A Boolean approach for disentangling the roles of submodules to the global properties of a biomodel

Elena Czeizler* Andrzej Mizera† Ion Petre‡

Abstract

To disentangle the numerical contribution of modules to the system-level behavior of a given biomodel, one often considers knock-out mutant models, investigating the change in the model behavior when modules are systematically included and excluded from the model architecture in all possible ways. We propose in this paper a Boolean approach for extracting conclusions about the role of each module from the systematic comparison of the numerical behavior of all knock-out mutants. We associate a Boolean variable to each module, expressing when the module is included in the architecture and when it is not. We can then express the satisfiability of system-level properties of the full model, such as efficiency, or economical use of resources, in terms of a Boolean formula expressing in a compact way which model architectures, i.e., which combinations of modules, give rise to the desired property. We demonstrate this method on a recently proposed computational model for the heat shock response in eukaryotes. We describe the contribution of each of its three feedback loops towards achieving an economical and effective heat shock response.

1 Introduction

There is a sustained experimental and computational effort nowadays towards building large, system-level models for biochemical processes, including regulatory networks, signaling pathways, metabolic pathways, etc. Models can encompass thousands of reactants and reactions, see [2]. On this scale, understanding the details of the network, especially its regulatory mechanisms, becomes a considerable challenge. Applying a control-theoretical analysis to a biological system can provide a systematic way to identify the main regulatory components of a biological system, including

*University of Helsinki, Computational Systems Biology Laboratory, elena.czeizler@helsinki.fi

†Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland, amizera@abo.fi

‡Department of Information Technologies, Åbo Akademi University, ipetre@abo.fi

its feedforward and feedback mechanisms, see [5]. This, in turn, contributes to the understanding of the reactivity, robustness and efficiency of the biological system. To disentangle the individual contribution of the various components to the network, knock-out mutants are often useful to consider, see [5]. The mutants are numerically compared to each other and to the reference model in an effort to extract the individual contribution of each mechanism to the overall behavior of the system.

We propose in this paper a novel approach for identifying the numerical contribution of a component to the system-level behavior of a model. The core technique we use throughout the paper is to associate a Boolean variable to each of the components. For each knock-out mutant we write a Boolean formula describing the presence or the absence of each component (using the conjunction and the negation of Boolean variables). The obtained Boolean formulas describe the architecture of the knock-out mutant models. We can write a Boolean formula characterizing all mutant architectures that exhibit a given property: we select all knock-out mutants that exhibit that property and construct the disjunction of their associated Boolean formulas. The formula thus obtained describes which components must be present/absent and in which configurations in order for the system to exhibit the desired property, thus helping elucidate the (qualitative) role of the various components in the system-level properties of the full model. In Section 5 we discuss the dependency of our analysis on the numerical setups (kinetic rate constants and initial values) of the knock-out mutant models.

We use as a case study in this paper a model for the heat shock response introduced in [17]. We apply our approach to identify the contribution of each of the three feedbacks of the model to having a response where the level of misfolded proteins remains low, with a relatively low cost in terms of transactivating the heat shock protein genes.

2 Method

Throughout this paper we consider a model M to consist of a list of m species $\Sigma = \{A_1, A_2, \ldots, A_m\}$ and n reactions $\rho = \{r_1, r_2, \ldots, r_n\}$ of the following form:

$$r_i \ : \ s_{i,1}A_1 + s_{i,2}A_2 + ... + s_{i,m}A_m \xrightarrow{k_i} s'_{i,1}A_1 + s'_{i,2}A_2 + ... + s'_{i,m}A_m,$$

where $s_{i,1}, ..., s_{i,m}, s'_{i,1}, ..., s'_{i,m}$ are non-negative integers called *the stoichiometric coefficients* of r_i and $k_i \geq 0$ is a real number called *the kinetic rate constant* of r_i. A number of different mathematical models can be associated to M e.g., in terms of continuous or discrete variables, deterministic or non-deterministic evolution, individual- or population-based, etc. We choose in this paper a continuous, mass-action formulation, where to each variable A_i,

$1 \leq i \leq n$, we associate a time-dependent function $[A_i] : \mathbb{R}_+ \to \mathbb{R}_+$, with $[A_i](t)$ describing its concentration at time t. The dynamics of the system is then described through a system of differential equations [14] in which, for each reaction, we assumed the principle of mass action, originally introduced in [8], [9]. In particular, for model M we obtain the following system of ordinary differential equations (ODE):

$$\frac{d[A_j]}{dt} = -\sum_{i=1}^{n}(k_i S_{i,j} \prod_{k=1}^{m}[A_k]^{S_{i,k}}) + \sum_{i=1}^{n}(k_i S'_{i,j} \prod_{k=1}^{m}[A_k]^{S_{i,k}}), \quad 1 \leq j \leq m.$$

Consider now a partition of the reaction set $\rho = \rho_0 \cup \rho_1 \cup \ldots \cup \rho_l$, $l \geq 2$, i.e., $\rho_i \neq \emptyset$ and $\rho_i \cap \rho_j = \emptyset$ for all $i \neq j$. Depending on the model under study and the model property that is analyzed, different partitions could be considered; for example, in the case-study discussed in this paper we consider a control-based partition of the reaction-set which identifies and isolates the three feed-back loops of the model, in addition to its basic, open-loop architecture. For any index set $I \subseteq \{1, 2, \ldots, l\}$ we say that M_I is the submodel of M consisting of the reaction set $\rho_0 \cup \bigcup_{i \in I} \rho_i$, where ρ_0 is a set forming some elementary core of M that is to be included in all submodels; an example here is the open-loop architecture of our case-study. M_I is also called a *knock-out mutant* of M in the sense that it is obtained from M by removing some blocks from it.

