Reliable Biological Circuit Design Including Uncertain Kinetic Parameters

Eva Sciacca and Salvatore Spinella

Abstract. In the context of possibilistic decision making, this work deals with biological design problems particularly important in the near future when it will be possible to produce biological entities and synthetic organisms for pharmacological and medical usage. The biological systems is investigated in terms of performances or main key features of the system. The analysis of the biological system is based on the idea that the set of parameters involved in the model can be classified into two different typologies: the uncertain kinetic parameters and the control design parameters. In order to design a robust and reliable biological system with respect to a target performance, the design parameter values are set up to balance the uncertainty of the kinetic parameters. To take into account these uncertainties arising from the estimations of the kinetic parameters, the function representing the feedback of the system is fuzzified and a measure of failure of the designed biological circuit is minimized to reach the required performance. An application of this methodology is illustrated on a case study of an autonomously oscillatory system: the Drosophila Period Protein which is a central component of the Drosophila circadian clocks. Finally, the results of the fuzzy methodology are compared with a deterministic method.

1 Introduction

In the context of possibilistic decision making, this work deals with biological design problems particularly important in the near future when it will be possible to produce biological entities and synthetic organisms for pharmacological and medical usage. Generally, the design frameworks that support Research & Development

Eva Sciacca · Salvatore Spinella
Department of Mathematics and Computer Science, University of Catania, Viale A. Doria 6, I 95125 Catania, Italy
e-mail: sciacca@dmi.unict.it, spins@unical.it

W.A. Lodwick & J. Kacprzyk (Eds.): Fuzzy Optimization, STUDFUZZ 254, pp. 457–477.
springerlink.com

technologies must face with uncertainty modelling. In this field the uncertainty arise from different sources related to:

- behavioral models, that connect uncertain model parameters to observed evolution of system state;
- equivalent models, that connect uncertain model parameters to system feedbacks;
- approximated models, that approximate various aspects of a system in a computational tractable manner.

One possible way to model biological networks is to employ Ordinary Differential Equations (ODE) systems. These models involve several parameters such as reaction rate constants or protein concentrations. A precise measurement of these parameters is difficult to experimentally estimate and their values are quite uncertain because they vary from one cell to another or throughout the lifetime of any individual.

The aim of this work is to set up a methodology to perform a reliable target satisfaction in the framework of biological systems design that include uncertain parameters. Biological pathways modeled by mass action reactions include uncertain parameters that are usually related to the equilibrium of the system; feedback loops between tumor suppressors and oncogene agents are some examples of these kinds of systems.

This chapter includes the following sections. A general survey of optimization in biological design is outlined in Sect. 2. In Sect. 3 the design optimization problem under uncertainty is formulated. The biological system was investigated in terms of performances or main key features of the model. The analysis of the biological system is based on the idea that the set of parameters involved in the model can be classified into different typologies, two different classes of parameters were identified: the uncertain kinetic parameters and the control design parameters. In order to design a robust and reliable biological system with respect to a target performance, the design parameter values were set up to balance the uncertainty of the kinetic parameters. To take into account these uncertainties arising from the estimations of the kinetic parameters, the function representing the feedback of the system was fuzzified as described in Sect. 4. One way to deal with design specifications is to compare the fuzzified performance with a crisp number representing a reasonable threshold and give a measure of satisfaction of these constraints. For this purpose the possibility measures of failure with respect to the specification constraints can give a useful information to improve the design. The measures of failure are then minimized using an appropriate optimization algorithm to reach the required biological design. Sect. 5 gives a description of the main concepts of the possibility theory. A summary of the whole methodology and a clear diagram of the numerical algorithms and components involved in the design simulation are illustrated in Sect. 6. Among all the possible biological circuit designs, autonomously oscillating systems are the most investigated because they provide invaluable resources for controlling and orchestrating the main biological functions. These kinds of systems are introduced in Sect. 7 where also a summary of the typical biochemical rate functions involved in biological mathematical modelling is given. The particular case

study of the Drosophila Period Protein (PER) and the involved model parameters are described in Sect. 8. Some results of the application of the methodology and the comparison with a deterministic method on the case study are illustrated in Sect. 9. Finally, the conclusions on this study and some perspectives for future research are outlined in Sect. 10.

2 Biological Design Optimization

The concept of optimization is certainly not new in biology [Banga, 2008]. Model-based optimization is a key methodology in engineering, helping in the design, analysis, construction and operation of all kind of devices. Optimization methods have been applied in both metabolic control analysis [Heinrich and Schuster, 1998] and biochemical systems theory [Torres and Voit, 2002]. Further, optimization (in particular, linear programming) has been the engine behind metabolic flux balance analysis, where the optimal flux distributions are calculated using linear optimization, and are used to represent the metabolic phenotype for certain conditions. This flux balance methodology provides a guide to metabolic engineering and a method for bio-process optimization [Banga, 2008].

Coupling constraint-based analysis with optimization has been used to generate a consistent framework for the generation of hypotheses and the testing of functions of microbial cells using genome-scale models. Extensions and modifications of flux balance analysis continue to use optimization methods extensively [Segre *et al*., 2002]. Constrained evolutionary optimization has been used to understand optimal circuit design. Moreover, optimization principles have also been used to explain the complexity and robustness found in biochemical networks [Stelling *et al*., 2004]. Reverse engineering in systems biology aims to reconstruct the biochemical interactions from data sets of a particular biological system. Optimization has been used for inferring important biomolecular networks, such as transcriptional regulatory networks [Wang *et al*., 2007], gene regulatory networks, signaling pathways and protein interaction networks.

System identification is a methodology widely used in engineering for building mathematical models of dynamical systems based on measured data. This methodology involves the selection of the model structure and the parameter estimation for the model from the available experimental data. The parameter estimation problem in biochemical pathways is formulated as a nonlinear programming problem subject to the pathway model acting as constraints. Since these problems are frequently multi-modal, global optimization methods are needed in order to avoid local solutions. A local solution can be very misleading when calibrating models: it would indicate a bad fit even for a model which could potentially match perfectly a set of experimental data [Banga, 2008].

In general, for biological design optimization, it is assumed that the mathematical model and the involved species are already chosen by the biologist and are fixed. Let r be the number of specifications of the biological system to be optimized, the

desired response $R^* \in \mathbb{R}^r$ is expressed in terms of design specifications or design goals. The problem of biological model design then can be formulated as:

$$x^* = \arg\min_{x \in X} U(R(x)) \tag{1}$$

where x^* is the optimal design, X is the feasible region, U is a suitable objective function, and hopefully $R(x^*) = R^*$. In general, the above problem corresponds to a constrained nonlinear programming problem. The objective function U is typically a combination of multiple objectives with conflicting criteria.

