

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,900

Open access books available

124,000

International authors and editors

140M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Applications of Monte Carlo Simulation in Modelling of Biochemical Processes

Kiril Ivanov Tenekedjiev<sup>1</sup>, Natalia Danailova Nikolova<sup>1</sup>  
and Krasimir Kolev<sup>2</sup>

<sup>1</sup>*N. Y. Vaptsarov Naval Academy, Varna*

<sup>2</sup>*Semmelweis University – Budapest*

<sup>1</sup>*Bulgaria*

<sup>2</sup>*Hungary*

## 1. Introduction

The biochemical models describing complex and dynamic metabolic systems are typically multi-parametric and non-linear, thus the identification of their parameters requires non-linear regression analysis of the experimental data. The stochastic nature of the experimental samples poses the necessity to estimate not only the values fitting best to the model, but also the distribution of the parameters, and to test statistical hypotheses about the values of these parameters. In such situations the application of analytical models for parameter distributions is totally inappropriate because their assumptions are not applicable for intrinsically non-linear regressions. That is why, Monte Carlo simulations are a powerful tool to model biochemical processes. The classification of Monte Carlo approaches is not unified, so here we comply with the interpretation given in (Press et al., 1992), where the general Monte Carlo approach is to construct parallel virtual worlds, in which the experimental estimates will play the role of true parameters, if the way in which the true parameters generate a sample is known. Bootstrap is a modification of Monte Carlo, which uses very few premises imposed on the data, and does not need to know the mechanism by which the true parameters generate experimental samples. Instead, resampling with replacement from the experimental sample is used to construct synthetic samples.

As far as confidence intervals (CI) are concerned, literature offers multiple types, but each of them belongs to one of the two main groups: root (Politis, 1998) and percentile intervals (Efron & Tibshirani, 1993). The difference in the philosophy of those two CI types is substantial for the biochemical interpretation of results. The difference here is explained with the difference between classical statistics (where the parameters are fixed unknown quantities) and Bayesian statistics (where the parameters are random variables with unknown distributions), and also with the philosophical differences between objectivity and subjectivity of scientific research. The main conclusion is that root confidence intervals are confidence intervals of the investigated parameters, whereas percentile confidence intervals refer to the estimates of the investigated parameters.

---

Address for correspondence: K. Kolev, Semmelweis University, Department of Medical Biochemistry, 1094 Budapest, Tűzoltó u. 37-47

Our first application of Monte Carlo and Bootstrap simulation procedures is with a simulation platform for training students in medical biochemistry (Tenekedjiev & Kolev, 2002). In this system, students search for estimates and confidence intervals of parameters of a given biochemical system for different enzyme-substrate pairs. The platform applies Monte Carlo simulation on two stages. Initially, a Monte Carlo procedure is applied to emulate a biochemical experimental measurement setting along with given enzyme kinetic reactions as realistically as possible. The system is in position to simulate continuous enzyme assay (used for adjustment of the “experimental” conditions) and end-point enzyme assay “measurements” (suitable for parameter identification). We use an ordinary differential equation (ODE) as basis of the generation of pseudo-experimental data. The pseudo-real nature of the generated data is ensured by the random incorporation of three types of errors for each repetition of the experiments. The Briggs-Haldane steady-state model is fitted to the pseudo-measured and end-point assay data obtained by the system. The kinetic parameters can be calculated by  $\chi^2$ -minimization. The task is simplified by the existence of a good initial guess from a linearized Lineweaver-Burk model. The two-dimensional root confidence regions of the parameters can be calculated by either Monte Carlo or Bootstrap, following similar procedures. The best point estimate is identified using trimmed mean over the flipped parameters taking only the values from the identified root confidence region.

In the majority of biochemical reactions, parameters are unknown in very wide intervals, and may have different numerical order. Finding the root confidence regions (intervals) includes parameter flipping, which often generates results with an incorrect sign. That is why, in a second example (Tanka-Salamon et al., 2008) we propose a multiplicative modification for the estimation of root confidence regions and the best estimate of the parameters, which ensures that all estimates will have a physical meaning. The main assumption is that the ratio between the true parameter value and the optimal parameter value derived from the true data sample has the same distribution as the ratio between the optimal parameter value derived from the true data sample, and the optimal synthetic parameter value derived from the synthetic data sample. The assumption is equivalent to performing classical Bootstrap over the logarithms of the estimated parameters. This method is applied in a real experimental set-up for the estimation of root confidence regions of kinetic constants and root best estimates in amidolytic activity of plasmin under the influence of three fatty acids. By doing so, the inhibition effect of the three fatty acids can be proven and quantified. The measured data have the form of continuous reaction progress curves with several replicas. The product concentrations are predicted by three different models with increasing complexity. We model the instability of the inhibited enzyme and represent the resulting continuous assay model with concomitant inactivation of the enzyme as a system of two stiff ODE. From there, we derive the closed form of the progress curve. The four-dimensional root confidence regions are acquired by Monte Carlo simulation in every data point in each of the progress curves using an analytical model of the measured standard deviation, similarly to the first example.

## 2. Monte Carlo and Bootstrap confidence regions of parameters

Statistical simulation methods are a powerful tool in the analysis of complex systems. The most popular among them is Monte Carlo. The numerical techniques that stand behind this

method are based on statistical simulation, i.e. on any method that uses random number sequences to conduct a simulation. The essence of the method is that it provides integral measures of uncertainty of the simulated system based on the known uncertainties of its parts (Hertz & Thomas, 1983). The integral measures are calculated on the basis of a large number of system instances in different pseudo-realities, each defined by a specific set of randomly generated states of its parts. The Monte Carlo approach is successfully employed in finding estimates of parameters that define the behavior of different stochastic systems. The flexibility, the very few premises imposed on the data and the application in hypothesis tests and statistical parameter assessment are what makes simulation techniques widely accepted.

Following the interpretation of (Press et al., 1992), one can assume that there is an experiment that intends to assess a certain set of  $M$  parameters that define the behavior of a measurable stochastic system. The true values  $\vec{a}_{true}$  of those parameters are unknown to the observer, but they are statistically realized in a set of real measurements  $D_0$  available to the observer (called a learning sample), which incorporates some random error. Then the experimenter can assess the parameters of a given model so that the discrepancy between the modeled and the measured data is minimized (e.g. using  $\chi^2$ -minimization or some other method) and a set of real parameters  $\vec{a}_0$  is formed. Due to the random character of the sample generating process, repetitions of the experimental measurement would generate many other possible measurement sets -  $D_1, D_2, D_3, \dots$  - with identical structure as  $D_0$ . Those in turn would generate sets of real parameters respectively  $\vec{a}_1, \vec{a}_2, \vec{a}_3, \dots$  that slightly differ from each other. So the estimate  $\vec{a}_0$  is just an instance of the  $M$ -dimensional random variable  $\vec{a}$ .

The task is to find the distribution of the deviation of the real parameters from the true ones, when just  $D_0$  is available. This random variable is called a *root* (Beran, 1986; DiGiccio & Romano, 1988). As long as  $\vec{a}_{true}$  is not known, the general approach is to create a fictitious world, where the true parameters are substituted with the real ones. The main assumption is that the root in the real world has the same distribution as in the fictitious world.

If we know the process that generates data under a given set of parameters  $\vec{a}_0$ , then we can simulate synthetic data sets  $D_1^S, D_2^S, \dots, D_Q^S$  with the same structure as the learning sample. The Bootstrap method (Efron & Tibshirani, 1993) applies to cases where the process that generates the data and/or the nature of the measurement error are unknown, and the learning sample  $D_0$  is formed out of  $N$  independent and identically distributed (*iid*) measurements. The Bootstrap generates synthetic samples  $D_1^S, D_2^S, \dots, D_Q^S$  with the same structure as  $D_0$  by drawing with replacement from  $D_0$ . In (Press et al., 1992) this type of Bootstrap is called *quick-and-dirty Monte Carlo*. There is another Bootstrap version, called *exact Bootstrap*, which forms all synthetic samples that can be generated by drawing with replacement from  $D_0$ . However, this method is rather impractical in a real problem so we shall not stress it here.

The parameters  $\vec{a}_1^S, \vec{a}_2^S, \dots, \vec{a}_Q^S$  can be identified from the synthetic samples in exactly the same way as  $\vec{a}_0$  was identified from  $D_0$ , which in turn would generate instances  $\vec{r}_q^f = \vec{a}_q^S - \vec{a}_0$  ( $q=1, 2, \dots, Q$ ) of the root  $\vec{r}^f$  in the fictitious world. If sufficiently large number of such instances  $Q$  is created, then it is possible to construct the empirical distribution of the root  $\vec{r}^f$  in the fictitious world, which according to the main assumption coincides with the distribution of the root  $\vec{r}^{true} = \vec{a} - \vec{a}_{true}$ .

The ultimate step in the simulation is to assess confidence intervals or confidence regions of the investigated  $M$ -dimensional parameter. There are different methods to assess one-dimensional CIs (e.g. percentile- $t$ , bootstrap- $t$ , bias-correction, simple bootstrap interval, etc.; see (Davison & Hinkley, 1997; MacKinnon & Smith, 1998)), yet for multi-dimensional CIs practically there are only two methods – root and percentile methods. The distinction of those two stems from the different approaches to probabilities in general.

