $k_2 = b$ will have the same C_C profile as a reation with $k_1 = b$ and $k_2 = a$, for any a and b. In other words, no matter how many measurements you take, you will never be able to distinguish k_1 from k_2 if you use only information from C_C. Another major consideration when using information from just the concentration of C is that it is possible that profiles generated using different reaction rates will be very similar (see Figure 2).

The reaction profile for C_B does not possess the symmetry $k_1 \leftrightarrow k_2$, so the choice to measure C_B at two distinct times may be reasonable. However, again two measurements will not give unique values for the rate constants. Figure 3 shows two different profiles for C_B. If we were to take measurements at the points where these curves intersect, we would have at least two different solutions for k_1 and k_2.

These difficulties are generic in the sense that they exist for any values of k_1, k_2, t_1, and t_2. Consider a two-stage reaction with concentrations of B = C_{B1} and C_{B2} at times t_1 and t_2, respectively. If we make an initial guess of the rate constants, k_1 and k_2 as k_1^0 and k_2^0, we obtain

$$k_1 = k_1^{\,0} + \delta k_1 \tag{4a}$$

$$k_2 = k_2^{\,0} + \delta k_2 \tag{4b}$$

(13) Moore, W. J. *Physical Chemistry*, 4th ed.; Longman: London, 1976; pp 345−347.

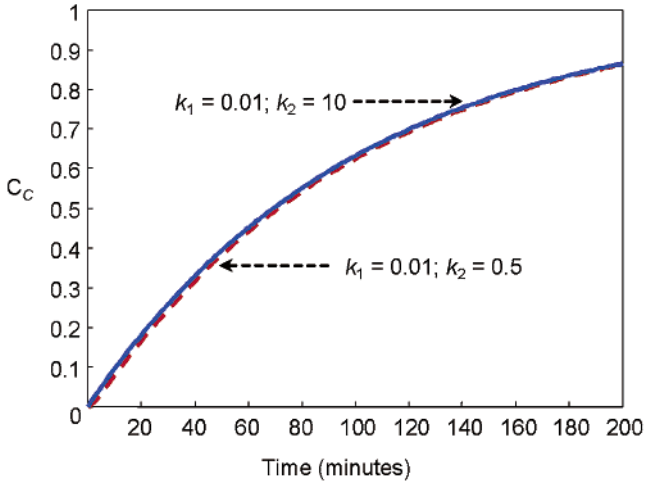


Figure 2. Two C_C profiles with rate constants $k_1 = 0.01$, $k_2 = 0.5$ (broken line) and $k_1 = 0.01$, $k_2 = 10$ (solid line). Even though the values of k_2 differ by a factor of 20, the curves are almost indistinguishable.

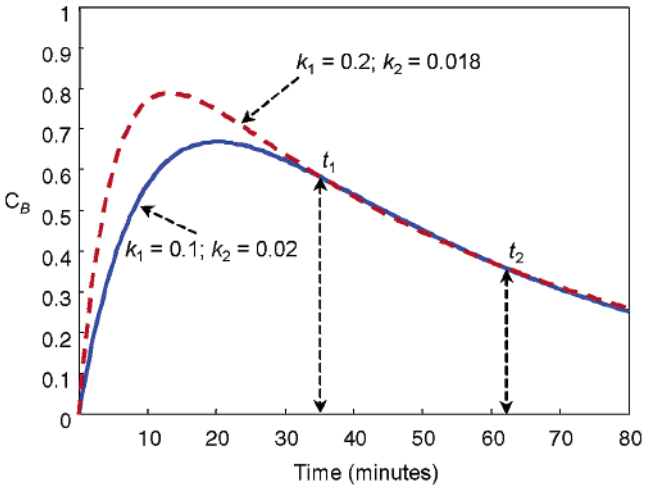


Figure 3. Broken line curve representing the concentration profile for B with $k_1 = 0.2$ and $k_2 = 0.018$. The solid line curve represents the concentration for B with $k_1 = 0.1$ and $k_2 = 0.02$. Sampling compound B at points t_1 and t_2 will render these two profiles indistinguishable.