Generally, "classical methods to solve equation 1 include Line Search and Trust Region strategies, based on methods such as Conjugate Gradient, Newton and Quasi-Newton methods. Usually methods that use only function evaluations are more suitable for problems that are very nonlinear or have many discontinuities (Search Methods) while methods that use derivative information are more effective when the function is continuous in the first and second derivatives (Gradient Methods).

For nominal design it is assumed that the design parameters are not subject to statistical fluctuations. When uncertain parameters are considered, one of the most common deterministic design methodology, used also in microelectronic industry, is named "Nominal Over-Design". The Nominal Over-Design fixes every objective to a secure value with regard to the nominal target specifications. The design specifications are increased of a certain percentage in the case of minimum thresholds, while in cases of maximum thresholds they are decreased. The main drawback of this scheme is often to point deterministically at unfeasible over-designs which could have the opposite effect blocking the optimization process at initial stages.

The methodology introduced in this study allows to combine optimization and uncertainty analysis for target satisfaction design problem. The treatment of uncertainty is an unavoidable step because the lack of precision in the model and among its parameter values can invalidate the results. The target satisfaction under uncertainty gives a robust framework to use optimization in a design context.

3 Biological Design Optimization under Uncertainty

In this work a methodology was formulated to perform a target satisfaction in the framework of biological systems design that incorporates reliable feedback. The proposed methodology is based on the scheme described in [Sciacca *et al.*, 2007] which was applied to a sizing problem of an electronic circuit. The biological system was investigated in terms of performances or main key features of the system. For example, in autonomously oscillating biochemical systems, the key features of oscillatory trajectories of the species concentrations can be taken into consideration. Once the performances of the system are chosen, the system can be designed in a reliable manner with respect to the uncertainties related to the estimation of the parameters. Our methodology combines an analysis of the system parameters and their

relationships with the uncertainties arising from the model describing the biological system.

The analysis of the biological system is based on the idea that the set of parameters involved in the model can be classified into different typologies. Referring for instance to the case study of the circadian oscillations of the PER gene product in Drosophila (introduced in Sect. 8) two different classes of parameters were identified:

- The uncertain kinetic parameters K_U, that are known in terms of confidence intervals. Those parameters are for example the Michaelis constants or the constant rates for the kinases and the phosphatases.
- The control design parameters K_{CD}, that can be defined in a reliable way and determine the behavior of the biological system. Those are for example the maximum rate of degradation of an enzyme or the first order transportation rate parameters.

Since this biological design problem includes elements of the input data in a real-valued confidence interval, we deal with an optimization problem under uncertainty [Lodwick and Jamison, 2007]. In particular, the following programming constraint satisfaction problem is considered:

$$g_i(K_{CD}, K_U) \leq t_i \quad i = 1, \ldots, n \tag{2}$$

The constraint set is denoted as $\Omega = \{K_{CD} | g_i(K_{CD}, K_U) \leq t_i \;\; i = 1, \ldots, n, K_{CD} \in X\}$. The values of K_U are input parameters of the programming problem and are subject to uncertainty arising from different sources. Depending on the nature of the uncertainty, they may be probability distributions, intervals, fuzzy sets, or possibilistic distributions. In our case, these parameters are intervals which are particular cases of fuzzy numbers. The values t_i are the maximum or minimum thresholds for the constraints and can also be considered as uncertain.

In order to design a robust and reliable biological system with respect to a target performance, the design parameters were set up to balance the uncertainty of the kinetic parameters. To take into account these uncertainties arising from the estimations of the kinetic parameters, the function representing the feedback of the system was fuzzified. Finally, a measure of failure of the designed biological circuit to reach the required performance was analyzed by means of the possibility theory. The possibility measure is a consistent alternative to the statistical and probabilistic hypothesis. In fact, especially in fields such as the biological one, statistic and probabilistic assumptions are difficult to justify and moreover they could not include all the possible phenomena involved in biological processes. For such problems, little information regarding the uncertainty is known, and the uncertainty is typically modelled depending on expert opinions and assumptions made by the biologist. Fuzzy set theory is able to compensate the fact that uncertainty is modelled based on subjective opinions and assumptions. In contrast, probabilistic methods require large amount of data and the results obtained are, in same cases, sensitive to both the accuracy of the data as well as the assumptions made during the design process.

The interpretation of the possibility failure stands for a measure for the worst case design in the uncertainty context of the system. An optimization methodology adopted to minimize the possibility failure leads to a reliable configuration which guarantee the desired design.

4 Fuzzification of the Objective Function Representing the Performance

Fuzzy sets have been introduced by [Zadeh, 1965] as an extension of the classical notion of set. In classical set theory, the membership of elements in a set is assessed in binary terms according to a bivalent condition in which an element either belongs or does not belong to the set. By contrast, fuzzy set theory permits the gradual assessment of the membership of elements in a set; this is described with the aid of a membership function valued in the real unit interval $[0,1]$.

A fuzzy set is a pair (F,m) where F is a set and $m : F \to [0,1]$. For each $x \in F$, $m(x)$ is the grade of membership of x. An element mapping to the value 0 means that the member is not included in the fuzzy set while the mapping value 1 describes a fully included member in the fuzzy set. Values strictly between 0 and 1 characterize the fuzzy members. The set $\{x \in F \mid m(x) > 0\}$ is called the support of the fuzzy set (F,m) and the set $\{x \in A \mid m(x) = 1\}$ is called the Core [Klir and Yuan, 1995] of the fuzzy set (F,m).

A fuzzy number is a convex, normalized fuzzy set $\tilde{F} \subseteq \mathbb{R}$ whose membership function is at least segmentally continuous and has the functional value $\mu_F(x) = 1$ at precisely one element [Klir and Yuan, 1995].

In order to model with fuzzy numbers [Zadeh, 1968] the uncertainty arising from simulation design, a response surface of the function representing the feedback of the system was used as suitable approximation. The response surface was fitted with respect to the uncertain parameters sampled using a Latin Hypercube methodology.