If the frequentist definition of probabilities is adopted (Von Mises, 1946), then the likelihood inference has to be applied in parameter identification (Berger & Wolpert, 1988). Here the parameters  $\vec{a}_{true}$  are considered unknown, but deterministic values. In that sense, the estimate  $\vec{a}_0$  is the only value, which is random, so its confidence region can be calculated. Since the distribution of  $\vec{r}^f$  and  $\vec{r}^{true}$  coincide, then the following is true for the random variable  $\vec{a}$ :  $\vec{a} = \vec{a}_{true} + \vec{r}^{true} = \vec{a}_{true} + \vec{r}^f$ . Then, the instances  $\vec{a}_q$  of  $\vec{a}$  may be generated by replacing  $\vec{a}_{true}$  with  $\vec{a}_0$  as the only available estimate:  $\vec{a}_i = \vec{a}_0 + \vec{r}_q^f = \vec{a}_0 + \vec{a}_q^S - \vec{a}_0 = \vec{a}_q^S$ . The so-called *percentile confidence interval (or region)* is identified from the instances of the synthetic parameters  $\vec{a}_q^S$  and it is in fact the confidence region of the estimate.

If the subjective definition of probabilities is adopted (Jeffrey, 2004), then Bayesian statistics can be used in the parameter identification (Berry, 1996). Here the parameters  $\vec{a}_{true}$  are considered to be random variables, whose distribution can be assessed, and its confidence region may be calculated. Since the distributions of  $\vec{r}^f$  and  $\vec{r}^{true}$  coincide, then the following is true for the random variable  $\vec{a}$ :  $\vec{a}_{true} = \vec{a} - \vec{r}^{true} = \vec{a} - \vec{r}^f$ . Then, the instances  $\vec{a}_{true,q}$  of  $\vec{a}_{true}$  may be generated by replacing  $\vec{a}$  with  $\vec{a}_0$  as the only available estimate:  $\vec{a}_{true,q} = \vec{a}_0 - \vec{r}_q^f = \vec{a}_0 - (\vec{a}_q^S - \vec{a}_0) = 2\vec{a}_0 - \vec{a}_q^S$ . The so-called *root confidence interval (or region)* is identified from the instances of the flipped (around the real parameters) synthetic parameters  $\vec{a}_q^{S,f} = 2\vec{a}_0 - \vec{a}_q^S$  and it is in fact the confidence region of the true parameter. Furthermore, using this approach it is possible to find a point estimate  $\vec{a}_{best}$  of  $\vec{a}_{true}$  that is better than  $\vec{a}_0$ . One possibility is to find  $\vec{a}_{best}$  as the mean value of the flipped synthetic parameters  $\vec{a}_q^{S,f}$ . Since the method is sensitive to errors (Davidson & MacKinnon, 1999), it is better to use trimmed mean value (Hanke & Reitsch, 1991). Regardless of the type of mean value, the resulting point estimate  $\vec{a}_{best}$  is unbiased unlike  $\vec{a}_0$ .

### 3. Generation and processing of data in enzyme kinetics

Metabolomics (Strohman, 2002) deals with the evaluation of the dynamic metabolic networks and links the genetic information to the phenotype through adequate metabolic control analysis. It is a prerequisite in the understanding of the cellular phenotype and its pathological alterations. For that purpose, one not only requires the technical developments that allow stringent monitoring of the metabolic fluctuations induced *in vivo* by biological signals and environmental changes, but also powerful mathematical models capable of treating dynamic metabolic systems in their variability. The principles of metabolomics are well known in biochemistry (Newsholme & Leech, 1984; Voet & Voet, 1995), but the training of biomedical and clinical researchers is still insufficient to exploit the opportunities provided by the up-to-date computer-intensive statistical procedures applicable in the field of enzyme kinetics. For that purpose, our earlier work (Tenekedjiev & Kolev, 2002) introduces a computer-simulated experimental setting, in which the user (a graduate medical student, who is familiar with the basic ideas in enzyme kinetics and the structure of metabolic pathways) acquires skills in adjusting experimental conditions to conform model

assumptions, in identification of kinetic parameters and determination of their confidence intervals, in application of these parameters for metabolic predictions in context-dependent *in vivo* situations. In the proposed system, students search for estimates and confidence intervals of the parameters  $k_p$  and  $K_M = \frac{k_{-1} + k_p}{k_1}$  of an enzyme-catalyzed reaction



for different enzyme-substrate pairs, where the substrate degrades into product under the influence of the enzyme. The enzyme and substrate bind reversibly to form enzyme-substrate complex, which in turn irreversibly transforms into product.

The proposed system applies Monte Carlo simulations at two levels.

### 3.1 Simulation of the experimental equipment data generation

Initially, the system uses Monte Carlo simulations to emulate enzyme kinetic reactions as close as possible to the real lab setting. The user can perform continuous enzyme assay and end-point assay “measurements”.

The continuous enzyme assay simulates an expensive experiment, where the product concentration is measured at equally spaced time intervals (sampling time), and the mean reaction rate is calculated for different conditions. Here it is not envisaged to make repetitions and perform identification of the parameters. The user sets up the enzyme-substrate pair, the designed concentrations of the total enzyme  $E_t^{des}$  (the free enzyme concentration plus the enzyme-substrate complex concentration) and the initial substrate  $S_0^{des}$ , the temperature  $T$ , the  $pH$ , the overall experimental time  $t^{over}$  and the sampling time  $\Delta t$  (see Fig. 3.1). As a result, the user gets the time course of the pseudo-measured product  $P_{mes}(i, \Delta t)$ , for  $i=1, 2, \dots, t^{over}/\Delta t$ . The time course of the substrate, the product, the free enzyme and the enzyme-substrate complex are recalculated from  $P_{mes}$  (see Fig. 3.2). These results form a single replica of the process.

conditions of the continuous assay for the study of effects on reaction rate

reaction E2-B ; in vivo concentrations: total enzyme- 0.0003(mM) substrate- 0.024(mM)  
to study the time-course of the concentrations of E, ES, S and P  
fill in the conditions for the continuous assay.

overall experiment time in (min)	60
temperature in (degrees C)	25
pH	5
total enzyme concentration in (mM)	0.00034
initial substrate concentration in (mM)	0.24
sampling time in (sec)	40

OK Cancel Help

Fig. 3.1. Setting up the designed concentrations of the total enzyme  $E_t^{des}$  and the initial substrate  $S_0^{des}$ , the temperature  $T$ , the acidity  $pH$ , the overall experimental time  $t^{over}$  and the sampling time  $\Delta t$  for a given enzyme-substrate pair

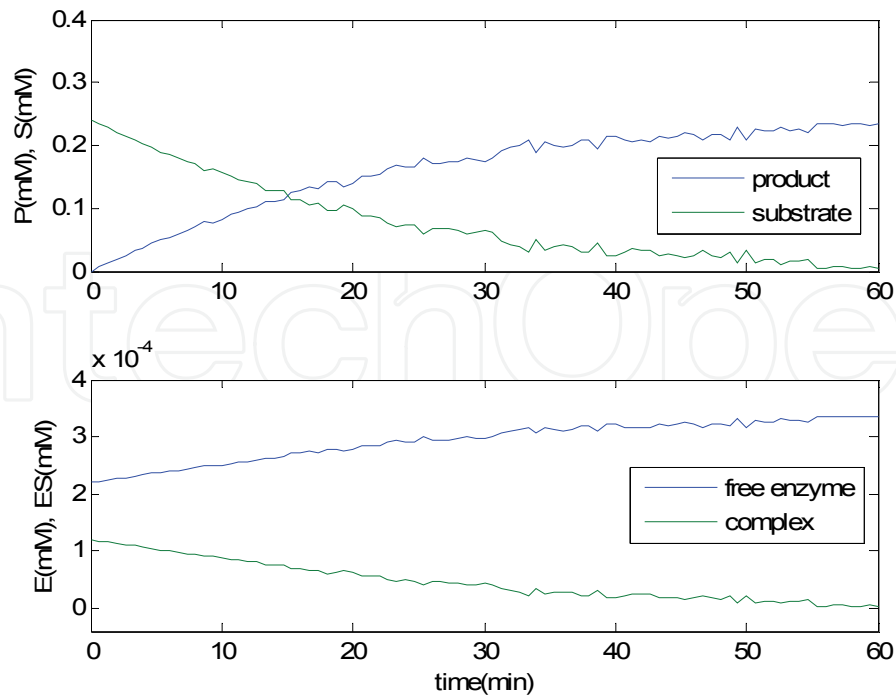


Fig. 3.2. Time course of the substrate, the product, the free enzyme and the enzyme-substrate complex from the setting in Fig. 3.1

For each replica we adopted the Briggs-Haldane steady-state approach (Segel, 1993) to model the transformation of the substrate into product in the biochemical system (3.1) using the ODE:

$$\frac{dP^{true}}{dt} = \frac{k_{p,app}(t, T, pH) \cdot E_t^{true} \cdot (S_0^{true} - P^{true})}{K_{M,app}(pH) + S_0^{true} - P^{true}}, \quad (3.2)$$

In (3.2), the true concentration of the product  $P^{true}$  is a function of the time  $t$ ;  $E_t^{true}$  is the true concentration of the total enzyme;  $S_0^{true}$  is the true concentration of the substrate;  $k_{p,app}$  and  $K_{M,app}$  are the apparent constants  $k_p$  and  $K_M$  in this replica. The initial condition of (3.2) is that the true concentration of the product is zero at the beginning:  $P^{true}(0)=0$ .

