Stochastic modelling of the eukaryotic heat shock response

Andrzej Mizera[*,a,b], Barbara Gambin[a]

[a] *Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawińskiego 5B, 02-106 Warsaw, Poland*
[b] *Department of Information Technologies, Åbo Akademi University & Turku Centre for Computer Science, Joukahaisenkatu 3-5A, FIN-20520 Turku, Finland*

Abstract

The heat shock response (HSR) is a highly evolutionarily conserved defence mechanism allowing the cell to promptly react to elevated temperature conditions and other forms of stress. It has been subject to intense research for at least two main reasons. First, it is considered a promising candidate for deciphering the engineering principles underlying regulatory networks. Second, heat shock proteins (main actors of the HSR) play crucial role in many fundamental cellular processes. Therefore, profound understanding of the heat shock response would have far-reaching ramifications for the cell biology.

Recently, a new deterministic model of the eukaryotic heat shock response has been proposed in the literature. It is very attractive since it consists of only the minimum number of components required by any functional regulatory network, while yet being capable of biological validation. However, it admits small molecule populations of some of the considered metabolites. In this paper a stochastic model corresponding to the deterministic one is constructed and the outcomes of these two models are confronted. The aim with this comparison is two show that, in the case of the heat shock response, the approximation of a discrete system with a continuous model is a reasonable approach. This is not always the truth, especially when the numbers of

[*]Corresponding author: A. Mizera, e-mail: amizera@ippt.gov.pl, amizera@abo.fi, tel. +48 22 8261281(414)

Email addresses: amizera@ippt.gov.pl, amizera@abo.fi (Andrzej Mizera), bgambin@ippt.gov.pl (Barbara Gambin)

Preprint submitted to Journal of Theoretical Biology

molecules of the considered species are small. By making the effort of performing and analysing 1000 stochastic simulations, we investigate the range of behaviour the stochastic model is likely to exhibit. We demonstrate that the obtained results agree well with the dynamics displayed by the continuous model, which strengthens the trust in the deterministic description. A proof of the existence and uniqueness of the stationary distribution of the Markov chain underlying the stochastic model is given. Moreover, the obtained view of the stochastic dynamics and the performed comparison to the outcome of the continuous formulation provide more insight into the dynamics of the heat shock response mechanism.

Key words:
Mathematical modelling, Stochastic model, Computer simulations, Markov chain, Gillespie algorithm, Stationary distribution, Steady-state, Clustering

---

1. Introduction

The heat shock response is the most highly evolutionarily conserved defence mechanism (Lindquist and Craig, 1988). It exists in all eukaryotic cells, protects them from the damaging influence of elevated temperature and allows them to promptly react to other forms of environmental stress. The heat shock response has been subject to intense research recently, see (Chen et al., 2007; Powers and Workman, 2007; Voellmy and Boellmann, 2007), for at least two reasons. On one hand, as a well-conserved mechanism, it is considered a promising candidate for deciphering the engineering principles underlying regulatory networks. On the other hand, heat shock proteins play crucial roles in many fundamental cellular processes such as protein biogenesis, dismantling of damaged proteins, activation of immune responses and signalling, see (Kampinga, 1993; Pockley, 2003). Therefore, profound understanding of the heat shock response would have far-reaching ramifications for the cell biology and could potentially allow for treatment of a number of diseases, such as neurodegenerative and cardiovascular disorders, cancer, ageing, see (Balch et al., 2008; Liu et al., 2002; Lukacs et al., 2000; Morimoto, 2008; Workman and de Billy, 2007).