The focus of our method is on describing the contribution of each of the partition blocks to a "system-level" property of the full model. In more details, we consider a property $\mathcal{P}$ that is decidable for model M and all of its submodels. To describe which combinations of partition blocks are necessary and sufficient for $\mathcal{P}$ to hold, we employ the following simple approach. For each submodel M_I we denote by μ_I its structure in terms of which partition blocks are included/excluded in model M_I, i.e.

$$\mu_I = \bigwedge_{i \in I} \varrho_i \wedge \bigwedge_{j \notin I} \overline{\varrho_j},$$

where the presence of a Boolean variable ϱ_i indicates that the reaction set in the partition block ρ_i are included in model M_I, while that of a negation $\overline{\varrho_j}$ indicates that the reaction set in ρ_j are not included in M_I. Note that this is a *discrete*, Boolean description of the architecture of a model M_I that is *continuous* in terms of its species dynamics. With these notations we can now describe property $\mathcal{P}$ in terms of the partition blocks that are necessary and sufficient for it to hold in the following way:

$$\mathcal{P} = \bigvee_{I \in I_{\mathcal{P}}} \mu_I,$$

where $I_{\mathcal{P}} = \{I \subseteq \{1, 2, \ldots, l\} \mid \mathcal{P} \text{ holds for model } M_I\}$. In other words, $\mathcal{P}$ is described through a Boolean formula detailing which combinations of

including some of the modules of M and excluding others are necessary and sufficient for $\mathcal{P}$ to hold. We apply this techniques in Section 4 to disentangle the role of the three feedbacks of the heat shock response model being efficient and economical.

3 Models

3.1 The eukaryotic heat shock response: a molecular model

The heat shock response (HSR) is an evolutionary-conserved global regulatory network found in virtually all living cells. It allows the cell to quickly react to elevated temperatures by the induction of some proteins called *heat shock proteins* (hsp). Exposure to raised temperature leads to protein misfolding, which then accumulate and form aggregates with disastrous effect for the cell. Stress conditions can be caused not only by increased temperature but also by other forms of environmental, chemical or physical stress, such as addition of ethanol, heavy metals, pollutants, high osmolarity, starvation, etc.

The central role in the heat shock response is played by the *heat shock proteins* (hsp), which act as chaperones for the *misfolded proteins* (mfp) by forming hsp: mfp complexes and helping them to refold. In the model presented in [17], the regulation of the heat shock response is done by controlling the transactivation of the hsp-encoding genes. The transcription of these genes is initiated by some specific proteins called *heat shock factors* (hsf) that first dimerize (hsf_2), then trimerize (hsf_3) to finally bind to the promoters of the hsp-encoding genes, called *heat shock elements* (hse). After the trimers bind to the promoter sites (hsf_3: hse) the transcription and translation of the hsp-encoding genes starts, ultimately producing new hsp molecules.

Once the level of hsp molecules is high enough, the transcription process is turned off through a self-regulating mechanism. The hsp molecules sequestrate the heat shock factors (hsp: hsf), thus preventing them to trimerize and bind to the heat shock elements. The sequestration of the heat shock factors by the heat shock proteins can be done in three different ways: by binding to free hsf, by breaking dimers and trimers, and by unbinding hsf_3 from the DNA promoter sites with simultaneous breaking of the trimer. Once the temperature increases, some of the proteins (prot) start to misfold, driving hsp away from hsf. Thus, the heat shock response is quickly switched on since the heat shock factors are again free and able to promote the synthesis of more heat shock proteins. The reaction rules of the molecular model introduced in [17] are presented in Table 1.

The model in Table 1 includes three mass conservation relations, see [17], for the total amount of hsf, the total amount of proteins (other than hsp and hsf) in the model, as well as for the total amount of hse: $[\mathsf{hsf}] + 2 \times [\mathsf{hsf}_2] +$

Table 1: The molecular model for the eukaryotic heat shock response proposed in [17].