To model the biological diversity of the substrate-enzyme pair in the real experiment,  $k_{p,app}$  and  $K_{M,app}$  are instances of random variables. The apparent constant  $K_{M,app}$  depends also on  $pH$ , whereas  $k_{p,app}$  depends on  $pH$ ,  $T$ , and  $t$ :

$$K_{M,app} = K_M \cdot \frac{\left(1 + \frac{1}{10^{pH} \cdot Ke_1} + 10^{pH} \cdot Ke_2\right) \cdot \left(1 + \frac{1}{10^{pH} \cdot Ka_1} + 10^{pH} \cdot Ka_2\right)}{1 + \frac{1}{10^{pH} \cdot Kes_1} + 10^{pH} \cdot Kes_2} \cdot R_m \quad (3.3)$$

$$k_{p,app} = \frac{10^{\frac{A_1}{T+273.16} - \frac{A_2}{T+273.16}}}{1 + \frac{1}{10^{pH} \cdot Kes_1} + 10^{pH} \cdot Kes_2} \cdot 10^{-Bt} \cdot R_p \quad (3.4)$$

$$\text{where, } B = \begin{cases} A_3 \cdot (T - T_d)^g \cdot R_B & \text{if } T > T_d \\ 0 & \text{if } T \leq T_d \end{cases} \quad (3.5)$$

In (3.3), (3.4) and (3.5),  $R_p$ ,  $R_m$  and  $R_B$  are instances of a positive continuous uniformly distributed random variable on the interval  $[1 - \Delta_{par}; 1 + \Delta_{par}]$ , where  $\Delta_{par} < 1$  is given. The constants  $T_d$ ,  $A_1$ ,  $A_2$ ,  $A_3$ ,  $g$ ,  $Ke_1$ ,  $Ke_2$ ,  $Kes_1$ ,  $Kes_2$ ,  $Ka_1$ ,  $Ka_2$  are known typical constants for each enzyme-substrate pair, with the following meaning:  $T_d$  - the temperature over which the enzyme degradation begins (in degrees C);  $A_1$  - the logscale factor of the apparent  $k_p$  constants at  $t=0$  (in 1/min);  $A_2$  - the heat acceleration factor of the apparent  $k_p$  constants at  $t=0$  (in 1/min);  $A_3$  - the scale factor of the temporal enzyme degradation constant  $B$ ;  $g$  - the power in the temporal enzyme degradation constant  $B$ ;  $Ke_1$  - the pK value for the first H<sup>+</sup> dissociation constant of the free enzyme;  $Ke_2$  - pK value for the second H<sup>+</sup> dissociation constant of the free enzyme;  $Kes_1$  - the pK value for the first H<sup>+</sup> dissociation constant of the enzyme-substrate complex;  $Kes_2$  - the pK value for the second H<sup>+</sup> dissociation constant of the enzyme-substrate complex;  $Ka_1$  - the pK value of the first acidic dissociation group in the diprotic substrate;  $Ka_2$  - the pK value of the second acidic dissociation group in the diprotic substrate.

To model the imperfection of the setup in the real experiment,  $S_0^{true}$  and  $E_t^{true}$  are instances of random variables, which slightly deviate from the designed  $S_0^{des}$  and  $E_t^{des}$ :

$$E_t^{true} = E_t^{des} \cdot R_E \quad (3.6)$$

$$S_0^{true} = S_0^{des} \cdot R_S \quad (3.7)$$

In (3.6) and (3.7),  $R_E$  and  $R_S$  are instances of positive continuous normally distributed random variables with unit mean values and standard deviations respectively  $c \cdot E_t^{des} + d$ , and  $c \cdot S_0^{des} + d$ , where  $c$  and  $d$  are given.

To model the measurement error in the real experiment, each of the pseudo-measured values of product concentration  $P^{mes}(i, \Delta t)$ , for  $i=1, 2, \dots, t^{over}/\Delta t$  is an instance of a random variable, which slightly deviates from the true product concentration  $P^{true}(i, \Delta t)$ , for  $i=1, 2, \dots, t^{over}/\Delta t$  that results from integrating the ODE (3.2) from 0 to  $t^{over}$ :

$$P^{mes}(i, \Delta t) = P^{true}(i, \Delta t) \cdot R_{P,i}, \text{ for } i=1, 2, \dots, t^{over}/\Delta t \quad (3.8)$$

In (3.8),  $R_{P,i}$  are instances of positive continuous normally distributed random variables with unit mean values and standard deviations  $b \cdot (P^{true}(i, \Delta t))^a$ , where  $a$  and  $b$  are given.

After finding  $P^{mes}$  it is possible to find the measured time course of the substrate, of the enzyme-substrate complex, and of the free enzyme:

$$\begin{aligned} S^{mes}(i, \Delta t) &= S_0^{true} - P^{mes}(i, \Delta t) \\ ES^{mes}(i, \Delta t) &= \frac{E_t^{true} \cdot S^{mes}(i, \Delta t)}{K_{M,app} + S^{mes}(i, \Delta t)}, \text{ for } i=1, 2, \dots, t^{over}/\Delta t \\ E^{mes}(i, \Delta t) &= E_t^{true} - ES^{mes}(i, \Delta t) \end{aligned} \quad (3.9)$$

The quantities in (3.8) and (3.9) are shown on Fig. 3.2. The measured velocity of the process can be approximated with the formula:



$$V_{mes} = P_{mes}(t_{over})/t_{over} \quad (3.10)$$

The purpose of the end-point assay experiment is to teach the user how to manually setup "optimal" experimental conditions. First of all, the steady-state assumptions have to be checked (that the concentration of the enzyme-substrate complex should stay approximately the same). For example, Fig. 3.2 shows that this condition is not met under the selected experimental conditions. An experiment that meets the requirements of the steady-state is shown on Fig. 3.3. Students can visually check the validity of the empirical "criteria" for the steady-state model (the degraded substrate should be up to 10% of the initial one, and the total enzyme concentration should be far less than the concentration of the initial substrate plus the constant  $K_{M,app}$ ), which otherwise should be proven with complex mathematical procedures (Segel, 1988). The influence of temperature, of acidity, of the substrate and enzyme concentrations, and the overall time of the experiment can be estimated by constructing the velocity graphics of the reaction as a function of the investigated condition. For example, Fig. 3.4 shows the influence of temperature on the product formation. The first graphics shows a continuous monitoring of the product generation in the course of the enzyme-catalysed reaction at various temperatures. The continuous assay illustrates also the intrinsic errors of the sampling procedure, which impose the necessity for repeated sampling. The second graph shows the end-point enzyme activity assay for various sampling time. The data are cross-section of the continuous assay. for 1- and 10-min incubation. Comparison of the two curves illustrates the apparent nature of the "optimal" temperature caused by the time-dependence of the heat denaturation.

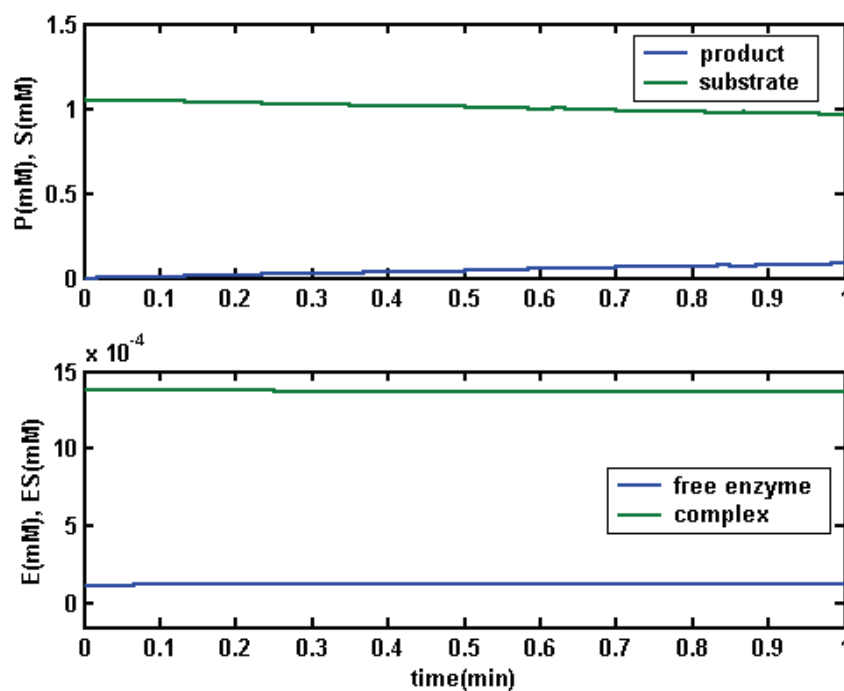


Fig. 3.3. Adjustment of the steady-state experimental conditions

As a whole, the instrumental model for a continuous-time assay is used as a manual Monte Carlo system to find suitable experimental conditions, appropriate for the end-point assay experiment, where the values of the kinetic constants and their confidence regions are identified. The end-point assay simulates a cheap experiment, where the user sets up  $T$ ,  $pH$ ,

$t_{over}$ ,  $E_t^{des}$ ,  $J$  predetermined initial substrate concentrations  $S_{0,1}^{des}, S_{0,2}^{des}, \dots, S_{0,J}^{des}$ , and the number  $K$  of replicas for any substrate concentration. The product concentration is measured  $K$  times just at time  $t_{over}$  for each  $S_{0,j}^{des}$  (for  $j=1, 2, \dots, J$ ).