Designing for uncertainty is computationally intensive and typically requires at least an order of magnitude more computational cost as compared to a corresponding deterministic design. Response surface approximations reduce the high computational cost associated with designing for uncertainty by using approximations that are accurate over the entire design space [Venter and Haftka, 1999] to replace costly stiff ODE system integrations. In the present work, the scheme to generate the response surface [Gavin and Yau, 2008] approximates the function with an approximate polynomial response surface of arbitrary order:

$$\tilde{g}(X) = a + \sum_{i=1}^{n} \sum_{j=1}^{h_i} (b_{ij} X_i^j) + \sum_{q=1}^{m} c_q \prod_{i=1}^{n} X_i^{p_{iq}} \tag{3}$$

where the coefficients b_{ij} correspond to terms involving only one random variable, and the coefficients c_q correspond to mixed terms, involving the product of two or more random variables. The polynomial order, h_i, the total number of mixed terms, m, and the order of a random variable in a mixed term, p_{iq}, are determined in the

algorithm described below. The algorithm used in this methodology, makes use of the last three stages of the High Order Stochastic Response Surface Method (HO-SRSM) [Gavin and Yau, 2008]. In the first stage, the number and types of mixed terms are determined. This stage results in the formulation of the higher order polynomial to be used for the response surface. After the formulation of the higher order polynomial, the coefficients of the higher order response surface polynomial are estimated in the second stage, using singular value decomposition to perform least squares on Latin Hypercube samples in the uncertain parameters space. Finally, from a Monte Carlo Simulation, the fuzzification of the performances are carried out with an acceptable approximation of the objective function given by the response surface.

The fuzzy representation of the performance is constructed enveloping the fitted data by intervals. The fuzzy map is built by α-level considering the minimum median interval which envelopes a fraction $(1-\alpha)$ of the performance values [Spinella and Anile, 2004]. In formal terms:

Definition 1 (Median Interval I_α). *Given n samples $X_1,\ldots,X_n$, and a reorder of them $X_{i_1},\ldots,X_{i_n}$, then the median interval at level α is:*

$$I_\alpha = [X_{i_j}, X_{i_k}] \tag{4}$$

where $j = \lfloor \frac{\alpha}{2} \rfloor + 1$ and $k = n - \lfloor \frac{\alpha}{2} \rfloor$.

The pseudocode in Algorithm 1 describes the procedure to fuzzify the performances. The method *RespSurfBuild* generates the response surface using equation 3 and return the coefficients C. Then, a set of samples s of uncertain parameters K_U is generated using a Latin Hypercube technique. Finally, the response surface is evaluated on the set s and its output, the performance approximation, is fuzzified.

Algorithm 1. $Fuzzify(\mathbf{K_{CD}}, \mathbf{inf_u}, \mathbf{sup_u}, sys_{sim})$

$\mathbf{C} := RespSurfBuild(\mathbf{K_{CD}}, \mathbf{inf_u}, \mathbf{sup_u}, model_{sim})$
$\mathbf{s} := LatinHypercube(\mathbf{inf_u}, \mathbf{sup_u}, N_{samples})$ {random variables between **inf** and **sup** generated using a Latin Hypercube sampling}
$\mathbf{F_s} = RespSurfEval(\mathbf{s}, \mathbf{C})$;
$\tilde{F} := FuzzyfyFun(\mathbf{F_s})$

5 Possibility of Failure

One way to deal with design specifications is to compare the fuzzy numbers representing the performances with the crisp numbers representing a reasonable threshold and give a measure of satisfaction of these constraints. For this purpose the possibility measure of failure with respect to the specification constraint can give a useful information to improve the design . Note that a fuzzy number may also be considered as the trace of a possibility measure Π on the singletons (single elements) x of the universal set X [Zadeh, 1978].

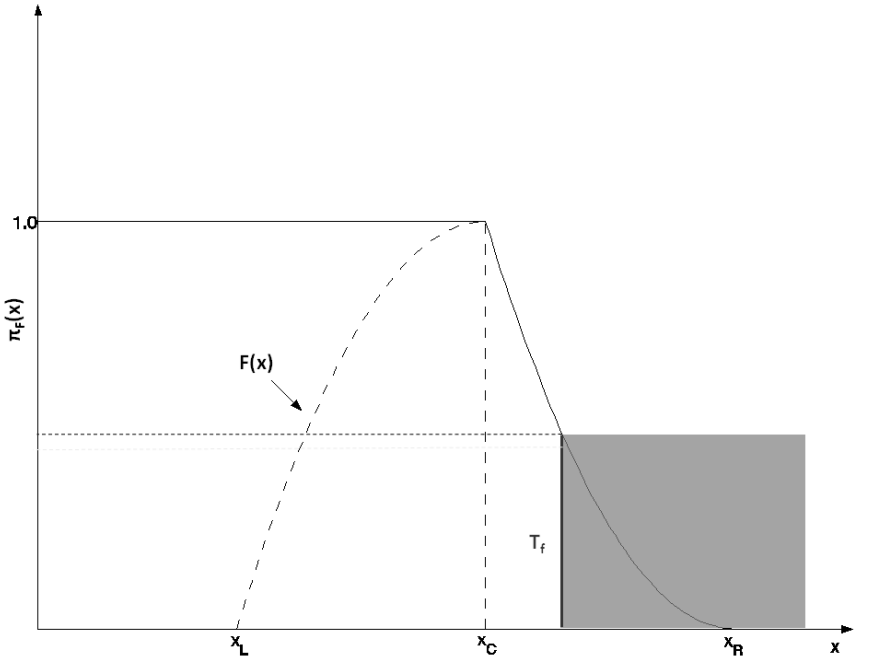

Fig. 1 Possibility distribution of $\tilde{F} \geq x$ for the nonlinear membership function $\mathbf{F}(x)$, with support in (x_L, x_R) and graphical description for the event of possibility measure of failure when $F > T_f$

When a possibility measure is considered, its possibility distribution π is then interpreted as the membership function of a fuzzy number $\tilde{F}$ describing the event that Π focuses on, as follows:

$$\Pi(\{x\}) = \pi(x) = \tilde{F}(x), \quad \forall x \in X \tag{5}$$

The possibility measure of a crisp number being smaller or equal to a fuzzy number $\tilde{F}$ is then defined [Dubois and Prade, 1988] as follows:

$$\Pi_{\tilde{F}}([x, +\infty)) = \sup_{y \geq x} \tilde{F}(y), \quad \forall x \tag{6}$$

The possibility distribution function $\Pi_{\tilde{F}}$ of the possibility measure $\pi_{\tilde{F}}$ can be seen in Fig. 1 for the general case of a nonlinear membership function.

