Computational methods for quantitative submodel comparison

Andrzej Mizera* † Elena Czeizler* ‡ Ion Petre§ †

{amizera,ipetre}@abo.fi, elena.czeizler@helsinki.fi

Abstract

Comparing alternative models for a given biochemical system is in general a very difficult problem: the models may focus on different aspects of the same system and may consist of very different species and reactions. The numerical setups of the models also play a crucial role in the quantitative comparison. When the alternative designs are submodels of a reference model, e.g. knockdown mutants of a model, the problem of comparing them becomes simpler: they all have very similar, although not identical, underlying reaction networks, and the biological constraints are given by the ones in the reference model. In the first part of our study we review several known methods for model decomposition and for quantitative comparison of submodels. In the second part of the paper, we consider as a case study the eukaryotic heat shock response, an evolutionary well conserved defence mechanism against the accumulation of misfolded proteins.

Keywords: Model comparison; model decomposition; computational knockdown analysis; control-based decomposition; quantitative model refinement; heat shock response.

1 Introduction

Much experimental and theoretical effort is invested nowadays in analysing large biochemical systems, e.g., metabolic pathways, regulatory networks, signal transduction networks, aiming to obtain a holistic perspective providing a comprehensive, system-level understanding of cellular behavior. This often results in the creation and analysis of very large and complex models, often encompassing hundreds of reactions and reactants, see e.g. [8]. Therefore, obtaining a global picture of the system's architecture, in particular understanding the interactions between various components, or distinguishing a high-level functional decomposition of the network, constitutes a significant challenge. An

*Authors with equal contribution.

†Department of Information Technologies and Turku Centre for Computer Science, Åbo Akademi University, FIN-20520 Turku, Finland

‡Faculty of Medicine, University of Helsinki, Finland

§Corresponding author

important insight here is that the architecture of biological systems is a consequence of their functional requirements. Even though evolution is driven by random events, some designs, such as having an extra feedback loop helping the system to correlate better the response of the system with its trigger, may offer a selective advantage and in time, may get to dominate the population, see [54]. Thus, comparing the performance of different alternative designs in terms of sub-components being on or off, one aims to formulate general principles for how functional requirements correlate biologically with various designs.

Similar problems have been encountered for instance in engineering sciences, see [11], and a variety of strategies and approaches for solving such problems have been already developed in this framework. Thus, when aiming to obtain a system-level understanding of such large biochemical networks, one possible approach is to adapt to systems biology some of the methods originating from engineering sciences, especially from control theory, see e.g. [20, 27, 32, 60, 61, 63, 67]. Such methods have been used, as we also do in this paper, to identify various functional modules of a model, including feedback and feedforward mechanisms. To identify the quantitative contribution of each of the modules to the global behavior of the model, the general approach is to consider knockdown mutants of the initial model, missing one or several of the modules. The main problem then becomes an objective quantitative comparison of several alternative submodels for the same biological process. We focus on this problem in our study, i.e. we concentrate on the comparison of submodels of a given reference model. This issue is a special case of the general problem of alternative model comparison.

The first part of our paper contains a review of existing techniques for model decomposition and for quantitative comparison of submodels. We describe the knockdown mutants, elementary flux modes, control-based decomposition, mathematically controlled comparison, local submodel comparison, a parameter-independent submodel comparison and a discrete approach for comparing continuous submodels. We discuss a quantitative measure for the goodness of a model's fit against experimental data, as well as a technique for quantitative model refinement, and we show how both can be used for model comparison. Finally, we discuss how quantitative model comparison can be used for pathway identification.

In the second part of this study we consider as a case study the eukaryotic heat shock response, which is an evolutionary conserved mechanism protecting the cell against protein misfolding. In particular, we consider a model recently introduced in [42] for this biological process. The model was analyzed in [13] using control-driven methods where it was decomposed into several modules, including three feedback loops. We focus in the case study on identifying the numerical contribution of each of these feedback loops to the global behavior of the model. For this, we show how we can apply various model comparison methods achieving either a local, point-wise view or a global, parameter-independent analysis of their individual contributions.

2 Methods for model decomposition

We discuss in this section a number of methods for decomposing a (large) biomodel into components. The criteria on which to decide what makes a

component can be very different depending on the focus of the analysis: a metabolic pathway, a regulatory component, a feedback loop, etc. We consider in this section three main approaches. The first one is that of knockdown mutants, obtained from a reference model by dropping out some of its components (defined separately). The second one is based on elementary flux models, a well-established concept that captures the steady-state dynamics of a system. The third one is based on control theory and aims to identify how a certain part of the model is controlled through various feedback or feedforward loops. We illustrate some of these approaches in our case-study in Section 4.

2.1 Knockdown mutants

To identify the quantitative role of a component C of a model M, it is often useful to consider the model $M \setminus C$, where all interactions in C are removed from M, and compare its behavior to that of M. Any differences in the behavior of M and that of the knockdown mutant $M \setminus C$, such as loss of functionality, delayed or non-optimal response, higher energy consumption, etc. may give a hint towards the role of C in M.

If the analysis considers several disjoint components $C_1, C_2, \ldots, C_n$ of a model M, then all combinations of knockdown mutant models should be considered: all models $M \setminus \{C_{i_1}, \ldots, C_{i_k}\}$, for all $1 \leq k \leq n$ and all $1 \leq i_1 < i_< \ldots < i_k \leq n$. The technique becomes computationally challenging for a large number of components: one should compare $2^n - 1$ knockdown mutants to each other and to the reference model.

2.2 Elementary flux modes

A well-established decomposition method for biochemical models appears in the context of the analysis of metabolic pathways. An intuitive definition of a pathway is a sequence of reactions linked by common metabolites ([29]). Examples of metabolic pathways are glycolysis or amino acid synthesis. Discovering new pathways in a large model driven only by biological intuition is very difficult. An attempt to formalize the notion of pathway has been proposed in [21, 44, 56, 57, 58, 59] in the form of elementary flux modes. The intuitive meaning of an elementary flux mode is a set of reactions whose combined quantitative contribution to the system is zero. In other words, the net loss of substance caused by any reaction in that set is compensated by a net gain in the same substance incurred by some other reactions in the set. A formal definition of elementary flux modes is beyond the scope of this paper; instead we refer to [21, 29, 56, 57, 58, 59] for details. For any given metabolic network, the full set of elementary fluxes can be determined using methods of linear algebra or dedicated software such as METATOOL ([44]) or COPASI ([48]). The identification of the elementary flux modes allows the detection of the full set of nondecomposable steady-state flows that the network can support, including cyclic flows. Any steady-state flux pattern can be expressed as a non-negative linear combination of these modes ([56, 57, 58]). The identified elementary flux modes should have clear biological interpretation: a flux mode is a set of enzymes that operate together at a steady state and a flux mode is elementary if the set of enzymes is minimal, i.e. complete inhibition of any of the enzymes would result in a termination of this flux ([56, 57, 58]). The lack of possibility to

interpret the modes in this way is a signal that the model under consideration may not be correct.

2.3 Control-based decomposition

A control-driven approach to model decomposition enables the identification of the main functional modules of a system and of their individual contribution to the emergent, complex behavior of the system as a whole. In turn, this can provide great insight about various properties of a given biochemical system, e.g., robustness, efficiency, reactivity, adaptation, regulation, synchronization, etc. Through this approach, one usually aims to identify the main regulatory components of a given biochemical system: the process to be regulated, referred to as the *plant*, the *sensors* which monitor the current state of the process and send the collected information to a decision-making module, i.e. the *controller*, and the *actuator* that modifies the state of the process in accordance with the controller's decisions, thus influences the activity of the plant. One of the fundamental concepts in control theory is the *feedback mechanism*, which provides the means to cope with the uncertainties: the information about the current state of the process is sent back to the controller, which reacts accordingly to facilitate a dynamic compensation for any disturbance from the intended behavior of the system. In the case of a complex system this decomposition can be performed in different ways depending on what is considered to be the main role of that system, i.e. there may be a few reasonable choices for the plant, and the remaining components are recognized with respect to the choice of the plant.