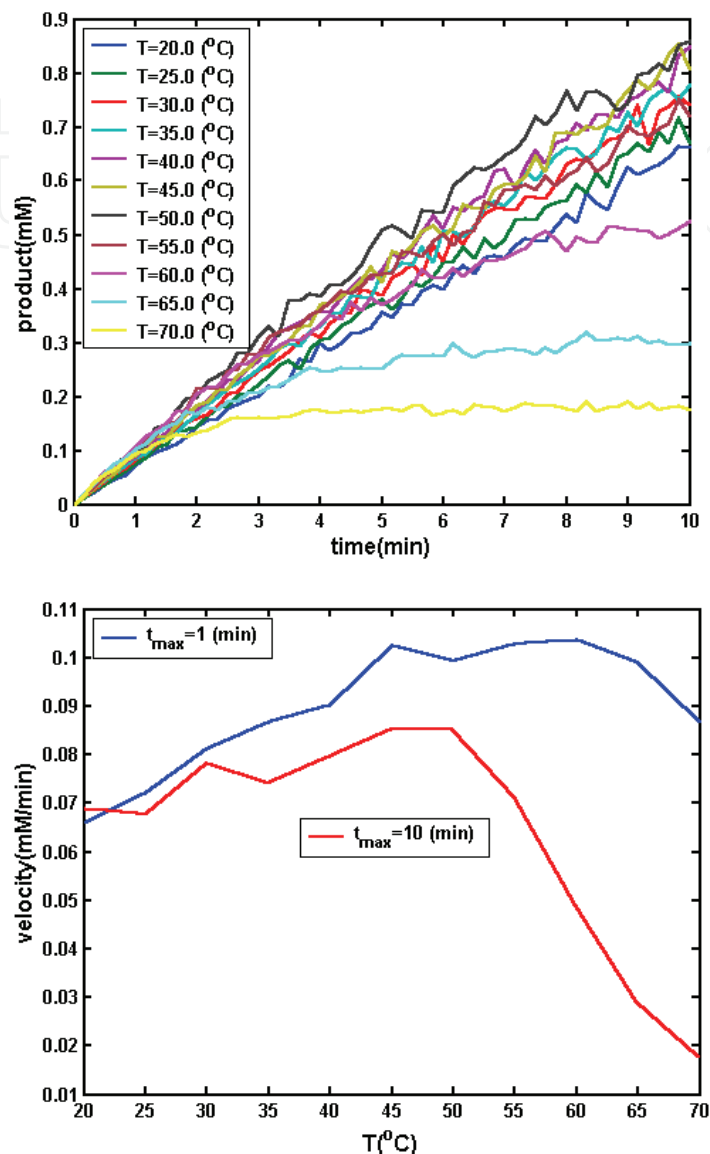


Fig. 3.4. General conditions for optimal enzyme action

### 3.2 Kinetic parameter identification with Monte Carlo or Bootstrap simulation

Each replica is simulated in the same way as in the continuous-time assay, but the learning sample  $D_0$  only consists of the product concentrations at time  $t_{over}$ :

$D_0 = \{P_{j,k}^{mes} \mid j=1,2,\dots,J; k=1,2,\dots,K\}$ , where  $P_{j,k}^{mes}$  is the  $k$ -th measurement of the product concentration  $P^{mes}$  at substrate concentration  $S_{0,j}^{des}$ . As long as all the designed experimental conditions are identical for  $P_{j,k}^{mes}$  ( $k=1, 2, \dots, K$ ), they can be referred as a process. Let

$P_j^{mes,mean}$  and  $P_j^{mes,std}$  are the mean value and the standard deviation for the end-point

product concentration of the  $j^{\text{th}}$  process, calculated by the  $K$  instances  $P_{j,k}^{\text{mes}}$  ( $k=1, 2, \dots, K$ ). A non-linear regression model of the standard deviation of the measured final product concentration is created as a function of the mean value of the measured final product concentration:

$$P_j^{\text{mod, std}} = P_j^{\text{mes, mean}} \cdot 10^{a_1 \cdot \lg(P_j^{\text{mes, mean}}) + a_2}, \text{ for } j=1, 2, \dots, J \quad (3.11)$$

The constants  $a_1$  and  $a_2$  are determined with  $\chi^2$ -minimization of:

$$\chi^2(a_1, a_2) = \sum_{j=1}^J \left( \lg(P_j^{\text{mes, std}}) - (a_1 + 1) \cdot \lg(P_j^{\text{mes, mean}}) - a_2 \right)^2 \quad (3.12)$$

The final product concentrations are predicted by a model different from (3.2), which takes the form of  $J$  number of ODEs, which can be solved separately:

$$\frac{dP_j^{\text{mod}}}{dt} = \frac{k_p \cdot E_t^{\text{des}} \cdot (S_{0,j}^{\text{des}} - P_j^{\text{mod}})}{K_M + S_{0,j}^{\text{des}} - P_j^{\text{mod}}}, \text{ for } j=1, 2, \dots, J \quad (3.13)$$

The initial conditions of (3.13) are  $P_j^{\text{mod}}(0) = 0$ . In (3.13),  $k_p$  and  $K_M$  are the kinetic constants. After integrating the ODE (3.13) from time 0 to  $t^{\text{over}}$ , the value at  $t^{\text{over}}$  would depend on the values of  $k_p$  and  $K_M$ . At the same time,  $P_j^{\text{mes, mean}}$  and  $P_j^{\text{mod, std}}$  would only depend on the sample  $D_0$ . Then the optimal parameters  $k_{p,0}$  and  $K_{M,0}$  can be found with  $\chi^2$ -minimization of

$$\chi^2(k_p, K_M, D_0) = \sum_{j=1}^J \left( \frac{P_j^{\text{mod}}(t^{\text{over}}, k_p, K_M) - P_j^{\text{mes, mean}}(D_0)}{P_j^{\text{mod, std}}(D_0)} \right)^2 \quad (3.14)$$

$$\vec{a}_0 = (k_{p,0}, K_{M,0}) = \arg \left( \min_{k_p, K_M} \chi^2(k_p, K_M, D_0) \right) \quad (3.15)$$

Solving (3.15) is simplified by the presence of a good initial guess from a linearized model of Lineweaver-Burk (Lineweaver & Burk, 1934).

The two-dimensional confidence region of  $k_p$  and  $K_M$  is again calculated by Monte Carlo or Bootstrap simulation. The synthetic samples in the Monte Carlo simulation  $D_q^S$  ( $q=1, 2, \dots, Q$ ) are formed so that for the  $j$ -th process, the final product concentrations are generated as  $K$  number of instances of a normally distributed random variable with mean value  $P_j^{\text{mes, mean}}$  and standard deviation  $P_j^{\text{mod, std}}$ . The synthetic samples in the Bootstrap simulation  $D_q^S$  ( $q=1, 2, \dots, Q$ ) are formed so that for the  $j$ -th process, the final product concentrations are generated by  $K$  drawing with replacement from the set  $\{P_{j,1}^{\text{mes}}, P_{j,2}^{\text{mes}}, \dots, P_{j,K}^{\text{mes}}\}$ . Whatever the method for generating  $D_q^S$ ,

$$\vec{a}_q^S = (k_{p,q}^S, K_{M,q}^S) = \arg \left( \min_{k_p, K_M} \chi^2(k_p, K_M, D_q^S) \right) \quad (3.16)$$

In order to find the root confidence interval, the resulting parameters are flipped in accordance with the discussion in section 2:

$$\vec{a}_q^{S,f} = 2\vec{a}_0 - \vec{a}_q^S = (2.k_{p,0} - k_{p,q}^S, 2.K_{M,0} - K_{m,q}^S) = (k_{p,q}^{S,f}, K_{M,q}^{S,f}) \quad (3.17)$$

Let  $\chi_q^{2,S} = \chi^2(k_{p,q}^{S,f}, K_{M,q}^{S,f}, D_0)$  be the discrepancy measure (3.14), calculated over the original sample with the flipped synthetic parameters. Let's renumerate the acquired discrepancy measures  $\chi_q^{2,S}$  in ascending order so that  $\chi_1^{2,S} \leq \chi_2^{2,S} \leq \dots \leq \chi_Q^{2,S}$ . Then the first  $Q \cdot p$  sorted vectors belong to the two-dimensional (2-D) confidence region with probability  $p$ . These vectors are called inside simulated points, and the rest are the outside simulated points. The area of the inside simulated points determines the root confidence region and its borders have constant  $\chi^2$  discrepancy measure. The projections of the confidence region on the  $K_M$  and  $k_p$  axes are the 2-D confidence intervals (Fig. 3.5.). The best point estimate of the parameters is calculated as trimmed mean of  $\vec{a}_q^{S,f}$  using just the inside simulated points. The percentile confidence region can be calculated likewise, but instead of the flipped synthetic parameters, one should use just the synthetic parameters.