Based on equations 5 and 6, given a representation of the function F and a maximal failure threshold T_f then the possibility measure of failure of this function is a measure for the event $F > T_f$, hence

$$\Pi_{\tilde{F}-T_f}([0, +\infty)) = \sup_{y \geq 0} (\tilde{F} - T_f)(y), \quad \forall x \tag{7}$$

Fig. 1 provides a graphical description of the event for the possibility measure of failure in the case when $F > T_f$. From each possibility measure related to the specification performances, it is possible to deduce a vector of measures $(p_1, \ldots, p_n)$ for a given design. A suitable metric that summarizes all can be used as objective for an optimization process. In this work the chosen function to optimize was:

$$\Sigma_{i \in I} \Pi_{\tilde{F}_i - T_{f_i}}([0, +\infty]) \tag{8}$$

where I is the set of performance $\tilde{F}_i$ to guarantee and T_{f_i} are the respective specification failures. This formulation of the function to optimize allows one to define the design goals through a single function that summarize all of them. Moreover, the sum of possibilities of failure avoids a conventional ideal design often based on unfeasible attainments while it identifies a more realistic design keeping the events of failure at a reasonable distance. From the mathematical point of view, the sum of possibilities is the L_1 norm inside the space of the problem objectives possibilities. This choice allows the characterization of convex regions of the multi-objective problem with a suitable merit function.

6 Methodology

Our methodology can be schematized in the diagram in Fig. 2. The two main interacting components are identified as Analyzer and Optimizer. The Analyzer

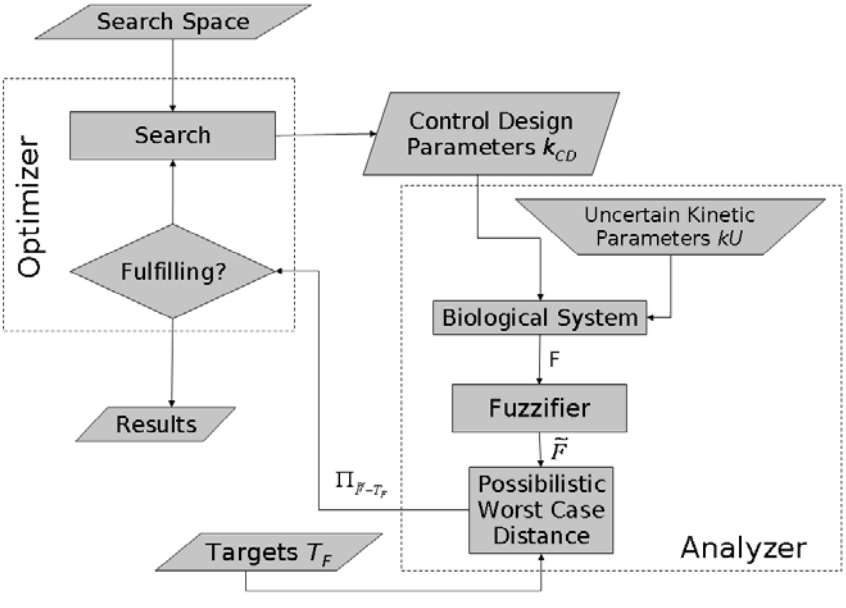

Fig. 2 Schema of the methodology

component contains the model describing the biological system, the Fuzzifier module and the Possibilistic Worst Case Distance (PWCD) module. This component can be seen as a black box taking as inputs the control design parameters k_{CD} and the chosen targets of the performances T_F of the biological system.

The Fuzzifier module samples the space of the uncertain kinetic parameters k_U through a Latin Hypercube methodology and approximates the performances F of the system using a response surface approximation. From this approximation, the fuzzification is carried out using the sampled uncertain kinetic parameters k_U while the control design parameters k_{CD} are considered as crisp numbers.

The PWCD module takes as input the fuzzified functions of the performances $\tilde{F}$ and the target of the performances T_F and gives as output the measure of failure of the biological system with respect to the given targets; this possibility measure of failure are estimated with the equation 7.

The Optimizer component searches for a set of control design parameters k_{CD} that minimizes the objective function (see equation 8) sum of all possibilities of failure for the biological system. The k_{CD} parameters will be given as new input for the Analyzer component. The optimization will run until a stop criterion is fulfilled (the sum of all possibilities of failure equals to zero), or alternatively until the unsatisfiability for the target is detected. Generally, the satisfiability of this stop criterion is not guaranteed for all the problems and a compromise between the targets should be chosen by the designer.

7 Biochemical Modeling

7.1 Oscillatory Biological Networks

Among the numerous biological networks models, autonomously oscillating systems are the most investigated. Such systems underly many of the periodic phenomena which have been identified in biology such as in processes describing glycolytic oscillations, the cell cycle, circadian rhythms, periodic neuronal signals [Goldbeter, 1996].

A first attempt at providing a mathematical model of the mechanism underlying these oscillations was presented by [Goldbeter, 1995] describing the circadian oscillations of the PER gene product in Drosophila. The PER gene was shown to play a role in circadian rhythms. This model was built in the absence of detailed molecular descriptions of the reactions involved, and was proposed as a minimal model which was able to reproduce experimental observations of wild type and mutant behavior. This model suffices for investigations of the core behavior of the system, and its simplicity recommends it for an illustrative example. Further analysis on how the kinetic parameters influence the extreme and the period of the oscillations can be found in [Ingalls, 2004, Bagheri *et al*., 2007] where a detailed sensitivity analysis was presented. In comparison with the methods described in [Ingalls, 2004, Bagheri *et al*., 2007], our methodology allows us to find a set of parameters that satisfy a fixed target without making use of an analytical approach.

Quantitative mechanism-based models could allow researchers to predict the comprehensive behavior of the specified system over time and to track its dynamics for each set of fixed system parameters [van Riel, 2006, Zi *et al*., 2005]. However, all of the parameters including rate constants and initial component concentrations in the mathematical models must be experimentally measured or inferred to specify the model. Even for those models with experimentally estimated parameters, it is still uncertain whether the particular set of parameters closely approximates the corresponding biological system because some of the kinetic parameters are usually taken or estimated from measurements reported by different laboratories using different in vitro models and conditions.

7.2 Kinetic of Biochemical Systems

A typical chemical or biochemical rate function relates the temporal change in a chemical compound to the concentration itself. In a simple first order degradation reaction which does not involve any enzyme, the change in concentration of the substrate over time is directly proportional to its concentration. This rate function can be mathematically formulated as follows:

$$\frac{dX}{dt} = -kX \tag{9}$$

where k (which is positive by definition) is the turnover per time unit and X denotes the concentration of the metabolite X. The negative sign indicates that material X is actually lost from existing pool. This mathematical form of the equation results from considerations of statistical thermodynamics.

