Methods for biochemical model decomposition and quantitative submodel comparison

Andrzej Mizera* Elena Czeizler* Ion Petre†

Department of Information Technologies,
Åbo Akademi University, FIN-20520 Turku, Finland

amizera, elczeizl, ipetre@abo.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. We describe the knockdown mutants, elementary flux modes, control-based decomposition, mathematically controlled comparison and its extension, local submodels comparison and a discrete approach for comparing continuous submodels. In the second part of the paper we present a new statistical method for comparing submodels that complements the methods presented in the review. The main difference between our approach and the known methods is related to the important question of how to chose the numerical setup in which to perform the comparison. In the case of the reviewed methods, the comparison is made in the numerical context of the reference model, i.e., in each of the alternative models both the kinetics of the reactions and the initial values of all variables are chosen to be identical to those from the reference model. We propose in this paper a different approach, better suited for response networks, where each alternative model is assumed to start from its own steady state under basal conditions. We demonstrate our approach on a case study focusing on the heat shock response in eukaryotes.

*Authors with equal contribution.

†Address correspondence to: Ion Petre, Department of Information Technologies, Åbo Akademi University, Joukahaisenkatu 3-5 A, FIN-20520 Turku, Finland. Phone: +358 2 215 3361; Fax: +358 2 215 4732; E-mail: ipetre@abo.fi

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 behaviour. This often results in the creation and analysis of very large and complex models, often encompassing hundreds of reactions and reactants, see e.g. [5]. Therefore, obtaining a global picture of the system's architecture, in particular understanding the interactions between various components, or even just distinguishing a high-level functional decomposition of the network, constitutes a significant challenge. An important insight here is that the architecture of some biological systems, e.g. some regulatory networks, is a consequence of functional requirements of the entire system. 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 [39]. 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 [7], 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. [12, 18, 21, 46, 47, 49, 53]. 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 behaviour 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 paper, 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. In the general case it is a very difficult issue and is not in the scope of this study.

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 and its extension, local submodels comparison and a discrete approach for comparing continuous submodels. In the second part of the paper we introduce a new approach to quantitative submodel comparison. A main difference in our approach with respect to previous methods is that we allow the alternative models to start from different initial states, rather than to assume the initial state of the reference model. We argue that this is a better approach at least in the case of response networks, where the system is assumed to be in a steady state under basal conditions and exhibit a response only as an effect of an external trigger. To treat each model as a genuine alternative for the biological process under study, we allow it to

start from its own steady state under basal conditions. Finally, we illustrate our approach on a case study focusing on the heat shock response in eukaryotes.

The numerical behaviour of any model is clearly sensitive to the numerical setup, i.e., the numerical values of the kinetic constants and of the initial values of the model variables. In our approach for quantitative comparison of alternative submodels we adopt some statistical, parameter-independent methods introduced in [1, 2]. These methods aim to sample the numerical behaviour of the model through a sampling of the parameter space. We adopt in this paper the latin hypercube sampling method of [14] that gives uniformly distributed samples over each parameter, of size independent of the number of parameters. We briefly survey this method and apply it to the heat shock response in eukaryotes.

The heat shock response is an evolutionary conserved mechanism protecting the cell against protein misfolding. In the case study for our new approach to quantitative submodel comparison we consider a model recently introduced in [28]. The model was analysed in [8] 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 behaviour of the model. A local, point-wise comparison of the three feedbacks was already done in [8], in the kinetic setup of the reference model. In this paper we do a global, parameter-independent analysis of the numerical role of each feedback, through a sampling of the whole parameter space.

2 Methods for model decomposition

2.1 Knockdown mutants

A simple model decomposition consists of isolating a single process or mechanism in the considered system. In this way the model is split into two parts: the first one comprising the process of interest and the second containing all the remaining elements of the system. Although such decomposition might seem unsophisticated, this approach is often very useful in discovering the role of a single mechanism in a larger system. It is widely exploited in reverse engineering, a process aiming at revealing the technological principles of a device, object or system. In Section 3 we shortly describe the method of mathematically controlled comparison ([39]), where this simple decomposition approach is at the basis of the method.

2.2 Elementary flux modes

Another well-established decomposition method for biochemical models appears in the context of the analysis of metabolic pathways. It is not easy to define a pathway in a given metabolic network. An intuitive definition of a pathway is a sequence of reactions linked by common metabolites ([19]). Examples of metabolic pathways are glycolysis or amino acid synthesis. Discovering new pathways in a large model driven only by biological intuition is even more difficult. An attempt to formalize the notion of pathway has been proposed in [13, 30, 41, 42, 43, 44] 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 [13, 19, 41, 42, 43, 44] 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 ([30]). The recognition 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 ([41, 42, 43]). 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 ([41, 42, 43]). 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 recognition of the main functional modules of a system and their individual contribution to the emergent, complex behaviours 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. In particular, by applying 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 behaviour 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 [7, 10, 18, 46, 47]. Here we shortly

describe the approach taken in [10]. 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 both feedbacks. Moreover, the case without the feedforward element is also considered, see [10] for details. 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, thereby 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 behaviour of the system. In consequence the authors succeed to perform an in-depth comparison between different model variants.

3 Known methods for submodel comparison

Comparing alternative models for a given biochemical system is in general a very difficult problem, involving a deep analysis of both the underlying network of reactions, the biological assumptions as well as 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, [39], 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, [36], [37], and [38]. This canonical nonlinear representation, developed within the power-law formalism, is a system of non-linear ordinary differential equations with a well-defined structure. Moreover, this framework allows the alternative models to differ from the reference design in only one process, e.g., the existence or not of some feedback mechanisms, which is actually the focuss of the comparison. Then, in each of the alternative models one sets the numerical values of the parameters to be identical with 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 alterna-

tives. 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, [16], [40], in gene circuits, [15], in immune networks, [4]. 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 [2], the method of mathematically controlled comparison was extended to include some statistical methods, [1], [3], that allow the use of numerical values for the parameters while still preserving the generality of the conclusions.