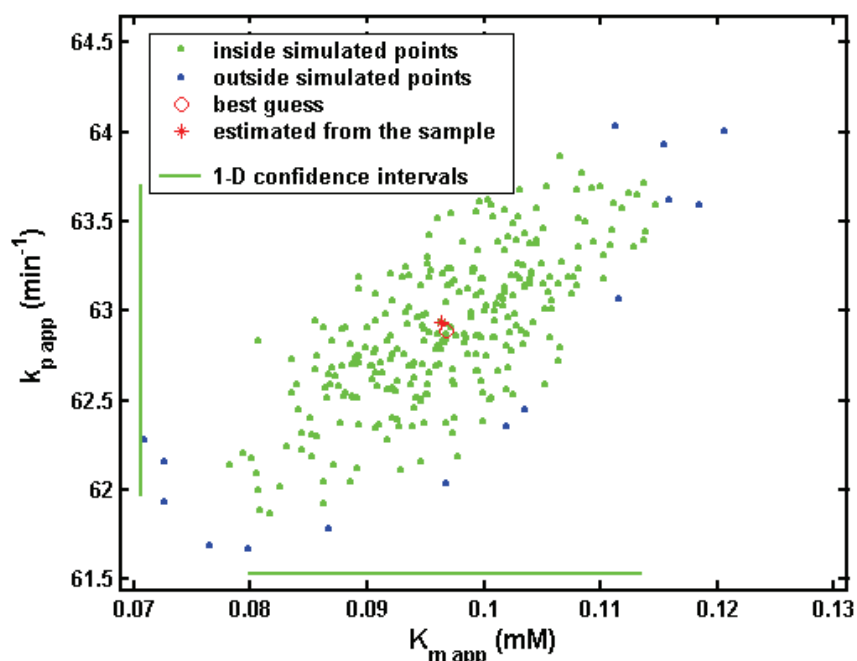


Fig. 3.5. Confidence area of the model parameters based on a Monte Carlo simulation procedure.

The described simulation system has been employed for over 8 years to train medical students in interpretation of the stochastic nature of experimentally estimated model-parameters at the Department of Medical Biochemistry of the Semmelweis University in Budapest (Hungary). After this training, students (and the users in general) are able to perceive the imperative requirements for multiple sampling replicas in experimentation, to interpret the stochastic nature of experimentally estimated model-parameters, as well as to gain insights into the application of in vitro determined kinetic parameters for the modeling of in vivo metabolic events.

#### 4. Influence of fatty acids on the amidolytic activity of plasmin

The dissolution of intravascular thrombi is performed through the hydrolytic degradation of their fibrin matrix catalyzed by the serine protease plasmin (Kolev et al., 2005). Arterial thrombi enclose millimolar concentrations of phospholipids (Varadi et al., 2004) and free fatty acids (Rabai et al., 2007). These lipid constituents of thrombi are reported to modulate the fibrinolytic process (Varadi et al., 2004, Rabai et al., 2007; Hazari et al., 1992; Hazari et al., 1994; Huet et al., 2004). The paper (Tanka-Salamon et al., 2008) investigates the effects of the three most abundant fatty acids in the structure of platelet membranes – arachidonic acid, stearic acid and oleic acid representing respectively 22.0, 19.5 and 18.8 % of the total fatty acid content of platelet phosphoglycerolipids (Rabai et al., 2007).

Plasmin ( $E_t=20$  nM) was incubated with sodium salts of fatty acids for 15 min at 37 °C. Then 180  $\mu$ l of this mixture was added to 20  $\mu$ l Spectrozyme-PL (H-D-norleucyl-hexahydroxytyrosyl-lysine-p-nitroanilide, American Diagnostica, Stamford, CT) at 7 different concentrations in the range 0.05 – 6 mM yielding final concentration  $S_{0j}$  ( $j=1,2,\dots,7$ ) in the volume of the reaction mixtures. The light absorbance at 405 nm ( $A_{405}$ ), which reflects the release of *p*-nitraniline, was measured continuously at  $t_i$  ( $i=1,2,\dots,60$ ) time points in the course of 10 min at 37 °C, 4 parallel measurements were done for each  $S_{0j}$ . The main problem in the initial data processing is the proper assessment of the baseline absorbance and the delay time of the measurement, which affect profoundly the absolute values of the pNA product. An original algorithm was employed to convert the measured absorbance into product concentration  $P_{i,j,k}^{mes}$  (the notation indicates the product concentration at time  $t_i$

for the  $k$ -th replica with  $S_{0j}$ ) which form the learning sample  $D_0$ :  $D_0 = \{P_{i,j,k}^{mes} | i=1,2,\dots,60; j=1,2,\dots,7; k=1,2,3,4\}$ . As long as all the designed experimental conditions are identical for  $P_{i,j,k}^{mes}$  ( $k=1, 2, 3, 4$ ), they can be referred as a process. Let

$P_{i,j}^{mes,mean}$  and  $P_{i,j}^{mes,std}$  are the mean value and the standard deviation for the product concentration of the  $j$ <sup>th</sup> process at time  $t_i$ , calculated by four instances  $P_{i,j,k}^{mes}$  ( $k=1, 2, 3, 4$ ). A non-linear regression model of the standard deviation of the measured product concentration for each process is created as a function of the mean value of the measured product concentration:

$$P_{i,j}^{mod,std} = b_j \cdot (P_{i,j}^{mes,mean})^{a_j}, \text{ for } j=1, 2, \dots, 7 \quad (4.1)$$

The constants  $a_j$  and  $b_j$  are determined with  $\chi^2$ -minimization of:

$$\chi_j^2(a_j, b_j) = \sum_{j=1}^7 \sum_{i=1}^{60} \left( \lg(P_{i,j}^{mes,std}) - a_j \cdot \lg(P_{i,j}^{mes,mean}) - \lg(b_j) \right)^2 \quad (4.2)$$

The product concentrations are predicted by three different models with increasing complexity. In the simplest case (*Model I*) the scheme  $E + S \xrightleftharpoons[k_{-1}]{k_1} ES \xrightarrow{k_2} E + P$  is assumed, where  $E$  is plasmin,  $S$  is Spectrozyme-PL,  $P$  is *p*-nitroaniline,  $k_1$ ,  $k_2$  and  $k_{-1}$  are the respective reaction rate constants. With the quasi-steady-state assumption the differential rate equation for this scheme is

$$\frac{dP_j^{mod}}{dt} = \frac{k_p \cdot E_{t0} \cdot (S_{0,j} - P_j^{mod})}{K_M + S_{0,j} - P_j^{mod}}, \text{ for } j=1,2, \dots, 7 \quad (4.3)$$

where  $E_{t0}$  and  $S_0$  are the initial concentrations of plasmin and its substrate,  $K_M = \frac{k_{-1} + k_2}{k_1}$  is the Michaelis constant, whereas  $k_p = k_2$  is the catalytic constant (Cornish-Bowden, 2004). The integration of (4.3) under initial condition  $P_j^{mod}(0) = 0$  for  $j=1,2, \dots, 7$  gives

$$t = \frac{1}{k_p \cdot E_{t0}} P_j^{mod} + \frac{K_M}{k_p \cdot E_{t0}} \ln \frac{S_{0,j}}{S_{0,j} - P_j^{mod}} \quad (4.4)$$

The time  $t$  in (4.4) is a strictly increasing function of  $P_j^{mod}$  for any combination of  $K_M$  and  $k_p$  and therefore it has an inverse function  $P_j^{mod} = P_j^{mod}(t, K_M, k_p, S_{0,j}, E_{t0})$ , which can be numerically estimated for all measured time points  $t_i$  by a look-up table procedure and the results can be denoted as  $P_{i,j}^{mod}$ ,  $i=1,2, \dots, 60$ ;  $j=1,2, \dots, 7$ .