An easy example illustrating these concepts and their interactions is given by the functioning principles of a motion activated spotlight. Here, the controller module is an electronic unit which receives an input from the motion sensor and then determines whether there are any changes in the environment. The actuator is a relay switch that operates the lighting system. This actuator is activated by the controller depending on the input sent by the sensor. Then, the switch is kept on by the controller as long as movement is detected by the sensor.

How this control-driven approach can be exploited to investigate and understand regulatory networks can be seen in [11, 16, 27, 60, 61]. Here we shortly describe the approach taken in [16]. The authors make a thorough study of the heat shock response mechanism in *Escherichia coli* based on modular decomposition. A model for the system is built and functional modules, i.e. the plant, sensors, controller, and actuator are identified. The decomposition reveals the underlying design of the heat shock response mechanism and its level of complexity, which, as the authors show, is not justified if only the functionality of an operational heat shock system is required. Further, this observation leads to the introduction and analysis of hypothetical design variants (mutants) of the original heat shock response model. In the original model one feedforward (temperature sensing) and two feedback elements (σ^{32} factor sequestration feedback loop and σ^{32} degradation feedback loop) can be isolated. The variants are obtained through the elimination of either the σ^{32} degradation feedback loop or of both feedbacks. One by one the variants in order of increasing complexity are considered starting from the simplest architecture containing just the feedforward element (the *open-loop design*). Based on numerical simulations,

the authors demonstrate how the addition of subsequent layers of regulation, (i.e., increase in the complexity of the model), improves the performance of the response in terms of systemic properties such as robustness, noise reduction, speed of response and economical use of cellular resources. Moreover, this systematic approach enables the identification of the role of each of the regulatory layers to the overall behavior of the system.

3 Methods for submodel comparison

Comparing alternative models for a given biochemical system is in general a very difficult problem, involving a deep analysis of the underlying network of reactions, of the biological assumptions, as well as of the numerical setup. To decide what are the benefits of one design over another, or to understand what are the selection requirements involved in an evolutionary design, one needs some unbiased methods to objectively compare the alternative designs.

3.1 Mathematically controlled model comparison

One such method is the mathematically controlled comparison, [54], which provides a structured approach for comparing alternative regulatory designs with respect to some chosen measures of functional effectiveness. Under this approach, mathematical models for both the reference design and the alternatives are first developed in the framework of canonical nonlinear modelling referred to as S-systems, [51], [52], and [53]. This canonical nonlinear representation, developed within the power-law formalism, is a system of non-linear ordinary differential equations (ODE) with a well-defined structure. Moreover, this framework allows the alternative models to differ from the reference design in only one process (e.g., only one feedback mechanism), which is the focus of the comparison. Then, in each of the alternative models one sets the numerical values of the parameters to be identical to those from the reference model for all processes other than the process of interest. This leads to a so-called internal equivalence between the reference model and the alternatives. Next, various systemic properties are selected and used to impose some constraints for all the other parameters in the alternative designs. In general in this approach, one imposes that some steady state values or logarithmic gains are equal in the reference model and its alternatives. This provides a way to express the parameters of the process of interest in the alternative models as functions of the parameters of the reference model. Thus, one obtains a so-called external equivalence between the reference model and the alternative designs, meaning that to an external observer the considered models are equivalent with respect to the selected systemic properties. Finally, one chooses various measures of functional effectiveness depending on the particularities of the biological context of these models and uses them to compare the alternative designs with the reference model. By doing this, one usually aims to determine analytically the qualitative differences between the compared models. This method was successfully used to compare alternative regulatory designs in, e.g., metabolic pathways, [25], [55], gene circuits, [23], immune networks, [6]. Moreover, by introducing specific numerical values for the parameters of the models, one is also able to quantify these differences but, at the same time, the generality of the results

is lost. Thus, in [3], the method of mathematically controlled comparison was extended to include some statistical methods, [2], [4], that allow the use of numerical values for the parameters while still preserving the generality of the conclusions. We discuss this extension in the following.

3.2 An extension of the mathematically controlled comparison

The first step of this extension is to generate a representative ensemble of sets of parameter values. Since the exact statistical distribution of the parameters values is often not known in practice, the most appropriate approach is to sample uniformly a given range of values. There exist different methods for scanning a given interval of values, ranging from various types of random samplings to some systematic deterministic scanning methods, see e.g., [49]. Using this ensemble of sets of parameters, we can then construct a large class of numerical models both for the reference and for the alternative designs. There are two different methods to construct such a class of systems for which we can then investigate some statistical properties. A *structural class* consists of systems having the same network topology, i.e., generated by the sampling of the parameter space. A *behavioral class* consists of systems that exhibit a particular systemic behavior, e.g., exhibiting a steady state behavior under given conditions, or low concentrations of intermediary products, or small values for the parameter sensitivity, see, e.g., [4]. The members of such a class are obtained in two steps: first generate a set of parameters by sampling the parameter space, then test the sample for the desired systemic behavior and keep only those systems that fulfil the conditions.

After constructing this large class of numerical models both for the reference and the alternative architectures, one can start comparing the values of a given systemic property P between the reference model and its alternative designs. One way to do this is by using density plots of the ratio $R = P_{reference}/P_{alternative}$ versus the values $P_{reference}$, where the subscript indicates in which model the property P was measured. Such density plots can be used for instance to compute rank correlations between the considered property P (measured in the reference model) and the values of the ratio R. However, this is not easy to do if the density plots are very scattered. Then, one can construct secondary density plots by using the moving median technique as follows. The density plot can be interpreted as a list of N pairs of values $(P_{reference}, R)$, which can be arranged in an ordered list L with respect to the first component, $P_{reference}$. Then, we pick a window size W, usually much smaller than the sample size N and we compute the median $< R >$ of the ratio values and the median $< P >$ of the values $P_{reference}$, for the first W pairs in the list L. Then, we advance the window by one, we collect the ratios and the values $P_{reference}$ from the second until the $(W+1)$st pair and compute the corresponding median values $< R >$ and $< P >$. This process is continued until the last pair of the list L is used. In the secondary density plot, we will pair the computed values $< R >$ with the corresponding $< P >$ values. This moving median technique is very useful since for a finite ordered sample of size N, the moving median tends to the median of the samples as the value W approaches N. These secondary density plots can be used to compare the efficiency of two classes of models from the point of view of a given systemic property.

3.3 Local submodel comparison

Another approach for comparing alternative designs that are actually submodels of a reference architecture was proposed in [13]. The focus of such a comparison of various submodels could be, for instance, a functional analysis of various modules of a large system. Then, the underlying reaction networks in the alternative designs are very similar (although not identical), and both the biological constraints and the kinetics of the reactions are given by those of the reference model. The only remaining question regards the initial distribution of the variables in the alternative models. In the mathematically controlled comparison they are usually taken from the reference model. However, for some biochemical systems this choice might lead to biased comparisons. For instance, in the case of regulatory networks, models should be in a steady state in the absence of the trigger of the response and indeed the initial values of the reference model are usually chosen in such a way to fulfil this condition. However, this will not imply in general that also a submodel will be in its steady state if it uses the same initial values as the reference model. Thus, the dynamic behavior of the submodel will be the result of two intertwined tendencies: migrating from a possible unstable state and the response to a trigger. If the focus of the comparison is exactly the efficiency of the response of various submodels to a trigger, then the approached proposed in [13] is more appropriate, yielding biologically unbiased results. In this approach, the initial distribution of the reactants is chosen in such a way that the initial setup of each submodel constitutes a steady state of that design in the absence of a trigger.

3.4 A quantitative measure for the goodness of model fit against experimental data

Another method for comparing alternative models is to analyze how well their predictions fit available experimental data. However, estimating the set of parameters of a given computational model in such a way that its predictions fit some experimental data is a computationally difficult problem, see e.g., [5, 36, 39]. A common approach for this problem is to minimize a cost function quantifying the differences between the experimental measurements and the values predicted by the model. For instance, one of the functions used for this purpose is the sum of squared deviations $SS = \sum_{i=1}^{n}(x_i - y_i)^2$, where n is the number of experimental data points and for each i, x_i and y_i represent the experimental value and the corespondent value predicted by the model. In particular in this case a smaller value of SS indicates a better fit.