where $\delta k_{1,2}$ is the error in the guess of $k_{1,2}$. Linearizing eq 2 about this guess gives us eq 5,

$$\begin{pmatrix}\delta C_{B1}\\ \delta C_{B2}\end{pmatrix} = J\begin{pmatrix}\delta k_1\\ \delta k_2\end{pmatrix} \tag{5}$$

Here $\delta C_{B1,B2}$ are the errors in the concentrations due to $\delta k_{1,2}$ and J is the Jacobian matrix,

$$J = \begin{pmatrix}\dfrac{\partial C_B}{\partial k_1}(k_1^0,k_2^0,t_1) & \dfrac{\partial C_B}{\partial k_2}(k_1^0,k_2^0,t_1)\\ \dfrac{\partial C_B}{\partial k_1}(k_1^0,k_2^0,t_2) & \dfrac{\partial C_B}{\partial k_2}(k_1^0,k_2^0,t_2)\end{pmatrix} \tag{6}$$

Substituting eq 2 into J, we find that the determinant of J is 0 for all values of k_1, k_2, t_1, and t_2. This means that the Jacobian is not invertible and we cannot solve for $\delta k_{1,2}$. In general, nonlinear systems with a Jacobian that cannot be inverted, cannot be inverted either. Even if we had very good estimates for the rate constants, any numerical solver will need to invert the Jacobian. As the Jacobian is not invertible, this will make finding an accurate estimate impossible. Thus, measuring C_B at two different times is not the best choice.

Measuring C_B and C_C is, at least in theory, a better option. As long as we do not choose to sample both concentrations at the same time (i.e., $t_1 = t_2$) (Note: this means that we cannot use the appealing choice of taking our measurements at the same time and thus save a measurement per rate constant calculation.), the Jacobian will be invertible and the implicit function theorem guarantees that we will have unique values of k_1 and k_2 for any pair of measurements of C_B and C_C. So at least in theory, this is an acceptable option. In practice, however, reactions with very different rate constants can have similar concentration profiles (see Figure 4).

This problem of instability is solved if we choose the concentration of A as one of the measurements. If C_A is measured, then we immediately have a value for k_1 from eq 3. The second measurement may then be used to obtain the value of k_2. Since we are only solving for one variable now, we may use a scalar version of Newton's root finding method to obtain an approximate value of k_2. The scalar version of Newton's method is much more robust then the vector version. If we use C_A and C_B then, as well as increasing the robustness of the root finder, we also make the problem of finding k_2 well defined by ensuring a unique solution. Once we have k_1 from the measurement of C_A, eq 2b becomes a monotonic function of k_2. Thus, by the inverse function theorem, invertible; given any $C_B(t_2)$, we can then find a unique value of k_2. Even though multiple solutions may exist if we measure C_A and C_C, this is not as much of a problem as it was in the case where C_A was not sampled. In this case, the problem is not generic. There are only isolated time points where we cannot invert the system. Note that in this case it *is* possible to determine the rate constants uniquely with measurements taken at a single point in time.

To summarize: Considering only invertability, the best options are to measure the concentrations of A and either B or C. Measuring the concentrations of B and C is possible, but less favorable. Determining the rate constants from measurements of a single compound is not possible. The choice to measure C_A is usually a logical choice for experimental reasons as well. As A is the starting material, pure A is usually readily available, and it is much easier to calibrate the equipment for accurate measurements. The chemist is often faced with a problem trying to calibrate for the intermediate B. Thus, for the general situation, the best results will be obtained when the concentrations of A and C are measured.

Minimizing the Effects of the Errors in Concentration Measurements. We will now consider how errors in the concentration measurements affect the calculation of k_1 and k_2. For now, we will assume that, given two appropriate concentration measurements, the exact corresponding values of k_1 and k_2 can be found. The measurement errors will be composed of systemic errors, due to faults in the equipment or the calibration model, and random errors. We will assume the systemic errors are small

(14) Press: W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed.; Cambridge University Press: Cambridge, U.K., 1993.