Because in the course of certain experiments the reaction rate declined faster than predicted by *Model I*, the more general scheme  $E + S \xrightleftharpoons[k_{-1}]{k_1} ES \xrightarrow{k_2} EP \xrightleftharpoons[k_{-3}]{k_3} E + P$  was also tested (*Model II*), which accounts for the accumulation of the product and its complex with the enzyme. Assuming steady-state for both  $ES$  and  $EP$  complexes the differential rate-equation is

$$\frac{dP_j^{mod}}{dt} = \frac{k_p \cdot E_{t0} \cdot (S_{0,j} - P_j^{mod})}{K_M \cdot (1 + K_i \cdot P_j^{mod}) + S_{0,j} - P_j^{mod}} \quad (4.5)$$

where  $K_M = \frac{(k_{-1} + k_2) \cdot k_3}{k_1(k_2 + k_3)}$  is the Michaelis constant,  $k_p = \frac{k_2 \cdot k_3}{k_2 + k_3}$  is the catalytic constant,

and  $K_i = \frac{k_{-3}}{k_3}$  is the equilibrium association constant for the product. Although the

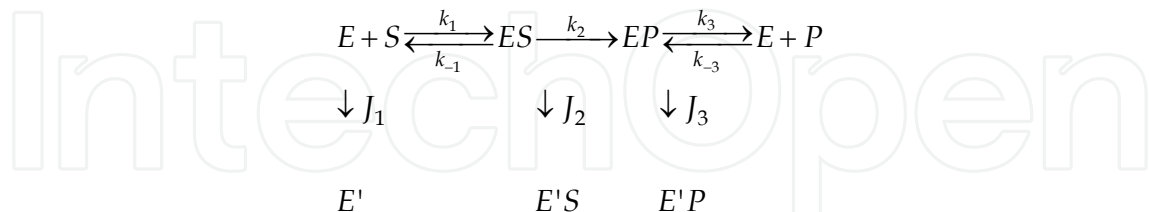
Michaelis constant and the catalytic constant derived for *Model I* and *Model II* have different algebraic form, their meaning within the context of the specific catalytic mechanism is identical; the  $K_M$  is the substrate concentration at which the initial reaction rate is half of the maximal rate possible for given enzyme concentration, whereas  $k_p$  has the properties of a first-order rate constant defining the capacity of the enzyme-substrate complex to form product (Cornish-Bowden, 2004). Integrating (4.5) under initial condition  $P_j^{mod}(0) = 0$  for  $j=1,2, \dots, 7$  gives

$$t = \frac{1 - K_M \cdot K_i}{k_p \cdot E_{t0}} P_j^{mod} + \frac{K_M(1 + S_{0,j} \cdot K_i)}{k_p \cdot E_{t0}} \ln \frac{S_{0,j}}{S_{0,j} - P_j^{mod}} \quad (4.6)$$

The time  $t$  in (4.6) is a strictly increasing function of  $P_j^{mod}$  for any combination of  $K_M$ ,  $k_p$  and  $K_i$  and therefore it has an inverse function  $P_j^{mod} = P_j^{mod}(t, K_M, k_p, K_i, S_{0,j}, E_{t0})$ , which can be

numerically estimated for all measured time points  $t_i$  by a look-up table procedure and the results can be denoted as  $P_{i,j}^{mod}$ ,  $i=1,2,\dots, 60; j=1,2,\dots, 7$ .

Because in certain cases the product inhibition could not model the progress curve of the reaction satisfactorily, the instability of the enzyme in the assay system was also considered according to the scheme suggested in (Duggleby, 1995; Duggleby, 2001) (*Model III*):



in which  $E'$  indicates the inactive form of the enzyme,  $J_1$ ,  $J_2$  and  $J_3$  are the rate constants for inactivation of the respective forms of the enzyme ( $J_1$  and  $J_2$  are decay rate constants of the enzyme-substrate complex;  $J_3$  is the decay rate constant of the enzyme product complex). In independent measurements with fatty acids we showed that the inactivation of free plasmin for the duration of the amidolytic assay was negligible (data not shown) and consequently the differential equation for the changes in enzyme concentration was derived only for the scheme with  $J_2$  and  $J_3$  (i.e.  $J_1=0$ ) yielding seven separately solvable systems of two ODEs:

$$\left. \begin{aligned}
 \frac{dP_j^{mod}}{dt} &= \frac{k_p \cdot E_{t,j} \cdot (S_{0,j} - P_j^{mod})}{K_M \cdot (1 + K_i \cdot P_j^{mod}) + S_{0,j} - P_j^{mod}} \\
 \frac{dE_{t,j}}{dt} &= - \frac{J_2 \cdot E_{t,j} \cdot (S_{0,j} - P_j^{mod}) / K_M + J_3 \cdot E_{t,j} \cdot K_i \cdot P_j^{mod}}{1 + K_i \cdot P_j^{mod} + (S_{0,j} - P_j^{mod}) / K_M}
 \end{aligned} \right\} \text{for } j=1,2,\dots,7 \quad (4.7)$$

The initial conditions of (4.7) are  $P_j^{mod}(0) = 0$ ,  $E_{t,j}(0) = E_{t0}$ . In (4.7),  $k_p$  and  $K_M$  have the same meaning as in *Model II*. The integration of ODE systems (4.7) from time 0 to  $t_{60}$  was done by quasi-constant step size implementation in terms of backward differences of the Klopfenstein-Shampine family of Numerical Differentiation Formulas of orders 1-5 and the initial steps were determined so that the solution would stay in its domain ( $0 \leq P_j^{mod} \leq S_{0,j}$ ,  $0 \leq E_{t,j} \leq E_{t0}$ ) during the whole integration (Shampine et al., 2003). The values of the product concentrations at time points  $t_i$  for *Model III* can be found from the first component of the solution:  $P_{i,j}^{mod} = P_j^{mod}(t_i, K_M, k_p, K_i, J_2, J_3, S_{0,j}, E_{t0})$  for  $i=1, 2, \dots, 60$  and for  $j=1,2,\dots, 7$ .

Since *Model I* and *Model II* are special cases of *Model III* for given values of  $J_2$ ,  $J_3$  and  $K_i$ , then further discussion only refers to *Model III*.

As in section 3,  $P_j^{mod}$  depends only on the kinetic parameters  $K_M, k_p, K_i, J_2, J_3$ ;  $P_{i,j}^{mes,mean}$  and  $P_{i,j}^{mod,std}$  would only depend on the sample  $D_0$ . Then the optimal parameters  $k_{p,0}$ ,  $K_{M,0}$ ,  $K_{i,0}$ ,  $J_{2,0}$ ,  $J_{3,0}$  can be found with  $\chi^2$ -minimization of

$$\chi^2(k_p, K_M, K_i, J_2, J_3, D_0) = \sum_{j=1}^7 \sum_{i=1}^{60} \left( \frac{P_{i,j}^{mod}(k_p, K_M, K_i, J_2, J_3) - P_{i,j}^{mes,mean}(D_0)}{P_{i,j}^{mod,std}(D_0)} \right)^2 \quad (4.8)$$

$$\bar{a}_0 = (k_{p,0}, K_{M,0}, K_{i,0}, J_{2,0}, J_{3,0}) = \arg \left( \min_{k_p, K_M, K_i, J_2, J_3} \chi^2(k_p, K_M, K_i, J_2, J_3, D_0) \right) \quad (4.9)$$

The minimization (4.9) was performed using the Nelder-Mead simplex direct search method (Lagarias et al., 1998). Since the optimization assigns values of less than  $10^{-12}$  ( $\text{sec}^{-1}$ ) to  $J_2$ , from now on only  $J_3$  shall be considered.

The four-dimensional confidence region of  $k_p$ ,  $K_M$ ,  $K_i$ , and  $J_3$  is calculated by Monte Carlo simulation. The synthetic samples  $D_q^S$  ( $q=1,2,\dots,Q$ ) are formed so that for the  $i$ -th time point of the  $j$ -th process, the product concentrations are generated as four instances of a normally distributed random variable with mean value  $P_{i,j}^{mes,mean}$  and standard deviation  $P_{i,j}^{mod,std}$ . Then

$$\bar{a}_q^S = (k_{p,q}^S, K_{M,q}^S, K_{i,q}^S, J_{3,q}^S) = \arg \left( \min_{k_p, K_M, K_i, J_3} \chi^2(k_p, K_M, K_i, J_3, D_q^S) \right) \quad (4.10)$$

Let's find the root confidence region of the parameters. That requires flipping the parameters as in (3.17). However, the resulting values may have no physical meaning (could be negative) unlike the case in section 3, because here we operate with real measurements, and there is no appropriate initial guess for the optimization. This situation is a rule rather than an exception in biochemical analysis, where parameters are supposed to be strictly positive. Therefore, one possible solution to find the root confidence region in such cases is to use a modification of the classical Monte Carlo procedure, called *multiplicative Monte Carlo*. The main assumption here, in accordance with section 1, is that the ratio between the true parameter value and the optimal parameter value derived from the true data sample  $\frac{\bar{a}_{true}}{\bar{a}}$  has the same distribution as the ratio between the optimal parameter value derived from the true data sample, and the optimal synthetic parameter value derived from the synthetic data sample  $\frac{\bar{a}_0}{\bar{a}_q^S}$ . The assumption is equivalent to performing classical Bootstrap

over the logarithms of the estimated parameters. Then, the flipped parameters are derived as follows:

$$\bar{a}_q^{S,f} = \frac{2\bar{a}_0}{\bar{a}_q^S} = \left( \frac{2 \cdot k_{p,0}}{k_{p,q}^S}, \frac{2 \cdot K_{M,0}}{K_{M,q}^S}, \frac{2 \cdot K_{i,0}}{K_{i,q}^S}, \frac{2 \cdot J_{3,0}}{J_{3,q}^S} \right) = (k_{p,q}^{S,f}, K_{M,q}^{S,f}, K_{i,q}^{S,f}, J_{3,q}^{S,f}) \quad (4.11)$$

Then the root confidence region is derived in the same way as in section 3. Let  $\chi_q^{2,S} = \chi^2(k_{p,q}^{S,f}, K_{M,q}^{S,f}, K_{i,q}^{S,f}, J_{3,q}^{S,f}, D_0)$  be the discrepancy measure (4.8), calculated over the original sample with the flipped synthetic parameters. The discrepancy measures  $\chi_q^{2,S}$  are renumerated in ascending order so that  $\chi_1^{2,S} \leq \chi_2^{2,S} \leq \dots \leq \chi_Q^{2,S}$ . The inside and outside simulated points are identified as in section 3. Again, the area of the inside simulated points determines the root confidence region. The best point estimate of the parameters is calculated as trimmed geometrical mean of  $\bar{a}_q^{S,f}$  using just the inside simulated points. Since it is impossible to plot the 4D confidence regions, they are visualized in pairs (e.g. Fig. 4.1 shows the confidence region of  $K_M$  and  $k_p$ , and of  $K_i$  and  $J_3$ ). The kinetic parameters of



plasmin in the presence of the three fatty acids at different concentrations are given in Table 4.1. It also gives the 2D confidence intervals (2D) and best estimates, whereas the multi-dimensional confidence regions are given on Fig. 4.2.