There are many methods for tackling such optimization problems either locally or globally, each of them having its own advantages and disadvantages. For example, local methods converge faster to a solution but they tend to find local optima while global optimization methods are slower but they to converge to global optima. Furthermore, there are two types of global optimization methods: deterministic [18, 24] and stochastic ones [1, 19]. Although the convergence to the global optimum is guaranteed when using a deterministic method, the termination of the search process within a finite time interval is not ensured [39]. On the other hand the stochastic optimization methods usually locate quite efficiently a good approximation of the solution, i.e., located in a vicinity of the global solution, within an acceptable time interval [39]. Thus, the stochastic

global optimization methods tend to be usually preferred for parameter estimation problems. Many parameter estimation methods are currently available in various software packages, such as COPASI [48].

One of the demanding tasks within this iterative process of finding a suitable parameter set is the identification of a measure to quantify the quality of the fit for each set of parameters. This measure can be also used to compare alternative models and decide which one fits best against the experimental data. One solution for this problem, proposed in [30], is to use a dimensionless number representing the deviation of the model from the experimental data, normalized by the mean of the predicted values, i.e.,

$$fq = \frac{\sqrt{SS/n}}{\text{mean of predicted values}} \cdot 100\%, \tag{1}$$

where n is the number of experimental data points. In particular, it was argued in [30] that a low value of fq, e.g., lower than 15%, indicates a successful fit. Thus, when comparing several alternative models we could actually compare their associated fit quality values indicating how well the predictions of these models fit existing experimental data.

3.5 Quantitative refinement

The two alternative models that we aim to compare are very often built on different levels of details. Thus, in order to ease the comparison, a preliminary refinement of one (or even both) of the models could be useful. In fact, model refinement is part of the complex process of model building. Indeed, this process includes a series of iterative steps including hypothesis generation, experimental design, experimental analysis and model refinement, [7], [27]. The first step when developing a model is to create an abstraction of the biological process, i.e., to select a relatively small number of biochemical reactions which succeed to describe the main mechanisms of the considered process. The reactions included in this model are abstract representations of some particular subprocess, and can encapsulate many biochemical reactions from the modeled system. A mathematical model is then associated to this molecular model. For this, we have to choose an appropriate kinetic law, e.g., mass-action law or Michaelis-Menten kinetics, based on which we can then write the mathematical equations describing the dynamics of the system. The only thing left is to chose a numerical setup of this mathematical model, which is either obtained from the literature, or is derived through various computational model fit procedures using available experimental data.

The refinement of a given model can be done in several different ways. For instance, within *data refinement* of a model we replace one (or more) of its species with several subspecies. This way, the new refined model will include more details about its subspecies and will illustrate various differences in their behavior. Another type of refinement, called *process refinement*, focuses on the model reactions. In this case, the model is refined by replacing a generic reaction which describes a particular process with several reactions describing in more details the intermediary steps of that process. In order to include all the intended changes of the initial models, one possibility is to simply repeat the whole model development procedure. However, this can lead to extremely

inefficient processes since it requires to re-fit the model, a step which is both time-consuming and computationally-intensive, [9]. Another approach, not so much investigated so far, see [38] and [14] for some recent case studies, is to refine the initial model in such a way such that the initial model fit is preserved.

3.6 Parameter-independent submodel comparison

Although by choosing particular numerical setups we can quantify the differences between various submodels, a major downside is that we lose the generality of the results. This can be avoided if we employ some statistical methods to scan the parameter space, see [37]. Since all alternative models are obtained by eliminating reactions from the reference model, the parameters of the alternative architecture are in fact a subset of the parameter set of the reference model. Thus, we only need to scan the parameter value space of the reference model. This provides us with a set of parameter value vectors. Each coordinate of these vectors is associated with one of the parameters in the reference model and determines the value of the corresponding parameter. We consider each of the vectors one by one. We set the parameters of the reference model and the submodel in accordance with the considered vector. The initial values of the variables of the reference model and of the submodel are determined independently of each other by a systemic property such as the system being in a steady state in a given setup. For example, in the general case of stress response, we expect in accordance with biological observations that a feasible mathematical model is in a steady state under the unstressed, physiological conditions. (We call steady state a numerical configuration of the model such that starting from that configuration, the model shows no change in the level of any of the variables; in other words, the net loss per unit of time in every variable is exactly compensated by the net gain per unit of time in that variable.) Now, assuring that all mathematical submodels satisfy such systemic properties makes them suitable to be considered as viable alternative formal descriptions of the biological mechanism being analysed. As a result we obtain the instantiations of the reference model and the submodels and we can run numerical simulations for all of them in order to evaluate their functional effectiveness. Finally, having done this for all sampled vectors, the obtained results for the variants are used to compare the models by use of some statistical measures.

3.7 Model comparison for pathway identification

Very often, the exact pathway through which a particular biochemical process is regulated is not known. Instead, several alternative pathways underlying different mechanisms are proposed in the literature. Then, model comparison is employed in order to decide which of them is better in terms of fitting the experimental data better, or explaining better an observed qualitative behavior. For instance, in [12] a hybrid quantum mechanics (QM) - molecular mechanics (MM) approach was used to compare three alternative mechanisms proposed for the triosephosphate isomerase catalyzed reactions. This approach, introduced in 1976 by Warshel and Levitt [66], is a molecular simulation method that combines the accuracy of the QM computations and the speed of the MM calculations. This way, it allows the study of chemical processes in solution and in proteins. In [15], a detailed comparison of several phosphorylation-driven

control mechanisms for the regulation of the eukaryotic heat shock response (HSR) was done by analysing the goodness of the fit of the alternative models. In particular, three different phosphorylation pathways were investigated in an attempt to uncover the contribution of each of these pathways in controlling the HSR process. A detailed computational model was created for each of these pathways, which was then subjected to parameter estimation in order to be fit to the existing experimental data.

4 Case study

4.1 A biochemical model for the heat shock response

The heat shock response (HSR) is a highly evolutionary-conserved defence mechanism among organisms ([33]). It serves to prevent and repair protein damage induced by elevated temperature and other forms of environmental, chemical or physical stress. Such conditions induce the misfolding of proteins, which in turn accumulate and form aggregates with disastrous effect for the cell. In order to survive, the cell has to abruptly increase the expression of heat shock proteins. These proteins operate as intra-cellular chaperons, i.e. play a crucial role in folding of proteins and re-establishment of proper protein conformation. They prevent the destructive protein aggregation. We discern two main reasons that account for the strong interest in the heat shock response mechanism observed in recent years, see e.g. [10, 46, 65]. First, as a well-conserved mechanism among organisms, it is considered a promising candidate for disentangling the engineering principles being fundamental for any regulatory network ([16, 17, 31, 64]). Second, besides their functions in the HSR, heat shock proteins have fundamental importance to many key biological processes such as protein biogenesis, dismantling of damaged proteins, activation of immune responses and signalling, see [26, 45]. In consequence, a thorough insight into the HSR mechanism would have significant implications for the advancement in understanding the cell biology.

In order to coherently investigate the HSR a number of mathematical models has been proposed in the literature, see e.g. [16, 34, 40, 47, 62]. In this study we consider a recently introduced model of the eukaryotic heat shock response ([42] and [43]). In this model the central role is played by the heat shock proteins (hsp), which act as chaperons for the misfolded proteins (mfp): the heat shock proteins sequester the misfolded proteins ($\mathsf{hsp}\colon \mathsf{mfp}$) and help the misfolded proteins to regain their native conformation (prot). The defence mechanism is controlled through the regulation of the transactivation of the hsp-encoding genes. The transcription is initiated by heat shock factors (hsf), some specific proteins which first form dimers (hsf_2), then trimers (hsf_3) and in this configuration bind to the heat shock elements (hse), i.e. certain DNA sequences in the promotor regions of the hsp-encoding genes. Once the trimers bind to the promoter elements ($\mathsf{hsf}_3\colon \mathsf{hse}$), the transcription and translation of the hsp-encoding genes boosts and, in consequence, new heat shock protein molecules get synthesized at a substantially augmented rate.