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 usually for biological systems the exact statistical distribution of the parameters values is not known, 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 (more or less sophisticated) random samplings to some systematic deterministic scanning methods, see e.g. [34]. 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 *behavioural class* consists of systems that exhibit a particular systemic behaviour, e.g., exhibiting a steady state behaviour under given conditions, or low concentrations of intermediary products, or small values for the parameter sensitivity, see, e.g., [3]. 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 behaviour 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. Basically, the density plot can be interpreted as a list of N pairs of values $(P_{reference}, R)$, which can be arranged in a 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 for the first time. 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 submodels comparison

When the alternative designs are actually submodels of the reference architecture, there is also another approach, see [8], for performing the comparison. This is the case when, for instance, one is interested in 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 behaviour 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 [8] 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 discrete approach for comparing continuous submodels

The application of the control-theoretical analysis described in Section 2 enables the identification of the main functional modules, their interconnections and control strategies of a biochemical network. In particular, this approach can

be very useful for identifying the main regulatory components of a biochemical network, including its feed-forward and feedback mechanisms. Then, in order to identify and quantify the exact role of each of these regulatory mechanisms, one usually uses knockdown mutants, see [10], lacking one or more of these components. In particular, the knockdown mutant models are submodels of the reference architecture. The approached proposed in [9], associates to each knockdown mutant a Boolean formula describing its control architecture in the following way. First, a Boolean variable is associated to each of the regulating mechanisms. Then, using the negation and conjunction of Boolean variables, one can write a Boolean formula for each of the knockdown mutants describing which of the regulating mechanisms are present in their architecture. In particular, these Boolean formulas describe a property of the alternative designs which is independent of time, i.e., their regulatory network. Moreover, one can go one step further and write a Boolean formula describing all those mutant architectures that show a given behavioural property, e.g., a high level of a given reactant or a given correlation between two reactants. This formula is actually the conjunction of all Boolean formulas characterizing the architectures of the mutants exhibiting the required property. The numerical comparison of the mutants is then performed by analysing the Boolean formulas associated to various behavioural properties.

4 A new approach for quantitative submodel comparison

Here we propose a new approach for quantitative comparison of biological models. Before presenting the method itself, we clarify the adopted terminology which is used in the description of our new approach. Usually biological models are expressed in terms of biochemical reactions. We will refer to a list of such reactions describing a biological mechanism as its *biochemical model*. From the biochemical model an associated *mathematical model* is derived by choosing one of the two commonly used frameworks: either a deterministic or a stochastic formulation. In the first case the biochemical reaction kinetics rely on the assumption that the reaction rate at a certain point in time and space can be expressed as a unique function of the concentrations of all substances at this point in time and space, see [19]. It is governed by the *mass action law*, which can be shortly summarized as follows: the rate of each reaction is proportional to the product of the reactant masses, with each mass raised to the power equal to the corresponding stoichiometric coefficient ([19]). With this assumption, the mathematical formulation of a biochemical model results in a system of ordinary differential rate equations constituting the associated deterministic mathematical model. In the second case, single molecules and their interactions are considered and the changes in molecular populations are described in terms of stochastic processes. In the stochastic framework the associated mathematical model is a continuous-time Markov chain defined by a chemical master equation describing the time evolution of the probability of the biochemical system to be in a certain state. Our new approach for model comparison is designed and presented for the deterministic framework, however we notice that it can be easily adapted for the stochastic formulation.

As mentioned above, the assumption of the mass action kinetics leads to a system of ordinary differential equations (ODE) constituting the mathematical model. The ODE system contains a certain number of parameters representing the kinetic rate constants of the biochemical reactions. By assigning numerical values to the parameters and setting the initial conditions for the equations, we obtain an *instantiation* of the mathematical model.

Our model comparison method can be outlined as follows. First, having a biochemical model of some biological mechanism, referred to as the *reference model* (or *reference architecture*) of this system, we construct a *submodel* (or *alternative architecture*) by eliminating certain reactions from the list of biochemical reactions of the reference model. In this stage, we can for example apply control-based decomposition techniques to identify a number of modules, and then study them separately by considering a number of knockdown mutants lacking one or more of the modules. Second, the associated mathematical models both for the reference and alternative architecture are formulated. Notice that this procedure assures that all the parameters of the alternative architecture are common with a subset of parameters of the reference model. Next, we perform the statistical sampling of the reference model and mutant behaviour. To this aim we 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. Since, as mentioned above, the alternative architecture contains only a subset of the reference model parameters, only the values of certain coordinates are used when setting the parameters of the submodel. Further, the initial values of the variables of the reference model and 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 (given by numerical values for all variables and parameters 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. The steady states of a model are defined by the values of its parameters and by the initial values of its variables. Now, assuring that both 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 submodel and we run numerical simulations for both of them in order to evaluate their functional effectiveness. Finally, having done this for all sampled vectors, we summarize the obtained results for the variants and compare the models by use of some statistical measures. Moreover, the methodology allows us to consider more than one submodel and thus the obtained results provide a basis for comparison between the different potential architectural designs underlying the analysed biological mechanism.