Although a number of mathematical models describing the heat shock response both in eukaryotes as well as in bacteria have been presented in the literature, see Donati et al. (1990); Jones et al. (1993); Parsell and Lindquist (1993); Peper et al. (1997); Petre et al. (2009b); Remondini et al. (2006);

Rieger et al. (2005); Szymańska and Żylicz (2009), still a comprehensive mechanistic understanding of this process is lacking. In Petre et al. (2009b) a new model of the eukaryotic heat shock response together with an associated continuous mathematical model based on ordinary differential equations have been discussed. The novelty of the model in Petre et al. (2009b) is due to the fact that, unlike other previous models, it is based solely on well-documented reactions and does not incorporate modelling "blackboxes" such as hypothetical, experimentally unsupported cellular mechanisms whose only purpose is to enforce appropriate behaviour. The simplified version of the model (see Petre et al. (2009a) for details) includes the temperature-induced protein misfolding, all three forms of heat shock factors: monomers, dimers and trimers, the backregulation of the transactivation of the heat shock protein encoding gene and the chaperone activity of heat shock proteins. At the same time, it contains as few reactions and reactants as possible. It is worth noticing that the model consists of only the minimum number of components required by any functional regulatory network: an activation mechanism and a feedback mechanism. Nonetheless, the associated continuous model predictions correlate well with experimental observations on the heat-induced transactivation of the hsp-encoding genes at different temperatures from the range $37\,^{\circ}\mathrm{C} - 43\,^{\circ}\mathrm{C}$ (in particular, the prolonged transcription at $43\,^{\circ}\mathrm{C}$ is confirmed) and the return to the original level of hsp production once the stress is removed (publication in preparation). Moreover, the model perfectly illustrates the experimentally observed process of "self-learning" of the HSR system: the response to a second consecutive heat shock is significantly lower. This is due to a transient increase in the free hsp level caused by the preliminary heat shock. In other words, the increase is a form of temporary memory of the fact that the cell was recently exposed to heat shock conditions.

However, the undertaken modelling approach that utilises ordinary differential equations is just one of many other modelling paradigms (e.g. stochastic formulation, process calculi, Petri nets, etc.), which could be exploited in the context of the heat shock response. In this paper we follow one of the other formalisms: we develop a stochastic model associated with the simplified version of the model from Petre et al. (2009b) which has been described in Petre et al. (2009a). According to current scientific knowledge, ignoring quantum mechanical effects, biological systems can be viewed as deterministic of their very nature, with their dynamics entirely specified, given sufficient information on the state of the system (position, orientation and momentum of every single molecule) and a complete understanding of the chemistry and

physics of the interactions between biomolecules. Unfortunately, we are still unable to model biological systems of realistic complexity and size using such a molecular dynamic approach (Wilkinson (2006)). Therefore the current models admit far-reaching simplifications, which result in a higher level view of the system being modelled. However, these abstractions change the character of the dynamics, which becomes intrinsically stochastic and requires consideration of statistical physics to describe the stochastic process governing it. Especially at low concentrations of the involved reactants, random fluctuations may have a significant impact on the reaction dynamics, but the deterministic approach to chemical kinetics fails to capture such phenomena, see McAdams and Arkin (1999); Srivastava et al. (2002). For example, let us consider the famous Lotka-Volterra system of coupled ordinary differential equations describing an ecological predator-prey model. The solutions of this system are known to be periodic (except for the stationary point) independently of the initial size of predator and prey populations. However, in the stochastic formulation there exists a "catastrophic" sequence of events which leads to depletion of preys by predators and, in consequence, to the extinction of predators as well. When running the model long enough, the probability of not executing this catastrophic sequence drops to zero. This leads to radical qualitative differences in the trajectories obtained by these two approaches: in the deterministic case the trajectory in the predator versus prey phase space is an ellipse, while in the stochastic case the trajectory eventually reaches the trivial steady state of no predators and no prey individuals in the system. The expected time it takes to reach this state depends on the initial number of species. Such discrepancy in the trajectories is especially easily observed when the initial population sizes are small.

Another significant impact of random fluctuations can be observed in the model of T cell receptor signalling presented in Lipniacki et al. (2008), where it is shown that, because of bistability of the system and the fact that the T cell activation is due to a small number of foreign peptides, the responses are highly stochastic. This results in stochastic trajectories not following the deterministic trajectory, which converges to a steady state. Instead, the stochastic realisations may occasionally jump between the basins of attraction of two possible states. In particular, as was shown in Lipniacki et al. (2008), stochastic noise can cause a transition from the higher stable state to the lower one and most of the stochastic trajectories are trapped in the basin of attraction of the latter steady state in contrast to the deterministic case. As a result, the qualitative behaviour revealed by the stochastic approach

differs significantly from the behaviour obtained from the deterministic description. For details we refer the reader to Lipniacki et al. (2008).