When the amount of the heat shock proteins reaches a high enough level that enables coping with the stress conditions, the production of new chaperon molecules is switched off by the excess of the heat shock proteins. To this aim

hsp form complexes with the heat shock factors (hsp: hsf) in three independently and concurrently running processes: 1) by binding to the free hsf, 2) by breaking the dimers and trimers, and 3) by breaking the hsf_3: hse, in result of which the trimer gets unbound from the DNA and decomposed into free hsf molecules. This terminates the enhanced production of new heat shock protein molecules and blocks the formation of new hsf trimers. As soon as the temperature increases, proteins present in the cell start misfolding. The misfolded proteins titrate hsp away from the hsp: hsf complexes. This enables the accumulation of free hsf molecules, which in turn form trimers and promote the production of new chaperons. In consequence the response mechanism gets switched on. The full list of biochemical reactions constituting the biochemical model from [42] is presented in Table 1. The model is based only on well-documented reactions without introducing any hypothetical mechanisms or experimentally unsupported biochemical reactions. Also, it assumes three conservation relations: for the total amount of heat shock factors, for the total amount of proteins (except heat shock proteins and heat shock factors), and for the total amount of the heat shock elements:

- $[\mathsf{hsf}] + 2 \times [\mathsf{hsf}_2] + 3 \times [\mathsf{hsf}_3] + 3 \times [\mathsf{hsf}_3\colon \mathsf{hse}] + [\mathsf{hsp}\colon \mathsf{hsf}] = C_1$,
- $[\mathsf{prot}] + [\mathsf{mfp}] + [\mathsf{hsp}\colon \mathsf{mfp}] = C_2$,
- $[\mathsf{hse}] + [\mathsf{hsf}_3\colon \mathsf{hse}] = C_3$,

for some constants C_1, C_2, C_3 called *mass constants*. For a full presentation and discussion of this model we refer the reader to [42].

Based on the assumption of mass-action law for all the reactions (2)-(13) an associated mathematical model of the eukaryotic heat shock response is obtained. The resulting mathematical model is expressed in terms of ten, first order, ordinary differential equations. The mathematical model comprises 16 independent kinetic parameters and 10 initial conditions. We refer to [42] for the details of the ODE model and its numerical setup.

4.2 Control-based decomposition

In [13] a control-driven modular decomposition of the heat shock response model has been performed. In result, the model has been divided into four main functional submodules usually distinguished in control engineering: the plant, the sensor, the controller and the actuator. In the case of the HSR model the plant is the misfolding and refolding of proteins, the actuator consists of the synthesis and degradation of the chaperons, the sensor measures the level of hsp in the system and the controller regulates the level of DNA binding. Moreover, within the controller we distinguish three feedback mechanisms. The feedback loops are responsible for sequestering the heat shock factors in different forms by the chaperons. In this way the feedback loops are decreasing the level of DNA binding. The three identified feedback mechanisms are the following:

- FB1: sequestration of free hsf, i.e. reaction $(6)^+$ (the 'left-to-right' direction of reaction (6));
- FB2: breaking of hsf dimers and trimers, i.e. reactions (7) and (8);
- FB3: unbinding of hsf_3 from hse and breaking the trimers, i.e. reaction (9).

The control-driven functional decomposition of the eukaryotic heat shock response model is shown in Figure 1, where the reaction numbers refer to the reactions in Table 1. In Figure 2 a graphical illustration of the control structure, i.e. the three feedback loops and their points of interactions with the mainstream process, is presented.

4.3 The knockdown mutants

In [13, 37] the reference architecture and seven knockdown mutants (alternative architectures) were considered. The mutants were obtained by eliminating from the reference architecture all possible combinations of the three feedback loops FB1, FB2 and FB3. The mutants were denoted as M_X, where $X \subset \{1, 2, 3\}$ is the set of numbers of the feedback mechanisms present in M_X:

- M_0 is determined by reactions (2)-(5), (10)-(13) and, in the terminology of control theory, is characterized by the *open-loop design*;
- M_1 is determined by reactions (2)-(6), (10)-(13);
- M_2 is determined by reactions (2)-(5), (7)-(8), (10)-(13), and the 'right-to-left' direction of reaction (6);
- M_3 is determined by reactions (2)-(5), (9)-(13), and the 'right-to-left' direction of reaction (6);
- $M_{1,2}$ is determined by reactions (2)-(8), (10)-(13);
- $M_{1,3}$ is determined by reactions (2)-(6), (9)-(13);
- $M_{2,3}$ is determined by reactions (2)-(5), (7)-(13), and the 'right-to-left' direction of reaction (6);
- $M_{1,2,3}$ is the reference architecture consisting of all reactions (2)-(13).

By looking at the associated differential equations, it is easy to see that if the mutant M_0 starts from a steady state at physiological conditions, i.e. 37 °C, then it is non-responsive, i.e., it shows no increase in DNA binding for any arbitrarily high temperature. Thus, we remove M_0 from further considerations.

4.4 Local comparison of the knockdown mutants

In order to identify the individual contributions of each of the three feedback mechanisms to the regulation of HSR, we can first locally compare the knockdown mutants by using the method described in Section 3.3, see [13]. The numerical setups for each of the eight knockdown mutants are chosen such that they satisfy the following two constraints:

(i) The kinetic rate constants of the reactions included in the mutants are chosen to be identical to those of the corresponding reactions in the reference model. Also, the values of the mass constants C_1, C_2, C_3 of each of the mutants are identical to those of the reference model.

(ii) The initial distribution of the variables of each knockdown mutant are chosen such that they form a steady state of that model at $37°C$. Note that the initial values of all variables for the reference model were chosen in the same way in [41].

Then, the comparison is focused on the dynamical behavior of the mutants at $42°C$. This particular temperature was chosen for the analysis since at $42°C$ in the experimental data we can notice both a pronounced heat shock response in terms of increased levels of misfolded proteins, as well as an explicit response in terms of increased, transient DNA binding of hsf_3, see [41].

While the main task of the heat shock response is to keep the level of misfolded proteins under control, it is also very important for the cell to use efficiently its materials and energy. Thus, the analysis of the mutants focused on two main aspects: the level of misfolded proteins and the level of chaperons needed to achieve a response at $42°C$. Based on these two aspects, the contribution of the three feedbacks was analyzed in [13] with respect to the following four performance indicators:

- (A) *The system makes economical use of the cellular resources* (i.e., the hsp-encoding gene is only transactivated for a short while when exposed to heat shock): FB_1 was found to play the major role here, while FB_2 and FB_3 are also important. In the absence of FB_1, gene transcription is at the 100% level even without heat shock.

- (B) *The system is fast to respond to a heat shock* (i.e., the hsp-encoding gene is quickly transactivated in response to heat shock): FB_3, FB_1 were found to play the major role here. In particular, model $M_{1,3}$ reacts to heat shock as fast as the reference model, although the level of its response is lower than that of the reference model.

- (C) *The response is effective* (i.e., the mfp concentration is kept low for mild heat shocks): the open-loop structure of the mutant M_0 is enough to achieve this property. Although none of the feedback mechanisms is needed for maintaining a low [mfp] for a heat shock at $42°C$, they play an important role in minimizing the cost of the response.

- (D) *The response is scalable* (i.e., we notice a higher response for a higher temperature): FB_3, FB_1 were found to play the major role here. The mutant $M_{1,3}$ is the only one that scales its response to higher temperatures, in a similar way as the reference model, i.e., for higher temperatures, gene transactivation raises faster and to a higher level.

For more details about this analysis we refer to [13].