Concentration of added fatty acid ( $\mu M$ )	$K_M$ ( $\mu M$ )		$k_p$ (sec <sup>-1</sup> )		$K_i$ ( $\mu M^{-1}$ )		$J_3$ (sec <sup>-1</sup> )	
	BE	CI	BE	CI	BE	CI	BE	CI
None	5.89	5.43-6.38	5.81	5.70-5.93	NA	NA	NA	NA
Oleate								
10	12.58	11.68-13.41	4.54	4.48-4.60	NA	NA	NA	NA
25	20.09	18.76-21.26	3.65	3.56-3.73	NA	NA	NA	NA
45	27.49	26.43-28.57	2.63	2.58-2.68	NA	NA	NA	NA
65	131.09	115.33-146.13	2.75	2.57-2.95	NA	NA	NA	NA
Arachidonate								
10	23.71	22.91-24.58	6.10	6.06-6.15	NA	NA	NA	NA
25	42.65	38.28-47.66	3.57	3.43-3.71	NA	NA	NA	NA
45	57.51	55.24-59.60	3.68	3.62-3.73	NA	NA	NA	NA
65	59.85	56.76-62.73	2.40	2.35-2.44	NA	NA	NA	NA
Stearate								
65	8.17	7.35-9.10	7.03	6.93-7.12	0.025	0.011-0.053	0.026	0.018-0.038
115	11.33	9.42-13.41	7.48	7.37-7.61	0.012	0.008-0.402	0.056	0.001-0.158
175	23.37	20.98-26.88	12.39	11.92-12.73	0.007	0.005-0.009	0.062	0.052-0.069
230	72.96	69.86-75.86	14.77	13.94-23.85	0.002	0.001-0.003	0.074	0.001-0.356

Table 4.1. Kinetic parameters of plasmin in the presence of fatty acids. Numerical values of the best estimates (BE) and their 95% confidence intervals (CI).

The described study implemented a novel numerical procedure based on Monte Carlo to give a quantitative characterization of the modulation of plasmin activity by the presence of three fatty acids. All three fatty acids caused a 10–20-fold increase in the Michaelis constant of plasmin. Based on the ratio of the catalytic and Michaelis constants, all three fatty acids acted as inhibitors of plasmin with various degrees of potency - oleate and arachidonate can be defined as mixed-type inhibitors of plasmin, whereas stearate has a rather unusual effect; the increase in the Michaelis constant is coupled to higher values of the catalytic constant. At saturating concentrations of the substrate this effect is seen as apparent activation of plasmin in the amidolytic assay. Our findings illustrate the general possibility for a modulator to change the kinetic parameters of an enzyme in an independent and controversial manner, so that the overall catalytic outcome may vary with the concentration of the substrate. In physiological context the reported results indicate that acting as mixed-type inhibitors, unsaturated fatty acids stabilize fibrin against plasmin, whereas through its discordant effects on the catalytic and Michaelis constants stearate may destabilize clots in the process of their formation.

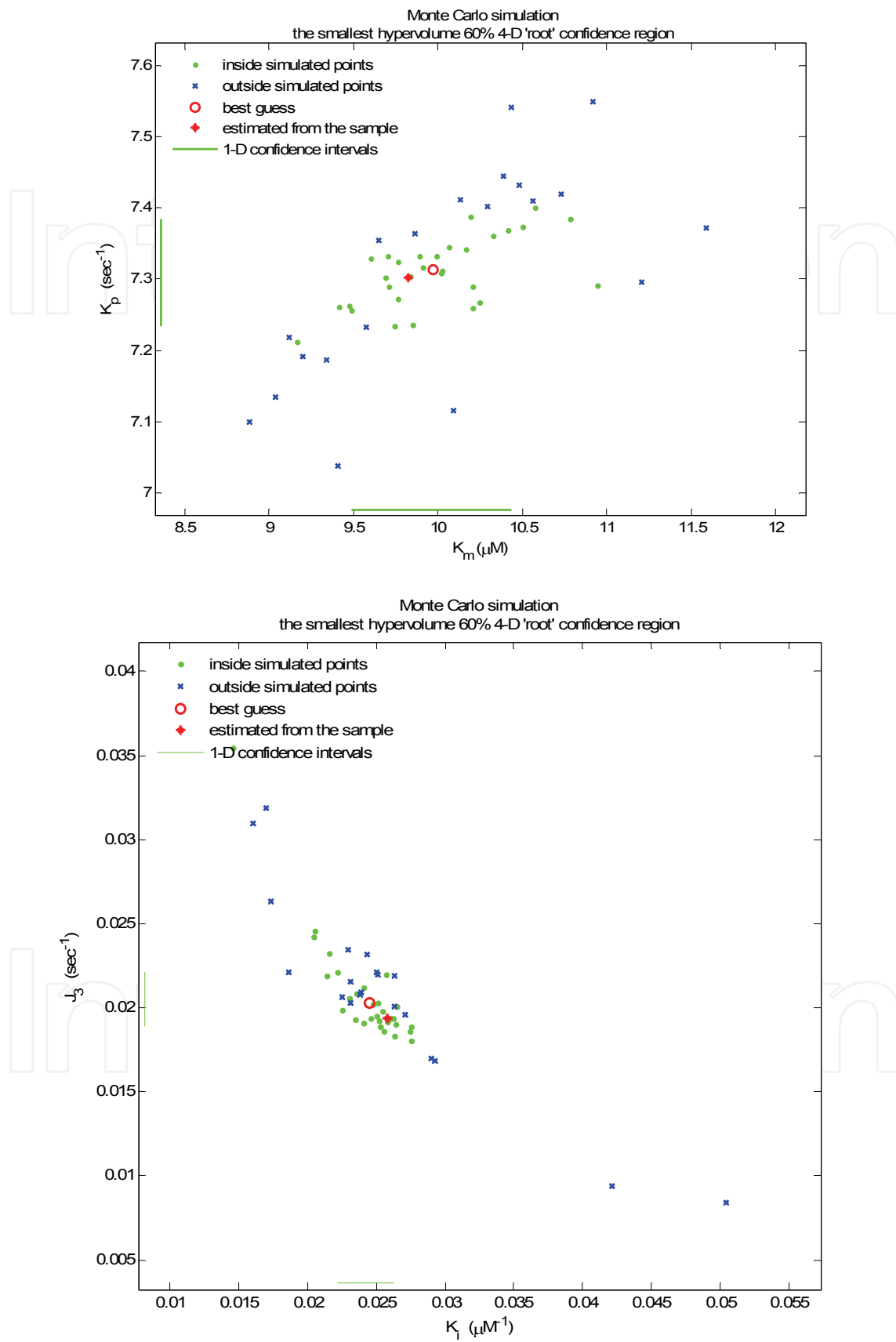


Fig. 4.1. Confidence regions of  $K_M$  and  $k_p$  (in the first plot) and of  $K_i$  and  $J_3$  (in the second plot)

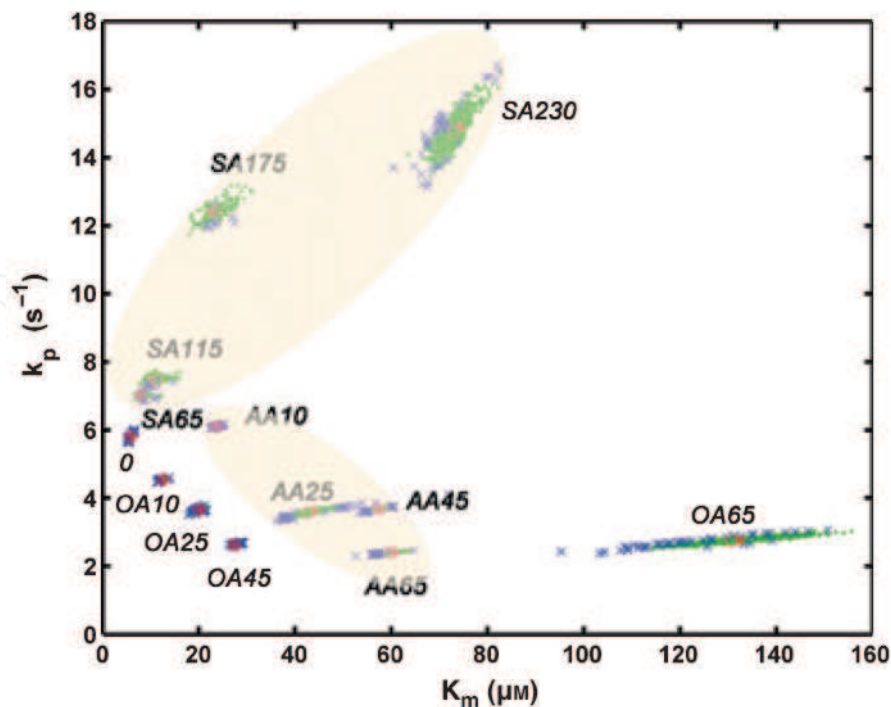


Fig. 4.2. Kinetic parameters of plasmin in the presence of fatty acids (oleate OA, arachidonate AA, and stearate SA), and their multi-dimensional confidence regions

The investigated biochemical problem was a suitable setup to demonstrate the multiplicative Monte Carlo, which provides a reasonable physical meaning of resulting parameters. Its advantages over the classical methods in tasks, where strictly positive parameter values are required, makes it a necessary tool in most biological and biochemical parameter identification studies.