For the parameter scanning, in the above procedure we use the *latin hypercube sampling* method (LHS), originally introduced in [26]. 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 [14, 26], see also [28] for an example of the application of this sampling method in the context of model identifiability problem.

In the next sections we show how the described method, where the sampling is performed with the LHS approach, can be utilized in the case of a recently introduced mathematical model for the eukaryotic heat shock response. In particular, we present how this method allows to discriminate between different variants of the model and to determine the roles of certain control mechanisms of the response system.

5 Case study

5.1 A biochemical model for the heat shock response

The heat shock response (HSR) is a highly evolutionarily conserved defence mechanism among organisms ([24]). 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 chaperones, 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. [6, 32, 51]. 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 ([10, 11, 20, 50]). 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 [17, 31]. 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. [10, 25, 27, 33, 48]. In this study we consider a recently introduced model of the eukaryotic heat shock response ([28] and [29]). In this model the central role is played by the heat shock proteins (hsp), which act as chaperones for the misfolded proteins (mfp):

the heat shock proteins sequester the misfolded proteins (hsp: 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 (hsf_3: 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 chaperone 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 chaperones. In consequence the response mechanism gets switched on. The full list of biochemical reactions constituting the biochemical model from [28] is presented in Table 1. The model is based only on well-documented reactions without introducing any hypothetical mechanisms or experimentally unsupported biochemical reactions. For a full presentation and discussion of this model we refer the reader to [28].

Based on the assumption of mass-action law for all the reactions (1)-(12) 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 full ODE system is shown in Table 2, where by k_i we denote the reaction rate constant of the irreversible reaction (i) in Table 1, by k_i^+ the rate constant associated with the 'left-to-right' direction of the reversible reaction (i), while k_i^- denotes the rate constant corresponding to its 'right-to-left' direction. By T we denote the numerical value of the temperature of the environment in Celsius degrees. The rate coefficient of protein misfolding with respect to the temperature ($\varphi(T)$) in reaction (10) is given by the following formula:

$$\varphi(T) = (1 - \frac{0.4}{e^{T-37}}) \cdot 1.4^{T-37} \cdot 1.45 \cdot 10^{-5}\ s^{-1},$$

which is valid for T in the range from 37 to 45. The formula was obtained based on experimental investigations described in [22, 23], was originally proposed in [27] and adapted for use in the mathematical model of HSR in [28]. The mathematical model comprises 16 independent kinetic parameters and 10 initial conditions. In the case of our method we do not fix the parameter values as was done in [28]: we neither fit nor validate the model with respect to experimental data. Instead, we sample the HSR model behaviour by randomly choosing different sets of parameter values. This results not in one, but in a collection

of instances of the HSR model. Notice that in the process of obtaining these instances no experimental data are considered. Thus, the instances are not required to confirm any experimental results. We discuss in details how the parameter values for the HSR model are obtained in subsection 5.4.

5.2 Control-based decomposition

In [8] 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 chaperones, 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 chaperones. 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 $(5)^+$ (the 'left-to-right' direction of reaction (5));
- FB2: breaking of hsf dimers and trimers, i.e. reactions (6) and (7);
- FB3: unbinding of hsf_3 from hse and breaking the trimers, i.e. reaction (8).

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.

5.3 The knockdown mutants

In [8] and [9] 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 (1)-(4), (9)-(12) and, in the terminology of control theory, is characterized by the *open-loop design*;
- M_1 is determined by reactions (1)-(5), (9)-(12);
- M_2 is determined by reactions (1)-(4), (6)-(7), (9)-(12), and the 'right-to-left' direction of reaction (5);
- M_3 is determined by reactions (1)-(4), (8)-(12), and the 'right-to-left' direction of reaction (5);
- $M_{1,2}$ is determined by reactions (1)-(7), (9)-(12);
- $M_{1,3}$ is determined by reactions (1)-(5), (8)-(12);

- $M_{2,3}$ is determined by reactions (1)-(4), (6)-(12), and the 'right-to-left' direction of reaction (5);
- $M_{1,2,3}$ is the reference architecture consisting of all reactions (1)-(12).

5.4 Statistical sampling of the mutant behaviour

We apply our model comparison method described in Section 3 to the presented model of eukaryotic heat shock response in order to investigate the functional role of the feedback mechanisms. It is easy to see that M_0 is non-responsive: starting from a steady state at physiological conditions, i.e. 37 °C, M_0 shows no increase in DNA binding for any arbitrarily high temperature, see [9]. We remove M_0 from further considerations. In our study we analyse the six knock-down mutants M_1, M_2, M_3, $M_{1,2}$, $M_{1,3}$ and $M_{2,3}$ as the variants of the reference architecture $M_{1,2,3}$. Our comparison method is applied in the following way. First, a sample of 10.000 vectors of parameter values for the reference architecture is obtained by the latin hypercube sampling described above. In our case 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 [35]. We choose here their half-life 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 ($M_{1,2,3}$). It follows from the construct of the mutant 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 results in a trend curve summarizing the data of the scatter plot and revealing the overall dependency between the considered quantities. Analogical 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.

5.5 Results

In our analysis of the obtained results we assume 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.

We base our analysis of the architecture properties of the six mutants with respect to the reference architecture on the following plots: R_1 vs Θ_1^{r}, R_2 vs Θ_2^{r}, Θ_3 vs Θ_4 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 chaperones 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 7. The trend lines of all mutants, despite some more or less pronounced fluctuations in the region of small Θ_4 values, can be seen as increasing (Figure 5), which is in agreement with our characterization of the heat shock response.