4.5 Parameter-independent comparison of the mutant behavior

The previous analysis is clearly heavily dependent on the parametric setups chosen for the compared models. In order to avoid this, one approach, see [37], would be to employ some statistical methods to scan the parametric space, as explained in Section 3.6. For this, we use the *Latin hypercube sampling* method (LHS), originally introduced in [35]. It provides samples which are uniformly

distributed over each parameter while the number of samples is independent of the number of parameters. The sampling scheme can be briefly described as follows. First, the desired size N of the sampling set is chosen. Next, the range interval of each parameter is partitioned into N non-overlapping intervals of equal length. For each parameter, N numerical values are randomly selected, one from each interval of the partition according to a uniform distribution on that interval. Finally, the N sampled values for the i-th parameter of the model are collected on the i-th column of a $N \times p$ matrix, where p is the number of model parameters and the values on each column are shuffled randomly. As a result, each of the N rows of the matrix contains numerical values for each of the p parameters. For a detailed description of this sampling scheme we refer the reader to [22, 35], see also [42] for an example of the application of this sampling method in the context of model identifiability problem.

In particular, in [37] a sample of 10.000 vectors of parameter values for the reference architecture was obtained through the Latin hypercube sampling described above. For the heat shock response model the sampled vectors are of length 15, i.e. the number of the unknown reference architecture parameters. The value of the 16th remaining parameter, i.e. the hsp degradation rate constant is assumed to be known and is obtained based on the fact that heat shock proteins are generally long-lived proteins, see [50]. In particular, their half-life was chosen to be 6 hours. Then, the procedure described next is repeated separately for each of the six mutants. To begin with, each sampled vector of parameter values is used to setup the parameters in the mathematical models of the considered mutant and the reference architecture ($M_{1,2,3}$). From the construction of each mutant it follows that the corresponding mathematical model contains only a subset of the parameters of the reference model, so this step can be performed. Next, the steady state concentrations at 37 °C both for the mutant and the reference model are numerically computed and set as their respective initial states. In this way we obtain two instances of the mathematical models, i.e. one for the mutant and the second for the reference model. Further, the temperature is increased to 42 °C and the quantities

$$\begin{aligned}
\Theta_1 &= \max_{t\in[0s,1800s]}(\text{total}\,\mathsf{mfp}(t)),\\
\Theta_2 &= \max_{t\in[0s,1800s]}(\mathsf{hsf}_3 : \mathsf{hse}(t)) - \mathsf{hsf}_3 : \mathsf{hse}(0),\\
\Theta_3 &= \frac{1}{T}\int_0^T (\text{total}\,\mathsf{hsp}(t))\mathrm{d}t,\\
\Theta_4 &= \frac{1}{T}\int_0^T (\text{total}\,\mathsf{mfp}(t))\mathrm{d}t
\end{aligned}$$

are computed both for the mutant and the reference instance. The initial 30 min. of the response are considered for the computation of Θ_1 and Θ_2. In the case of Θ_3 and Θ_4 the time range of 4 hours ($T = 14400s$) is taken into account. These quantities are used to evaluate the functional effectiveness of the mutant. Having these quantities computed for all the 10.000 sampled parameter values, the scatter plot of the $R_1 = \Theta_1^{\mathrm{m}}/\Theta_1^{\mathrm{r}}$ against Θ_1^{r} values is made, where the superscripts m and r indicate the instance for which Θ_1 was computed, i.e. the instance of the mutant or the reference model, respectively. Finally, the moving median technique is applied to the scatter plot with the window size set

to 500. These result in a trend curve summarizing the data of the scatter plot and revealing the overall dependency between the considered quantities. Similar plots are computed for $R_2 = \Theta_2^{\mathrm{m}}/\Theta_2^{\mathrm{r}}$. Moreover, scatter plots of Θ_3 versus Θ_4 both for the mutant and the reference architecture are made and the moving median technique is applied to each of these plots.

The mutants represent six different potential architectures of the heat shock response mechanism and the sampling procedure, as explained above, provides us with 10.000 different instantiations of each of the mutants and the reference architecture.

In the analysis performed in [37] it was assumed that the heat shock response at raised temperatures is accompanied, and hence characterized, by the following three phenomena:

1. increase in DNA-binding with respect to the steady-state level at 37 °C,
2. increase in the level of mfp, and
3. increase in the level of hsp as the effect of the response to the higher level of mfp in the cell.

Thus, the analysis of the architecture properties of the six mutants with respect to the reference architecture was based on the following plots: R_1 vs Θ_1^{r} (Figure 5 of [37]), R_2 vs Θ_2^{r} (Figure 7 of [37]), Θ_3 vs Θ_4 (Figure 3) made for each of the mutants and for the reference architecture. We refer to the Θ_3 vs Θ_4 plot as the cost plot (or simply the cost) of the corresponding architecture. This is motivated by the fact that the efficiency of the heat shock response mechanism could be measured by the amount of chaperons needed to cope with the intensified misfolding of proteins. Hypothetically, a cell which produces smaller amounts of hsp than some other cell to cope with the heat shock would be considered the one which manages with stress conditions at a lower cost in terms of its resources than the latter one. Notice however that in our case we are not assessing the ability of particular models to cope with heat shock, i.e. the sampled models are neither validated against experimental data nor classified by any other means whether they enable the cell to survive or not in the stress conditions. Hence the cost plots reflect just the general tendency of the models instantiating a particular architecture to keep certain average in time amounts of hsp in response to different average levels of mfp present in the system. The reference trend line indicates a clear linear dependency between the average levels of hsp and mfp, see Figure 4.

Based on these aspects, it was noted in [37] that all the mutants lacking two feedbacks exhibit no heat shock response in the sense of the above definition: as observed previously, there is no increase in the DNA-binding. This is in agreement with the results presented in [13], where the models with only one feedback kept the DNA-binding at the maximum possible level both at 37 °C and 42 °C throughout the simulation time of 50.000s. The HSR can be observed however in the mutants $M_{1,3}$ and $M_{1,2}$. In the case of the $M_{2,3}$ mutant the HSR is still observed, but only for a fraction of the 10.000 sampled models, i.e. only those parameter values for which the reference architecture displays the maximal possible increase in the peak of DNA-binding with respect to the steady-state level at 37 °C. This is in complete agreement with previous observations that FB1 is the most powerful feedback, see [13]. Since FB2 and FB3 include hsf

sequestration as one of their features, they compensate partially for the lack of FB1. However, only FB2 or only FB3 are not enough to enforce the system's behavior to have the HSR characteristics. Despite its power, FB1 alone is also not enough and one of the other feedbacks is also needed in order to implement a response mechanism with the features describing the heat shock response.

4.6 Pathway identification for the phosphorylation-driven control of the heat shock response

In [15] an extension of the biochemical model for the eukaryotic heat shock response from Section 4.1 was introduced where phosphorylation of hsfs was considered. The synthesis of hsps is highly regulated at the transcription level and phosphorylation of hsfs constitutes one of the regulation mechanisms. The introduced extended model allowed the authors of [15] to investigate the positive role of phosphorylation in the up-regulation of hsf transcriptional activity. In particular, a number of computational models associated with the extended biochemical model were derived by assuming the mass-action kinetics and used to investigate three plausible (de)phosphorylation pathways. First, it was analyzed whether the kinase and phosphates dynamics for the heat-induced misfolding and refolding can lead to the experimentally observed evolution of the total level of phosphorylated hsf molecules. Second, on top of the heat-induced misfolding and refolding of both kinase and phosphates, it was supposed that the hsf molecules can be both phosphorylated and dephosphorylated while they form hsp : hsf complexes. Finally, with respect to the second pathway, the hsf molecules forming hsp : hsf complexes could only be dephosphorylated. For each of these three pathways two different mechanisms for modeling the phosphorylation-dependent transcription were considered. For the first mechanism, it was assumed that the transcription proceeds linearly depending on the level of phosphorylation of the hsf_3 : hse complex. For the second one, the assumption was that hsp synthesis is activated only by the hyper-phosphorylation of the hsf_3 : hse compound, i.e., that the transcription of the hsp-encoding gene proceeds only after at least two of the three hsf molecules from this compound become phosphorylated. Moreover, in order to verify and strengthen the results, for each computational model implementing one of the six scenarios (3 pathways, each with 2 different mechanisms of phosphorylation-dependent transcription) a corresponding reduced model employing the Michaelis-Menten kinetics was considered, where the kinase and phosphatase enzymes were not modeled explicitly anymore. All in all, twelve computational models of phosphorylation mediated transcription of hsp were considered in [15] and the analysis regarding possible (de)phosphorylation pathways was based on the assessment of repetitive parameter estimation procedures of the computational models, without the use of any quantitative measurement of the results. The models were fitted to two experimental measurements reported in [28] regarding the DNA binding activity and the total number of phosphorylated hsf proteins in HeLa cells under 42 °C heat-shock. For more details on the extended biochemical model, the associated twelve computational models, parameter estimation procedure and performed analysis regarding possible (de)phosphorylation pathways, we refer to [15].