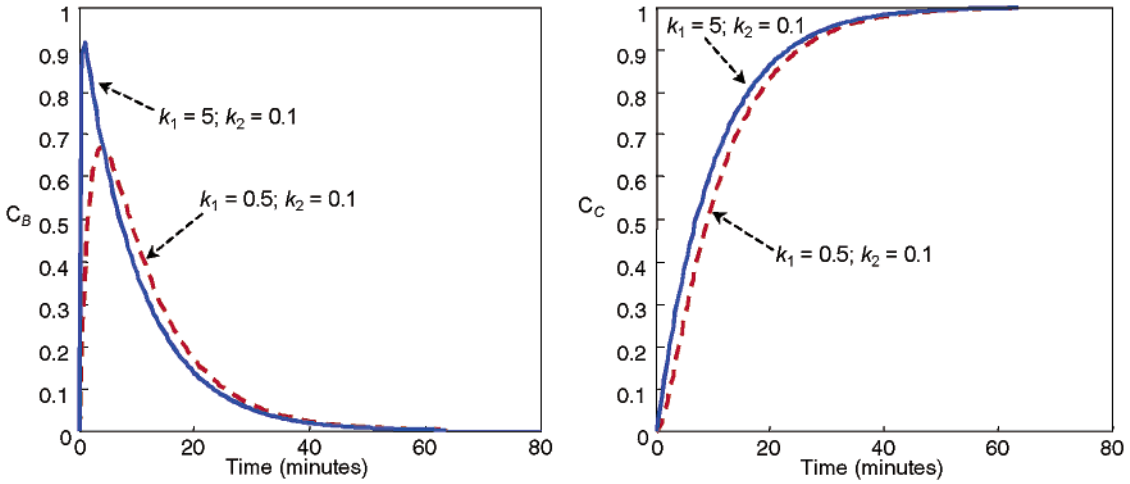


Figure 4. Reaction profiles for C_B (left) and C_C (right). The broken line curves refer to a reaction with rate constants $k_1 = 0.5$ and $k_2 = 0.1$, while the solid line curves represent a reaction with rate constants $k_1 = 5.0$ and $k_2 = 0.1$. If measurements are taken after $t = 10$ min, then any small experimental errors will render these profiles indistinguishable.

compared to the random errors and thus may be ignored. The time at which the measurements are made is crucial. If the measurements are taken close to the start of the reaction or near its completion, a small measurement error may result in a large error in the calculated value of k. Here we will use a D-optimal approach to find the best times to measure the concentrations.

The criterion for determining D-optimality is to find sampling times that minimize $\mathrm{Var}(k_1) \times \mathrm{Var}(k_2)$. Var is the variance that corresponds to the error in the absence of systemic effects.

We will assume that we are monitoring a cascade reaction such as in eq 1. We may choose any of the concentration pairs discussed above, but to simplify the notation, we will use C_A and C_B. Each measurement contains errors, so we will write the ith set of measurements as in eq 7, where $\delta_{Ai,Bi}$ represent the errors in the ith set of measurements. Although these measurements have the same index i, we do not necessarily intend this to mean that they are taken at the same time, but to mean that this pair of measurements is used to determine one set of $k_{1,2}$ values.

$$C_A + \delta_{Ai}, C_B + \delta_{Bi} \tag{7}$$

We denote as t_{i1} and t_{i2} the sampling times for the ith measurements of C_A and C_B, respectively. The measurement errors will lead to errors in the calculated values of $k_{1,2}$. We write the ith set of calculated k values as,

$$k_1 + \delta_{k1i} \tag{8a}$$

$$k_2 + \delta_{k2i} \tag{8b}$$

where δ_{k1i} and δ_{k2i} represent the errors, in the calculated values of k_1 and k_2, respectively, from the ith set of measurements. Since we expect the relative error to be small, we may substitute all of the concentration measurements and calculated rate constants into eq 2, and then linearize to obtain eq 9.

$$\begin{pmatrix} \delta_{Ai} \\ \delta_{Bi} \end{pmatrix} = J_i \begin{pmatrix} \delta_{k1i} \\ \delta_{k2i} \end{pmatrix} \tag{9}$$