Reaction		Reaction	
$2\,\mathsf{hsf} \leftrightarrows \mathsf{hsf}_2$	[r1]	$\mathsf{hsp} + \mathsf{hsf}_3 \to \mathsf{hsp{:}\,hsf} + 2\,\mathsf{hsf}$	[r7]
$\mathsf{hsf} + \mathsf{hsf}_2 \leftrightarrows \mathsf{hsf}_3$	[r2]	$\mathsf{hsp} + \mathsf{hsf}_3{:}\,\mathsf{hse} \to \mathsf{hsp{:}\,hsf} + 2\,\mathsf{hsf} + \mathsf{hse}$	[r8]
$\mathsf{hsf}_3 + \mathsf{hse} \leftrightarrows \mathsf{hsf}_3{:}\,\mathsf{hse}$	[r3]	$\mathsf{hsp} \to \emptyset$	[r9]
$\mathsf{hsf}_3{:}\,\mathsf{hse} \to \mathsf{hsf}_3{:}\,\mathsf{hse} + \mathsf{hsp}$	[r4]	$\mathsf{prot} \to \mathsf{mfp}$	[r10]
$\mathsf{hsp} + \mathsf{hsf} \leftrightarrows \mathsf{hsp{:}\,hsf}$	[r5]	$\mathsf{hsp} + \mathsf{mfp} \leftrightarrows \mathsf{hsp{:}\,mfp}$	[r11]
$\mathsf{hsp} + \mathsf{hsf}_2 \to \mathsf{hsp{:}\,hsf} + \mathsf{hsf}$	[r6]	$\mathsf{hsp{:}\,mfp} \to \mathsf{hsp} + \mathsf{prot}$	[r12]

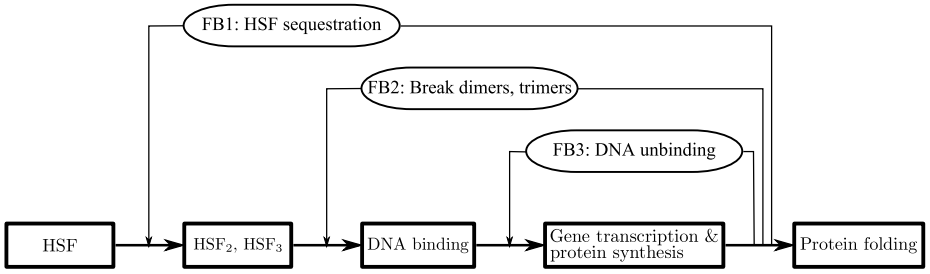


Figure 1: The control structure of the heat shock response network.

$3 \times [\mathsf{hsf}_3] + 3 \times [\mathsf{hsf}_3{:}\,\mathsf{hse}] + [\mathsf{hsp{:}\,hsf}] = C_1$, $[\mathsf{prot}] + [\mathsf{mfp}] + [\mathsf{hsp{:}\,mfp}] = C_2$, $[\mathsf{hse}] + [\mathsf{hsf}_3{:}\,\mathsf{hse}] = C_3$, for some mass constants C_1, C_2, C_3.

3.2 The mathematical model

We associate with the molecular model in Table 1 a mathematical model in terms of ordinary differential equations, where for each reaction we assume the principle of mass action, see, e.g., [15]. We associate with each reactant a continuous, time-dependant variable that gives its concentration level. For each variable, its differential equation gives the cumulated consumption and production rates of the reactant corresponding to it in the molecular model. Thus, the dynamic behavior of the molecular model is described through the set of all resulting differential equations. We list them in Table 2 and refer to [17] for more details. For the numerical values of the kinetic rate constants and of the initial values of the model we refer to [17] and [4].

3.3 A control-based modularization of the heat shock response model

A control-driven analysis of the heat shock response model of [17] was introduced in [3] to decompose the heat shock response model. The model was divided into the following submodules: *the plant*, i.e., the process to be regulated, *the controller*, i.e., the decision-making module, and *the actuator*, i.e., the module which modifies the current state of the system, thus

Table 2: The differential equations of the associated mathematical model of the heat shock response ([17]).

$$
\begin{aligned}
d[\mathsf{hsf}]/dt &= -2k_1^+[\mathsf{hsf}]^2 + 2k_1^-[\mathsf{hsf}_2] - k_2^+[\mathsf{hsf}][\mathsf{hsf}_2] + k_2^-[\mathsf{hsf}_3] - k_5^+[\mathsf{hsf}][\mathsf{hsp}] \\
&\quad + k_5^-[\mathsf{hsp{:}\ hsf}] + k_6[\mathsf{hsf}_2][\mathsf{hsp}] + 2k_7[\mathsf{hsf}_3][\mathsf{hsp}] + 2k_8[\mathsf{hsf_3{:}\ hse}][\mathsf{hsp}] \\
d[\mathsf{hsf}_2]/dt &= k_1^+[\mathsf{hsf}]^2 - k_1^-[\mathsf{hsf}_2] - k_2^+[\mathsf{hsf}][\mathsf{hsf}_2] + k_2^-[\mathsf{hsf}_3] - k_6[\mathsf{hsf}_2][\mathsf{hsp}] \\
d[\mathsf{hsf}_3]/dt &= k_2^+[\mathsf{hsf}][\mathsf{hsf}_2] - k_2^-[\mathsf{hsf}_3] - k_3^+[\mathsf{hsf}_3][\mathsf{hse}] + k_3^-[\mathsf{hsf_3{:}\ hse}] - k_7[\mathsf{hsf}_3][\mathsf{hsp}] \\
d[\mathsf{hse}]/dt &= -k_3^+[\mathsf{hsf}_3][\mathsf{hse}] + k_3^-[\mathsf{hsf_3{:}\ hse}] + k_8[\mathsf{hsf_3{:}\ hse}][\mathsf{hsp}] \\
d[\mathsf{hsf_3{:}\ hse}]/dt &= k_3^+[\mathsf{hsf}_3][\mathsf{hse}] - k_3^-[\mathsf{hsf_3{:}\ hse}] - k_8[\mathsf{hsf_3{:}\ hse}][\mathsf{hsp}] \\
d[\mathsf{hsp}]/dt &= k_4[\mathsf{hsf_3{:}\ hse}] - k_5^+[\mathsf{hsf}][\mathsf{hsp}] + k_5^-[\mathsf{hsp{:}\ hsf}] - k_6[\mathsf{hsf}_2][\mathsf{hsp}] - k_7[\mathsf{hsf}_3][\mathsf{hsp}] \\
&\quad - k_8[\mathsf{hsf_3{:}\ hse}][\mathsf{hsp}] - k_{11}^+[\mathsf{hsp}][\mathsf{mfp}] + (k_{11}^- + k_{12})[\mathsf{hsp{:}\ mfp}] - k_9[\mathsf{hsp}] \\
d[\mathsf{hsp{:}\ hsf}]/dt &= k_5^+[\mathsf{hsf}][\mathsf{hsp}] - k_5^-[\mathsf{hsp{:}\ hsf}] + k_6[\mathsf{hsf}_2][\mathsf{hsp}] + k_7[\mathsf{hsf}_3][\mathsf{hsp}] \\
&\quad + k_8[\mathsf{hsf_3{:}\ hse}][\mathsf{hsp}] \\
d[\mathsf{mfp}]/dt &= \phi_T[\mathsf{prot}] - k_{11}^+[\mathsf{hsp}][\mathsf{mfp}] + k_{11}^-[\mathsf{hsp{:}\ mfp}] \\
d[\mathsf{hsp{:}\ mfp}]/dt &= k_{11}^+[\mathsf{hsp}][\mathsf{mfp}] - (k_{11}^- + k_{12})[\mathsf{hsp{:}\ mfp}] \\
d[\mathsf{prot}]/dt &= -\phi_T[\mathsf{prot}] + k_{12}[\mathsf{hsp{:}\ mfp}]
\end{aligned}
$$