5 Discussion

Very often, various experimental investigations of a given biochemical system generate a large variety of alternative molecular designs, thus raising questions about comparing their functionality, efficiency, and robustness. Comparing alternative models for a given biochemical system is, in general, a very difficult problem which involves a deep analysis of various aspects of the models: the underlying networks, the biological constraints, and the numerical setup. The problem becomes somewhat simpler when the alternative designs are actually submodels of a larger model: the underlying networks are similar, although not identical, and the biological constrains are given by the larger model. It only remains to decide how to chose the numerical setup for each of the alternative submodels, i.e., the initial conditions and the kinetics.

In the first part of our study we review several known methods for model decomposition and for quantitative comparison of submodels. We describe the knockdown mutants, elementary flux modes, control-based decomposition, mathematically controlled comparison and its extension, local submodels comparison, a parameter-independent submodel comparison and a discrete approach for comparing continuous submodels. We also show how the goodness of model fit against experimental data and quantitative model refinement can be used for comparing alternative designs as well as how to use model comparison for pathway identification. In the second part of this study we consider as a case study the eukaryotic heat shock response. First, by using a control-based approach, the main functional modules of this system are identified, see [13]. In particular, we underline three feedback mechanisms regulating this response. Then, in order to identify the individual role these mechanisms play in the regulation of the heat shock response, we construct knockdown mutants by eliminating from the reference architecture all possible combinations of the three feedbacks. These mutants are then compared both with themselves and with the reference model by using various methods described in the first part of this study.

Acknowledgments. The work of Elena Czeizler, Andrzej Mizera and Ion Petre was supported by Academy of Finland, grants 129863, 108421, and 122426. Andrzej Mizera is on leave of absence from the Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland.

References

[1] M. M. Ali, C. Storey, and A. Törn. Application of stochastic global optimization algorithms to practical problems. *J. Optim. Theory Appl.*, 95(3):545–563, 1997.

[2] R. Alves and M. A. Savageau. Comparing systemic properties of ensembles of biological networks by graphical and statistical methods. *Bioinformatics*, 16(6):527–533, 2000.

[3] R. Alves and M. A. Savageau. Extending the method of mathematically controlled comparison to include numerical comparisons. *Bioinformatics*, 16(9):786–798, 2000.

[4] R. Alves and M. A. Savageau. Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways. *Bioinformatics*, 16(6):534–547, 2000.

[5] S. M. Baker, K. Schallau, and B. H. Junker. Comparison of different algorithms for simultaneous estimation of multiple parameters in kinetic metabolic models. *J Integr Bioinform*, 7(3):133, 2010.

[6] R. J. D. Boer and P. Hogeweg. Stability of symmetric idiotypic networks-a critique of hoffmann's analysis. *Bull. Math. Biol.*, 51:217–222, 1989.

[7] F. J. Bruggeman and H. V. Westerhoff. The nature of systems biology. *Trends in Microbiology*, 15(1):45–50, 2007.

[8] W. W. Chen, B. Schoeberl, P. J. Jasper, M. Niepel, U. B. Nielsen, D. A. Lauffenburger, and P. K. Sorger. Inputoutput behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Molecular Systems Biology*, 5:239.

[9] W. W. Chen, B. Schoeberl, P. J. Jasper, M. Niepel, U. B. Nielsen, D. A. Lauffenburger, and P. K. Sorger. Input-output behavior of erbb signaling pathways as revealed by a mass action model trained against dynamic data. *Molecular Systems Biology*, 5(239), 2009.

[10] Y. Chen, T. S. Voegeli, P. P. Liu, E. G. Noble, and R. W. Currie. Heat shock paradox and a new role of heat shock proteins and their receptors as anti-inflammation targets. *Inflammation & Allergy - Drug Targets*, 6(2):91–100, 2007.

[11] M. E. Csete and J. C. Doyle. Reverse Engineering of Biological Complexity. *Science*, 295:1664–1669, 2002.

[12] Q. Cui and M. Karplus. Triosephosphate isomerase: A theoretical comparison of alternative pathways. *J. Am. Chem. Soc.*, 123:2284–2290, 2001.

[13] El. Czeizler, E. Czeizler, R.-J. Back, and I. Petre. Control strategies for the regulation of the eukaryotic heat shock response. In P. Degano and R. Gorrieri, editors, *Computational Methods in Systems Biology*, volume 5688 of *Lecture Notes in Computer Science*, pages 111–125, Heidelberg, 2009. Springer-Verlag.

[14] E. Czeizler, E. Czeizler, B. Iancu, and I. Petre. Quantitative model refinement as a solution to the combinatorial size explosion of biomodels. In *Proceedings of The Second International Workshop on Static Analysis and Systems Biology*, ENTCS, 2011.

[15] E. Czeizler, V. Rogojin, and I. Petre. The phosphorylation of the heat shock factor as a modulator for the heat shock response. In F. (ed.), editor, *Proceedings of the 9th International Conference on Computational Methods in Systems Biology*, volume to appear, 2011.

[16] H. El-Samad, H. Kurata, J. C. Doyle, C. A. Gross, and M. Khammash. Surviving heat shock: Control strategies for robustness and performance. *PNAS*, 102(8):2736–2741, 2005.

[17] H. El-Samad, S. Prajna, A. Papachristodoulou, M. Khammash, and J. C. Doyle. Model validation and robust stability analysis of the bacterial heat shock response using SOSTOOLS. In *Proceedings of the 42th IEEE Conference on Decision and Control*, volume 4, pages 3766–3771, Dec. 2003.

[18] I. E. Grossmann. *Global optimization in engineering design.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[19] C. Guus, E. Boender, and H. E. Romeijn. Stochastic methods. In R. Horst and P. M. Pardalos, editors, *Handbook of Global Optimization.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.

[20] B. A. Hawkins and H. V. Cornell, editors. *Theoretical Approaches to Biological Control.* Cambridge University Press, 1999.

[21] R. Heinrich and S. Schuster. *The regulation of cellular systems.* Chapman & Hall, New York, 1996.

[22] J. C. Helton and F. J. Davis. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety*, 81(1):23–69, 2003.

[23] W. S. Hlavacek and M. A. Savageau. Rules for coupled expression of regulator and effector genes in inducible circuits. *J. Mol. Biol.*, 255:121–139, 1996.

[24] R. Horst and H. Tuy. *Global optimization: Deterministic approaches.* Springer-Verlag, Berlin, 1990.

[25] A. Hunding. Limit-cycles in enzyme-systems with nonlinear negative feedback. *Biophys. Struct. Mech.*, 1:47–54, 1974.

[26] H. K. Kampinga. Thermotolerance in mammalian cells: protein denaturation and aggregation, and stress proteins. *Journal of Cell Science*, 104:11–17, 1993.

[27] H. Kitano. Systems biology: A brief overview. *Science*, 295(5560):1662–1664, 2002.

[28] M. P. Kline and R. I. Morimoto. Repression of the heat shock factor 1 transcriptional activation domain is modulated by constitutive phosphorylation. *Molecular and Cellular Biology*, 17(4):2107–2115, 1997.

[29] E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach. *Systems Biology in Practice. Concepts, Implementation and Application*. Wiley-VCH, 2005.

