# Computational heuristics for simplifying a biological model

Ion Petre, Andrzej Mizera, and Ralph-Johan Back

Department of Information Technologies, Åbo Akademi University,
Joukahaisenkatu 3-5A, FIN-20520 Turku, Finland
ipetre, amizera, backrj@abo.fi

**Abstract.** Computational modeling efforts for biological processes adopt either of the following strategies: strive to build large, as complete as possible models in an effort to obtain very realistic models, or on the contrary, build as simple as possible models focusing only on the core aspects of the process, in an effort to obtain a model that is easier to analyze, fit, and validate. When the latter strategy is adopted, the aspects that are left outside the models are very often up to the subjective options of the modeler. We show in this paper a heuristic method to simplify an already fit model in such a way that the numerical fit to the experimental data is not lost. We focus in particular on eliminating some of the variables of the model and the reactions they take part in, while also modifying some of the remaining reactions. We illustrate the method on a computational model for the eukaryotic heat shock response.

**Keywords:** Model reduction, heat shock response, mathematical model.

## 1 Introduction

When designing a new molecular model for some biological process or network, the choice one has to make early on in the modeling process is whether to strive for a rich model, capturing many details, or on the contrary, to focus on a more abstract model, capturing only a few, main actors of interest. The choice is not obvious and depends heavily on the goals of the modeling project and on the modeler itself. On one hand, a rich model has the potential of being more realistic but it leads to a more complex mathematical model that may be difficult to fit to experimental data, to analyze, and ultimately may be less apt to answer to biological queries. On the other hand, a less finely grained molecular model leads to a smaller mathematical model (in terms of the number of variables and equations) that may be easier to work with, but it pays a price in ignoring a number of details. A main difficulty in choosing between a rich and a simplified molecular model is that the potential cost of starting off with a rich model only becomes transparent at a latter stage, in the process of analyzing the corresponding mathematical model. Moreover, in the case of choosing a simplified model, the selection of the aspects to be ignored in the model is left up to the subjective choice of the modeler. We discuss in this paper

an intermediate approach where we start with a (potentially large, rich) model that has already been fit and validated against experimental data and we aim to simplify it in such a way that its numerical behavior remains largely unchanged. In this way, the simplified model is the result of a systematic, numerical analysis of the larger model that preserves its validation. We illustrate the approach on a computational model for the eukaryotic heat shock response and discuss the biological relevance of the simplifications we operate on the model. We also discuss the strong dependency of this approach on the numerical setup of the model; we show that our approach in the case of the heat shock response model is robust to some changes in the numerical values of the parameters, but it is sensitive to others.

## 2 The heat shock response model

The heat shock response is a well-conserved defence mechanism across all eukaryotic cells that enables them to survive under conditions of elevated temperatures. When exposed to heat shock, proteins inside cells tend to misfold. In turn, as an effect of their hydrophobic core being exposed, misfolded proteins form bigger and bigger aggregates with disastrous consequences for the cell, see [1]. In order to survive, the cell has to immediately react by increasing the level of chaperones (proteins that assist other proteins in the process of folding or refolding). Once the heat shock is removed, the defence mechanism is turned off and the cell eventually re-establishes the original level of chaperones, see [7, 11, 17].

The heat shock response has been intensively investigated in recent years for at least three main reasons. First, as a well-conserved mechanism in all eukaryotes, it is considered a promising candidate for investigating the engineering principles of gene regulatory networks, see [3, 4, 8, 18]. Second, heat shock proteins (hsp) act as main components in a large number of cellular processes such as signaling, regulation and inflammation, see [6, 16]. Moreover, their contribution to the resilience of cancer cells makes them an attractive target for cancer treatment, see [2, 9, 10, 19].

We consider in this paper the molecular model proposed in [14] for the eukaryotic heat shock response. This model consists of only the minimum number of components that any regulatory network must contain: an activation mechanism and a feedback mechanism. Moreover, the model consists of only well-documented reactions, without using any hypothetical, unknown cellular mechanism. The control over the cellular defence mechanism against protein misfolding is implemented through the regulation of the transactivation of the hsp-encoding gene. The transcription of the gene is activated by heat shock factors (hsf) which trimerize (the trimerization includes a transient dimerization phase) and in this form bind to the heat shock element (hse), which is the promoter of the hsp-encoding gene. Once the hsf trimer is bound to the specific DNA sequence, the gene is transactivated and the transcription and translation take place. As a result, new hsp molecules are eventually synthesized. When the level of hsp is high enough, the synthesis is switched off by the following mechanism:

**Table 1.** The list of variables in the mathematical model, their initial concentration values and their concentration values in one of the steady states of the system, for $T = 42$. Note that the initial state of the model is a steady state for $T = 37$. All concentrations are in $\frac{\#}{\text{cell}}$, where # denotes the number of molecules. The values should be interpreted as an average of a population of cells. [14, 15]

| Metabolite | Variable | Initial conc. |
|---|---|---|
| hsf | $X_1$ | 0.67 |
| $\mathsf{hsf}_2$ | $X_2$ | $8.73 \cdot 10^{-4}$ |
| $\mathsf{hsf}_3$ | $X_3$ | $1.22 \cdot 10^{-4}$ |
| $\mathsf{hsf}_3$: hse | $X_4$ | 3 |
| hse | $X_5$ | 30 |
| hsp | $X_6$ | 766.92 |
| hsp: hsf | $X_7$ | 1403.26 |
| hsp: mfp | $X_8$ | 71.65 |
| prot | $X_9$ | $1.14915 \cdot 10^{8}$ |
| mfp | $X_{10}$ | 517.32 |
| mhsf | $X_{11}$ | $3.01 \cdot 10^{-6}$ |
| mhsp | $X_{12}$ | 0.02 |
| hsp: mhsf | $X_{13}$ | $4.17 \cdot 10^{-7}$ |
| hsp: mhsp | $X_{14}$ | $2.24 \cdot 10^{-3}$ |

hsp bind to free hsf as well as break the hsf trimers (both free and those bound to DNA). This turns off DNA transcription and blocks the forming of new hsf trimers. The whole defense mechanism is turned on again when, as a result of raised temperature, the proteins (prot) in the cell begin misfolding again. The heat shock proteins become involved in refolding and free the hsf, which in turn trimerize and activate the synthesis of hsp, etc. What drives the heat shock response is the race to keep under control the level of misfolded proteins, in such a way that they are not able to accumulate, form aggregates, and eventually lead to cell death. The model consists of the following molecular reactions:

1. $2\ \mathsf{hsf} \leftrightarrows \mathsf{hsf}_2$
2. $\mathsf{hsf} + \mathsf{hsf}_2 \leftrightarrows \mathsf{hsf}_3$
3. $\mathsf{hsf}_3 + \mathsf{hse} \leftrightarrows \mathsf{hsf}_3\mathsf{:\ hse}$
4. $\mathsf{hsf}_3\mathsf{:\ hse} \rightarrow \mathsf{hsf}_3\mathsf{:\ hse} + \mathsf{mhsp}$
5. $\mathsf{hsp} + \mathsf{hsf} \leftrightarrows \mathsf{hsp:\ hsf}$
6. $\mathsf{hsp} + \mathsf{hsf}_2 \rightarrow \mathsf{hsp:\ hsf} + \mathsf{hsf}$
7. $\mathsf{hsp} + \mathsf{hsf}_3 \rightarrow \mathsf{hsp:\ hsf} + 2\ \mathsf{hsf}$
8. $\mathsf{hsp} + \mathsf{hsf}_3\mathsf{:\ hse} \rightarrow \mathsf{hsp:\ hsf} + 2\ \mathsf{hsf} + \mathsf{hse}$
9. $\mathsf{hsp} \rightarrow \emptyset$
10. $\mathsf{prot} \rightarrow \mathsf{mfp}$
11. $\mathsf{hsp} + \mathsf{mfp} \leftrightarrows \mathsf{hsp:\ mfp}$
12. $\mathsf{hsp:\ mfp} \rightarrow \mathsf{hsp} + \mathsf{prot}$
13. $\mathsf{hsf} \rightarrow \mathsf{mhsf}$
14. $\mathsf{hsp} \rightarrow \mathsf{mhsp}$
15. $\mathsf{hsp} + \mathsf{mhsf} \leftrightarrows \mathsf{hsp:\ mhsf}$

16. $\mathsf{hsp{:}\ mhsf} \rightarrow \mathsf{hsp} + \mathsf{hsf}$
17. $\mathsf{hsp} + \mathsf{mhsp} \leftrightarrows \mathsf{hsp{:}\ mhsp}$
18. $\mathsf{hsp{:}\ mhsp} \rightarrow 2\ \mathsf{hsp}$

When designing this molecular model, several criteria were followed, see [14], including that only well-documented reactions should be included and that the model should explicitly consider the temperature-induced protein misfolding as the trigger of the response. The model was also designed in such a way that is consistent with itself and with the kinetic principles of biochemistry. E.g., although hsf dimers are not experimentally detectable, they should be included in the model to account as a transient step in the formation of hsf trimers. Also, since hsp and hsf are themselves proteins, they should be subject to temperature-induced misfolding just like the regular proteins prot. Moreover, the refolding of mhsf and mhsp is controlled by the same kinetic constants as the refolding of mfp. The proper folding of newly synthesized hsp is supported by chaperoning as in the case of most proteins, see [1]. The degradation of hsf, prot, and mfp was on the other hand not included in the model so that compensating mechanisms of protein synthesis would not be explicit, see [14].

**Table 2.** The numerical values of parameters for the fitted model [14, 15].

| Kinetic constant | Reaction | Numerical value | Unit |
|---|---|---|---|
| $k_1^+$ | (1), forward | 3.49 | $\frac{\text{cell}}{\#\cdot s}$ |
| $k_1^-$ | (1), backward | 0.19 | $s^{-1}$ |
| $k_2^+$ | (2), forward | 1.07 | $\frac{\text{cell}}{\#\cdot s}$ |
| $k_2^-$ | (2), backward | $10^{-9}$ | $s^{-1}$ |
| $k_3^+$ | (3), forward | 0.17 | $\frac{\text{cell}}{\#\cdot s}$ |
| $k_3^-$ | (3), backward | $1.21 \cdot 10^{-6}$ | $s^{-1}$ |
| $k_4$ | (4) | $8.3 \cdot 10^{-3}$ | $s^{-1}$ |
| $k_5^+$ | (5), forward | 9.74 | $\frac{\text{cell}}{\#\cdot s}$ |
| $k_5^-$ | (5), backward | 3.56 | $s^{-1}$ |
| $k_6$ | (6) | 2.33 | $\frac{\text{cell}}{\#\cdot s}$ |
| $k_7$ | (7) | $4.31 \cdot 10^{-5}$ | $\frac{\text{cell}}{\#\cdot s}$ |
| $k_8$ | (8) | $2.73 \cdot 10^{-7}$ | $\frac{\text{cell}}{\#\cdot s}$ |
| $k_9$ | (10) | $3.2 \cdot 10^{-5}$ | $s^{-1}$ |
| $k_{11}^+$ | (14), forward | $3.32 \cdot 10^{-3}$ | $\frac{\text{cell}}{\#\cdot s}$ |
| $k_{11}^-$ | (14), backward | 4.44 | $s^{-1}$ |
| $k_{12}$ | (11) | 13.94 | $s^{-1}$ |

The mathematical model associated with the molecular model 1–18 is in terms of ordinary differential equations and it is obtained by assuming for all reactions the law of mass-action. The reasons for this choice is so that the explicit contribution of each reaction to the overall behavior could be followed. Let us denote the reactants occurring in the molecular model 1–18 according to the convention in Table 1. We use $\kappa \in \mathbb{R}_+^{25}$ to denote the vector with all reaction rate con-

stants as its components, see Table 2: $\kappa = (k_1^+, k_1^-, k_2^+, k_2^-, k_3^+, k_3^-, k_4, k_5^+, k_5^-, k_6, k_7, k_8, k_9, \phi(T), k_{11}^+, k_{11}^-, k_{12}, \phi(T), \phi(T), k_{11}^+, k_{11}^-, k_{12}, k_{11}^+, k_{11}^-, k_{12})$.

The corresponding mathematical model consists of the following differential equations:

$$
\begin{aligned}
dX_1/dt &= -k_2^+ X_1 X_2 + k_2^- X_3 - k_5^+ X_1 X_7 + k_5^- X_9 + 2 k_8 X_4 X_7 + k_6 X_2 X_7 \\
&\quad -\varphi(T) X_1 + k_{14} X_{10} + 2 k_7 X_3 X_7 - 2 k_1^+ X_1^2 + 2 k_1^- X_2 \\
dX_2/dt &= -k_2^+ X_1 X_2 + k_2^+ X_3 - k_6 X_2 X_7 + k_1^+ X_1^2 - k_1^- X_2 \\
dX_3/dt &= -k_3^+ X_3 X_6 + k_2^+ X_1 X_2 - k_2^- X_3 + k_3^- X_4 - k_7 X_3 X_7 \\
dX_4/dt &= k_3^+ X_3 X_6 - k_3^- X_4 - k_8 X_4 X_7 \\
dX_5/dt &= \varphi(T) X_1 - k_{13}^+ X_5 X_7 + k_{13}^- X_{10} \\
dX_6/dt &= -k_3^+ X_3 X_6 + k_3^- X_4 + k_8 X_4 X_7 \\
dX_7/dt &= -k_5^+ X_1 X_7 + k_5^- X_9 - k_{11}^+ X_7 X_{14} + k_{11}^- X_{12} - k_8 X_4 X_7 - k_6 X_2 X_7 \\
&\quad -k_{13}^+ X_5 X_7 + (k_{13}^- + k_{14}) X_{10} - (\varphi(T) + k_9) X_7 - k_{15}^+ X_7 X_8 \\
&\quad -k_7 X_3 X_7 + (k_{15}^- + 2 k_{16}) X_{11} + k_{12} X_{12} \\
dX_8/dt &= k_4 X_4 + \varphi(T) X_7 - k_{15}^+ X_7 X_8 + k_{15}^- X_{11} \\
dX_9/dt &= k_5^+ X_1 X_7 - k_5^- X_9 + k_8 X_4 X_7 + k_6 X_2 X_7 + k_7 X_3 X_7 \\
dX_{10}/dt &= k_{13}^+ X_5 X_7 - (k_{13}^- + k_{14}) X_{10} \\
dX_{11}/dt &= k_{15}^+ X_7 X_8 - (k_{15}^- + k_{16}) X_{11} \\
dX_{12}/dt &= k_{11}^+ X_7 X_{14} - (k_{11}^- + k_{12}) X_{12} \\
dX_{13}/dt &= k_{12} X_{12} - \varphi(T) X_{13} \\
dX_{14}/dt &= -k_{11}^+ X_7 X_{14} + k_{11}^- X_{12} + \varphi(T) X_{13}
\end{aligned}
$$

The rate coefficient of protein misfolding $\varphi(T)$ with respect to temperature $T$ has been investigated experimentally in [12, 13], and a mathematical expression describing the relation has been proposed in [11]. After adapting this formula in [11] to the time unit of our mathematical model (second), we obtain the following misfolding rate coefficient:

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

where $T$ is the numerical value of the temperature of the environment in Celsius degrees. The formula is valid for $37 \leq T \leq 45$.

For the numerical fit of the model, data of [7] on DNA binding at $42°C$ was used to relate it to $\mathsf{hsf_3}$: $\mathsf{hse}$. Moreover, the initial values of the model were sought so that they give a steady state of the model at $37°C$. This latter restriction was imposed since the heat shock response is absent at $37°C$. Once suitable numerical values for the parameters were found, the model was subjected to a number of other validation tests. For a detailed discussion on the fit and the validation of the model we refer to [14] and [15]. The final numerical setup of the model is shown in Tables 1 and 2.

## 3 Simplifying the model

We discuss in this section a series of numerical observations leading to several simplifications we can operate on our model, without changing its numerical behavior, in particular without losing its fit and validation against experimental data. We then discuss the extent to which these simplifications are dependent on the numerical values of our parameters.

The first observation is that the variables mhsf and hsp: mhsf both assume negligible numerical values throughout numerical simulations for temperatures from $37°C$ to $45°C$. Even when their initial values are increased to higher values, e.g. to 100 each, their numerical convergence towards their steady state values is very fast. Moreover, if the increase in the initial values of mhsf and hsp: mhsf is so that the total amount of hsf remains unchanged, then the experimental fit and validation of the model remain largely unchanged. The reason for this behavior is that the reactions having mhsf as a product, i.e. reactions 15 and the reverse reaction 12, have a negligible flux rate, primarily due to the form of the protein misfolding law, see (1). Consequently, the reaction producing hsp: mhsf, i.e. reaction 12, also has negligible flux rate. On the other hand, the reactions having mhsf and hsp: mhsf as reactants potentially have much higher flux rates because of larger kinetic constants and high levels of hsp, a co-reactant in reaction 12. We decide then to eliminate both mhsf and hsp: mhsf from the model, along with the reactions where they take part in, i.e., reactions 15, 12, and 16.

Note now that the situation is somewhat similar for hsf, $\mathsf{hsf}_2$ and $\mathsf{hsf}_3$: they all assume small (albeit not negligible) values throughout numerical simulations. There is however a crucial difference which points to the significance of having all three variables in the model: when increasing the initial level of $\mathsf{hsf}_3$, even in such a way that the total level of hsf is unchanged, the fit to experimental data (with respect to $\mathsf{hsf}_3$: hse) is drastically changed.

The observation that the flux of the hsf misfolding reaction is negligible was the main rationale behind eliminating mhsf and hsp: mhsf from the model. This leads to the observation that the flux of the hsp misfolding reaction, leading to the formation of mhsp is also negligible. The case of mhsp is however different because it is the end product of a second reaction, 4. Moreover, mhsp plays a central role in our model, being the source of all induced hsp through reactions 4, 18 and 9. The numerical values assumed by mhsp throughout simulations for temperatures between $37°C$ and $45°C$ are small, but not negligible. They are however negligible relative to the total level of hsp. Moreover, the numerical convergence of mhsp towards its steady state value is very fast, even in the case when the initial level of mhsp is increased several folds. This points to the observation that mhsp plays the role of a transient state towards hsp, having a very high turnover rate. As such, it could be eliminated from the model if only mhsp were replaced in reaction 4 with hsp. Consequently, we eliminate mhsp from the model, along with reactions 17, 18 and 9. At the same time, we replace reaction 4 with

$4'.\ \mathsf{hsf}_3\mathsf{:\,hse} \rightarrow \mathsf{hsf}_3\mathsf{:\,hse} + \mathsf{hsp}$

The simplified molecular model has only 10 variables and 12 reactions, compared to 14 variables and 18 reactions in the initial model. Moreover, the numerical simulations of the simplified model for temperatures between $37°C$ and $45°C$ are indistinguishable from those of the initial model.

Regarding the biological relevance, the simplified model differs from the initial model in ignoring the misfolded form of hsf and hsp, as well as ignoring that newly synthesized proteins often need chaperons to form their native fold. Excluding the misfolding of hsf and hsp is reasonable because the levels of misfolded hsf and hsp are negligible with respect to the level of mfp and thus, their competition for the chaperon resources of the cell is insignificant. Excluding the role of chaperons in assisting the formation of the native fold of newly synthesized proteins is justified by the high speed of the reaction, relative to the speed of the other reactions in our model. As such, the complex chaperon - newly synthesized protein is a very fast transient stage in the model and can be ignored.

It should be noted that the simplifications we have made on the model are based on numerical arguments and so, in principle, they are dependant on the numerical values of the parameters of the model. To test the robustness of the model reductions against changes in the numerical setup of the model, we perform several tests. In each test, we either change the initial values of some variables, or we change the values of some kinetic rate constants. For the new numerical setup we set the initial values all variables to their steady state values at $37°C$, similarly as done in [15] (to underline that the heat shock response is missing at $37°C$). Finally, we compare the numerical behavior of the model with that of its simplified version obtained as above, for temperatures between $37°C$ and $45°C$.

We first consider a numerical setup where the total level of hsf is increased by 1000 to a value of around 2400. In a second test, we increase both the total level of hsf by 1000 and the total level of hse by 100. In both test, the numerical behaviors of the models and those of their simplified versions are undistinguishable. In a third test, we increase the total level of hsp by 1000. When estimating the steady state values of the model at $37°C$, we note that they are identical with those of the initial model, summarized in Table 1. This raises an intriguing problem of independent interest: is the steady state of the model independent of the initial total level of hsp?

A test where the complex chaperon - misfolded protein is made more unstable by increasing the kinetic rate constant $k_{11}^{-}$ to 25 yields a numerically equivalent simplified model. In a final test, we decrease the value of the kinetic rate constant $k_{12}$ of the refolding reaction 11 from almost 14 to 1. In this way, we induce a great increase in the values of misfolded proteins of all types to test whether eliminating mhsf and mhsp is still possible in this context. It turns out that eliminating mhsf and hsp: mhsf is possible and yields a numerically equivalent simplified model. On the other hand, eliminating mhsp and hsp: mhsp changes the behavior of the model pronouncedly. E.g., mfp peaks at a lower value showing that the simplified model, where hsp is not subject to misfolding, is more efficient in fighting off the accumulation of mfp. A main reason why the elimination of misfolded hsp fails is because, unlike in the previous tests, the change in the

refolding rate is not accounted for when setting the initial values of the variables to the steady state values at $37°C$, since the refolding reaction has a negligible flux at that temperature. At $42°C$ however, protein refolding, in particular that of mhsp, becomes very important and removing it from the model makes a big difference.

## 4 Discussion

Having simple models is very important for analyzing their mathematical properties and for their integration into larger models. In the case of the heat shock response, adding the phosphorylation of hsf and all its polymers, and their influence on gene transcription leads to a combinatorial explosion in the number of variables of the model. As such, decreasing the number of variables, in particular the elimination of mhsf and hsp: mhsf reduces the difficulty of the problem. The main difficulty in designing a simple biomodel is that the decision to exclude variables and reactions form the model is most often done at the early stage of considering the molecular basis of the model. At that stage it is crucial to make sure that all aspects of potential interest are included in the model. Appreciating the potentially insignificant contribution of some of the aspects is very difficult at that stage, without having first a well-validated numerical setup for the model. The approach we have discussed in this paper takes an intermediate view: one may start with a rich model that is first numerically fit and validated against experimental data and then is subject to a numerical analysis that can point to some components that can be safely eliminated. The result is a model that remains faithful to the biological data and soundly ignores a number of aspects of the biological reality.

Several aspects contribute to the model simplification in a given numerical setup. Most importantly is that we eliminate variables that have a fast convergence to their steady state values. This procedure is often referred to as a time-separation principle, see []. Another factor is the flux rate of the reactions producing certain variables of the model. If the total flux contributing to producing a given variable is very small, then that variable will converge fast to a negligible value and can be eliminated from the model. There are at least two reasons why a flux rate can be small: a small kinetic constant, or much higher kinetic constant in reactions using some of the same reactants. In the context of the heat shock response model, one more factor plays an important role: the condition that the initial values of all variables are a steady state of the model at $37°C$. It turns out that the model has an interesting property, formulated as a theorem in the appendix: the steady state values of most of its variables are independent of the temperature. In this way, several of the variables of the model start from their steady state values and witness only minor numerical disturbances before returning to the same values.

The model simplification discussed in this paper is dependant on the numerical setup of the model: on the initial values of the variables and on the numerical values of the kinetic constants. Even if the initial and the simplified

models appear to be numerically equivalent in one particular setup, they may be very different in other setups. To evaluate the robustness of the model simplifications, one should compare the two models in several numerical setups, spanning the domain of expected values for the model parameters. Some of the simplifications may turn out to be robust against numerical variations, as it is the case with eliminating hsf and mhsf in the heat shock model, while others may be valid only in special numerical setups.

*Acknowledgments* This work has been partially supported by projects 108421 and 203667 (to I.P.) from Academy of Finland. The numerical simulations and estimations discussed in this paper were performed with Copasi, see [5].

## References

1. Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Essential Cell Biology, 2nd edition*. Garland Science, 2004.
2. Daniel R. Ciocca and Stuart K. Calderwood. Heat shock proteins in cancer: diagnostic, prognostic, predictive, and treatment implications. *Cell Stress and Chaperones*, 10(2):86–103, 2005.
3. H. El-Samad, H. Kurata, J. Doyle, C.A. Gross, and M. Khamash. Surviving heat shock: control strategies for robustness and performance. *PNAS*, 102(8):2736–2741, 2005.
4. H. El-Samad, S. Prajna, A. Papachristodoulu, M. Khamash, and J. Doyle. Model validation and robust stability analysis of the bacterial heat shock response using sostools. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, pages 3766–3741, 2003.
5. Stefan Hoops, Sven Sahle, Ralph Gauges, Christine Lee, Jrgen Pahle, Natalia Simus, Mudita Singhal, Liang Xu, Pedro Mendes, and Ursula Kummer. Copasi – a COmplex PAthway SImulator. *Bioinformatics*, 22(24):3067–3074, 2006.
6. Harm K. Kampinga. Thermotolerance in mammalian cells: protein denaturation and aggregation, and stress proteins. *J. Cell Science*, 104:11–17, 1993.
7. Michael P. Kline and Richard I. Morimoto. Repression of the heat shock factor 1 transcriptional activation domain is modulated by constitutive phosphorylation. *Molecular and Cellular Biology*, 17(4):2107–2115, 1997.
8. H. Kurata, H. El-Samad, T.M. Yi, M. Khamash, and J. 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.
9. Bei Liu, Anna M. DeFilippo, and Zihai Li. Overcomming immune toerance to cancer by heat shock protein vaccines. *Molecular cancer therapeutics*, 1:1147–1151, 2002.
10. Katalin V. Lukacs, Olivier E. Pardo, M.Jo Colston, Duncan M. Geddes, and Eric WFW Alton. Heat shock proteins in cancer therapy. In Habib, editor, *Cancer Gene Therapy: Past Achievements and Future Challenges*, pages 363–368. 2000.
11. A. Peper, C.A. Grimbergent, J.A.E. Spaan, J.E.M. Souren, and R. van Wijk. A mathematical model of the hsp70 regulation in the cell. *Int. J. Hyperthermia*, 14:97–124, 1997.
12. James R. Lepock, Harold E. Frey, and Kenneth P. Ritchie. Protein denaturation in intact hepatocytes and isolated cellular organelles during heat shock. *The Journal of Cell Biology*, 122(6):1267–1276, 1993.

13. James R. Lepock, Harold E. Frey, A. Michael Rodahl, and Jack 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.
14. Ion Petre, Andrzej Mizera, Claire L. Hyder, Andrey Mikhailov, John E. Eriksson, Lea Sistonen, and Ralph-Johan Back. A new mathematical model for the heat shock response. In J. Kok, editor, *Algorithmic bioprocesses*, Natural Computing. Springer, 2008.
15. Ion Petre, Claire L. Hyder, Andrzej Mizera, Andrey Mikhailov, John E. Eriksson, Lea Sistonen, and Ralph-Johan Back. A simple mathematical model for the eukaryotic heat shock response. *Manuscript*, 2009.
16. A. Graham Pockley. Heat shock proteins as regulators of the immune response. *The Lancet*, 362(9382):469–476, 2003.
17. Theodore R. Rieger, Richard I. Morimoto, and Vassily Hatzimanikatis. Mathematical modeling of the eukaryotic heat shock response: Dynamics of the hsp70 promoter. *Biophysical Journal*, 88(3):1646–58, 2005.
18. Claire J. Tomlin and Jeffrey D. Axelrod. Understanding biology by reverse engineering the control. *PNAS*, 102(12):4219–4220, 2005.
19. Paul Workman and Emmanuel de Billy. Putting the heat on cancer. *Nature Medicine*, 13(12):1415–1417, 2007.

## 5 Appendix

The next theorem formulates an interesting property of the heat shock response model. We formulate the property for the simplified model of the heat shock response.

**Theorem 1.** *Let* $c^1 = (c^1_1, c^1_2, c^1_3, \ldots, c^1_{10})$ *be a steady state of the simplified model at temperature* $T_1$ *and* $c^2 = (c^2_1, c^2_2, c^2_3, \ldots, c^2_{10})$ *a steady state at temperature* $T_2$*, where* $c^1_i$ *and* $c^2_i$ *for* $i = 1, \ldots, 10$ *are steady state concentrations of metabolite* $X_i$ *at temperatures* $T_1$ *and* $T_2$ *respectively. Then* $c = (c^1_1, \ldots, c^1_7, c^2_8, c^2_9, c^2_{10})$ *is a steady state of the system at temperature* $T_2$*.*

*Proof.* Let $c^1$ and $c^2$ be steady states at temperatures $T_1$ and $T_2$, respectively. Further, let us split the system of differential equations (1)-(1) into two subsystems: one containing equations (1)-(1) and the other consisting of equations (1)-(1). Eq. (1) is the only equation of the first subsystem with right-hand side containing functions defined by the second subsystem, i.e. $X_8(t)$, $X_9(t)$ and $X_{10}(t)$, and can be by (1) rewritten in the following form:

$$\begin{aligned} dX_6/dt = {} & k_4\, X_4 - k_5^+\, X_1\, X_6 + k_5^-\, X_7 - k_8\, X_4\, X_6 - k_6\, X_2\, X_6 \\ & - k_7\, X_3\, X_6 - k_9\, X_6 - dX_8/dt. \end{aligned} \tag{2}$$

When considering the steady states, the left-hand sides of (1)-(1) are set to 0 and in consequence equation (2) can be written as

$$0 = k_4\, X_4 - k_5^+\, X_1\, X_6 + k_5^-\, X_7 - k_8\, X_4\, X_6 - k_6\, X_2\, X_6 - k_7\, X_3\, X_6 - k_9\, X_6.$$

This algebraic relation does not contain any of functions $X_8(t)$, $X_9(t)$ or $X_{10}(t)$ and hence the steady state algebraic relations of subsystem (1)-(1) become independent of them. As a consequence, the relations do not contain temperature as a parameter and are the same both for $T_1$ and $T_2$. Since the same equations have the same solutions, it follows that $c = (c_1^1, \ldots, c_7^1, c_8^2, c_9^2, c_{10}^2)$ is a steady state of the whole system at temperature $T_2$.

The biological significance of Theorem 1 deserves some comments. Even though the cell approaches similar steady state levels regardless of the temperature values, the *time* it takes to arrive in a certain neighborhood of the steady state is longer for higher temperature values. Even if one starts in the steady state, the *effort* required of the cell is higher for higher temperatures: the fluxes of all reactions are higher for higher temperatures. The intuitive reason for this is that the misfolding rate is vastly accelerated for higher temperatures, eventually accelerating all other reactions.