influencing the activity of the plant. A *sensor* which measures the current state of the system and sends this information to the controller. Further, as presented in [16], based on the control-driven analysis three *feedback mechanisms* responsible for the regulation of the process were distinguished within the controller. This decomposition of the heat shock model is presented in Table 3, where the reaction numbers refer to the reactions in Table 1. The three identified feedback mechanisms and their points of interaction with the mainstream process are depicted in Figure 1. We investigate in the continuation the role of the three feedback mechanisms identified with the control-driven approach in the overall behavior of the full model by applying the method introduced in Section 2.

3.4 Knock-out mutant models

In order to disentangle the role of the feedback mechanisms within the full model, we consider eight knock-out mutants obtained by eliminating from the basic model all combinations of the feedbacks FB_1, FB_2, and FB_3. In accordance with the notation introduced in Section 2, we will denote each of these mutants as M_I, where $I \subseteq \{1,2,3\}$ represents the set of indexes of the feedbacks included in the model M_I, i.e. $FB_1 = \rho_1 = \{[\mathsf{r5}]^+\}$, $FB_2 = \rho_2 = \{[\mathsf{r6}], [\mathsf{r7}]\}$, $FB_3 = \rho_3 = \{[\mathsf{r8}]\}$, and $l = 3$. To identify the individual contributions of the three feedback mechanisms, we compare the dynamics

Table 3: The control-based decomposition of the model in Table 1. We denote the 'left-to-right' direction of reaction [r5] by $[\mathsf{r5}]^+$ and by $[\mathsf{r5}]^-$ its 'right-to-left' direction.

Module and main task	Reactions
Plant (protein misfolding and refolding)	[r10], [r11], [r12]
Actuator (regulate the level of hsp)	[r4], [r9]
Sensor (measure the level of hsp)	—
Controller (modulate level of DNA binding)	[r1], [r2], [r3], $[\mathsf{r5}]^-$
Feedback FB_1 (sequestration of hsf)	$[\mathsf{r5}]^+$
Feedback FB_2 (dimer and trimer breaking)	[r6], [r7]
Feedback FB_3 (hsp-forced DNA unbinding)	[r8]

of these eight models at $42°C$. We choose this temperature since at $42°C$ the experimental data shows a heat shock response both in terms of increased level of misfolded proteins and in terms of transcription activity of the hsp-encoding genes, see [17].

In our comparison of the numerical knock-out mutant models we aim to focus on the differences stemming from the intrinsic dissimilarities in their architectures and eliminate as much as possible differences coming from unfavorable numerical setups chosen for the various models. For example, we consider all knock-out mutants as viable alternatives for the heat shock response model. We impose the following three constraints:

(1) The kinetic rate constants for the reactions of each of the eight knock-out mutants should be chosen in such a way that the numerical prediction for the time evolution of the level of $\mathsf{hsf_3{:}\,hse}$ fits in with the experimental data given in [13] on DNA binding of hsf_3.

(2) The initial distribution of the reactants of each mutant should be chosen in such a way that they form a steady state at $37°C$ for that particular model.

(3) For all knock-out mutants, the values of the mass constants C_1, C_2, C_3 are chosen to be identical to those of the reference model $M_{1,2,3}$.