When two substrates, say X_1 and X_2, go in to reaction to yield product X_3, the change of concentration of X_3 with respect to time can be written as:

$$\frac{dX_3}{dt} = kX_1X_2 \quad \frac{dX_1}{dt} = -kX_1X_2 \quad \frac{dX_2}{dt} = -kX_1X_2 \tag{10}$$

These formulas are written as the product of the concentrations of the species X_1 and X_2. This product form of the rate law is partly due to thermodynamics and partly due to the fact that the two metabolites have to come to the physical contact which is a matter of probability. That implies the simplest formulation as a product.

Michaelis-Menten rate equations describe the kinetics of many enzymes. This kinetic model is relevant in situations where very simple kinetics can be assumed. A substrate S and its catalyzing enzyme E form an intermediate complex (ES) in a reversible reaction. This complex can either break apart and return into substrate and enzyme or generate product P while releasing enzyme E unchanged.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \overset{k_2}{\rightarrow} E + P \tag{11}$$

In order to obtain the Michaelis-Menten rate equations, some assumptions about the system should be made. The first assumption states that the total enzyme

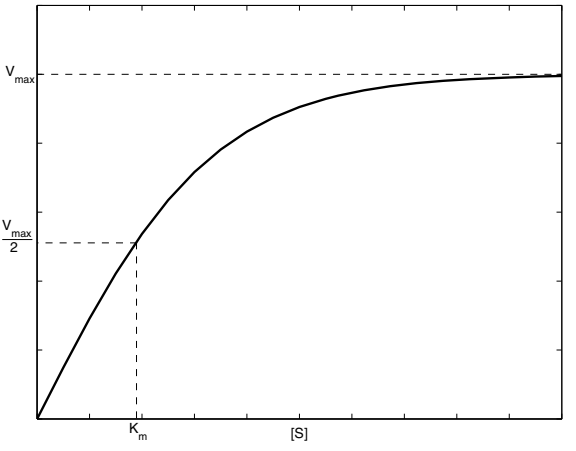

Fig. 3 Michaelis-Menten plot relating the reaction rate $\frac{dP}{dt}$ to the substrate concentration [S]

concentration can be divided into free enzyme and enzyme bound in the intermediate complex. The second assumption indicates that the total substrate concentration is much larger than the total enzyme concentration. The last assumption, which is called quasi-steady-state assumption, states that no enzyme is formed or lost during the reaction and the concentration of the intermediate complex is constant. Then the Michaelis-Menten rate is defined with the following equation:

$$\frac{dP}{dt} = \frac{V_{max}S}{K_m + S} \tag{12}$$

where the value of Michaelis-Menten rate constant is $K_m = \frac{k_{-1}+k_2}{k_1}$ and the maximum velocity $V_{max} = k_2 E_T$ occurs when the enzyme is saturated, i.e., when all enzyme molecules are tied up with S, or $(ES) = E_T$. Fig. 3 displays the variation of the product with respect to the concentration of the substrate.

The binding of a ligand to a macromolecule is often enhanced if there are already other ligands present on the same macromolecule. The Hill rate equation, provides a way to quantify this effect. It describes the fraction of the macromolecule saturated by ligand as a function of the ligand concentration; it is used in determining the degree of cooperativity of the ligand binding to the enzyme or receptor.

$$\theta = \frac{L^n}{K_d + L^n} = \frac{L^n}{K_A^n + L^n}, \tag{13}$$

where θ is the fraction of ligand binding sites filled, L denotes ligand concentration, K_d is the apparent dissociation constant derived from the law of mass action (equilibrium constant for dissociation), K_A is the ligand concentration producing half occupation (ligand concentration occupying half of the binding sites), that is also the microscopic dissociation constant and finally n is called the Hill coefficient and it describes the cooperativity. A coefficient $n = 1$ indicates completely independent binding, regardless of how many additional ligands are already bound.

Numbers greater than one indicate positive cooperativity, while numbers less than one indicate negative cooperativity.

8 Drosophila Circadian Rhythm Case Study

8.1 The Period Protein Model

Drosophila melanogaster is a two-winged fly and it is also one of the most commonly used model organisms in biology, especially in genetics and physiology. Some reasons for the choice of Drosophila as the most studied organism in biological research are: the simplicity of its morphology, the short generation time and the high fecundity. Period proteins are central components of the Drosophila circadian clock. Circadian clock generates circadian rhythms that are 24-hour activity cycles exhibited by the organisms during their life time. The model structure introduced in [Goldbeter, 1995] for the Period Protein (PER) of Drosophila is depicted in Fig. 4.

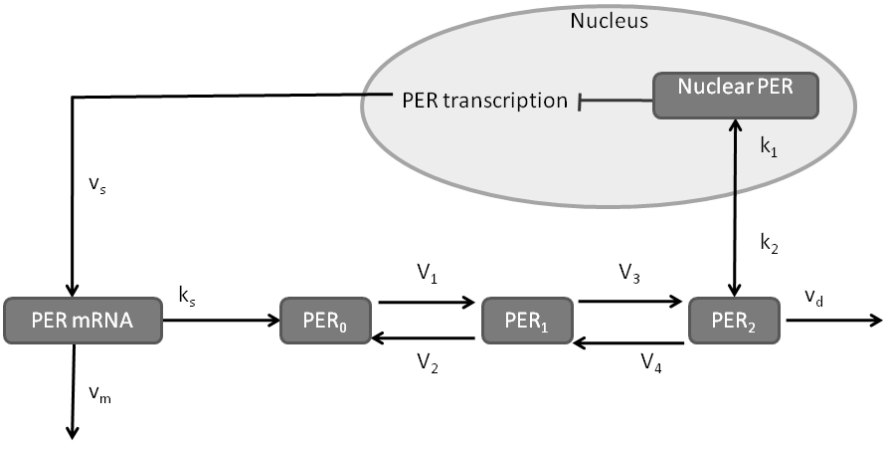

Fig. 4 Model of the period protein (PER) of Drosophila

The mechanism can be described as follows: the protein PER (P0) is produced in the cytosol at a rate determined by the concentration of PER mRNA (M). It is then reversibly phosphorylated at two sites (producing species P1 and P2). The fully phosphorylated protein can then be degraded or can migrate across the nuclear membrane. The variable PN describes the concentration of nuclear PER, which inhibits transcription of PER mRNA. This mRNA is subsequently degraded.