[30] M. Kühnel, L. S. Mayorga, T. Dandekar, J. Thakar, R. Schwarz, E. Anes, G. Griffiths, and J. Reich. Modelling phagosomal lipid networks that regulate actin assembly. *BMC Systems Biology*, 2:107, 2008.

[31] H. Kurata, H. El-Samad, T.-M. Yi, M. Khammash, and J. C. Doyle. Feedback regulation of the heat shock response in E. coli. In *Proceedings of the 40th IEEE Conference on Decision and Control*, pages 837–842, 2001.

[32] Y. Lazebnik. Can a Biologist Fix a Radio? – or, What I Learned while Studying Apoptosis. *Cancer Cell*, 2(3):179–182, 2002.

[33] S. Lindquist and E. A. Craig. The heat-shock proteins. *Annual Review of Genetics*, 22:631–677, 1988.

[34] O. Lipan, J.-M. Navenot, Z.Wang, L. Huang, and S. Peiper. Heat shock response in cho mammalian cells is controlled by a nonlinear stochastic process. *PLoS Computational Biology*, 3(10):1859–1870, 2007.

[35] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[36] P. Mendes and D. Kell. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, 14(10):869–883, 1998.

[37] A. Mizera, E. Czeizler, and I. Petre. Methods for biochemical decomposition and quantitative submodel comparison. *Israel Journal of Chemistry*, 51:151–164, 2011.

[38] A. Mizera, E. Czeizler, and I. Petre. Self-assembly models of variable resolution. *(submitted)*, 2011.

[39] C. G. Moles, P. Mendes, and J. R. Banga. Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Res*, 13:2467–2474, 2003.

[40] A. Peper, C. Grimbergent, J. Spaan, J. Souren, and R. van Wijk. A mathematical model of the hsp70 regulation in the cell. *International Journal of Hyperthermia*, 14:97–124, 1997.

[41] I. Petre, C. L. Hyder, A. Mizera, A. Mikhailov, J. E. Eriksson, L. Sistonen, and R.-J. Back. A simple mathematical model for the eukaryotic heat shock response. *submitted*, 2009.

[42] I. Petre, A. Mizera, C. L. Hyder, A. Meinander, A. Mikhailov, R. I. Morimoto, L. Sistonen, J. E. Eriksson, and R.-J. Back. A simple mass-action model for the eukaryotic heat shock response and its mathematical validation. *Natural Computing*, 2010. doi: http://dx.doi.org/10.1007/s11047-010-9216-y.

[43] I. Petre, A. Mizera, C. L. Hyder, A. Mikhailov, J. E. Eriksson, L. Sistonen, and R.-J. Back. A new mathematical model for the heat shock response. In A. Condon, D. Harel, J. N. Kok, A. Salomaa, and E. Winfree, editors, *Algorithmic Bioprocesses*, Natural Computing Series, pages 411–425. Springer, 2009.

[44] T. Pfeiffer, I. Sanchez-Valdenebro, J. C. Nuno, F. Montero, and S. Schuster. METATOOL: for studying metabolic netwroks. *Bioinformatics*, 15:251–257, 1999.

[45] A. G. Pockley. Heat shock proteins as regulators of the immune response. *The Lancet*, 362(9382):469–476, 2003.

[46] M. V. Powers and P. Workman. Inhibitors of the heat shock response: Biology and pharmacology. *FEBS Letters*, 581(19):3758–3769, 2007.

[47] T. R. Rieger, R. I. Morimoto, and V. Hatzimanikatis. Mathematical modeling of the eukaryotic heat shock response: Dynamics of the hsp70 promoter. *Biophysical Journal*, 88(3):1646–58, 2005.

[48] R. G. C. L. J. P. N. S. M. S. L. X. P. M. S. Hoops, S. Sahle and U. Kummer. COPASI–a COmplex PAthway SImulator. *Bioinformatics*, 22:3067–3074, 2006.

[49] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons Ltd, Chichester, England, 2004.

[50] A. M. Sapozhnikov, G. A. Gusarova, E. D. Ponomarev, and W. G. Telford. Translocation of cytoplasmic HSP70 onto the surface of EL-4 cells during apoptosis. *Cell Proliferation*, 35(4):193–206, 2002.

[51] M. A. Savageau. Biochemical systems analysis: I. Some mathematical properties of the rate law for the component enzymatic reactions. *Journal of Theoretical Biology*, 25(3):365–369, 1969.

[52] M. A. Savageau. Biochemical systems analysis: II. The steady state solution for an n-pool system using a power law approximation. *Journal of Theoretical Biology*, 25(3):370–379, 1969.

[53] M. A. Savageau. Biochemical systems analysis: III. Dynamic solutions using a power-law approximation. *Journal of Theoretical Biology*, 26:215–226, 1970.

[54] M. A. Savageau. The behavior of intact biochemical control systems. *Current Topics in Cellular Regulation*, 6:63–130, 1972.

[55] M. A. Savageau. Optimal design of feedback control by inhibition: steady state considerations. *J. Mol. Evol.*, 4:139–156, 1974.

[56] C. H. Schilling, S. Schuster, B. O. Palsson, and R. Heinrich. Metabolic Pathway Analysis: Basic Concepts and Scientific Applications in the Post-genomic Era. *Biotechnological Progress*, 15(3):296–303, 1999.

[57] S. Schuster, T. Dandekar, and D. A. Fell. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in biotechnology*, 17(2):53–60, 1999.

[58] S. Schuster, D. A. Fell, and T. Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, 18:326–332, 2000.

[59] S. Schuster, C. Hilgetag, J. H. Woods, and D. A. Fell. Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. *Journal of Mathematical Biology*, 45(2):153–181, 2002.

[60] E. D. Sontag. Some new directions in control theory inspired by systems biology. *IEE Systems Biology*, 1(1):9–18, 2004.

[61] E. D. Sontag. Molecular systems biology and control. *European Journal of Control*, 11(4):396435, 2005.

[62] R. Srivastava, M. Peterson, and W. Bentley. Stochastic kinetic analysis of the escherichia coli stres circuit using σ^{32}-targeted antisense. *Biotechnology and Bioengineering*, 75(1):120–129, 2001.

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

[64] C. J. Tomlin and J. D. Axelrod. Understanding biology by reverse engineering the control. *PNAS*, 102(12):4219–4220, 2005.

[65] R. Voellmy and F. Boellmann. Chaperone regulation of the heat shock protein response. *Advances in Experimental Medicine and Biology*, 594:89–99, 2007.

[66] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2), 1976.

[67] O. Wolkenhauer. Systems biology: The reincarnation of systems theory applied in biology? *Briefings in Bioinformatics*, 2(3):258–270, 2001.

Reaction	(Reaction number)
$2\,\mathsf{hsf} \leftrightarrow \mathsf{hsf}_2$	(2)
$\mathsf{hsf} + \mathsf{hsf}_2 \leftrightarrow \mathsf{hsf}_3$	(3)
$\mathsf{hsf}_3 + \mathsf{hse} \leftrightarrow \mathsf{hsf}_3\colon \mathsf{hse}$	(4)
$\mathsf{hsf}_3\colon \mathsf{hse} \rightarrow \mathsf{hsf}_3\colon \mathsf{hse} + \mathsf{hsp}$	(5)
$\mathsf{hsp} + \mathsf{hsf} \leftrightarrow \mathsf{hsp}\colon \mathsf{hsf}$	(6)
$\mathsf{hsp} + \mathsf{hsf}_2 \rightarrow \mathsf{hsp}\colon \mathsf{hsf} + \mathsf{hsf}$	(7)
$\mathsf{hsp} + \mathsf{hsf}_3 \rightarrow \mathsf{hsp}\colon \mathsf{hsf} + 2\,\mathsf{hsf}$	(8)
$\mathsf{hsp} + \mathsf{hsf}_3\colon \mathsf{hse} \rightarrow \mathsf{hsp}\colon \mathsf{hsf} + \mathsf{hse} + 2\,\mathsf{hsf}$	(9)
$\mathsf{hsp} \rightarrow$	(10)
$\mathsf{prot} \rightarrow \mathsf{mfp}$	(11)
$\mathsf{hsp} + \mathsf{mfp} \leftrightarrow \mathsf{hsp}\colon \mathsf{mfp}$	(12)
$\mathsf{hsp}\colon \mathsf{mfp} \rightarrow \mathsf{hsp} + \mathsf{prot}$	(13)

Table 1: The list of reactions of the biochemical model for the heat shock response originally introduced in [42].