All three constraints come as natural consequences of the fact that we consider all knock-out mutants as viable alternatives for the heat shock model. As such, their dynamic behavior should be in agreement with the existent experimental data and, at the same time, they should be in a steady state in the absence of a heat shock, i.e., at $37°C$. Moreover, since they are all models for the same biological process, they should all assume the same values for the mass constants.

4 Results

When comparing the performance of the eight alternative models we focused on two aspects: the total amount of hsp and the total amount of mfp both at $37°C$ and at $42°C$. We were interested mainly in these two aspects since a very high level of mfp indicates a non-effective response while a very high level of hsp indicates a non-economical response.

We associated to each of the three feedback mechanisms a Boolean variable, denoted by F_1, F_2 and F_3, respectively. Then, for each knock-out mutant we wrote a Boolean formula expressing which of the feedback mechanisms are present in the model, where we denoted by $\wedge$ the conjunction operator and by $\overline{F_i}$ the negation of the variable F_i. For example, to knock-out mutant $M_{1,2}$ we associated the Boolean formula $F_1 \wedge F_2 \wedge \overline{F_3}$ to express that feedbacks FB_1, FB_2 are included in the model, while FB_3 is not. Similarly, we associated to mutant M_2 the Boolean formula $\overline{F_1} \wedge F_2 \wedge \overline{F_3}$, indicating that FB_2 is included in the model, while FB_1 and FB_3 are not.

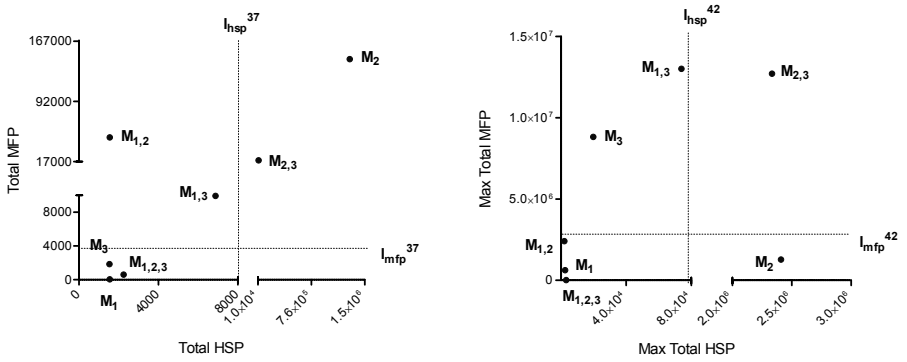
We considered all knock-out mutant models having a 'low' total amount of hsp at $37°C$ and at $42°C$, respectively. By writing the disjunction, denoted by $\vee$, of the formulas corresponding to these mutants we obtained a Boolean formula describing the contribution of each feedback to achieving the property: which feedbacks must be present in the model in order for it to exhibit the desired property. We applied the same technique to describe which models exhibit 'low' levels of mfp at $37°C$ or at $42°C$.

We start our analysis with the mutant M_0, which does not include any of the three feedback mechanisms. The ODE mass-action model associated to M_0 shows that if the mutant starts from its steady state at $37°C$, then at any temperature the differentials for [hsf], [hsf_2], [hsf_3], [hsf_3: hse], [hse] and [hsp: hsf] are zero. That is, those functions remain constant at their steady state values independent of temperature. In particular, the DNA binding level, i.e., hsf_3: hse, remains constant even when we increase the temperature. So, for no numerical setup, this mutant can provide numerical predictions in agreement with the data from [13] if it starts from its steady state at $37°C$. Thus, we discarded this knock-out mutant from our considerations.

For each of the mutants M_1, M_2, M_3, $M_{1,2}$, $M_{1,3}$, and $M_{2,3}$ we performed parameter estimation to identify a numerical setup, i.e., a set of values for the kinetic rate constants, that provides numerical predictions in accordance with the experimental data of [13]. The results are shown in Table 6A of [4]. We then numerically estimated the steady state of each model at $37°C$; the results are given in Table 6B of [4]. Finally, we numerically integrated the mathematical model corresponding to each knock-out mutant starting from its own steady state values. For the numerical integration we used the software COPASI [10].

We chose empirically four numerical thresholds separating the 'low' and 'high' values for the total amount of: (i) hsp proteins at $37°C$; (ii) mfp

proteins at $37°C$; (iii) hsp proteins at $42°C$; and (iv) mfp proteins at $42°C$. The thresholds we selected were the following: $l^{37}_{\mathsf{hsp}} = 8000$, $l^{37}_{\mathsf{mfp}} = 3000$, $l^{42}_{\mathsf{hsp}} = 8 \times 10^4$, and $l^{42}_{\mathsf{mfp}} = 2.5 \times 10^6$, respectively, all in terms of number of molecules. We plotted the behavior of each knock-out mutant model with respect to these thresholds in Figure 2.



(a) (b)

Figure 2: (a) The total amount of hsp and mfp for each of the seven models at $37°C$; (b) the maximal value for the total amount of hsp and mfp for each of the seven models at $42°C$. Values on the axes are in terms of number of molecules and should be interpreted as an average of a population of cells.