Here J_i is the Jacobian matrix defined in eq 10:

$$J_i = \begin{pmatrix} \dfrac{\partial C_A}{\partial k_1}(k_1,k_2,t_{i1}) & \dfrac{\partial C_A}{\partial k_2}(k_1,k_2,t_{i1}) \\ \dfrac{\partial C_B}{\partial k_1}(k_1,k_2,t_{i2}) & \dfrac{\partial C_B}{\partial k_2}(k_1,k_2,t_{i2}) \end{pmatrix} \tag{10}$$

The errors in the concentration measurements are amplified by the inverse of the Jacobian matrices. Assuming that the magnitudes of the concentration errors in eq 9 are of the same order, the D-optimal sampling times are determined by minimizing the amplification of the errors by the inverse Jacobian or by maximizing the information function $f = |\mathrm{Det}(J^T J)|$.[9] Larger values of f correspond to measurements that yield more accurate rate constant estimates.

We will now consider the D-optimal times for taking the minimum number of samples. There are five different possible minimal sample types. Instead of listing all possible formulas, we will label the concentration sampled at time t_i as C_i and the concentration sampled at t_j as C_j. *The pair* (C_i,C_j) may be one of (C_A,C_B), (C_A,C_C), or (C_B,C_C) values. With this notation, the function f is given by eq 11:

$$f = \left|\left(\frac{\partial C_i}{\partial k_1}\right)^2\left(\frac{\partial C_j}{\partial k_2}\right)^2 + \left(\frac{\partial C_i}{\partial k_2}\right)^2\left(\frac{\partial C_j}{\partial k_1}\right)^2 - 2\frac{\partial C_i}{\partial k_1}\frac{\partial C_j}{\partial k_2}\frac{\partial C_i}{\partial k_2}\frac{\partial C_j}{\partial k_1}\right| \tag{11}$$

The D-optimal sampling times are at the global maximum of f. For any given values of k_1 and k_2, we may find this maximum numerically. Table 1 gives the D-optimal sampling times for a two-stage reaction for all of the possible sampling combinations. In both cases where C_A is sampled, the results are identical. This is not a coincidence, as the function f will always be identical for these two cases (eq 12), and the minimum error in the calculation of k_1 by measuring C_A will be at $t = 1/k_1$.[10] It is not possible to predict f for a given $k_{1,2}$ pair a priori, because eq 12 is highly nonlinear (cf. entries 1−3 with entries 4−6). This means that f must always be computed (the computation itself is not costly).

Table 1. Sampling Times That Afford the Maximum *f* Values Depending on Compounds Sampled[a]

entry	concns sampled	t_1/min	t_2/min	*f*
1[b]	C_A and C_B	4.21	43.19	475.6
2[b]	C_A and C_C	4.21	43.19	475.6
3	C_B and C_C	3.89	42.74	420.4
4[b]	C_A and C_B	10.00	12.56	14.0
5[b]	C_A and C_C	10.00	12.56	14.0
6	C_B and C_C	8.85	14.80	17.11
7[c]	C_A and C_B	9.58	9.58	22.4
8[c]	C_A and C_C	9.58	9.58	22.4
9[c,d]	C_B and C_C	9.58	9.58	22.4

[a] The rate constants used for these calculations are $k_1 = 0.237$ and $k_2 = 0.026$ (entries 1−3 and 7−9) and $k_1 = 0.1$ and $k_2 = 0.2$ (entries 4−6). The first two values match the values of experimental data that is used to test the validity of the theory, and the second pair is used to demonstrate the nonlinearity of eq 12. [b] The first two choices result in the same values since the function *f* is the same for these cases. [c] The last three entries are the same because when $t_1 = t_2$ the expressions simplify to the same result in all cases. [d] Although the value of *f* can be calculated in this case, it is difficult to find the rate constants from data pertaining to C_B and C_C when t_1 approaches t_2.

Entries 7−9 show the results for the special case when only one sample is taken, at $t_1 = t_2$. This case is particularly interesting in high-throughput applications (vide infra).