Considering the three mutants with only one feedback, i.e. M_1, M_2 and M_3, we observe that the mfp level peak value in the first 30 min. of heat shock is smaller than in the reference architecture: the ratio R_1 in Figure 3(a), 3(b) and 3(c) is always smaller than 1. This is especially pronounced in mutant instances obtained with samples characterized by providing high mfp peak values in the case of the reference architecture. However, for all these mutants the cost is definitely higher than in the reference architecture, compare Figure 5(a), 5(b),

5(c) with Figure 7. Notice also that the M_2 mutant is more economic in terms of cost than the two other mutants with only one feedback.

In the mutants $M_{1,2}$, $M_{1,3}$ and $M_{2,3}$ the mfp level also peaks at a lower value than in the reference case, although this time the $M_{1,3}$ and $M_{2,3}$ mutants have the cost comparable with the one of the reference architecture. Both $M_{1,3}$ and $M_{2,3}$ reveal the same linear relationship between the average amounts of hsp and mfp as is observable in the reference case, however in both cases the trend line is slightly shifted upwards with respect to the reference. This indicates that the mutants have a tendency to keep a bit higher than the reference amount of hsp with a certain amount of misfolded proteins (Figure 5(e), 5(f) and Figure 7). The same is true also for the $M_{1,2}$ mutant. Although it admits an order of magnitude larger range of observable average amounts of misfolded proteins (Figure 5(d)), the cost plot restricted to Θ_4 moving median values less than or equal 4000 is basically identical with the cost plots of the two other mutants, see Figure 6(a).

Another thing which we observe for the three mutants missing one feedback is that the samples characterized by significant increase in DNA-binding in the reference architecture, i.e. by 15 and more, span a wide range of possible behaviours in the mutants: from almost no DNA-binding increase to an increase comparable with the one observed for the reference architecture. This feature is clearly visible in Figure 4(e) and 4(f) for the mutants $M_{1,3}$ and $M_{2,3}$, respectively. In the case of the $M_{1,2}$ mutant we need to zoom in Figure 4(d). To this aim we observe in the scatter plot R_2 vs Θ_2^r for the $M_{1,2}$ mutant that all points with $R_2 > 1000$ are concentrated in the range $[0, 0.0307]$ of Θ_2^r values (not shown). We exclude all samples with Θ_2^r in this range, irrespective of the R_2 value they admit in the mutant. All in all 2247 samples are filtered out and we apply the moving median technique to the remaining ones. The resulting plot is shown in Figure 6(b). It clearly illustrates that the discussed feature is also a characteristic of the $M_{1,2}$ mutant. This is not true for the three mutants with only one feedback. In these cases we do not observe any substantial increase in the DNA-binding with respect to the steady-state levels at 37 °C for samples which generate such increase in the reference case (Figure 4(a), 4(b) and 4(c)).

On the basis of the presented results, we notice 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 [8], 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 [8]. 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 behaviour 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.

6 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 and a discrete approach for comparing continuous submodels. In the second part of the paper we present a new statistical method for comparing submodels that complements the methods presented in the review. When choosing the initial setup for the alternative submodels, i.e., the initial values of all variables, one approach is to take them from the reference model. This approach is based on the technique of mathematically controlled comparison, [39], see also [10] and [52] for some case studies using this method. However, in the case of biological systems this approach may lead to biased conclusions. For instance, regulatory networks exhibit a steady state behaviour in the absence of stimulus. In general for the reference model, the initial values of the variables are chosen such that it exhibits a steady state behaviour in the absence of a trigger. However, this will not insure that also the submodels of the reference model will exhibit the same property if they start from the same initial values. Thus, the dynamical behaviours of the considered submodels will exhibit the interwind influences of two tendencies: the migration from a (possibly) unstable state and the response to the stimulus. In this context, an analysis of the efficiency of the response and the robustness of the alternative models may lead to erroneous conclusions. In alternative, we propose in this paper to chose the initial values in such a way that each alternative design starts from its own steady state. Our main motivation for this is that we considered all submodels to be viable alternatives for the biological system and, as such, they should exhibit (some of) its main properties. Regarding the values of the kinetic parameters in each of the alternative submodes, there are several approaches in the literature. In the mathematically controlled comparison approach, the values of the kinetic parameters in each of the alternative designs are uniquely determined from the parameters of the reference model, see e.g., [10], [39]. Another approach is to chose in each alternative submodel independent values for the kinetic parameters, e.g., through parameter estimation and validation against experimental data, see e.g., [8]. However, restricting to some particular values for the kinetic rate constants, will also confine the conclusions of our analysis to that particular system. Instead, we take the approach proposed in [1] and [2] and we sample a large set of parameter values from a given range of values. Then we use some statistical techniques to analyse various properties of a general class of systems

which includes the considered system. In particular, for each sampled parameter vector, various functional effectiveness measures are computed both in the reference and in the alternative models. Then by analysing both the density of ratios plots and the moving median plots one can identify and quantify the differences in the dynamical behaviours of the considered models, see e.g., [3], [45] for some case studies where these methods were applied.

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] 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.

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

[9] El. Czeizler, A. Mizera, and I. Petre. A Boolean logic-based approach for comparing biomodels. *submitted*, 2010.

[10] 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.

[11] 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.

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

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

[14] 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.

[15] 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.

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

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

[18] H. Kitano. Systems Biology: A Brief Overview. *Science*, 295:1662–1664, 2002.

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

[20] 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.

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

[22] J. R. Lepock, H. E. Frey, and K. P. Ritchie. Protein denaturation in intact hepatocytes and isolated cellular organelles during heat shock. *The Journal of Cell Biology*, 122(6):1267–1276, 1993.