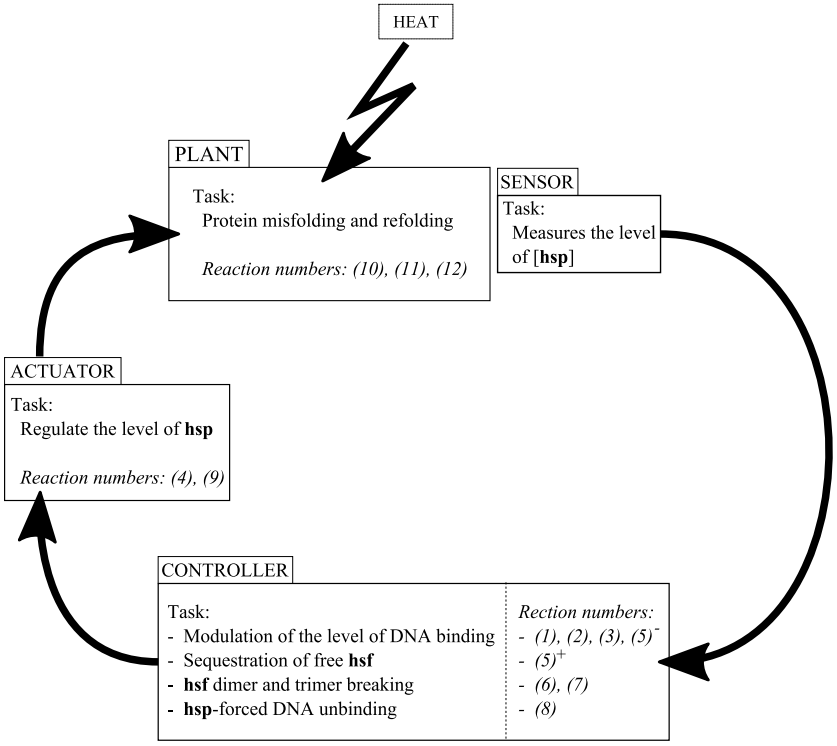


Figure 1: The control-based decomposition of the heat shock response network. The reaction numbers refer to the reactions in Table 1. We denote the 'left-to-right' direction of reaction (6) by $(6)^+$ and by $(6)^-$ its 'right-to-left' direction.

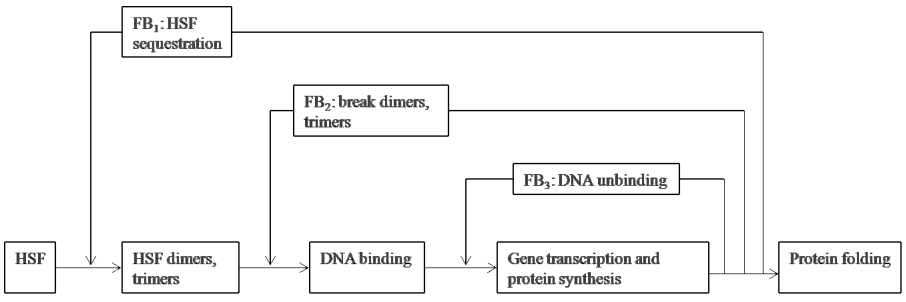
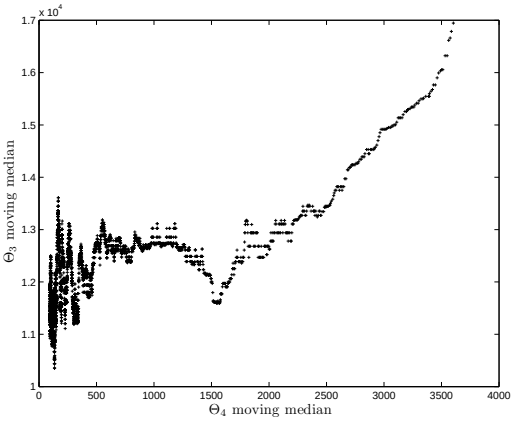
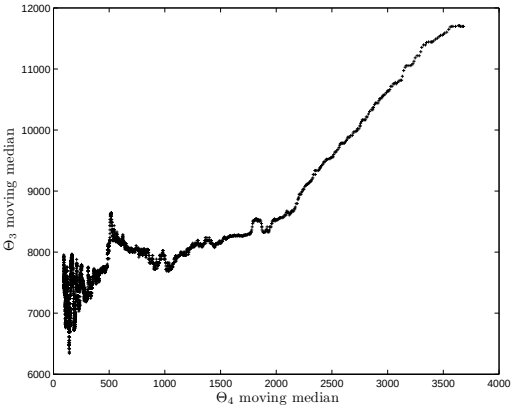


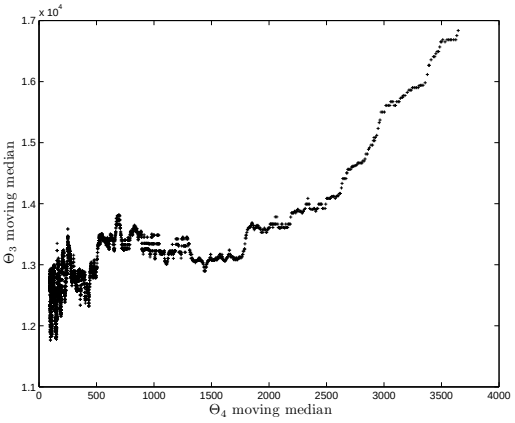
Figure 2: The control structure of the heat shock response network. The three identified feedback loops and their points of interaction with the mainstream process are depicted.


x 10^4
Θ3 moving median
Θ4 moving median

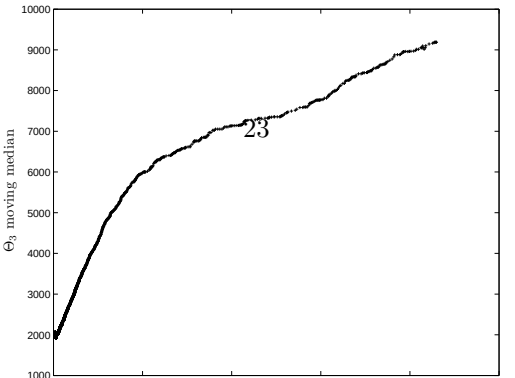
(a) M_1


Θ3 moving median
Θ4 moving median

(b) M_2


x 10^4
Θ3 moving median
Θ4 moving median

(c) M_3


Θ3 moving median

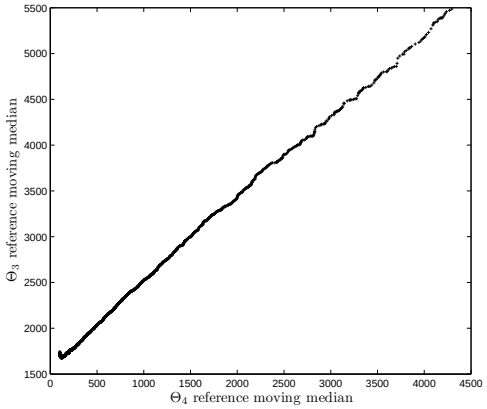


Figure 4: The plot shows the result of applying the moving median technique to the scatter plots of the cost, i.e. Θ_3 vs Θ_4, obtained for the reference architecture. For each sampled vector of parameters, the values of Θ_3 and Θ_4 were computed and plotted against each other. Then, the moving median technique was applied to discern the overall trend in the data depicted in the obtained scatter plot. The window size of the moving median was set to 500 and the sample size of the vectors of parameter values was 10.000.