We considered the following four properties:

Property P_1: Low level for the total amount of hsp *at* $37°C$. This property is exhibited only by the mutants M_1, M_3, $M_{1,2}$, $M_{1,3}$, and $M_{1,2,3}$. Using the Boolean formulas expressing each mutant in terms of their feedback structure, we constructed a Boolean formula for property P_1. This is easily obtained as a disjunctive formula (logical OR) among the Boolean formulas for M_1, M_3, $M_{1,2}$, $M_{1,3}$, and $M_{1,2,3}$: $(F_1 \wedge \overline{F_2} \wedge \overline{F_3}) \vee (\overline{F_1} \wedge \overline{F_2} \wedge F_3) \vee (F_1 \wedge F_2 \wedge \overline{F_3}) \vee (F_1 \wedge \overline{F_2} \wedge F_3) \vee (F_1 \wedge F_2 \wedge F_3)$, which can be rewritten in a compact form as:

$$F_1 \vee (\overline{F_1} \wedge \overline{F_2} \wedge F_3). \tag{1}$$

Thus, property P_1 can be satisfied if and only if either feedback FB_1 is present (regardless of whether FB_2 and FB_3 are included or not) or feedback FB_3 is present while feedbacks FB_1 and FB_2 are absent.

Property P_2: Low level for the maximal value of the total amount of hsp *at* $42°C$. This property is exhibited again only by mutants M_1, M_3, $M_{1,2}$, $M_{1,3}$, and $M_{1,2,3}$. So, we obtained the Boolean formula

$$F_1 \vee (\overline{F_1} \wedge \overline{F_2} \wedge F_3). \tag{2}$$

Property P_3: Low level for the total amount of mfp *at* $37°C$. This property is exhibited only by the mutants M_1, M_3, and $M_{1,2,3}$. So, in this case we

obtained the Boolean formula

$$(F_1 \wedge \overline{F_2} \wedge \overline{F_3}) \vee (\overline{F_1} \wedge \overline{F_2} \wedge F_3) \vee (F_1 \wedge F_2 \wedge F_3). \quad (3)$$

Property P_4: Low level for the maximal value of the total amount of mfp *at* $42°C$. This property is exhibited by the mutants M_1, M_2, $M_{1,2}$, and $M_{1,2,3}$. In this case, we obtained the Boolean formula

$$(F_1 \wedge \overline{F_3}) \vee (F_1 \wedge F_2 \wedge F_3) \vee (\overline{F_1} \wedge F_2 \wedge \overline{F_3}). \quad (4)$$

Note that when defining the properties P_1 and P_3, we consider the total amounts of hsp and mfp without referring to the maximal values, as in the case of P_2 or P_4. This is due to the fact that P_1 and P_3 are considered at $37°C$, a temperature at which the system is in a steady state. Thus, the total amounts are not changing in time.

To investigate which knock-out mutants can be both effective and economic, we looked at the models that exhibit low levels for both hsp and mfp. For a temperature of $37°C$, we considered the models that verify simultaneously properties P_1 and P_3. The Boolean formula describing these architectures was easily obtained as a conjunctive formula (logical AND) among the formulas for properties P_1 and P_3, which could then be rewritten in a compact form as $(F_1 \wedge \overline{F_2} \wedge \overline{F_3}) \vee (\overline{F_1} \wedge \overline{F_2} \wedge F_3) \vee (F_1 \wedge F_2 \wedge F_3)$. Since this was identical with (3), we concluded that at $37°C$, once a mutant achieved a low level for the total amount of mfp, it would also exhibit a low level for the total amount of hsp. For the similar analysis at $42°C$ we were interested in the models that verify simultaneously properties P_2 and P_4. In this case, the Boolean formula describing these architectures is $F_1 \wedge (F_2 \vee (\overline{F_2} \wedge \overline{F_3}))$. This shows that to obtain low values for both hsp and mfp at $42°C$ the first feedback is essential. Moreover, only two types of mutant architectures predicted this outcome: if both FB_1 and FB_2 were present in the model (regardless of whether FB_3 is included or not), or if FB_1 was included while FB_2 and FB_3 were not. Furthermore, it showed that the second feedback, in addition to the first one, has a role in decreasing the levels of both hsp and mfp at $42°C$. The second type of architecture, i.e., when FB_1 was present in the model while FB_2 and FB_3 were absent, showed that the first feedback alone is sufficient to ensure a low enough level of both hsp and mfp at $42°C$. However, when we compared the values predicted by M_1 and $M_{1,2,3}$, see Figure 2(b), we noticed that the cumulative effect of the second and the third feedbacks added to the first one is to further reduce the total level of mfp.

We noticed that the Boolean formulas corresponding to properties P_1 and P_2 were identical. This means that once a knock-out mutant is able to keep a low level of hsp at $37°C$, it will also be able to respond to heat shock with a relatively low level of hsp. Moreover, this was the case only for two types of mutant architectures: either when the feedback FB_1 was

present (regardless of whether FB_2 and FB_3 were included or not) or when feedback FB_3 was present while feedbacks FB_1 and FB_2 were absent. This showed that the first and the third feedbacks have roles in lowering the level of hsp both at $37°C$ and at $42°C$. The first type of mutant architecture, having the feedback FB_1 present, was insensitive to the second and the third feedbacks: whether they were included in the model or not did not change the behavior of the model with respect to P_1 and P_2. The second type of mutant architecture that satisfies the Boolean formula (1) showed that in the absence of the first feedback, the third one is necessary to obtain low levels of hsp both at $37°C$ and at $42°C$.