The quantities involved in this model are the following:

M is the PER mRNA,
P_0 unphosphorylated PER form,

P_1 monophosphorylated PER form,
P_2 bisphosphorylated PER form,
P_N is the nuclear bisphosphorylated form of PER,
v_s is the rate of mRNA production,
v_m is the rate of mRNA degradation,
n is the Hill constant which describes the cooperativity of ligand binding,
K_I is mRNA repression constant,
K_m is the Michaelis constant for the mRNA, degradation,
k_s is the rate of PER production,
V_i is the constant for the kinase,
K_i is the constant for the phosphatase,
v_d is the rate of degradation of the bisphosphorylated PER form,
k_1 is the transportation rate of the bisphosphorylated PER form in the nucleus,
k_2 is the transportation rate of bisphosphorylated nuclear PER form in the cytosol,
K_d is the Michaelis constant for the degradation of bisphosphorylated PER form

The time evolution of the specie concentrations is governed by the following kinetic equations:

$$\begin{aligned}
\frac{dM}{dt} &= v_s \frac{K_I^n}{K_I^n + P_N^n} - v_m \frac{M}{K_m + M} \\
\frac{dP_0}{dt} &= k_s M - V_1 \frac{P_0}{K_1 + P_0} + V_2 \frac{P_1}{K_2 + P_1} \\
\frac{dP_1}{dt} &= V_1 \frac{P_0}{K_1 + P_0} - V_2 \frac{P_1}{K_2 + P_1} - V_3 \frac{P_1}{K_3 + P_1} + V_4 \frac{P_2}{K_4 + P_2} \\
\frac{dP_2}{dt} &= V_3 \frac{P_1}{K_3 + P_1} - V_4 \frac{P_2}{K_4 + P_2} - k_1 P_2 + k_2 P_N - v_d \frac{P_2}{K_d + P_2} \\
\frac{dP_N}{dt} &= k_1 P_2 - k_2 P_N
\end{aligned}$$

The PER mRNA is synthesized in the nucleus and transfers to the cytosol where it accumulates at a maximum rate v_s and it is degraded by an enzyme of maximum rate v_m and Michaelis constant K_m. The rate of synthesis of the PER protein is proportional to the concentration of the PER mRNA and is characterized by a first order rate constant k_s. The reversible phosphorylation of P_0 into P_1 and P_1 into P_2 involves the parameters V_i and K_i that denote the maximum rates and Michaelis constants of the kinases and the phosphatases. The bisphosphorylated PER form P_2 is degraded by an enzyme of maximum rate v_d and Michaelis constant K_d and it is transported into the nucleus at a rate characterized by the first order rate constant k_1. The transportation of the nuclear bisphosphorylated form of PER into the cytosol is characterized by the first order rate constant k_2. The negative feedback exerted by P_N on the trascription of the PER protein is described by an equation of the Hill type where K_I is the threshold constant for repression.

8.2 The Use of Optimization: Parameters and Target Performances

The nominal values of the parameters are given in [Goldbeter, 1995] and were chosen so as to yield a period of oscillations close to 24 hours. A value of n=4 was chosen because the model can produce sustained oscillation in a larger parameter domain. But in this case, it was instead chosen a value of $n = 1$ to explore unstable configurations for the system. The new calibrated system leads to the parameter values showed in Table 1 with respect to reference concentrations of the species given in [Goldbeter, 1995]. Table 1 also shows the classification of the uncertain and control parameters.

Table 1 Estimated values and classification of the parameters in the PER model

Parameter	Estimated Value	Type
v_s	3.84	Control
v_m	3.381	Control
k_s	0.383	Control
k_1	1.747	Control
k_2	1.194	Control
v_d	0.934	Control
K_I	4.599	Uncertain
K_m	0.05	Uncertain
K_d	0.184	Uncertain
K_1	1.109	Uncertain
K_2	0.541	Uncertain
K_3	1.267	Uncertain
K_4	0.781	Uncertain
V_1	2.887	Uncertain
V_2	1.239	Uncertain
V_3	4.24	Uncertain
V_4	1.825	Uncertain

In this case study, the target performances of the required design problem are:

- the period (measured in hours), and
- the amplitude (measured in μmol)

of the concentration of the total quantity of the PER protein (P_t) which is given by:

$$P_t = P_0 + P_1 + P_2 + P_N \tag{14}$$

These performances are optimized by the methodology and they are expressed in terms of possibility of failure. In this particular test case, the minimum threshold for the period of the PER protein oscillations is fixed to 24 hours while the minimum threshold of the amplitude is fixed to a rather large value in order to guarantee significant oscillations. The optimization algorithm searches for the set of control design parameters that minimizes the objective function represented by the sum of these two possibilities of failure for the PER protein model.

An example of the sustained oscillations of the concentration of P_t generated by the model are depicted in Fig. 5.

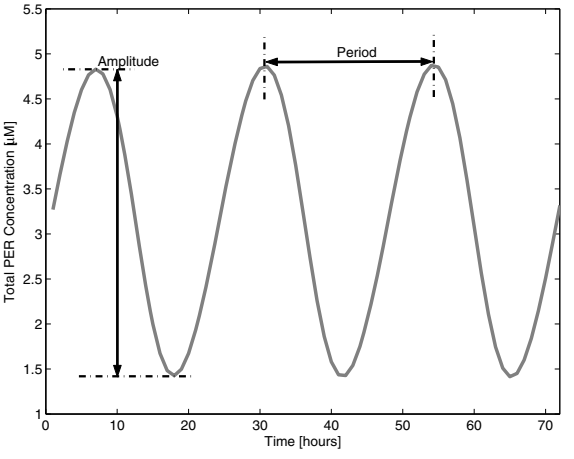

Fig. 5 Characteristics (period and amplitude) of the oscillations of the temporal variations of the total amount of PER protein

The initial conditions to solve the ODE system in this example are $\{M = 2.81, P_0 = 1.44, P_1 = 0.72, P_2 = 0.48, P_N = 0.63\}$ and the parameters are taken from Table 1. Starting from this set of initial conditions the system reaches a unique, closed curve, in the (M, P_t) plane, characterized by a period and amplitude that are fixed for the given set of parameter values (see Fig. 6).

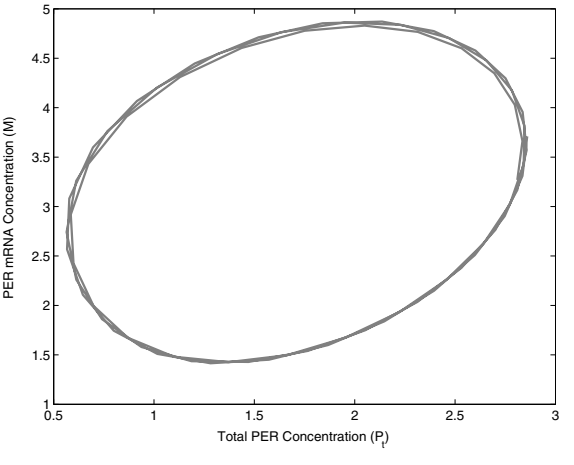

Fig. 6 The sustained oscillations of the temporal variations of the total amount of PER protein and of the per mRNA correspond to a system that reaches a unique, closed curve, in the (M, P_t) plane

9 Results

The fuzzified performances were approximated using 50 samples computed by means of a response surface of the second order. An example to explain the application of the methodology on the PER protein model can be seen in Fig. 7 and 8, where an uncertainty of a 10% in the kinetic parameters k_U was considered.