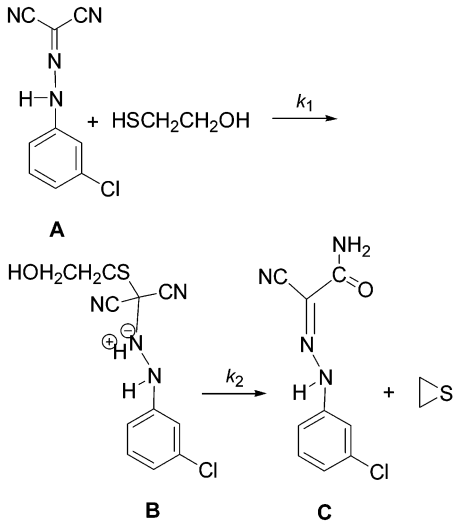
$$f = (t_1 \exp(-k_1 t_1)^2((\exp(-k_1 t_1) - k_1 t_2 \exp(k_2 t_2))/(k_2 - k_1) - (k_1 \exp(-k_1 t_2) - k_2 \exp(k_1 t_2))/(k_2 - k_1)^2)^2 \quad (12)$$

An Experimental Example. The above mathematical model was compared with experimental results from a two-stage cascade reaction. The test reaction should have well-defined cascade kinetics, and sufficient sample points must be measured to enable a good statistical analysis. We used the carbon−sulfur coupling of 3-chlorophenylhydrazonopropane dinitrile **A** with *β*-mercaptoethanol to give the adduct **B** that subsequently underwent unimolecular decomposition to 3-chlorophenylhydrazonocynaoacetamide **C** and ethylene sulfide (Scheme 1).[15,16] Pseudo-first-order conditions were realized for the step **A** → **B** by adding an excess of *β*-mercaptoethanol.

The reaction was followed using UV−visible spectroscopy. A total of 32 repetitions of this experiment were performed, the first of which was used to estimate the spectra of **A**, **B**, and **C**. A total of 271 UV−visible spectra were recorded for each experiment (one spectrum every 10 s), and using these spectra, a set of 31 concentration profiles was obtained for **A**, **B**, and **C**. The "true values" (i.e., best estimates) for k_1 and k_2 obtained by averaging all 271 sample points over the 31 profiles. The values were 0.237 and 0.026 min^{-1}, respectively, with standard deviations of 8.5×10^{-3} and 7.8×10^{-4} min^{-1}. These values were then compared with the respective k_1 and k_2 values obtained by sampling the pairs **AB**, **AC**, and **BC** at various points in time. In each case, two sample times t_1 and t_2 were chosen, and the resulting accuracy in k_1 and k_2 was calculated.

Figure 5 shows the comparison between the error amplification calculated by the theoretical model (top) and the determinant of the covariance matrix obtained from the experimental results (bottom). In both cases, the red areas pertain to sampling times that afford low error values. As the information function differs from the covariance matrix (the latter contains an additional source of variance, in the form of the vector $\{\delta_{A1}, \delta_{B1}.. \delta_{An}, \delta_{Bn}\}$, see eq 9), an outright quantitative comparison cannot be made. However, much can be inferred from a qualitative comparison of the two sets of results: There is a clear similarity between the columns of graphs. This means, essentially, that *f* values can be used to determine the best points in time to sample this cascade reaction (this is advantageous, as *f* is obtained by simple calculations, rather than from experiments). Table 2 gives the sampling times that correspond to the minimum in the covariance. The model predictions for the pair **AB** are much better than those for **AC** or **BC**, as can also be seen by comparing the optimal sampling times in Tables 1 and 2. The reason for this discrepancy may be that the experimental measurements of C_C have the highest noise in this case (in the theoretical studies, we assumed that all three compounds were sampled with the same accuracy). The strong change in the minimum covariance sampling times could also be due to a flatness of the error surface minima (cf. the rather flat minimum of the error curve in the case of the single-step reaction A → B, with the same data[8]).