[23] J. R. Lepock, H. E. Frey, A. M. Rodahl, and J. Kruuv. Thermal analysis of CHL V79 cells using differential scanning calorimetry: Implications for hyperthermic cell killing and the heat shock response. *Journal of Cellular Physiology*, 137(1):14–24, 1988.

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

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

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

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

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

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

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

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

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

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] J. H. Schwacke and E. O. Voit. Improved methods for the mathematically controlled comparison of biochemical systems. *Theoretical Biology & Medical Modelling*, 1(1), 2004.

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

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

[48] 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.

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

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

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

[52] M. E. Wall, W. S. Hlavacek, and M. A. Savageau. Design principles for regulator gene expression in a repressible gene circuit. *Journal of Molecular Biology*, 332(4):861–876, 2003.

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

Reaction	(Reaction number)
$2\,\mathsf{hsf} \leftrightarrow \mathsf{hsf}_2$	(1)
$\mathsf{hsf} + \mathsf{hsf}_2 \leftrightarrow \mathsf{hsf}_3$	(2)
$\mathsf{hsf}_3 + \mathsf{hse} \leftrightarrow \mathsf{hsf}_3\colon \mathsf{hse}$	(3)
$\mathsf{hsf}_3\colon \mathsf{hse} \rightarrow \mathsf{hsf}_3\colon \mathsf{hse} + \mathsf{hsp}$	(4)
$\mathsf{hsp} + \mathsf{hsf} \leftrightarrow \mathsf{hsp}\colon \mathsf{hsf}$	(5)
$\mathsf{hsp} + \mathsf{hsf}_2 \rightarrow \mathsf{hsp}\colon \mathsf{hsf} + \mathsf{hsf}$	(6)
$\mathsf{hsp} + \mathsf{hsf}_3 \rightarrow \mathsf{hsp}\colon \mathsf{hsf} + 2\,\mathsf{hsf}$	(7)
$\mathsf{hsp} + \mathsf{hsf}_3\colon \mathsf{hse} \rightarrow \mathsf{hsp}\colon \mathsf{hsf} + \mathsf{hse} + 2\,\mathsf{hsf}$	(8)
$\mathsf{hsp} \rightarrow$	(9)
$\mathsf{prot} \rightarrow \mathsf{mfp}$	(10)
$\mathsf{hsp} + \mathsf{mfp} \leftrightarrow \mathsf{hsp}\colon \mathsf{mfp}$	(11)
$\mathsf{hsp}\colon \mathsf{mfp} \rightarrow \mathsf{hsp} + \mathsf{prot}$	(12)

Table 1: The list of reactions of the biochemical model for the heat shock response originally introduced in [28].

Equation	(Equation number)

$$d[\mathsf{hsf}]/dt = -2k_1^+[\mathsf{hsf}]^2 + 2k_1^-[\mathsf{hsf}_2] - k_2^+[\mathsf{hsf}][\mathsf{hsf}_2] + k_2^-[\mathsf{hsf}_3] - k_5^+[\mathsf{hsf}][\mathsf{hsp}] + k_5^-[\mathsf{hsp{:}\ hsf}] + k_6[\mathsf{hsf}_2][\mathsf{hsp}] + 2k_7[\mathsf{hsf}_3][\mathsf{hsp}] + 2k_8(\mathsf{hsf_3{:}\ hse})\,\mathsf{hsp} \tag{13}$$

$$d[\mathsf{hsf}_2]/dt = k_1^+[\mathsf{hsf}]^2 - k_1^-[\mathsf{hsf}_2] - k_2^+[\mathsf{hsf}][\mathsf{hsf}_2] + k_2^-[\mathsf{hsf}_3] - k_6[\mathsf{hsf}_2][\mathsf{hsp}] \tag{14}$$

$$d[\mathsf{hsf}_3]/dt = k_2^+[\mathsf{hsf}][\mathsf{hsf}_2] - k_2^-[\mathsf{hsf}_3] - k_3^+[\mathsf{hsf}_3][\mathsf{hse}] + k_3^-[\mathsf{hsf_3{:}\ hse}] - k_7[\mathsf{hsf}_3][\mathsf{hsp}] \tag{15}$$

$$d[\mathsf{hse}]/dt = -k_3^+[\mathsf{hsf}_3][\mathsf{hse}] + k_3^-[\mathsf{hsf_3{:}\ hse}] + k_8[\mathsf{hsf_3{:}\ hse}][\mathsf{hsp}] \tag{16}$$

$$d[\mathsf{hsf_3{:}\ hse}]/dt = k_3^+[\mathsf{hsf}_3][\mathsf{hse}] - k_3^-[\mathsf{hsf_3{:}\ hse}] - k_8[\mathsf{hsf_3{:}\ hse}][\mathsf{hsp}] \tag{17}$$

$$d[\mathsf{hsp}]/dt = k_4[\mathsf{hsf_3{:}\ hse}] - k_5^+[\mathsf{hsf}][\mathsf{hsp}] + k_5^-[\mathsf{hsp{:}\ hsf}] - k_6[\mathsf{hsf}_2][\mathsf{hsp}] - k_7[\mathsf{hsf}_3][\mathsf{hsp}] - k_8[\mathsf{hsf_3{:}\ hse}][\mathsf{hsp}] - k_{11}^+[\mathsf{hsp}][\mathsf{mfp}] + (k_{11}^- + k_{12})[\mathsf{hsp{:}\ mfp}] - k_9[\mathsf{hsp}] \tag{18}$$