The starting point of the parameters (in Table 1) of the PER model led to a period of 23.7 hours and to an amplitude of $3.45\mu mol$ and the initial fuzzy numbers are shown in Fig. 7. The two graphs in Fig. 8 display the fuzzy numbers of the period and the amplitude of the oscillations of the PER protein concentration after the optimization process fixing the failure threshold of the period to the value of 24 hours (figure on the left) and the failure threshold of the amplitude to the value of $5\mu mol$ (figure on the right).

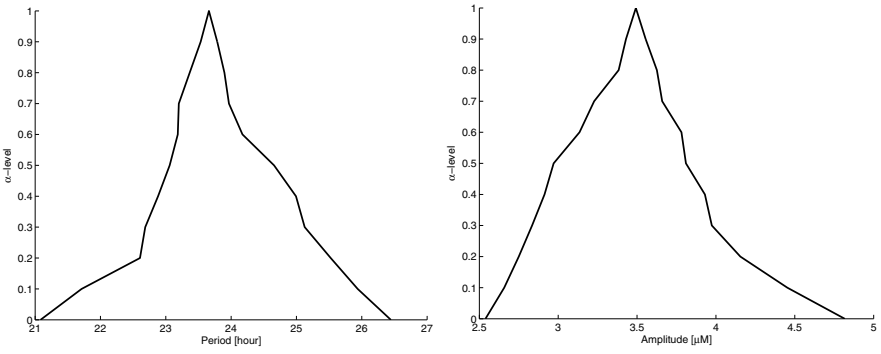

Fig. 7 Fuzzy numbers representing the period and the amplitude of the oscillations of the PER protein concentration before the optimization process

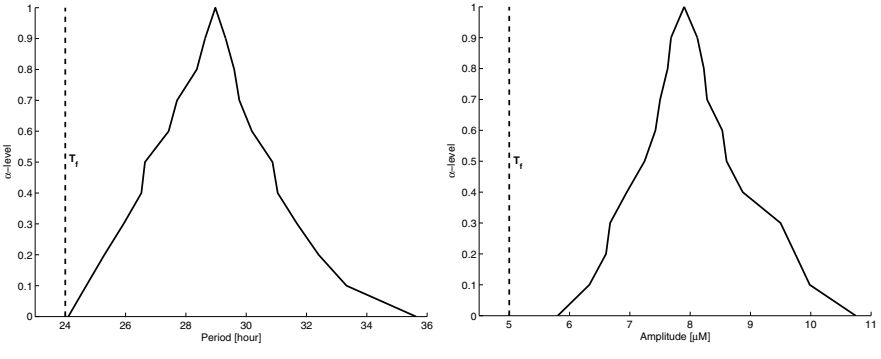

Fig. 8 Fuzzy numbers representing the period and the amplitude of the oscillations of the PER protein concentration after the optimization process

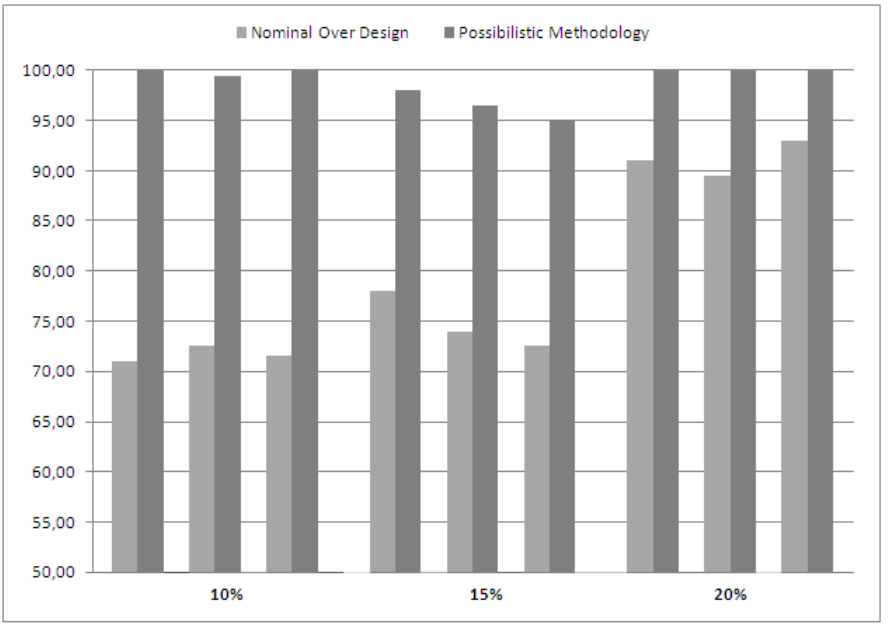

Fig. 9 Comparison between Nominal Over-Design and the Possibilistic methodology in terms of acceptable circuit varying the uncertainty of the kinetic parameters of 10%,15%,20% and 25% with regard to their nominal default value

The histogram in Fig. 9 describes in detail the comparison between the presented methodology and the "Nominal Over-Design", one of the most used deterministic design methodology. For the Nominal Over-Design, every objective was fixed to a given factor of safety with regard to the nominal target specifications. In this test case the target of the period and the amplitude were increased of a 10%, 15% and 20% in regard to the minimum threshold of 24 hours and $5\mu mol$ respectively, in order to counterbalance the uncertainty of the kinetic parameters.

Both methodologies used the Nelder and Mead Simplex [Nelder and Mead, 1965] optimization algorithm. Even thought the convergence of the algorithm is not always guaranteed, in this case the shrink steps were rare and when the Nelder-Mead iteration stagnated, a restart with the same best point and a different set of directions helped the convergence. The choice of this algorithm is due to the fact that it requires only function evaluations without derivatives estimation and it turns out to be the most appropriate in this application where the merit function is well shaped.

After the optimization process, an estimation of the percentage of "acceptable" biological circuits value was computed by means of Montecarlo simulations. The biological circuit is classified as "acceptable" if the performance specification is met i.e. if the resulting period is at least 24 hours and the amplitude of the oscillations was at least $5\mu mol$. Three independent tests were carried out considering a statistical

distribution of the uncertain kinetic parameters of 10%,15% and 20% with regard to the nominal default value. Every test performed 200 simulations.