Scheme 1



Application to High-Throughput Experimentation. In many high-throughput systems used for kinetic studies, for example, in catalysis, an array of reactors is interfaced to one (or two) analysis devices, such as a GC or HPLC. Such setups reflect the basic chemical requirements (i.e., different chemical reactions are carried out under different conditions) and the financial limitations (10-mL-scale autoclave reactors can cost as low as $10, but the price of an HPLC analyzer can be in the order of $50.000). In such cases, and especially when using robotic apparatus, the traditional practice is inverted: the analyzer time becomes the scarce commodity, but it is relatively inexpensive to discard a reaction and start a new one in its place.[17]

(15) Bijlsma, S.; Louwerse, D. J.; Smilde, A. K. *J. Chemom.* **1999**, *13*, 311−329.
(16) Bijlsma, S.; Boelens, H. F. M.; Hoefsloot, H. C. J.; Smilde, A. K. *Anal. Chim. Acta* **2000**, *419*, 197−207.
(17) Another option is to analyze several reactions in parallel, as was recently demonstrated using gas-phase IR image analysis; see: Hendershot, R. J.; Fanson, P. T.; Snively, C. M.; Lauterbach, J. A. *Angew. Chem., Int. Ed.* **2003**, *42*, 1152−1155.

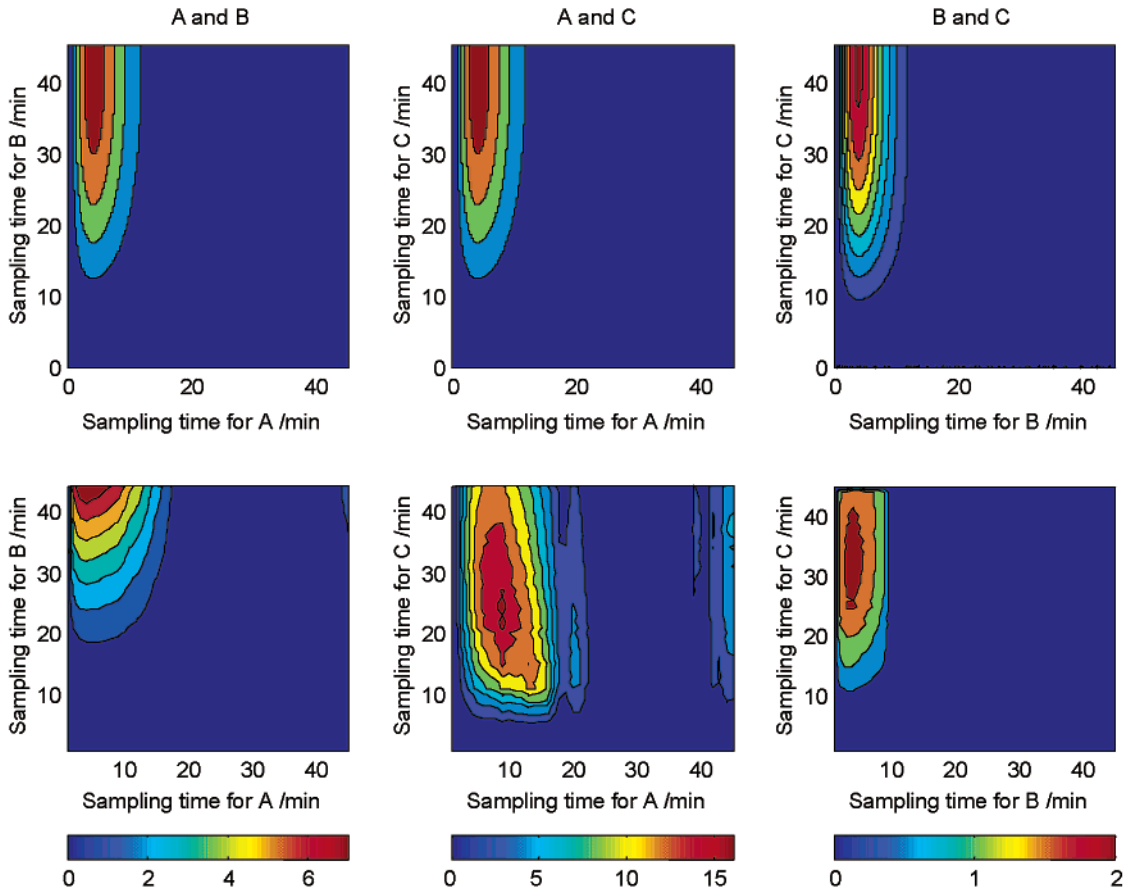


Figure 5. Qualitative comparison of *f* values (mathematical model, top) with the determinant of the covariance matrix for the experimental results (bottom) when sampling different reactant/product pairs at different times. The red areas pertain to sampling times that afford low error values. In all cases, and especially when **A** and **B** are sampled, a direct relationship is observed between *f* and det-[Cov(t_1,t_2)].