Although for a complex system detailed mathematical analysis based on the "chemical master equation" is intractable (Wilkinson (2006)), it is possible to gain insight into the system's dynamics by performing a series of stochastic simulations of the time-evolution of such system by so called Gillespie algorithm (Gillespie (1976)). The algorithm is a well-established procedure for generating a stochastic realisation of the system's temporal behaviour. However, due to reasons such as computational efficiency, availability of dedicated simulation software with analysis tools (steady-state, sensitivity analysis, etc.), and expertise in the theory of differential equations, the deterministic modelling approach is commonly used in examination of biological systems, although the stochastic formulation in many cases would be more justified.

Bearing in mind the above mentioned merits of the new simplified heat shock response model described in Petre et al. (2009a), the aim of this paper is to show that in this particular case approximating a discrete system with a continuous model is a valid approach. A stochastic model complementary to the deterministic one is developed. An effort to perform 1000 stochastic simulations is made in order to investigate whether the qualitative results of the stochastic model agree with the deterministic outcome. Having the problem of small number of molecules of some of the reactants in mind (initial number concentrations of hsf, hsf_2, hsf_3, hsf_3 : hse, hse, hsp : mfp, see Petre et al. (2009a) for details), as explained above, one could expect the time-course trajectories obtained with the stochastic model to be substantially different from the trajectories computed in the deterministic formalism. However, we show that the influence of the random fluctuations does not invalidate the continuous approach and the obtained results support the use of the deterministic formulation in this case. In particular, we investigate the number of steady states of the deterministic model and compare the obtained results with the dynamics demonstrated by the stochastic model. We show that the underlying stochastic process of our model has a unique stationary distribution and that the performed stochastic simulation results reveal no evidence of multistationarity, which is consistent with the deterministic description. Additionally, this analysis let us gain some more insight into the dynamics of the heat shock response mechanism. The question about the stationarity and stability, i.e. the number of steady-states and whether they are stable or unstable, is important in the examination of the dynamics of bi-

ological systems. For example, bistability in biological systems is, in general, accompanied by hysteresis, which in turn promotes robustness (Karmakar and Bose (2007); Lipniacki et al. (2008)).

The paper is organised as follows. In Section 2 we briefly describe the simplified deterministic model (named *deterministic model* for compactness in the continuation) of the heat shock response in eukaryotic cells which was proposed in Petre et al. (2009a). Next, in Section 3, we discuss the Markov jump process which constitutes the corresponding stochastic model and show that it has a unique stationary distribution. Further, in Section 4, the stochastic simulation results are discussed and a comparison between the deterministic and stochastic model is presented. Finally, we end with conclusions in Section 5.

2. Deterministic model

The model of the eukaryotic heat shock response consists of four main modules: the heat-induced protein misfolding, the dynamic transactivation of the genes encoding heat shock proteins, their backregulation and the chaperone activity of the heat shock proteins.

At elevated temperatures proteins tend to misfold and create aggregates, which has disastrous effects on the cell. In order to survive, the cell has to promptly increase the level of *heat shock proteins* (hsp), which is the main task of the heat shock response mechanism. Heat shock proteins act as chaperones: they interact with the *misfolded proteins* (mfp) and assist them in refolding to their native conformation (prot). The control over the defence mechanism against the temperature-induced harmful phenomena is implemented through the regulation of the transactivation of the hsp-encoding gene. Activation of the transcription proceeds along the following scheme: *heat shock factors* (hsf) trimerize (through a transient dimerization) and in this form bind to the *heat shock element* (hse), i.e. the promoter of the hsp-encoding gene. Once the hsf-trimer (hsf_3) is bound to the specific DNA sequence (hsf_3 : hse), the gene is transactivated and new hsp molecules are eventually synthesised. Finally, when the level of hsps is high enough to cope with the thermal stress, the production is switched off: hsps bind both to free hsfs and hsfs that occur in compound forms (hsf_2, hsf_3, hsf_3 : hse), which, in consequence, get disassembled. As a result, DNA transcription of hsp-encoding gene is turned off and the formation of new hsf trimers is blocked. The full list of molecular reactions constituting the model is presented in