If, on the other hand, we required low levels of mfp both at $37°C$ and at $42°C$, i.e., if we asked for properties P_3 and P_4 to be satisfied, then we would see that the first feedback has to be present in the model. Otherwise, i.e., if $F_1 = 0$, the two Boolean formulas (3) and (4) become $\overline{F_2} \wedge F_3$ and $F_2 \wedge \overline{F_3}$, respectively, which obviously cannot be simultaneously satisfied. This confirmed again our conclusion that the first feedback is essential for the model to satisfy all four properties P_1, P_2, P_3, and P_4, i.e., for the model to exhibit low levels for both hsp and mfp, both at $37°C$ and at $42°C$.

5 Discussion

In this paper, we proposed a novel approach for the knock-out mutant model comparison problem. There are several other methods for submodel comparison in the literature, e.g., the mathematically controlled comparison [20] or the local submodel comparison method [3]. The technique of mathematically controlled comparison provides a structured approach for comparing several alternative designs with respect to some chosen measures of functional effectiveness. However, this framework imposes one important constraint on the alternative designs: they are allowed to differ from the reference design in only one component. Moreover, the mathematical models both for the reference design and for the alternative architectures are developed in the framework of canonical nonlinear modeling referred to as S-systems, [18] and [19].

In the case of the local submodel comparison the alternative designs are considered submodels of the reference model. The underlying reaction networks of these submodels are very similar (although not identical), and both the biological constraints and the kinetics of the reactions are taken from the reference model. To assure an unbiased comparison, i.e., to avoid the situation where the submodel exhibits two intertwined tendencies of the migration from a possible unstable state and the response to a stimulus, the initial values of the reactants are chosen in such a way that they constitute a steady state of that design in the absence of a trigger. For a detailed review of these and other approaches we refer to [16].

In our new approach proposed in this study we first associated a Boolean variable to each of the three feedback mechanisms identified in [3] for the reference model of the eukaryotic heat shock response. Then, for each knock-out mutant we wrote a Boolean formula (using the conjunction and negation of the three introduced Boolean variables) characterizing its control architecture, i.e., which of the three feedback mechanisms are present in the model. As such, each of these formulas encompass time-independent properties of the models. This makes our approach very different from the Boolean network framework for modeling biological systems, see [1], [11], [12], [21], where one usually associates a Boolean variable to each species present in the system. Boolean formulas are then used to simulate the time evolution of the species. Moreover, in our approach the Boolean formulas associated to each knock-out mutant are parameter independent, i.e., they are not influenced by the parameters used to describe the compared models. We also introduced a Boolean formula characterizing all those mutant architectures that exhibit a given behavioral property, e.g., low levels of hsp or mfp. This can be easily obtained as a disjunctive formula (logical OR) of the Boolean formulas describing the architectures of the mutants exhibiting the required property. However, in order to perform numerical simulations of the models we needed numerical setups for each of the knock-out mutants, i.e., specific values both for the initial distribution of the reactants and for the kinetic rate constants of the models. For the initial values of the variables, we chose the approach proposed in [3] for the local submodel comparison method, i.e., we set them separately for each knock-out mutant in such a way that they form a steady state for that particular model. Regarding the kinetic rate constants in each of the knock-out mutants, one approach is to take them from the reference model, see [3]. The idea in this case is to make the whole comparison in the numerical setup of the reference model. Alternatively, we proposed here to separately estimate the kinetic constants of each alternative model with respect to available experimental data. In other words, we considered all models to be viable alternatives for the biological system and, as such, we took for each of them a most favorable numerical setup.

Since the numerical setup giving a good model fit is in general not unique, it means that our analysis is sensitive with respect to the choice of the values for the kinetic constants. This is often the case when model fitting is involved, see [2]. Repeating the analysis for several numerical setups (all of them as good in terms of fitting the model to the experimental data) would enrich the conclusions, by potentially showing that the same model architecture can exhibit different properties depending on the numerical setup. The conclusions of the analysis also depend on the numerical values chosen for the thresholds $l^{37}_{\mathsf{hsp}}, l^{37}_{\mathsf{mfp}}, l^{42}_{\mathsf{hsp}}$, and l^{42}_{mfp}.

It is crucial for our approach that all knock-out mutant models are considered in the analysis, i.e., all possible combinations ON/OFF of the model components are included in the comparison. In this way, we obtain a com-

plete characterization of the properties being analyzed in terms of *all* model architectures that can exhibit those properties. For a large number of components, this approach becomes quickly computationally challenging: for n components to be analyzed, there are 2^n knock-out mutant models to be compared. Including in the comparison only a part of those mutants is also possible but then the output of the method is partial: one only discovers *some* of the architectures exhibiting the property of interest. However, we notice that the exponential explosion problem is not caused by the size of the considered model – in fact, the model itself can be large (e.g. in terms of the number of reactants and reactions). The problem is about how many components (Boolean variables) one considers. In other words, the practical application of our method is not restricted by the size of the model but the number of considered components. In fact, the control-driven decomposition enables a split of any model into a relatively small number of modules, i.e., the following well-defined basic components: the plant, the sensor, the actuator and the controller. In the case of the heat hock response model discussed in this paper, three feedback mechanisms were distinguished within the controller and only their role was investigated. This provides a general hint how one could avoid the potential problem of computational intractability by taking the following hierarchical approach. In the first step, one would perform a basic decomposition leading to components of well-defined role in the whole system. In the second step, one would identify the elements of interest within one of the higher-level, previously distinguished components. In this way one could investigate the influence of each particular element on the behavior of the component, which role in the whole system is in turn well-defined. In consequence, our method would be applied to just some subpart of the whole system with a smaller number of components and in this way made computationally tractable.