$$d[\mathsf{hsp{:}\ hsf}]/dt = k_5^+[\mathsf{hsf}][\mathsf{hsp}] - k_5^-[\mathsf{hsp{:}\ hsf}] + k_6[\mathsf{hsf}_2][\mathsf{hsp}] + k_7[\mathsf{hsf}_3][\mathsf{hsp}] + k_8[\mathsf{hsf_3{:}\ hse}][\mathsf{hsp}] \tag{19}$$

$$d[\mathsf{mfp}]/dt = \phi_T[\mathsf{prot}] - k_{11}^+[\mathsf{hsp}][\mathsf{mfp}] + k_{11}^-[\mathsf{hsp{:}\ mfp}] \tag{20}$$

$$d[\mathsf{hsp{:}\ mfp}]/dt = k_{11}^+[\mathsf{hsp}][\mathsf{mfp}] - (k_{11}^- + k_{12})[\mathsf{hsp{:}\ mfp}] \tag{21}$$

$$d[\mathsf{prot}]/dt = -\phi_T[\mathsf{prot}] + k_{12}[\mathsf{hsp{:}\ mfp}] \tag{22}$$

$$d[\mathsf{prot}]/dt = -\phi_T[\mathsf{prot}] + k_{12}[\mathsf{hsp{:}\ mfp}] \tag{23}$$

Table 2: The system of differential equations of the mathematical model associated with the biochemical model in Table 1.

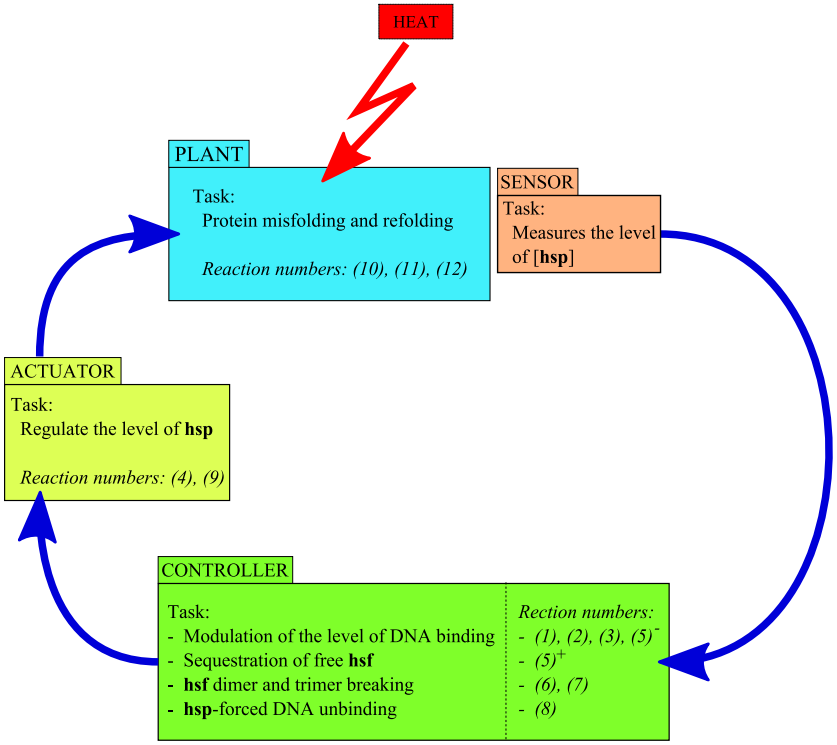


Figure 1: The control-based decomposition of the heat shock response network. The reaction numbers refer to the reactions in Table 1. We denote the 'left-to-right' direction of reaction (5) by $(5)^+$ and by $(5)^-$ its 'right-to-left' direction.

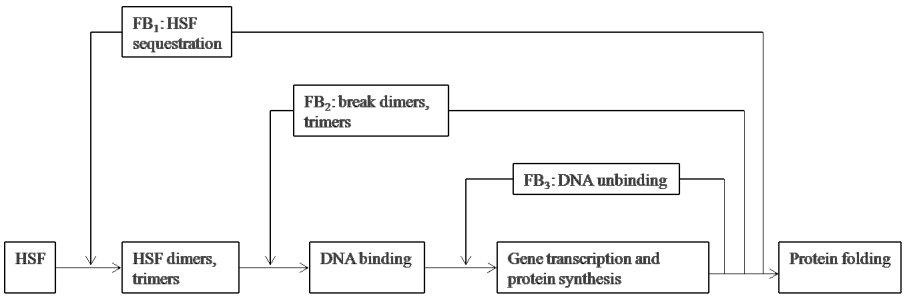
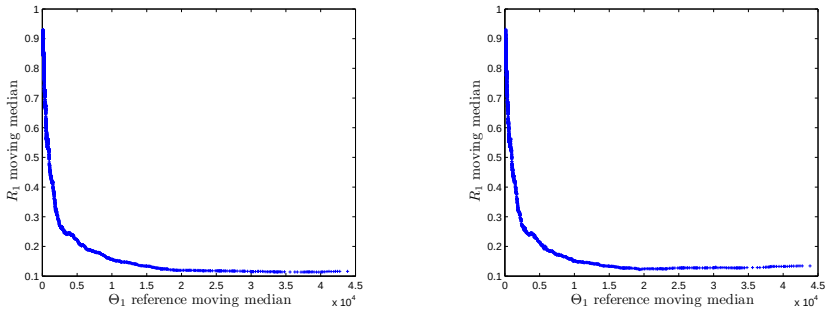
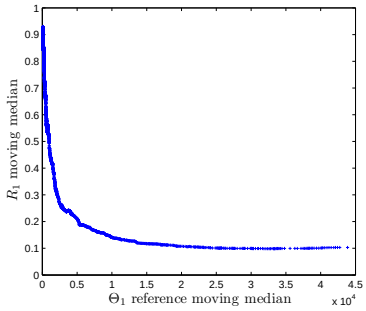


Figure 2: The control structure of the heat shock response network. The three identified feedback loops and their points of interaction with the mainstream process are depicted.

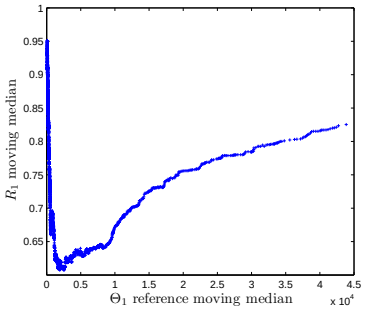


(a) M_1

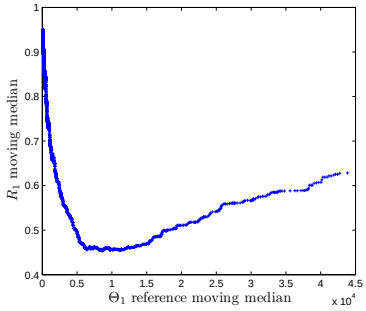
(b) M_2



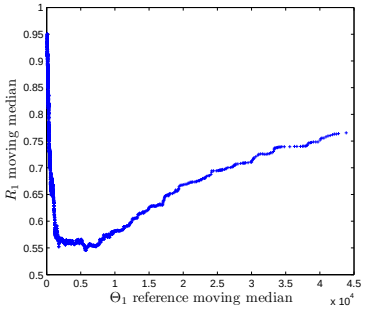
(c) M_3



(d) $M_{1,2}$

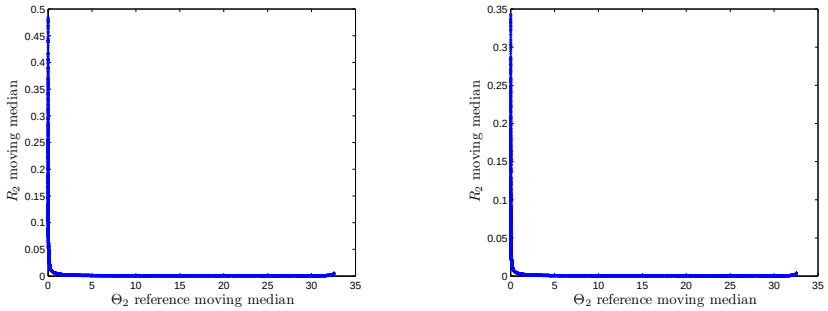


(e) $M_{1,3}$



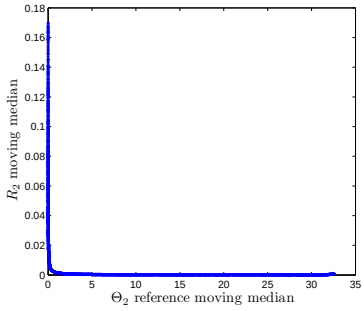
(f) $M_{2,3}$

Figure 3: The plots show the result of applying the moving median technique to the scatter plots of R_1 vs Θ_1^r obtained individually for each of the six considered mutants. For each mutant and each sampled vector of parameters, the value of R_1 was computed and plotted against the value of Θ_1 obtained for the reference architecture with the same parameter vector. Then, the moving median technique was applied to discern the overall trend in the data depicted in the obtained scatter plots. The window size of the moving median was set to 500 and the sample size of the vectors of parameter values was 10.000.

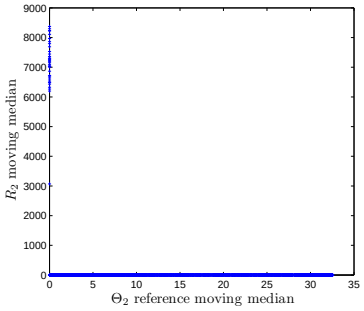


(a) M_1

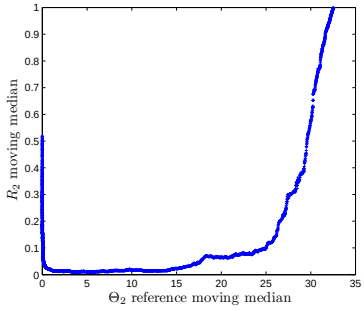
(b) M_2



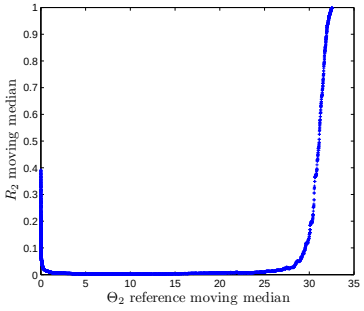
(c) M_3



(d) $M_{1,2}$

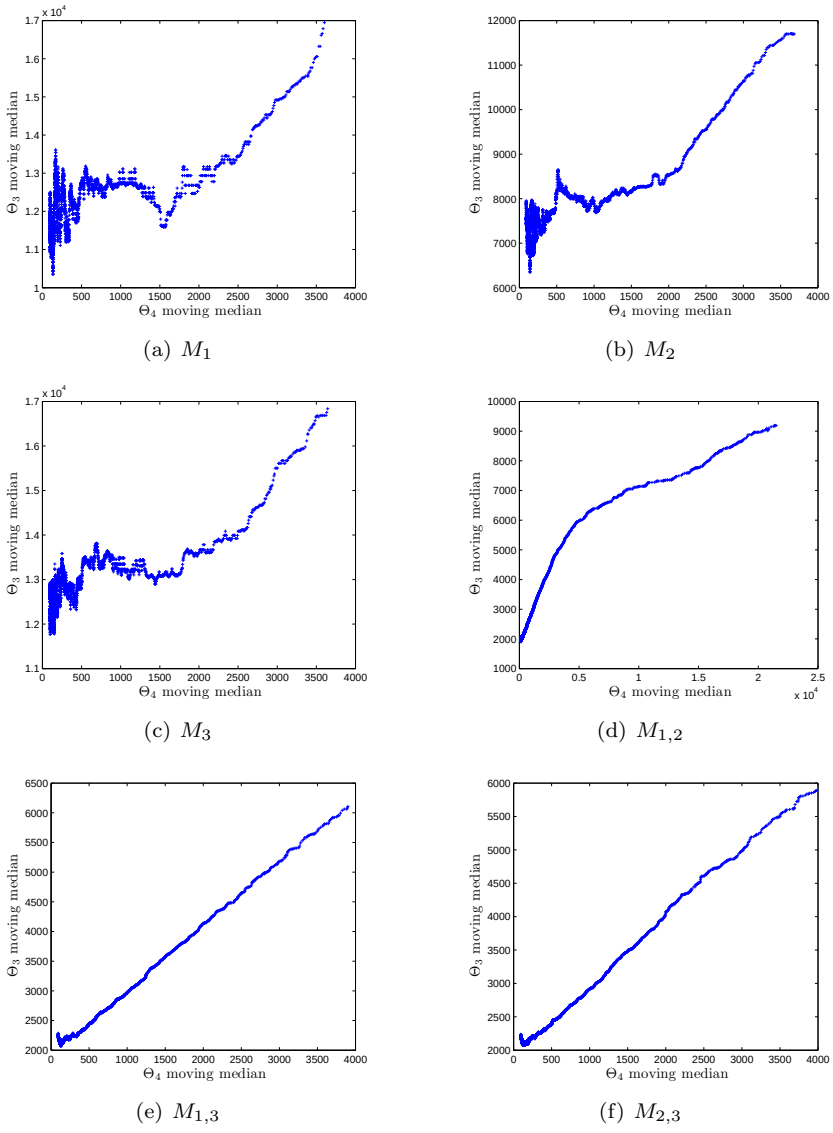


(e) $M_{1,3}$



(f) $M_{2,3}$

Figure 4: The plots show the result of applying the moving median technique to the scatter plots of R_2 vs Θ_2^r obtained individually for each of the six considered mutants. For each mutant and each sampled vector of parameters, the value of R_2 was computed and plotted against the value of Θ_2 obtained for the reference architecture with the same parameter vector. Then, the moving median technique was applied to discern the overall trend in the data depicted in the obtained scatter plots. The window size of the moving median was set to 500 and the sample size of the vectors of parameter values was 10.000.

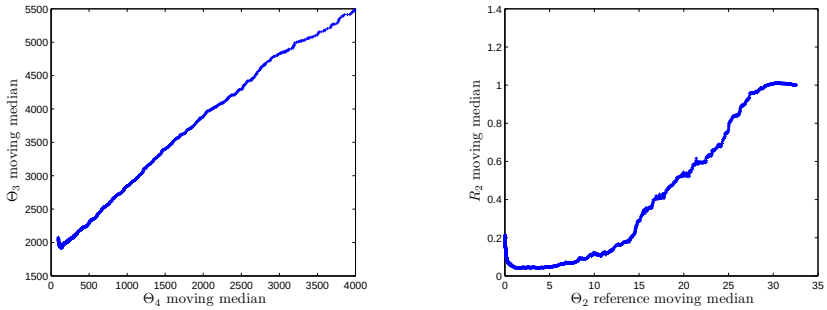


(a) M_1 (b) M_2

(c) M_3 (d) $M_{1,2}$

(e) $M_{1,3}$ (f) $M_{2,3}$

Figure 5: The plots show the result of applying the moving median technique to the scatter plots of the cost, i.e. Θ_3 vs Θ_4, obtained individually for each of the six considered mutants. For each mutant and 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 plots. The window size of the moving median was set to 500 and the sample size of the vectors of parameter values was 10.000.



(a) $M_{1,2}$ (Θ_4 moving median ≤ 4000) (b) $M_{1,2}$ (without samples with $\Theta_2^r \leq 0.0307$)

Figure 6: (a) The zoomed in version of Figure 5(d) where Θ_4 moving median is not greater than 4000. (b) A version of Figure 4(d) where samples with $\Theta_2^r \leq 0.0307$ were not considered. It shows that the samples characterized by significant increase in DNA-binding in the reference architecture (by 15 and more) span a wide range of possible behaviours in the $M_{1,2}$ mutant: from almost no DNA-binding increase (the moving median of $R_2 = 0.2$) to an increase comparable with the one observed for the reference architecture (the moving median of $R_2 \geq 1$).

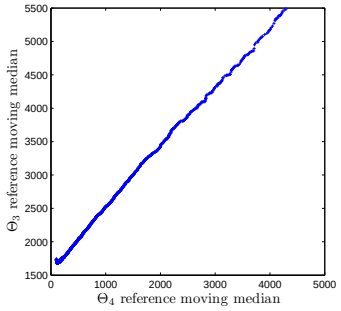


Figure 7: The plot shows the result of applying the moving median technique to the scatter plots of the cost, i.e. Θ_3 vs Θ_4, obtained for the reference architecture. For each sampled vector of parameters, the values of Θ_3 and Θ_4 were computed and plotted against each other. Then, the moving median 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.