Table 2. Sampling Times Having Lowest Variance (1−3) and Other Sampling Times

entry	concns sampled	t_1/min	t_2/min	Det(Cov(k_1,k_2)) $\times 10^{-11}$
1	C_A and C_B	4	44	1.26
2	C_A and C_C	9	24	6.01
3	C_B and C_C	4	31	4.54
4	C_A and C_B	13	13	47.8
5	C_A and C_C	13	13	7.7
6	C_B and C_C	9	9	57.1

Conventional analysis practices (i.e., following each reaction for a fixed period of time with periodic sampling) will result in many cases in sampling at suboptimal times. In practice, this would mean that the robots would be generating garbage data and may also miss out on sampling other reactions in the array at their optimal times. In contrast, if one uses the concept of the information function, a near-optimal analysis protocol can be reached.

First, all reactors are sampled twice according to initial guesses for the rate constants generated by the system. Then, an array of rate constant estimates is calculated based on these samples, and from these, the corresponding *f* values are calculated for each reaction. From these the optimal sampling time can be obtained. If according to this information vector, the optimum sampling time already passed, two possibilities exist to improve the estimation of the reaction rate constants. The first is to let the reaction run and analyze the concentrations at suboptimal sampling times (this only leads to a suboptimal improvement of the rate constant estimates). The second possibility is to discard the old reaction and start the same reaction again at the identical conditions. Then, it is possible to measure at the optimal sampling time and have an optimal improvement of the reaction rate estimates. The latter option is the better alternative as far as the costly sampling time is concerned. One assumes that the automated robotic system can reproduce the reaction conditions with high precision.

Finally, it is worthwhile to consider obtaining the so-called "initial guess estimates" for k_1 and k_2. If these guesses are good, it will be easy to reach the optimal sampling times, and vice versa. However, the methods for obtaining these initial guesses and for performing the first set of measurements will depend strongly on the experimental apparatus used and on the amount of chemical knowledge available. Different approaches should be used, for example, for parallel reactor arrays and for fast sequential systems. As we showed earlier for first-order reactions,[12] iterative algorithms may be useful. This problem will be the subject of future study in our laboratory.

CONCLUSIONS

By combining simple kinetic models and numeric analysis methods it is possible to formulate for cascade reactions an

information function that can predict the usefulness of sampling a given reaction at a given time. This approach can be used to minimize the experimental effort involved in kinetic studies and calibration experiments. The analysis process has three steps: first, you must invert the rate equations to find k_1 and k_2 from your data; second, you must decide when and which compounds to sample; finally, you must take into account the individuality of the experiment; that is, figure out how the chemistry and the experimental setup itself affect the measured concentrations. In principle, this method may be applied to reduce the experimental effort and the amount of "garbage data" in robotic reaction setups.

EXPERIMENTAL SECTION

All chemicals were commercially available (99% pure) and were used without further purification. KH_2PO_4 buffers were purchased from Acros (pro analysis 0.2 M). UV−visible spectra were recorded using a Hewlett-Packard 8453 spectrophotometer (quartz cuvettes, 1.00-cm path length). Data processing was performed using MATLAB.[18] A detailed description of the sample preparation methods and the experimental apparatus was published previously.[15]

A total of 32 identical experiments were performed and monitored using UV−visible. A stock solution of 3-chlorophenylhydrazonopropane dinitrile A (1.034 M in 0.1 N NaOH) was prepared. For each experiment, part of this stock solution was then diluted to 51.71 μM, buffered to pH 5.4 with KH_2PO_4, and mixed in the quartz cuvette with an excess (276:1 mol/mol) of β-mercaptoethanol solution (2.5 μL of β-mercaptoethanol in 7.5 μL of KH_2PO_4 buffer solution). UV−visible spectra of the reaction mixtures were recorded every 10 s at a wavelength range from 300 to 500 nm.

(18) MATLAB Version 6.1, 2001, is commercially available from MathWorks.

Received for review July 1, 2003. Accepted September 12, 2003.

AC034719B