This comparison points out that the possibilistic methodology always guarantees a better outcome of performance with respect to the Nominal Over-Design methodology. In fact, with regards to the Nominal Over-Design, it turns out to be difficult to set a suitable over-achievement which identify the best parametrization towards the uncertainty of the kinetic parameters. Note that there exist a fundamental difference between the deterministic approach and the fuzzy set based approach. The deterministic approach tends to equalize the failure load of each failure criterion, while the fuzzy set based design tends to equalize the possibility of failure of each failure criterion.

10 Conclusions

This study proposes the use of fuzzy numbers and possibility theory to perform a target satisfaction in the framework of future biological systems design that will demand reliable feedback. In order to assure a certain performance of a given biological circuit the presented methodology set the values of the design parameters of the model to balance the uncertainty of the kinetic parameters. The feedbacks of the system are represented as fuzzy numbers. By means of the possibility theory a failure value, which represents the worst case design of the system in the uncertainty context, was measured. This possibilistic failure value finally was minimized using an optimization methodology.

The proposed methodology was tested on the model describing the circadian oscillations of the PER gene product in Drosophila. The application over the oscillatory circuit of the PER gene has shown that the proposed methodology guarantees always a reliable outcome of the period and the amplitude of the oscillations of the PER gene. Moreover, a comparison of the methodology with a more diffused methodology named “Nominal Over-Design”, demonstrated an higher percentage of acceptable biological circuit in the possibilistic methodology.

Future works will apply this methodology on more then two performances of a biological system including also the phase [Bagheri *et al*., 2007] for example. In fact, usually, a biological system is characterize by multiple features that can also be in conflict with each other. The methodology could allow to design the biological system to assure the defined performance including the uncertainty of the parameters into the model. To further improve the proposed methodology other optimization algorithms will be tested in the framework, such as genetic algorithms or simulated annealing. Finally, more biological systems, even non oscillatory networks, will be tested and analyzed.

Acknowledgements. This work came from an idea given by our mentor Prof. A.M.Anile some years ago. We always keep his real paragon of scientist and his helpful encouragements in our memory.

References

[Bagheri *et al*., 2007] Bagheri, N., Stelling, J., Doyle III, F.J.: Quantitative performance metrics for robustness in circadian rhythms. Bioinformatics 23(3), 358–364 (2007)

[Banga, 2008] Banga, J.R.: Optimization in computational systems biology. BMC Systems Biology 2, 47 (2008)

[Dubois and Prade, 1988] Dubois, D., Prade, H.: Possibility theory: An approach to computerized processing of uncertainty. Plenum Press, New York (1988)

[Gavin and Yau, 2008] Gavin, H.P., Yau, S.C.: High-order limit state functions in the response surface method for structural reliability analysis. Structural Safety 2(30), 162–179 (2008)

[Goldbeter, 1995] Goldbeter, A.: A model for circadian oscillations in the Drosophila period protein (PER). Proc. R. Soc. Lond. B 261, 319–324 (1995)

[Goldbeter, 1996] Goldbeter, A.: Biochemical Oscillations and Cellular Rhythms. Cambridge University Press, Cambridge (1996)

[Heinrich and Schuster, 1998] Heinrich, R., Schuster, S.: The modelling of metabolic systems: Structure, control and optimality. Biosystems 12(47), 61–77 (1998)

[Ingalls, 2004] Ingalls, B.P.: Autonomously Oscillating Biochemical Systems: Parametric Sensitivity of Extrema and Period. Systems Biology 1(1), 62–70 (2004)

[Kholodenko *et al*., 1999] Kholodenko, B.N., Demin, O.V., Moehren, G., Hoek, J.B.: Quantification of short term signaling by the epidermal growth factor receptor. J. Biol. Chem. 274, 30169–30181 (1999)

[Klir and Yuan, 1995] Klir, G., Yuan, B.: Fuzzy sets and fuzzy logic: theory and applications. Prentice Hall, USA (1995)

[Lodwick and Jamison, 2007] Lodwick, W.A., Jamison, K.D.: Theoretical and semantic distinctions of fuzzy, possibilistic, and mixed fuzzy/possibilistic optimization. Fuzzy Sets Syst. 158(17), 1861–1872 (2007)

[Nelder and Mead, 1965] Nelder, J., Mead, R.: A simplex method for function minimization. Computer Journal 7, 308–313 (1965)

[Sciacca *et al*., 2007] Sciacca, E., Spinella, S., Anile, A.M.: Possibilistic Worst Case Distance and Applications to Circuit Sizing. Theoretical Advances and Applications of Fuzzy Logic and Soft Computing 42, 287–295 (2007)

[Segre *et al*., 2002] Segre, D., Vitkup, D., Church, G.M.: Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. USA 23(99), 15112–15117 (2002)

[Spinella and Anile, 2004] Spinella, S., Anile, A.M.: Modeling uncertain sparse data with fuzzy b-splines. Reliable Computing 10(5), 335–355 (2004)

[Stelling *et al*., 2004] Stelling, J., Sauer, U., Szallasi, Z., Doyle, F.J., Doyle, J.: Robustness of cellular functions. Cell 6(118), 675–685 (2004)

[Torres and Voit, 2002] Torres, N.V., Voit, E.O.: Pathway analysis and optimization in metabolic engineering. Cambridge University Press, New York (2002)

[van Riel, 2006] van Riel, N.A.W.: Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. Briefings in Bioinformatics 7(4), 364–374 (2006)

[Venter and Haftka, 1999] Venter, G., Haftka, R.T.: Using response surface approximations in fuzzy set based design optimization. Structural and Multidisciplinary Optimization 18(4), 218–227 (1999)

[Wang *et al*., 2007] Wang, R.S., Wang, Y., Zhang, X.S., Chen, L.: Inferring transcriptional regulatory networks from high-throughput data. Bioinformatics 22(23), 3056–3064 (2007)

[Zadeh, 1965] Zadeh, L.A.: Fuzzy sets. Information and Control 8(3), 338–353 (1965)
[Zadeh, 1968] Zadeh, L.A.: Fuzzy algorithms. Information and Control 12(2), 94–102 (1968)
[Zadeh, 1978] Zadeh, L.A.: Fuzzy sets as the basis for a theory of possibility. Fuzzy Sets and Systems 1, 3–28 (1978)
[Zi *et al*., 2005] Zi, Z., Cho, K.H., Sung, M.H., et al.: In silico identification of the key components and steps in IFN-λ induced JAK-STAT signaling pathway. FEBS Letters 579, 1101–1108 (2005)