## 5. Conclusions

The application of Monte Carlo and Bootstrap techniques in a multi-dimensional setup is principally similar to the one-dimensional case. However, instead of CI, we need to find confidence regions that contain certain percentage of the entire probability distribution. As demonstrated, only root and percentile confidence regions are of interest in the multi-dimensional case. The use of either of them stems from the two different approaches to probabilities, so researchers should first clarify their viewpoint to probabilities in general before choosing the type of confidence regions to exploit. Working with root confidence intervals (regions) requires parameter flipping, which in some cases may generate results with an inappropriate sign. This is particularly true for biochemical (biological) processes, where parameters are supposed to be positive most of the cases. Therefore we presented the multiplicative Monte Carlo as a procedure that ensures the physical meaning of parameters.

Since Monte Carlo and Bootstrap are computer intensive methods, all calculation and visualization procedures, discussed and demonstrated here were executed with MATLAB R2009a. Whenever we used root confidence intervals, we also provided a best point estimate, which as a rule is better than the sample estimate. Generally, the mean of the flipped parameters can serve for a best estimate, but having in mind the sensitivity to errors

(extreme values, outliers in the sample, etc.) it is recommended to use trimmed means. Regardless of the way to calculate this best estimate though, it will always be unbiased unlike the sample estimate.

The recent expansion of technological tools in biomedical research poses the requirement for appropriate modeling of the processes under investigation. The described examples from our work underscore the applicability of Monte Carlo simulations in biochemical models of variable complexity providing a robust tool to estimate the reliability of estimated parameters.

## 6. Acknowledgements

This work was supported by grants from the Wellcome Trust [083174] and OTKA [K83023].

## 7. References

- Beran, R. (1986). Discussion to Wu, C.F.J.: Jackknife, Bootstrap and other resampling methods in regression analysis. *Ann. Statist.*, Vol. 14, 1295-1298
- Berger, J. & Wolpert, R. L. (1988). *The Likelihood Principle: A Review, Generalizations, and Statistical Implications*, Second Edition, IMS, Hayward, CA
- Berry, D. A. (1996). *Statistics: A Bayesian Perspective*, Duxbury Press, CA, USA
- Cornish-Bowden, A. (2004). *Fundamentals of Enzyme Kinetics*, Third Edition, Portland Press, London
- Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*, Cambridge University Press, Cambridge, NY, USA
- Davidson, R. & MacKinnon J. G. (1999). The size distortion of bootstrap tests. *Econometric Theory*, Vol. 15, 361-376
- DiCiccio, T. & Romano, J. (1988). A review of bootstrap confidence intervals (with discussion). *J. Roy. Statist. Soc., Ser. B*, Vol. 50, 338-370
- Duggleby, R.G. (1995). Analysis of enzyme progress curves by nonlinear regression, *Methods Enzymol*, Vol. 249, 61-90
- Duggleby, R.G. (2001). Quantitative analysis of the time courses of enzyme-catalyzed reactions, *Methods*, 24(2), 168-174.
- Efron, B. & Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, NY, USA
- Hanke, J. E. & Reitsch A. G. (1991). *Understanding Business Statistics*. Irwin, USA
- Higazi A.A.R.; Finci-Yeheskel, Z.; Samara, A.A.R.. Aziza, R. & Mayer, M. (1992). Stimulation of plasmin activity by oleic acid, *Biochem J*, Vol. 282, 863-866.
- Higazi, A.A.R.; Aziza, R.; Samara, A.A.R. & Mayer, M. (1994). Regulation of fibrinolysis by non-esterified fatty acids, *Biochem J*, Vol. 330, 251-255.
- Hertz, D. & Thomas, H. (1983). *Risk Analysis and its Applications*. John Wiley, NY, USA
- Huet, E.; Cauchard, J.H.; Berton, A.; Robinet, A.; Decarme, M.; Hornebeck, W. & Bellon, G. (2004). Inhibition of plasmin-mediated prostromelysin-1 activation by interaction of long chain unsaturated fatty acids with kringle 5, *Biochem Pharmacol*, Vol. 67, 643-654.
- Jeffrey, R. (2004). *Subjective Probability - The Real Thing*. Cambridge University Press, Cambridge, UK

- Kolev, K.; Longstaff, C. & Machovich, R. (2005). Fibrinolysis at the fluid-solid interface of thrombi, *Curr Medic Chem Cardiovasc Hematol Agents*, Vol. 3, 341-355
- Lagarias, J.; Reeds, J.; Wright, M. & Wright, P. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions, *SIAM J Optimiz*, Vol. 9, 112-147
- Lineweaver, H. & Burk, D. (1934). The determination of enzyme dissociation constants, *J A Chem Soc*, Vol. 56: 658-666
- MacKinnon, J. G. & Smith, A.A. Jr. (1998). Approximate bias correction in econometrics, *Journal of Econometrics*, Vol. 85 205-230.
- Newsholme, E.A. & Leech A.R. (1984). *Biochemistry for the medical sciences*. John Wiley & Sons, Inc., Chichester.
- Politis, D.N. (1998). Computer-intensive methods in statistical analysis, *IEEE Signal Proc Mag*, Vol. 15, 39-55
- Press, W. H.; Teukolski, S. A.; Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes – The Art of Scientific Computing*. Cambridge University Press, NY, USA.
- Rábai, G.; Váradi, B.; Longstaff, C.; Sótonyi, P.; Kristóf, V.; Timár, F.; Machovich, R. & Kolev, K. (2007). Fibrinolysis in a lipid environment: modulation through release of free fatty acids, *J Thromb Haemost*, Vol. 5, 1265-1273
- Segel, L.A. (1988). On the validity of the steady state assumption of enzyme kinetics, *Bull. Math. Biol.*, Vol. 50, 579-593
- Segel, I.H. (1993). *Enzyme kinetics*. John Wiley & Sons, Inc., NY, USA, pp. 25-29
- Shampine, L.F.; Gladwell, I. & Thompson, S. (2003). *Solving ODEs with Matlab*. Cambridge University Press, Cambridge.
- Strohman, R. (2002). Maneuvering in the complex path from genotype to phenotype, *Science*, Vol. 296, 701-703
- Tanka-Salamon, A.; Tenekedjiev, K.; Machovich, R. & Kolev, K. (2008). Suppressed catalytic efficiency of plasmin in the presence of long-chain fatty acids: identification of kinetic parameters from continuous enzymatic assay with monte carlo simulation, *FEBS Journal*, Vol. 275 1274-1282
- Tenekedjiev, K. & Kolev, K. (2002). Introduction to interpretation of stochastic parameters: computer-intensive procedures for evaluation of data in enzyme kinetics, *Biochemistry and Molecular Biology Education*, Vol. 30, 414-418
- Váradi, B.; Kolev, K.; Tenekedjiev, K.; Mészáros, G.; Kovalszky, I.; Longstaff, C. & Machovich, R. (2004). Phospholipid-barrier to fibrinolysis: role for the anionic polar head charge and the gel-phase crystalline structure, *J Biol Chem*, Vol. 279, 39863-39871
- Voet, D. & Voet, J.G. (1995). *Biochemistry*. Second Edition, John Wiley & Sons, Inc., NY, USA.
- Von Mises, R. (1946). *Notes on the Mathematical Theory of Probability and Statistics*. Harvard University Press, USA.



## **Applications of Monte Carlo Methods in Biology, Medicine and Other Fields of Science**

Edited by Prof. Charles J. Mode

ISBN 978-953-307-427-6

Hard cover, 424 pages

**Publisher** InTech

**Published online** 28, February, 2011

**Published in print edition** February, 2011

This volume is an eclectic mix of applications of Monte Carlo methods in many fields of research should not be surprising, because of the ubiquitous use of these methods in many fields of human endeavor. In an attempt to focus attention on a manageable set of applications, the main thrust of this book is to emphasize applications of Monte Carlo simulation methods in biology and medicine.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Kiril Ivanov Tenekedjiev, Natalia Danailova Nikolova and Krasimir Kolev (2011). Applications of Monte Carlo Simulation in Modelling of Biochemical Processes, Applications of Monte Carlo Methods in Biology, Medicine and Other Fields of Science, Prof. Charles J. Mode (Ed.), ISBN: 978-953-307-427-6, InTech, Available from: <http://www.intechopen.com/books/applications-of-monte-carlo-methods-in-biology-medicine-and-other-fields-of-science/applications-of-monte-carlo-simulation-in-modelling-of-biochemical-processes>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen