TUCS

Ion Petre | Andrzej Mizera | Claire Hyder | Andrey Mikhailov |
John Eriksson | Lea Sistonen | Ralph-Johan Back

# A new mathematical model for the heat shock response

Turku Centre *for* Computer Science

TUCS Technical Report
No 883, April 2008

# A new mathematical model for the heat shock response

Ion Petre
> Åbo Akademi University, Department of Information Technologies,
> Turku Centre for Computer Science,
> Academy of Finland, (`ipetre@abo.fi`)

Andrzej Mizera
> Åbo Akademi University, Department of Information Technologies,
> Turku Centre for Computer Science, (`amizera@abo.fi`)

Claire Hyder
> Åbo Akademi University, Department of Biochemistry,
> Turku Centre for Biotechnology, (`chyder@btk.fi`)

Andrey Mikhailov
> Åbo Akademi University, Department of Biochemistry,
> Turku Centre for Biotechnology, (`andrey.mikhailov@btk.fi`)

John Eriksson
> Åbo Akademi University, Department of Biochemistry,
> Turku Centre for Biotechnology, (`john.eriksson@btk.fi`)

Lea Sistonen
> Åbo Akademi University, Department of Biochemistry,
> Turku Centre for Biotechnology,
> Academy of Finland, (`lea.sistonen@btk.fi`)

Ralph-Johan Back
> Åbo Akademi University, Department of Information Technologies,
> Turku Centre for Computer Science,
> Academy of Finland, (`backrj@abo.fi`)

**Abstract**

We present in this paper a novel molecular model for the gene regulatory network responsible for the eukaryotic heat shock response. Our model includes the temperature-induced protein misfolding, the chaperone activity of the heat shock proteins and the backregulation of their gene transcription. We then build a mathematical model for it, based on ordinary differential equations. Finally, we discuss the parameter fit and the implications of the sensitivity analysis for our model.

**TUCS Laboratory**
Computational Biomodelling Laboratory

# 1 Introduction

One of the most impressive algorithmic-like bioprocesses in living cells, crucial for the very survival of cells is the *heat shock response*: the reaction of the cell to elevated temperatures. One of the effects of raised temperature in the environment is that proteins get misfolded, with a rate that is exponentially dependent on the temperature. As an effect of their hydrophobic core being exposed, misfolded proteins tend to form bigger and bigger aggregates, with disastrous consequences for the cell, see [1]. To survive, the cell needs to increase quickly the level of chaperons (proteins that are assisting in the folding or refolding of other proteins). Once the heat shock is removed, the cell eventually re-establishes the original level of chaperons, see [10, 18, 22].

The heat shock response has been subject of intense research in the last few years, for at least three reasons. First, it is a well-conserved mechanism across all eukaryotes, while bacteria exhibit only a slightly different response, see [5, 12, 23]. As such, it is a good candidate for studying the engineering principle of gene regulatory networks, see [4, 5, 12, 25]. Second, it is a tempting mechanism to model mathematically, since it involves only very few reactants, at least in a simplified presentation, see [18, 19, 22]. Third, the heat shock proteins (the main chaperons involved in the eukaryotic heat shock response) play a central role in a large number of regulatory and of inflammatory processes, as well as in signaling, see [9, 20]. Moreover, they contribute to the resilience of cancer cells, which makes them attractive as targets for cancer treatment, see [3, 15, 16, 27].

We focus in this paper on a new molecular model for the heat shock response, proposed in [19]. We consider here a slight extension of the model in [19] where, among others, the chaperons are also subject to misfolding. After introducing the molecular model in Section 2, we build a mathematical model in Section 3, including the fitting of the model with respect to experimental data. We discuss in Section 4 the results of the sensitivity analysis of the model, including its biological implications.

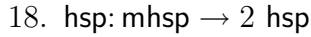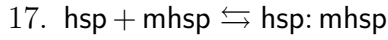# 2 A new molecular model for the eukaryotic heat shock response

The heat shock proteins (hsp) play the key role in the heat shock response. They act as chaperons, helping misfolded proteins (mfp) to refold. The response is controlled in our model through the regulation of the transactivation of the hsp-encoding genes. The transcription of the gene is promoted by some proteins called heat shock factors (hsf) that trimerize and then bind to a specific DNA sequence called heat shock element (hse), upstream of the hsp-encoding gene. Once the hsf trimer is bound to the heat shock element,

the gene is transactivated and the synthesis of hsp is thus switched on (for the sake of simplicity, the role of RNA is ignored in our model). Once the level of hsp is high enough, the cell has an ingenious mechanism to switch off the hsp synthesis. For this, hsp bind to free hsf, as well as break the hsf trimers (including those bound to hse, promoting the gene activation), thus effectively halting the hsp synthesis.

Under elevated temperatures, some of the proteins (prot) in the cell get misfolded. The heat shock response is then quickly switched on simply because the heat shock proteins become more and more active in the refolding process, thus leaving the heat shock factors free and able to promote the synthesis of more heat shock proteins. Note that several types of heat shock proteins exist in an eukaryotic cell. We treat them all uniformly in our model, with hsp70 as common denominator. The same comment applies also to the heat shock factors.

Our molecular model for the eukaryotic heat shock response consists of the following molecular reactions:

1. $2 \text{ hsf} \leftrightarrows \text{hsf}_2$

2. $\text{hsf} + \text{hsf}_2 \leftrightarrows \text{hsf}_3$

3. $\text{hsf}_3 + \text{hse} \leftrightarrows \text{hsf}_3\text{: hse}$

4. $\text{hsf}_3\text{: hse} \rightarrow \text{hsf}_3\text{: hse} + \text{mhsp}$

5. $\text{hsp} + \text{hsf} \leftrightarrows \text{hsp: hsf}$

6. $\text{hsp} + \text{hsf}_2 \rightarrow \text{hsp: hsf} + \text{hsf}$

7. $\text{hsp} + \text{hsf}_3 \rightarrow \text{hsp: hsf} + 2 \text{ hsf}$

8. $\text{hsp} + \text{hsf}_3\text{: hse} \rightarrow \text{hsp: hsf} + 2 \text{ hsf} + \text{hse}$

9. $\text{hsp} \rightarrow \emptyset$

10. $\text{prot} \rightarrow \text{mfp}$

11. $\text{hsp} + \text{mfp} \leftrightarrows \text{hsp: mfp}$

12. $\text{hsp: mfp} \rightarrow \text{hsp} + \text{prot}$

13. $\text{hsf} \rightarrow \text{mhsf}$

14. $\text{hsp} \rightarrow \text{mhsp}$

15. $\text{hsp} + \text{mhsf} \leftrightarrows \text{hsp: mhsf}$

16. $\text{hsp: mhsf} \rightarrow \text{hsp} + \text{hsf}$

17. $\mathsf{hsp} + \mathsf{mhsp} \leftrightarrows \mathsf{hsp} : \mathsf{mhsp}$

18. $\mathsf{hsp} : \mathsf{mhsp} \rightarrow 2\,\mathsf{hsp}$

It is important to note that the main addition we consider here with respect to the model in [19] is to include the misfolding of $\mathsf{hsp}$ and $\mathsf{hsf}$. This is, in principle, no minor extension since in the current model the repairing mechanism is subject to failure, but it is capable to fix itself.

Several criteria were followed when introducing this molecular model:

(i) as few reactions and reactants as possible;

(ii) include the temperature-induced protein misfolding;

(iii) include $\mathsf{hsf}$ in all its three forms: monomers, dimers, and trimers;

(iv) include the $\mathsf{hsp}$-backregulation of the transactivation of the $\mathsf{hsp}$-encoding gene;

(v) include the chaperon activity of $\mathsf{hsp}$;

(vi) include only well-documented, textbook-like reactions and reactants.

For the sake of keeping the model as simple as possible, we are ignoring a number of details. E.g., note that there is no notion of locality in our model: we make no distinction between the place where gene transcription takes place (inside nucleus) and the place where protein synthesis takes place (outside nucleus). Note also that protein synthesis and gene transcription are greatly simplified in reaction 4: we only indicate that once the gene is transactivated, protein synthesis is also switched on. On the other hand, reaction 4 is faithful to the biological reality, see [1] in indicating that newly synthesized proteins often need chaperons to form their native fold.

As far as protein degradation is concerned, we only consider it in the model for $\mathsf{hsp}$. If we considered it also for $\mathsf{hsf}$ and $\mathsf{prot}$, then we should also consider the compensating mechanism of protein synthesis, including its control. For the sake of simplicity and also based on experimental evidence that the total amount of $\mathsf{hsf}$ and of $\mathsf{prot}$ is somewhat constant, we ignore the details of synthesis and degradation for $\mathsf{hsf}$ and $\mathsf{prot}$.

# 3   The mathematical model

We build in this section a mathematical model associated to the molecular model 1–18. Our mathematical model is in terms of coupled ordinary differential equations and its formulation is based on the principle of mass-action.

## 3.1 The principle of mass-action

The mass-action law is widely used in formulating mathematical models in physics, chemistry, and engineering. Introduced in [6, 7], it can be briefly summarized as follows: *the rate of each reaction is proportional to the concentration of reactants.* In turn, the rate of each reaction gives the rate of consuming the reactants and the rate of producing the products. E.g., for a reaction

$$R_1 : \ A + B \rightarrow C,$$
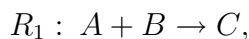
the rate according to the principle of mass action is $f_1(t) = kA(t)B(t)$, where $k \geq 0$ is a constant and $A(t)$, $B(t)$ are functions of time giving the level of the reactants $A$ and $B$, respectively. Consequently, the rate of consuming $A$ and $B$, and the rate of producing $C$ is expressed by the following differential equations:

$$\frac{dA}{dt} = \frac{dB}{dt} = -k \, A(t) \, B(t), \qquad \frac{dC}{dt} = k \, A(t) \, B(t).$$

For a reversible reaction

$$R_2 : \ A + B \leftrightarrows C,$$
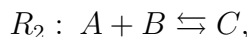
the rate is $f_2(t) = k_1 \, A(t) \, B(t) - k_2 \, C(t)$, for some constants $k_1$, $k_2 \geq 0$. The differential equations are written in a similar way:

$$\frac{dA}{dt} = \frac{dB}{dt} = -f_2(t), \qquad \frac{dC}{dt} = f_2(t). \tag{*}$$

For a set of coupled reactions, the differential equations capture the combined rate of consuming and producing each reactant as an effect of all reactions taking place simultaneously. E.g., for reactions

$$R_3 : \ A + B \leftrightarrows C, \qquad R_4 : \ B + C \leftrightarrows A, \qquad R_5 : \ A + C \leftrightarrows B,$$
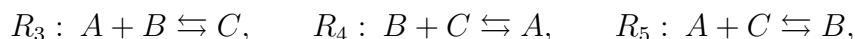
the associated system of differential equations is

$$
\begin{aligned}
dA/dt &= -f_3(t) + f_4(t) - f_5(t), \\
dB/dt &= -f_3(t) - f_4(t) + f_5(t), \\
dC/dt &= f_3(t) - f_4(t) - f_5(t),
\end{aligned}
$$

where $f_i(t)$ is the rate of reaction $R_i$, for all $3 \leq i \leq 5$, formulated according to the principle of mass action.

We recall that for a system of differential equations

$$
\begin{aligned}
\frac{dX_1}{dt} &= f_1(X_1, \ldots, X_n), \\
&\cdots \\
\frac{dX_n}{dt} &= f_n(X_1, \ldots, X_n),
\end{aligned}
$$

4

Table 1: The list of variables in the mathematical model, their initial values, and their values in one of the steady states of the system, for $T = 42$. Note that the initial values give one of the steady states of the system for $T = 37$.

| Metabolite | Variable | Initial value | A steady state (T=42) |
|---|---|---|---|
| hsf | $X_1$ | 0.669 | 0.669 |
| hsf$_2$ | $X_2$ | $8.73 \cdot 10^{-4}$ | $8.73 \cdot 10^{-4}$ |
| hsf$_3$ | $X_3$ | $1.23 \cdot 10^{-4}$ | $1.23 \cdot 10^{-4}$ |
| hsf$_3$: hse | $X_4$ | 2.956 | 2.956 |
| mhsf | $X_5$ | $3.01 \cdot 10^{-6}$ | $2.69 \cdot 10^{-5}$ |
| hse | $X_6$ | 29.733 | 29.733 |
| hsp | $X_7$ | 766.875 | 766.875 |
| mhsp | $X_8$ | $3.45 \cdot 10^{-3}$ | $4.35 \cdot 10^{-2}$ |
| hsp: hsf | $X_9$ | 1403.13 | 1403.13 |
| hsp: mhsf | $X_{10}$ | $4.17 \cdot 10^{-7}$ | $3.72 \cdot 10^{-6}$ |
| hsp: mhsp | $X_{11}$ | $4.78 \cdot 10^{-4}$ | $6.03 \cdot 10^{-3}$ |
| hsp: mfp | $X_{12}$ | 71.647 | 640.471 |
| prot | $X_{13}$ | $1.14 \cdot 10^8$ | $1.14 \cdot 10^8$ |
| mfp | $X_{14}$ | 517.352 | 4624.72 |

we say that $(x_1, x_2, \ldots, x_n)$ is a *steady states* (also called *equilibrium points*) if it is a solution of the algebraic system of equations $f_i(X_1, \ldots, X_n) = 0$, for all $1 \leq i \leq n$, see [24, 28]. Steady states are particularly interesting because they characterize situations where although reactions may have non-zero rates, their combined effect is zero. In other words, the concentration of all reactants and of all products are constant.

We refer to [11, 17, 29] for more details on the principle of mass action and its formulation based on ordinary differential equations.

## 3.2   Our mathematical model

Let $\mathbb{R}_+$ be the set of all positive real numbers and $\mathbb{R}_+^n$ the set of all $n$-tuples of positive real numbers, for $n \geq 2$. We denote each reactant and bond between them in the molecular model 1–18 according to the convention in Table 1. We also denote by $\kappa \in \mathbb{R}_+^{17}$ the vector with all reaction rate constants as its components, see Table 2: $\kappa = (k_1^+, k_1^-, k_2^+, k_2^-, k_3^+, k_3^-, k_4, k_5^+, k_5^-, k_6, k_7, k_8, k_9, k_{11}^+, k_{11}^-, k_{12}, k_{13}^+, k_{13}^-, k_{14}, k_{15}^+, k_{15}^-, k_{16})$.

The mass action-based formulation of the associated mathematical model in terms of differential equations is straightforward, leading to the following system of equations:

Table 2: The numerical values for the fitted model.

| Kinetic constant | Reaction | Numerical value |
|---|---|---|
| $k_1^+$ | (1), forward | 3.49091 |
| $k_1^-$ | (1), backward | 0.189539 |
| $k_2^+$ | (2), forward | 1.06518 |
| $k_2^-$ | (2), backward | $1 \cdot 10^{-9}$ |
| $k_3^+$ | (3), forward | 0.169044 |
| $k_3^-$ | (3), backward | $1.21209 \cdot 10^{-6}$ |
| $k_4$ | (4) | 0.00830045 |
| $k_5^+$ | (5), forward | 9.73665 |
| $k_5^-$ | (5), backward | 3.56223 |
| $k_6$ | (6) | 2.33366 |
| $k_7$ | (7) | $4.30924 \cdot 10^{-5}$ |
| $k_8$ | (8) | $2.72689 \cdot 10^{-7}$ |
| $k_9$ | (9) | $3.2 \cdot 10^{-5}$ |
| $k_{11}^+$ | (11), forward | 0.00331898 |
| $k_{11}^-$ | (11), backward | 4.43952 |
| $k_{12}$ | (12) | 13.9392 |
| $k_{13}^+$ | (15), forward | 0.00331898 |
| $k_{13}^-$ | (15), backward | 4.43952 |
| $k_{14}$ | (16) | 13.9392 |
| $k_{15}^+$ | (17), forward | 0.00331898 |
| $k_{15}^-$ | (17), backward | 4.43952 |
| $k_{16}$ | (18) | 13.9392 |

$$
\begin{aligned}
dX_1/dt &= f_1(X_1, X_2, \ldots, X_{14}, \kappa) & (1)\\
dX_2/dt &= f_2(X_1, X_2, \ldots, X_{14}, \kappa) & (2)\\
dX_3/dt &= f_3(X_1, X_2, \ldots, X_{14}, \kappa) & (3)\\
dX_4/dt &= f_4(X_1, X_2, \ldots, X_{14}, \kappa) & (4)\\
dX_5/dt &= f_5(X_1, X_2, \ldots, X_{14}, \kappa) & (5)\\
dX_6/dt &= f_6(X_1, X_2, \ldots, X_{14}, \kappa) & (6)\\
dX_7/dt &= f_7(X_1, X_2, \ldots, X_{14}, \kappa) & (7)\\
dX_8/dt &= f_8(X_1, X_2, \ldots, X_{14}, \kappa) & (8)\\
dX_9/dt &= f_9(X_1, X_2, \ldots, X_{14}, \kappa) & (9)\\
dX_{10}/dt &= f_{10}(X_1, X_2, \ldots, X_{14}, \kappa) & (10)\\
dX_{11}/dt &= f_{11}(X_1, X_2, \ldots, X_{14}, \kappa) & (11)\\
dX_{12}/dt &= f_{12}(X_1, X_2, \ldots, X_{14}, \kappa) & (12)\\
\\
dX_{13}/dt &= f_{13}(X_1, X_2, \ldots, X_{14}, \kappa) & (13)\\
dX_{14}/dt &= f_{14}(X_1, X_2, \ldots, X_{14}, \kappa) & (14)
\end{aligned}
$$

where

$$
\begin{aligned}
f_1 &= -k_2^+ X_1 X_2 + k_2^- X_3 - k_5^+ X_1 X_7 + k_5^- X_9 + 2\,k_8 X_4 X_7 + k_6 X_2 X_7 \\
&\quad -\varphi(T)\,X_1 + k_{14} X_{10} + 2\,k_7 X_3 X_7 - 2\,k_1^+ X_1^2 + 2\,k_1^- X_2 \\
f_2 &= -k_2^+ X_1 X_2 + k_2^+ X_3 - k_6 X_2 X_7 + k_1^+ X_1^2 - k_1^- X_2 \\
f_3 &= -k_3^+ X_3 X_6 + k_2^+ X_1 X_2 - k_2^- X_3 + k_3^- X_4 - k_7 X_3 X_7 \\
f_4 &= k_3^+ X_3 X_6 - k_3^- X_4 - k_8 X_4 X_7 \\
f_5 &= \varphi(T)\,X_1 - k_{13}^+ X_5 X_7 + k_{13}^- X_{10} \\
f_6 &= -k_3^+ X_3 X_6 + k_3^- X_4 + k_8 X_4 X_7 \\
f_7 &= -k_5^+ X_1 X_7 + k_5^- X_9 - k_{11}^+ X_7 X_{14} + k_{11}^- X_{12} - k_8 X_4 X_7 - k_6 X_2 X_7 \\
&\quad -k_{13}^+ X_5 X_7 + (k_{13}^- + k_{14}) X_{10} - (\varphi(T) + k_9) X_7 - k_{15}^+ X_7 X_8 \\
&\quad -k_7 X_3 X_7 + (k_{15}^- + 2\,k_{16}) X_{11} + k_{12} X_{12} \\
f_8 &= k_4 X_4 + \varphi(T)\,X_7 - k_{15}^+ X_7 X_8 + k_{15}^- X_{11} \\
f_9 &= k_5^+ X_1 X_7 - k_5^- X_9 + k_8 X_4 X_7 + k_6 X_2 X_7 + k_7 X_3 X_7 \\
f_{10} &= k_{13}^+ X_5 X_7 - (k_{13}^- + k_{14}) X_{10} \\
f_{11} &= k_{15}^+ X_7 X_8 - (k_{15}^- + k_{16}) X_{11} \\
f_{12} &= k_{11}^+ X_7 X_{14} - (k_{11}^- + k_{12}) X_{12} \\
f_{13} &= k_{12} X_{12} - \varphi(T)\,X_{13} \\
f_{14} &= -k_{11}^+ X_7 X_{14} + k_{11}^- X_{12} + \varphi(T)\,X_{13}
\end{aligned}
$$

The rate of protein misfolding $\varphi(T)$ with respect to temperature $T$ has been investigated experimentally in [13, 14], and a mathematical expression for it has been proposed in [18]. We have adapted the formula in [18] to obtain the following misfolding rate per second:

$$\varphi(T) = (1 - \frac{0.4}{e^{T-37}}) \cdot 0.8401033733 \cdot 10^{-6} \cdot 1.4^{T-37} \; s^{-1},$$

where $T$ is the temperature of the environment in Celsius degrees, with the formula being valid for $37 \leq T \leq 45$.

The following result gives three mass-conservation relations for our model.

**Theorem 3.1.** *There exists $K_1$, $K_2$, $K_3 \geq 0$ such that:*

*(i)* $X_1(t) + 2\,X_2(t) + 3\,X_3(t) + 3\,X_4(t) + X_5(t) + X_9(t) = K_1,$

*(ii)* $X_4(t) + X_6(t) = K_2,$

*(iii)* $X_{13}(t) + X_{14}(t) + X_{12}(t) = K_3,$

*for all $t \geq 0$.*

*Proof.* We only prove here part (ii), as the others may be proved analogously. For this, note that from equations (4) and (6), it follows that

$$\frac{d(X_4 + X_6)}{dt} = (f_4 + f_6)(X_1, \ldots, X_{14}, \kappa, t) = 0,$$

i.e., $(X_4 + X_6)(t)$ is a constant function. $\qquad \square$

The steady states of the model (1)-(14) satisfy the following algebraic relations, where $x_i$ is the numerical value of $X_i$ in the steady state, for all

$1 \le i \le 14$.

$$
\begin{aligned}
0 &= -k_2^+ x_1 x_2 + k_2^- x_3 - k_5^+ x_1 x_7 + k_5^- x_9 + 2 k_8 x_4 x_7 + k_6 x_2 x_7 \\
&\quad -\varphi(T) x_1 + k_{14} x_{10} + 2 k_7 x_3 x_7 - 2 k_1^+ x_1^2 + 2 k_1^- x_2 \tag{15}
\end{aligned}
$$

$$
0 = -k_2^+ x_1 x_2 + k_2^- x_3 - k_6 x_2 x_7 + k_1^+ x_1^2 - k_1^- x_2 \tag{16}
$$

$$
0 = -k_3^+ x_3 x_6 + k_2^+ x_1 x_2 - k_2^- x_3 + k_3^- x_4 - k_7 x_3 x_7 \tag{17}
$$

$$
0 = k_3^+ x_3 x_6 - k_3^- x_4 - k_8 x_4 x_7 \tag{18}
$$

$$
0 = \varphi(T) x_1 - k_{13}^+ x_5 x_7 + k_{13}^- x_{10} \tag{19}
$$

$$
0 = -k_3^+ x_3 x_6 + k_3^- x_4 + k_8 x_4 x_7 \tag{20}
$$

$$
\begin{aligned}
0 &= -k_5^+ x_1 x_7 + k_5^- x_9 - k_{11}^+ x_7 x_{14} + k_{11}^- x_{12} - k_8 x_4 x_7 - k_6 x_2 x_7 \\
&\quad -k_{13}^+ x_5 x_7 + (k_{13}^- + k_{14}) x_{10} - (\varphi(T) + k_9) x_7 - k_{15}^+ x_7 x_8 - k_7 x_3 x_7 \\
&\quad +(k_{15}^- + 2 k_{16}) x_{11} + k_{12} x_{12} \tag{21}
\end{aligned}
$$

$$
0 = k_4 x_4 + \varphi(T) x_7 - k_{15}^+ x_7 x_8 + k_{15}^- x_{11} \tag{22}
$$

$$
0 = k_5^+ x_1 x_7 - k_5^- x_9 + k_8 x_4 x_7 + k_6 x_2 x_7 + k_7 x_3 x_7 \tag{23}
$$

$$
0 = k_{13}^+ x_5 x_7 - (k_{13}^- + k_{14}) x_{10} \tag{24}
$$

$$
0 = k_{15}^+ x_7 x_8 - (k_{15}^- + k_{16}) x_{11} \tag{25}
$$

$$
0 = k_{11}^+ x_7 x_{14} - (k_{11}^- + k_{12}) x_{12} \tag{26}
$$

$$
0 = k_{12} x_{12} - \varphi(T) x_{13} \tag{27}
$$

$$
0 = -k_{11}^+ x_7 x_{14} + k_{11}^- x_{12} + \varphi(T) x_{13} \tag{28}
$$

It follows from Theorem 3.1 that only eleven of the relations above are independent. E.g., relations (15)-(17), (19), (21)-(27) are independent. The system consisting of the corresponding differential equations is called the *reduced system* of (1)-(14).

## 3.3 Fitting the model to experimental data

The experimental data available for the parameter fit is from [10] and reflects the level of DNA binding, i.e., variable $X_4$ in our model, for various time points up to 4 hours, with continuous heat shock at $42\,°C$. Additionally, we require that the initial value of the variables of the model is a steady state for temperature set to $37\,°C$. This is a natural condition since the model is supposed to reflect the reaction to temperatures raised above $37\,°C$.

Mathematically, the problem we need to solve is one of global optimization, as formulated below. For each 17-tuple $\kappa$ of positive numerical values for all kinetic constants, and for each 14-tuple $\alpha$ of positive initial values for all variables in the model, the function $X_4(t)$ is uniquely defined for a fixed temperature T. We denote the value of this function at time point $\tau$, with parameters $\kappa$ and $\alpha$ by $x_4^T(\kappa, \alpha, \tau)$. Note that this property holds for all the other variables in the model and it is valid in general for any mathematical model based on ordinary differential equations (one calls such models

*deterministic*). We denote the set of experimental data in [10] by

$$E_n = \{(t_i, r_i) \mid t_i, r_i > 0,\ 1 \leq i \leq N\},$$

where $N \geq 1$ is the number of observations, $t_i$ is the time point of each observation and $r_i$ is the value of the reading.

With this setup, we can now formulate our optimization problem as follows: find $\kappa \in \mathbb{R}_+^{17}$ and $\alpha \in \mathbb{R}_+^{14}$ such that:

(i) $f(\kappa, \alpha) = \frac{1}{N} \sum_{i=1}^{N} (x_4^{42}(\kappa, \alpha, t_i) - r_i)^2$ is minimal and

(ii) $\alpha$ is a steady state of the model for $T = 37$ and parameter values given by $\kappa$.

The function $f(\kappa, \alpha)$ is a cost function (in this case *least mean squares*), indicating numerically how the function $x_4^T(\kappa, \alpha, t)$, $t \geq 0$, compares with the experimental data.

Note that in our optimization problem, not all 31 variables (the components of $\kappa$ and $\alpha$) are independent. On one hand, we have the three algebraic relations given by Theorem 3.1. On the other hand, we have eleven more independent algebraic relations given by the steady state equations (15)-(17), (19), (21)-(27). Consequently, we have 17 independent variables in our optimization problem.

Given the high degree of the system (1)-(14), finding the analytical form of the minimum points of $f(\kappa, \alpha)$ is very challenging. This is a typical problem when the system of equations is non-linear. Adding to the difficulty of the problem is the fact that the eleven independent steady state equations cannot be solved analytically, given their high overall degree.

Since an analytical solution to the model fitting problem is often intractable, the practical approach to such problems is to give a numerical simulation of a solution. Several methods exist for this, see [2, 21]. The trade-off with all these methods is that typically they offer an estimate of a *local* optimum, with no guarantee of it being a *global* optimum.

Obtaining a numerical estimation of a local optimum for (i) is not difficult. However, such a solution may not satisfy (ii). To solve this problem, for a given local optimum $(\kappa_0, \alpha_0) \in \mathbb{R}_+^{17} \times \mathbb{R}_+^{14}$ one may numerically estimate a steady state $\alpha_1 \in \mathbb{R}_+^{14}$ for $T = 37$. Then the pair $(\kappa_0, \alpha_1)$ satisfies (ii). Unfortunately, $(\kappa_0, \alpha_1)$ may not be close to a local optimum of the cost function in (i).

Another approach is to replace the algebraic relations implicitly given by (ii) with an optimization problem similar to that in (i). Formally, we replace all algebraic relations $R_i = 0$, $1 \leq i \leq 11$, given by (ii) with the condition that

$$g(\kappa, \alpha) = \frac{1}{M} \sum_{j=1}^{M} R_i^2(\kappa, \alpha, \delta_j)$$
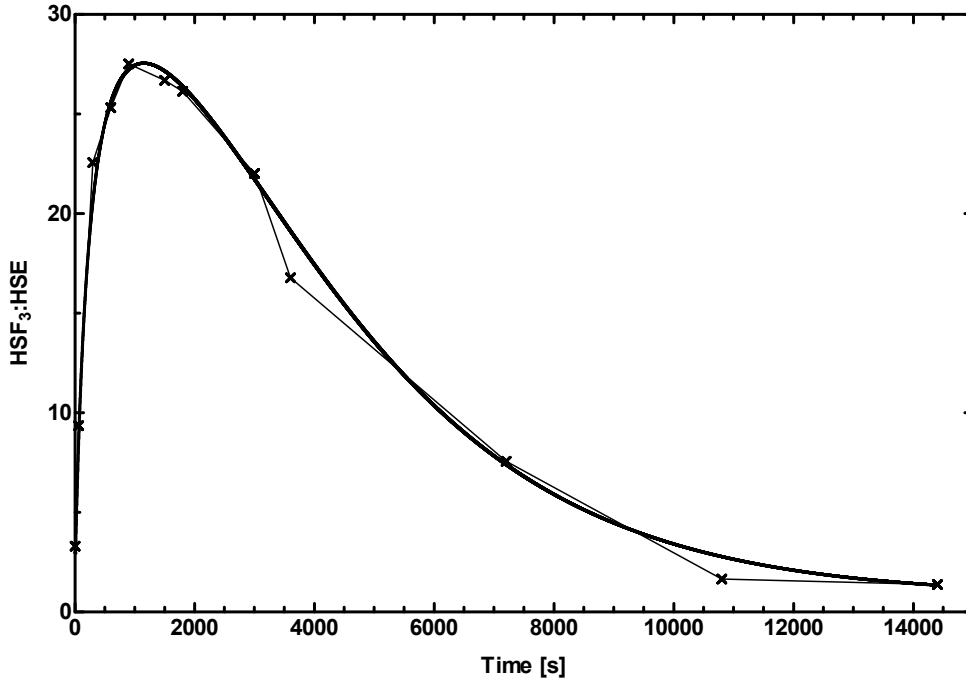
10

Figure 1: The continuous line shows a numerical estimation of function $X_4(t)$, standing for DNA binding, for the initial data in Table 1 and the parameter values in Table 2. With crossed points we indicated the experimental data of [10].

is minimal, where $0 < \delta_1 < \cdots < \delta_M$ are some arbitrary (but fixed) time points. Our problem thus becomes one of optimization with cost function $(f, g)$, with respect to the order relation $(a, b) \leq (c, d)$ if and only if $a \leq c$ and $b \leq d$. The numerical values in Table 2 give one solution to this problem obtained based on Copasi [8]. The plot in Figure 1 shows the time evolution of function $X_4(t)$ up to $t = 4$ hours, with the experimental data of [10] indicated with crosses.

The solution in Table 2 has been compared with a number of other available experimental data (such as behavior at $41\,^{\circ}\mathrm{C}$ and at $43\,^{\circ}\mathrm{C}$), as well as against qualitative, non-numerical data. The results were satisfactory and better than those of previous models reported in the literature, such as [18, 22]. For details on the model validation analysis we refer to [19].

Note that the steady state of the system of differential equations (1)-(14), for the initial values in Table 1 and the parameter values in Table 2 is *asymptotically stable*. To prove it, it is enough to consider its associated *Jacobian*:

11

$$J(t) = \begin{pmatrix} \partial f_1/\partial X_1 & \partial f_1/\partial X_2 & \dots & \partial f_1/\partial X_{14} \\ \partial f_2/\partial X_1 & \partial f_2/\partial X_2 & \dots & \partial f_2/\partial X_{14} \\ \vdots & \vdots & & \vdots \\ \partial f_{14}/\partial X_1 & \partial f_{14}/\partial X_2 & \dots & \partial f_{14}/\partial X_{14} \end{pmatrix}$$

As it is well-known, see [28, 24], a steady state is asymptotically stable if and only if all eigenvalues of the Jacobian at the steady state have negative real parts. A numerical estimation done with *Copasi* [8] shows that the steady state for $T = 42$, see Table 1, is indeed asymptotically stable.

# 4 Sensitivity analysis

Sensitivity analysis is a method to estimate the changes brought into the system through small changes in the parameters of the model. In this way one may estimate both the robustness of the model against small changes in the model, as well as identify possibilities for bringing a certain desired changed in the system. E.g., one question that is often asked of a biochemical model is what changes should be done to the model so that the new steady state satisfies certain properties. In our case we are interested in changing some of the parameters of the model so that the level of mfp in the new steady state of the system is smaller than in the standard model, thus presumably making it easier for the cell to cope with the heat shock. We also analyze a scenario in which we are interested in increasing the level of mfp in the new steady state, thus increasing the chances of the cell not being able to cope with the heat shock. Such a scenario is especially meaningful in relation with cancer cells that exhibit the properties of an excited cell, with increased levels of hsp, see [3, 15, 16, 27]. In this section we follow in part a presentation of sensitivity analysis due to [26].

We consider the partial derivatives of the solution of the system with respect to the parameters of the system. These are called *first-order local concentration sensitivity coefficients*. Second- or higher-order sensitivity analysis considering the simultaneous change of two or more parameters is also possible. If we denote $X(t, \kappa) = (X_1(t, \kappa), X_2(t, \kappa), \dots, X_{14}(t, \kappa))$ the solution of the system (1)-(14) with respect to the parameter vector $\kappa$, then the concentration sensitivity coefficients are the time functions $\partial X_i/\partial \kappa_j(t)$, for all $1 \le i \le 14$, $1 \le j \le 17$. Differentiating the system (1)-(14) with respect to $\kappa_j$ yields the following set of *sensitivity equations*:

$$\frac{d}{dt}\frac{\partial X}{\kappa_j} = J(t)\frac{\partial X}{\partial \kappa_j} + \frac{\partial f(t)}{\partial \kappa_j}, \qquad \text{for all } 1 \le j \le 17, \tag{29}$$

where $\partial X/\partial \kappa_j = (\partial X_1/\partial \kappa_j, \dots, \partial X_{14}/\partial \kappa_j)$ is the component-wise vector of partial derivatives, $f = (f_1, \dots, f_{14})$ is the model function in (1)-(14), and

12

$J(t)$ is the corresponding Jacobian. The initial condition for the system (29) is that $\partial X/\partial\kappa_j(0) = 0$, for all $1 \leq j \leq 17$.

The solution of the system (29) can be numerically integrated, thus obtaining a numerical approximation of the time evolution of the sensitivity coefficients. Very often however, the focus is on sensitivity analysis around steady states. If the considered steady state is asymptotically stable, then one may consider the limit $\lim_{t\to\infty}(\partial X/\partial\kappa_j)(t)$, called *stationary sensitivity coefficients*. They reflect the dependency of the steady state on the parameters of the model. Mathematically, they are given by a set of algebraic equations obtained from (29) by setting $d/dt(\partial X/\kappa_j) = 0$. We then obtain the following algebraic equations:

$$\left(\frac{\partial X}{\partial\kappa_j}\right) = -J^{-1}F_j, \qquad \text{for all } 1 \leq j \leq 17, \tag{30}$$

where $J$ is the value of the Jacobian at the steady state and $F_j$ is the $j$-th column of the matrix $F = (\partial f_r/\partial\kappa_s)_{r,s}$ computed at the steady state.

When used for comparing the relative effect of a parameter change in two or more variables, the sensitivity coefficients must have the same physical dimension or be dimensionless, see [26]. Most often, one simply considers the matrix $S'$ of (dimensionless) *normalized* (also called *scaled*) sensitivity coefficients:

$$S'_{ij} = \frac{\kappa_j}{X_i(t,\kappa)} \cdot \frac{\partial X_i(t,\kappa)}{\partial\kappa_j} = \frac{\partial\ln X_i(t,\kappa)}{\partial\ln\kappa_j}$$

Numerical estimations of the normalized sensitivity coefficients for a steady state may be obtained, e.g. with Copasi. For $X_{14}$ (standing for the level of mfp in the model), the most significant (with the largest module) sensitivity coefficients are the following:

- $\partial\ln(X_{14})/\partial\ln(T) = 14.24$,
- $\partial\ln(X_{14})/\partial\ln(k_1^+) = -0.16$,
- $\partial\ln(X_{14})/\partial\ln(k_2^+) = -0.16$,
- $\partial\ln(X_{14})/\partial\ln(k_5^+) = 0.49$,
- $\partial\ln(X_{14})/\partial\ln(k_5^-) = -0.49$,
- $\partial\ln(X_{14})/\partial\ln(k_6) = 0.16$,
- $\partial\ln(X_{14})/\partial\ln(k_9) = 0.15$,
- $\partial\ln(X_{14})/\partial\ln(k_{11}^+) = -0.99$,
- $\partial\ln(X_{14})/\partial\ln(k_{11}^-) = 0.24$,
- $\partial\ln(X_{14})/\partial\ln(k_{12}) = -0.24$.

These coefficients being most significant is consistent with the biological intuition that the level of mfp in the model is most dependant on the temperature (parameter $T$), on the rate of mfp being sequestered by hsp (parameters $k_{11}^+$ and $k_{11}^-$) and the rate of protein refolding (parameter $k_{12}$). However, the sensitivity coefficients also reveal less intuitive, but significant dependencies such as the one on the reaction rate of hsf being sequestered by hsp (parameters $k_5^+$ and $k_5^-$), on the rate of dissipation of hsf dimers (parameter $k_6$), or on the rate of dimer- and trimer-formation (parameters $k_1^+$ and $k_2^+$).

Note that the sensitivity coefficients reflect the changes in the steady state for *small* changes in the parameter. E.g., increasing the temperature from

13

42 with 0.1% yields an increase in the level of mfp with 1.43%, roughly as predicted by $\partial \ln(X_{14})/\partial \ln(T) = 14.24$. An increase of the temperature from 42 with 10% yields however an increase in the level of mfp of 311.93%.

A similar sensitivity analysis may also be performed with respect to the initial conditions, see [26]. If we denote by $X^{(0)} = X(0, \kappa)$, the initial values of the vector $X$, for parameters $\kappa$, then the *initial concentration sensitivity coefficients* are obtained by differentiating system (1)-(14) with respect to $X^{(0)}$:

$$\frac{d}{dt}\frac{\partial X}{\partial X^{(0)}} = J(t)\frac{\partial X}{\partial X^{(0)}}(t), \tag{31}$$

with the initial condition that $\partial X/\partial X^{(0)}(0)$ is the identity matrix. It follows then that the initial concentration sensitivity matrix is given by the following matrix exponential:

$$\frac{\partial X}{\partial X^{(0)}}(t) = e^{J(t)} = \sum_{k=0}^{\infty} \frac{J(t)^k}{k!}.$$

Similarly as for the parameter-based sensitivity coefficients, it is often useful to consider the normalized, dimensionless coefficients

$$\frac{\partial X_i}{\partial X^{(0)}{}_j}(t) \cdot \frac{X^{(0)}{}_j(t)}{X_i(t)} = \frac{\partial \ln(X_i)}{\partial \ln(X^{(0)}{}_j)}.$$

A numerical estimation of the initial concentration sensitivity coefficient of mfp around the steady state given in Table 2 for $T = 42$, shows that all are negligible except for the following two coefficients: $\partial \ln(X_{14})/\partial \ln(X_9^{(0)}) = -0.497748$ and $\partial \ln(X_{14})/\partial \ln(X_{13}^{(0)}) = 0.99$. While the biological significance of the dependency of mfp on the initial level of prot is obvious, its dependency on the initial level of hsp: hsf is perhaps not. Moreover, it turns out that several other variables have a significant dependency on the initial level of hsp: hsf:

- $\partial \ln(X_1)/\partial \ln(X_9(0)) = 0.49$,
- $\partial \ln(X_2)/\partial \ln(X_9(0)) = 0.49$,
- $\partial \ln(X_3)/\partial \ln(X_9(0)) = 1.04$,
- $\partial \ln(X_4)/\partial \ln(X_9(0)) = 0.49$,
- $\partial \ln(X_{10})/\partial \ln(X_9(0)) = 0.49$,
- $\partial \ln(X_6)/\partial \ln(X_9(0)) = -0.04$,
- $\partial \ln(X_7)/\partial \ln(X_9(0)) = 0.49$,
- $\partial \ln(X_9)/\partial \ln(X_9(0)) = 0.99$,
- $\partial \ln(X_{14})/\partial \ln(X_9(0)) = -0.49$,
- $\partial \ln(X_{11})/\partial \ln(X_9(0)) = 0.49$,

E.g., increasing $X_9^{(0)}$ by 1% increases the steady state values of $X_7$ by 0.49% and decreases the level of $X_{14}$ by 0.49%. Increasing $X_9^{(0)}$ by 10% increases the steady state values of $X_7$ by 4.85% and decreases the level of $X_{14}$ by 4.63%.

The biological interpretation of this significant dependency of the model on the initial level of hsp: hsf is based on two arguments. On one hand, the most significant part (about two thirds) of the initial available molecules of

hsp in our model are present in bonds with hsf. On the other hand, the vast majority of hsf molecules are initially bound to hsp. Thus, changes in the initial level of hsp:hsf have an immediate influence on the two main drivers of the heat shock response: hsp and hsf. Interestingly, the dependency of the model on the initial levels of either hsp or hsf is negligible.

# References

[1] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Essential Cell Biology, 2nd edition.* Garland Science, 2004.

[2] R.L. Burden and J. Douglas Faires. *Numerical Analysis.* Thomson Brooks/Cole, 1996.

[3] 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.

[4] 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.

[5] 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.

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

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

[8] 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.

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

[10] 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.

[11] E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach. *Systems Biology in Practice*. Wiley-VCH, 2006.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] David L. Nelson and Michael M. Cox. *Principles of Biochemistry, 3rd edition*. Worth Publishers, 2000.

[18] 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.

[19] Ion Petre, Claire L. Hyder, Andrzej Mizera, Andrey Mikhailov, John E. Eriksson, Lea Sistonen, and Ralph-Johan Back. Two metabolites are enough to drive the eukaryotic heat shock response.

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

[21] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flammery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge Universoty Press, 2007.

[22] 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.

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

[24] Clifford Henry Taubes. *Modeling Differential Equations in Biology*. Cambridge University Press, 2001.

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

[26] Tamás Turányi. Sensitivity analysis of complex kinetic systems. tools and applications. *Journal of Mathematical Chemistry*, 5:203–248, 1990.

[27] Paul Workman and Emmanuel de Billy. Putting the heat on cancer. *Nature Medicine*, 13(12):1415–1417, 2007.

[28] Dennis G. Zill. *A First Course in Differential Equations*. Thomson, 2001.

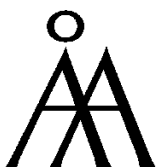[29] Dennis G. Zill. *A First Course in Differential Equations with Modeling Applications*. Thomson, 2005.

# Turku Centre *for* Computer Science

University of Turku
- Department of Information Technology
- Department of Mathematics

Åbo Akademi University
- Department of Computer Science

Turku School of Economics and Business Administration
- Institute of Information Systems Sciences