The usefulness of our new method for disentangling the roles of submodules in the overall behaviour of a biomodel is illustrated on the example of a model for the eukaryotic heat shock response. The heat shock response represents an exceptionally well-conserved regulatory network. It involves all mechanisms that any regulatory network must include: a stress-induced activations and a feedback regulation. As such, it can be regarded as an archetype for a cellular regulatory process. The heat shock response model presented in Section 3 is attractive to be considered as a case study for the reason that it is a simple model which captures in mechanistic details all key aspects of regulation: the heat-induced protein misfolding, the transactivation of the hsp-encoding genes, the chaperone activity of heat shock proteins, and the repression of hsp transactivation once the stress is removed. This justifies the claim that our approach is general, i.e., it is applicable to virtually any biomodel describing a regulatory mechanism.

When we compared the numerical behavior of the knock-out mutants, we chose a mathematical model formulation in terms of ordinary differential

equations. However, our approach is independent of this formulation and it would work equally well with other formulations, such as continuous-time Markov chains and their numerical simulations based on Gillespie's algorithm, see [6, 7].

Our approach can be easily extended to a more refined analysis, where the range of the properties to be analyzed is divided into more domains than just 'low' and 'high'. The range could in fact be divided into an arbitrarily high number of intermediate domains, depending on the details of the case study. A Boolean formula could be associated to characterize each of those domains in a manner similar to that demonstrated in this paper.

Acknowledgements This work was supported by Academy of Finland, grants 129863, 108421, and 122426.

References

[1] M. Chaves, R. Albert, and E. D. Sontag. Robustness and fragility of boolean models for genetic regulatory networks. *J Theor Biol*, 235:431–49, 2005.

[2] 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.

[3] E. 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, *Lecture Notes in Bioinformatics*, volume 5688, pages 111–125. Springer, Berlin, Heidelberg, 2009.

[4] E. Czeizler, A. Mizera, and I. Petre. A boolean approach for disentangling the numerical contribution of modules to the system-level behavior of a biomodel. Technical Report 997, Turku Centre for Computer Science, `http://tucs.fi/research/publication-view/?pub_id=tCzMiPe11a`, January 2011.

[5] H. El-Samad, H. Kurata, J. C. Doyle, C. A. Gross, and M. Khammash. Surviving heat shock: Control strategies for robustness and performance. *Proc Natl Acad Sci USA*, 102:2736–2741, 2005.

[6] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys*, 22:403–434, 1976.

[7] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem*, 81:2340–2361, 1977.

[8] C. M. Guldberg and P. Waage. Studies concerning affinity. *C. M. Forhandlinger: Videnskabs-Selskabet i Christiana*, 35, 1864.

[9] C. M. Guldberg and P. Waage. Concerning chemical affinity. *Erdmann's Journal fr Practische Chemie*, 127:69–114, 1879.

[10] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. Copasi – a COmplex PAthway SImulator. *Bioinformatics*, 22:3067–3074, 2006.

[11] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein. Random boolean network models and the yeast transcriptional network. *Proc Natl Acad Sci USA*, 100:14796–14799, 2003.

[12] G. Kervizic and L. Corcos. Dynamical modeling of the cholesterol regulatory pathway with boolean networks. *BMC Systems Biology*, 2:99, 2008.

[13] 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*, 17:2107–2115, 1997.

[14] E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach. *Systems Biology in Practice*. WILEY-VCH Verlag GmbH & Co, KGaA, Weinheim, 2005.

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

[16] A. Mizera, E. Czeizler, and I. Petre. Methods for biochemical model decomposition and quantitative submodel comparison. *Israel J Chem*, 51(1):151–164, 2011.

[17] I. Petre, A. Mizera, C. L. Hyder, A. Mikhailov, J. E. Eriksson, L. Sistonen, and R.-J. Back. A simple mass-action model for the eukaryotic heat shock response and its mathematical validation. *Natural Computing*, 10(1):595–612, 2011.

[18] M. A. Savageau. Biochemical systems analysis: I. Some mathematical properties of the rate law for the component enzymatic reactions. *J Theoret Biol*, 25:365–369, 1969.

[19] M. A. Savageau. Biochemical systems analysis: II. the steady-state solutions for an n-pool system using a power-law approximation. *J Theoret Biol*, 25:370–379, 1969.

[20] M. A. Savageau. The behavior of intact biochemical control systems. *Curr. Top. Cell. Reg.*, 6:63–130, 1972.

[21] I. Shmulevich, R. Dougherty, and W. Zhang. From boolean to probabilistic boolean networks as models of genetic regulatory networks. *Proc IEEE*, 90:1778–1792, 2002.