Table 1. By assuming the law of mass-action for all reactions (R_1)–(R_{17}), the associated mathematical model based on ordinary differential equations is obtained. The rate coefficient of protein misfolding with respect to the temperature ($\varphi(T)$) in reaction (R_{14}) is given by the following formula:

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

where T is the numerical value of the temperature of the environment in Celsius degrees. The formula is valid for $37 \leq T \leq 45$. It is based on experimental investigations in Lepock et al. (1993), Lepock et al. (1988) and was originally proposed in Peper et al. (1997). Expression (1) in its current form was obtained by adapting the original formula to the time unit of the discussed mathematical model (see Petre et al. (2009a)). In our survey the temperature is set to 42 °C, i.e. the cells are exposed to heat shock conditions.

As shown in Petre et al. (2009a), there are three mass-conservation relations in the model: the total number of heat shock factor molecules, heat shock elements and protein molecules (either misfolded or in native conformation) is conserved in time. This can be written formally as

$$\begin{aligned} C_1 &= \mathrm{hse}(t) + 3\ \mathrm{hsf}_3 : \mathrm{hse}(t) && (2)\\ C_2 &= \mathrm{hsf}(t) + 2\ \mathrm{hsf}_2(t) + 3\ \mathrm{hsf}_3(t) \\ &\quad +3\ \mathrm{hsf}_3 : \mathrm{hse}(t) + \mathrm{hsp} : \mathrm{hsf} && (3)\\ C_3 &= \mathrm{prot}(t) + \mathrm{mfp}(t) + \mathrm{hsp} : \mathrm{mfp}(t) && (4) \end{aligned}$$

for all $t \geq 0$, where C_1, C_2, $C_3 \geq 0$ are some constants determined by initial conditions, i.e. right-hand side expressions at $t = 0$ in the above equations (2)-(4).

The described model of eukaryotic heat shock response is based solely on well-documented reactions and does not include any "artificial" elements such as experimentally unsupported components or biochemical reactions. For a detailed discussion of the model, we refer the reader to Petre et al. (2009a).

3. Stochastic model

Stochastic modelling of biochemical networks is today well-established. The time-evolution of a reaction system can be regarded as a stochastic process (cf. Wilkinson (2006)). In particular, the dynamics of a biochemical

network can be viewed as a continuous-time Markov process. A continuous-time stochastic process $\{X(t),\, t \geq 0\}$ with discrete state space $\mathcal{S}$ is said to be a continuous-time Markov chain (CTMC for short) if

$$\begin{aligned} P\{X(t_n) = i_n \,|\, X(t_0) = i_0,\, \ldots,\, X(t_{n-1}) = i_{n-1}\} \\ = P\{X(t_n) = i_n \,|\, X(t_{n-1}) = i_{n-1}\} \end{aligned}$$

for all $0 \leq t_0 < \cdots < t_{n-1} < t_n$ and $i_0, \ldots, i_{n-1}, i_n \in \mathcal{S}$. The Markov property expresses that the conditional distribution of a future state given the present and past states depends only on the present state and is independent of the past.

We consider a time-homogeneous Markov chain for which the transition probability $P\{X(t+u) = j \,|\, X(u) = i\}$ is independent of u. Let $\mathbf{Q} = (q_{ij})_{i,j\in\mathcal{S}}$ be the infinitesimal transition rate matrix of the continuous-time Markov chain $\{X(t)\}$ such that the following assumption is satisfied.

Assumption 1. *$\nu_i = \sum_{j\neq i} q_{ij}$ are positive and bounded in $i \in \mathcal{S}$.*

As stated in Tijms (2003), if Assumption 1 is fulfilled, it can be shown that the infinitesimal transition rates determine a unique continuous-time Markov chain which is precisely a Markov jump process constructed in the following way:

a) if the system jumps to state i, it then stays in state i for an exponentially distributed time with mean $1/\nu_i$ independently of how the system reached state i and how long it took to get there (this explains the name *sojourn-time rates* used for $\{\nu_i\}$);

b) if the system leaves state i, it jumps to state j $(j \neq i)$ with probability p_{ij} independently of the duration of the stay in state i.

Further, let $\{X_n,\, n = 0, 1, \ldots\}$ be the *embedded Markov chain*, i.e. X_n is defined as the state of $\{X(t)\}$ just after the n-th transition with the convention that $X_0 = X(0)$. The one-step transition probabilities of the discrete-time Markov chain $\{X_n\}$ are given by

$$p_{ij} = \begin{cases} q_{ij}/\nu_i & j \neq i, \\ 0 & j = i, \end{cases}$$

for all $i, j \in \mathcal{S}$ (e.g. see Tijms (2003)).

The corresponding stochastic formulation of the HSR molecular model presented in Section 2 is a system of 10 chemically active species S_i, ($i = 1, \ldots, 10$), ($S_1 \equiv$ hse, $S_2 \equiv$ hsf, $S_3 \equiv \text{hsf}_2$, $S_4 \equiv \text{hsf}_3$, $S_5 \equiv \text{hsf}_3$: hse, $S_6 \equiv$ hsp, $S_7 \equiv$ hsp : hsf, $S_8 \equiv$ hsp : mfp, $S_9 \equiv$ mfp, $S_{10} \equiv$ prot) that participate in 17 chemical reactions (R_1) – (R_{17}) in some volume V. More specifically, in our case the abstract volume V is simply a eukaryotic cell. The system's state space $\mathcal{S} \subset \mathbb{N}^{10}$ is defined as

$$\begin{aligned} \mathcal{S} \;=\; \{&(s_1, \ldots, s_{10})^T \,:\, \forall_{i \in \{1,\ldots,10\}}\, s_i \in \mathbb{N}, \\ &s_1 + s_5 = C_1, \\ &s_2 + 2\,s_3 + 3\,s_4 + 3\,s_5 + s_7 = C_2, \\ &s_8 + s_9 + s_{10} = C_3, \\ &s_6 + s_7 + s_8 \leq K\}, \end{aligned} \tag{5}$$

where $C_1 \geq 1$, $C_2 \geq 3$ and $C_3 \geq 1$ are constants describing the fixed in time total number of hse, hsf and protein (either misfolded or in native conformation) molecules present in the system, respectively. The last inequality, i.e. $s_6 + s_7 + s_8 \leq K$, requires some comment. In the absence of it the model would make allowance for any unbounded number of free hsp molecules to co-exist. However, this is certainly contrary to the fact that any living cell has a limited volume and, in consequence, can only contain a finite number of hsp molecules. Thus, in order to make the model more realistic, a big enough (in the sense that it allows for the appropriate number of free hsp molecules to be present in the system) constant K is introduced and an upper bound on the value of the S_6 variable is imposed by the last inequality. The direct consequence of adding it is that the state space $\mathcal{S}$ becomes finite.

The system is in state $s = (s_1, \ldots, s_{10})^T$ at time t if and only if the number of molecules of species S_i at time t is s_i for all $i = 1, \ldots, 10$. The conditions posed on the constants C_1, C_2 and C_3 ensure that the HSR mechanism is operational, i.e. that at least one hse molecule is present in the system, that the system is able to produce at least one hsf_3 molecule that can bind to the DNA and, in consequence, initiate the transcription and translation of hsp. Finally, that at least one generic protein prone to misfolding exists in the system.

Each reaction R_μ is characterised by a stochastic rate constant c_μ ($\mu = 1, \ldots, 17$), see Table 3. The values were obtained from the deterministic model (Petre et al., 2009a). In the case of reaction R_1, the deterministic rate constant value was multiplied by 2 in order to obtain the value for the corresponding stochastic rate constant.

Let v_μ be the μ-th column of the stoichiometry matrix of the HSR system presented in Table 2. Reaction R_μ causes the system to make a transition from some state $i \in \mathcal{S}$ to state $j = i + v_\mu$. The *fundamental hypothesis* of the stochastic formulation of chemical kinetics (Gillespie, 1976) is that the reaction parameter c_μ can be defined as follows.

$$c_\mu \, \mathrm{d}t \quad \equiv \quad \begin{array}{l} \text{average probability, to first order in } \mathrm{d}t, \\ \text{that a particular combination of } R_\mu \text{ re-} \\ \text{actant molecules will react accordingly} \\ \text{in the next time interval } \mathrm{d}t. \end{array}$$

As shown by Gillespie in Gillespie (1976), c_μ is dependent on the radii of the molecules involved in the reaction and their average relative velocities, where the average relative velocity is a function of the temperature of the system and the individual molecular masses. Further, it is shown that the probability of reaction R_μ occurring in V in the time interval $(t, t+\mathrm{d}t)$, given that the system is in state i at time t, has the form $h_\mu^i \, c_\mu \, \mathrm{d}t$. h_μ^i denotes the number of possible combinations of reactant molecules involved in reaction R_μ when the system is in state i. However, since the total number of hsp molecules that might co-exist in a cell is limited, no further hsp production (reaction (R_7)) should take place when the system is in any of the states in which the limit is reached. Let us denote all these states by $\mathcal{S}_K$, i.e.

$$\mathcal{S}_K = \{s \,|\, s_6 + s_7 + s_8 = K\}. \tag{6}$$

Thus, the probability of reaction R_7 occurring in V in the time interval $(t, t + \mathrm{d}t)$ when the system is in $i \in \mathcal{S}_K$ should be 0. Hence

$$h_7^i = \begin{cases} \text{No. of combinations of} \\ \text{reactant molecules of } R_7 & i \in \mathcal{S} \setminus \mathcal{S}_K, \\ \\ 0 & i \in \mathcal{S}_K. \end{cases} \tag{7}$$

Due to the fact that the reaction hazards depend only on the current state of the system, the time-evolution of the state of the reaction system can be regarded as a CTMC. Since the state space $\mathcal{S}$ is finite, Assumption 1 is fulfilled and hence the chain is a Markov jump process constructed as described above. The infinitesimal transition rates of the Markov jump process are

$$q_{ij} = h_\mu^i \, c_\mu, \tag{8}$$

where $j = i + v_\mu$.

Now, let us consider the *reaction probability density function* $P^i(\tau, \mu)\,\mathrm{d}\tau$ of the HSR system, which forms the basis for the Gillespie's simulation algorithm. It is defined by

> $P^i(\tau, \mu)\,\mathrm{d}\tau \equiv$ probability at time t (when the system is in state $i \in \mathcal{S}$) that the next reaction in V will occur in the differential time interval $(t + \tau, t + \tau + \mathrm{d}\tau)$, and will be an R_μ reaction.

As shown in Gillespie (1976), the density function can be expressed as

$$P^i(\tau, \mu)\,\mathrm{d}\tau = h^i_\mu\, c_\mu \cdot \exp\left[-\sum_{\nu=1}^{17} h^i_\nu c_\nu \tau\right]\,\mathrm{d}\tau. \tag{9}$$

The superscript $i \in \mathcal{S}$ indicates that in fact we deal with a whole family of such functions. Which of them is considered at time t depends on the state of the system at time t. In the continuation, in order to lighten the language, "probability at time t" will be a shorthand for "probability at time t when the system is in state i". Let $P^i_1(\tau)\,\mathrm{d}\tau$ denote the probability at time t that the next reaction will occur between times $t + \tau$ and $t + \tau + \mathrm{d}\tau$, irrespective of which reaction it might be. By the definition of $P^i(\tau, \mu)$ we have that

$$\begin{aligned} P^i_1(\tau) &= \sum_{\mu=1}^{17} P^i(\tau, \mu) \qquad (10)\\ &= (\sum_{k=1}^{17} h^i_k\, c_k) \cdot \exp[-(\sum_{k=1}^{17} h^i_k\, c_k) \cdot \tau]. \end{aligned}$$

Hence, the sojourn-time rates ν_i of the Markov jump process are given by

$$\nu_i = \sum_{k=1}^{17} h^i_k\, c_k. \tag{11}$$

The probability of the transition from state i to state $j = i + v_\mu$ of the Markov jump process (and, in consequence, of the embedded Markov chain $\{X_n\}$) is the probability at time t that the next reaction in V will be an R_μ reaction. Using equation (9), it can be expressed as

$$p_{ij} = \int_0^\infty P^i(\tau, \mu)\,\mathrm{d}\tau = \frac{h^i_\mu\, c_\mu}{\sum_{k=1}^{17} h^i_k\, c_k} \tag{12}$$

if $j \neq i$ and $p_{ii} = 0$ for all $i \in \mathcal{S}$.

Lemma 1. *The embedded Markov chain* $\{X_n\}$ *is irreducible.*

Before presenting the proof let us divide the species into two groups. The first one, called *elementary species group* (denoted by $G_{elementary}$), contains hsf, hse, hsp, mfp and prot species. The other one, denoted by $G_{compound}$ and named *compound species group*, is made of all the remaining species.

Proof of Lemma 1. Let i, j be any two states from the state space $\mathcal{S}$. By $i \rightsquigarrow j$ we denote that the state j is reachable from the state i, i.e. that there exists a sequence of reactions (R_1)–(R_{17}) which leads the system from the state i to the state j.

In order to prove that $\{X_n\}$ is irreducible it is enough to show that $i \rightsquigarrow j$, since i and j are two arbitrarily chosen states. Let $\mathrm{hsp}(k)$, $\mathrm{mfp}(k)$ and $\mathrm{prot}(k)$ be the total number of hsp, mfp and prot molecules present in the system when in the state k, respectively. Further, let

$$z = (C_1, C_2, 0, 0, 0, \mathrm{hsp}(i), 0, 0, \mathrm{mfp}(i), \mathrm{prot}(i))^T.$$

z is obtained from i by disassembling all compound species from $G_{compound}$. Thus, in the state z the number of molecules of any species from $G_{compound}$ is 0 and the number of molecules of any $s \in G_{elementary}$ is equal to the total number of s molecules in the system in the state i. Clearly $z \in \mathcal{S}$ and $i \rightsquigarrow z$, since for any $s \in G_{compound}$ there exists a sequence of reactions (R_1)–(R_{17}) which disassembles s into elements from $G_{elementary}$.

Let

$$z' = (C_1, C_2, 0, 0, 0, \mathrm{hsp}(j), 0, 0, \mathrm{mfp}(i), \mathrm{prot}(i))^T$$

and

$$z'' = (C_1, C_2, 0, 0, 0, \mathrm{hsp}(j), 0, 0, \mathrm{mfp}(j), \mathrm{prot}(j))^T.$$

$z' \in \mathcal{S}$ and we continue to show that $z \rightsquigarrow z'$. There are three cases: $\mathrm{hsp}(i) = \mathrm{hsp}(j)$, $\mathrm{hsp}(j) < \mathrm{hsp}(i)$ or $\mathrm{hsp}(j) > \mathrm{hsp}(i)$. In the first case $z = z'$ and trivially $z \rightsquigarrow z'$. If $\mathrm{hsp}(j) < \mathrm{hsp}(i)$, z' can be reached from z by applying reaction (R_{13}) $\mathrm{hsp}(i) - \mathrm{hsp}(j)$ times. If finally $\mathrm{hsp}(j) > \mathrm{hsp}(i)$, first $\mathrm{hsf}_3 : \mathrm{hse}$ is produced (this is doable since $C_1 \geq 1$ and $C_2 \geq 3$). Next, by applying reaction (R_7) $\mathrm{hsp}(j) - \mathrm{hsp}(i)$ times the required number of additional hsp molecules is produced. Finally, $\mathrm{hsf}_3 : \mathrm{hse}$ is disassembled by applying a sequence of reactions $<(R_6), (R_3), (R_2)>$. Hence $z \rightsquigarrow z'$.

We continue to show that $z' \rightsquigarrow z''$. There are two cases. Either $\mathrm{mfp}(i) > \mathrm{mfp}(j)$ or $\mathrm{prot}(i) \geq \mathrm{prot}(j)$ since $\mathrm{mfp}(k) + \mathrm{prot}(k) = C_3$ for any state $k \in \mathcal{S}$.

In the first case, if $\mathrm{hsp}(j) = 0$, first one hsp molecule is produced by applying reaction sequence $<(R_1), (R_3), (R_5), (R_7), (R_6), (R_4), (R_2)>$, which leads to state $(C_1, C_2, 0, 0, 0, 1, 0, 0, \mathrm{mfp}(i), \mathrm{prot}(i))^T$. Then, by applying reaction sequence $<(R_{15}), (R_{17})>$ $\mathrm{mfp}(i) - \mathrm{mfp}(j)$ times the system reaches state:

$$(C_1, C_2, 0, 0, 0, 1, 0, 0, \mathrm{mfp}(j), \mathrm{prot}(j))^T$$

since, as mentioned before, $\mathrm{mfp}(k) + \mathrm{prot}(k) = C_3$ for any state $k \in \mathcal{S}$. Finally, the only hsp molecule is degraded by applying reaction (R_{13}) and the system arrives in state z''. If $\mathrm{hsp}(j)$ is greater than 0, state z'' can be reached by less steps since neither production nor degradation of the one additional hsp molecule is required.

In the second case, when $\mathrm{prot}(i) \geq \mathrm{prot}(j)$, state z'' can be reached by applying misfolding reaction (R_{14}) $\mathrm{prot}(i) - \mathrm{prot}(j)$ times. Hence $z' \rightsquigarrow z''$.

At last $z'' \rightsquigarrow j$. State j is reached by producing the appropriate numbers of molecules of all compound species. Since in state z'' the required number of molecules of all elementary species is already present, by applying appropriate reactions all compound species molecules can be produced. The mass-conservation law ensures that the numbers of elementary species molecules will be decreased appropriately and that the system will reach state j. Hence $\{X_n\}$ is irreducible.

□

The irreducibility of the embedded chain $\{X_n\}$ implies the irreducibility of the continuous-time Markov chain $\{X(t)\}$. Since the state space $\mathcal{S}$ is finite, it follows that the CTMC $\{X(t)\}$ is positive recurrent. In consequence, it has an invariant measure η which is unique up to multiplicative factors and can be found as the solution of the equation $\eta^T \mathbf{Q} = 0$. Moreover, $\sum_{i \in \mathcal{S}} \eta_i < \infty$ since $\mathcal{S}$ is finite and there exists a unique stationary distribution π of $\{X(t)\}$ given by

$$\pi = \left(\frac{\eta_i}{\sum_{k \in \mathcal{S}} \eta_k} \right)_{i \in \mathcal{S}}. \tag{13}$$

For the theoretical details we refer the reader to, e.g., Norris (1998); Resnick (1992).

4. Results and discussion

The deterministic approach, based on the law of mass action, yields a system of ordinary differential equations for molecular concentrations. In consequence, the biochemical system is modelled as being continuous. But such

description does not capture effects that occur due to either the discreteness of molecular quantities or the stochastic nature of chemical reactions (McAdams and Arkin (1999); Pahle (2009); Sandmann (2008); Wilkinson (2006)). As discussed in Section 1, random fluctuations may have a significant impact on the reaction dynamics, especially as the numbers of molecules of some reactants become smaller (McAdams and Arkin, 1999; Srivastava et al., 2002). This is the case of the deterministic heat shock response model being discussed: except for prot, hsp, hsp : hsf and mfp, all the other species have very small initial number of molecules (Table 3) and, as can be seen from the continuous simulation results, stay at the low level throughout the time of simulation. This might be the main objection to the continuous approach applied in Petre et al. (2009a,b). Since the stochastic modelling seems more reasonable in this case, we made the effort to run 1000 stochastic simulations in order to check whether the dynamics of the continuous description agrees qualitatively with the behaviour demonstrated by the discrete system. The results of 1000 independent stochastic simulation runs (blue and green points) for 5 species: hsf_3 : hse, hsp, mfp, hsp : hsf and hsp : mfp, overlaid with the deterministic outcome (yellow line) are shown in Figure 1. The mean together with the mean $+/-$ standard deviation are shown in Figure 2. The ratios of the sample standard deviation to the sample mean were computed for the 5 considered species and are depicted in Figure 3. According to Gillespie (1976), since the ratios are small (less than 0.12 in the case of mfp and hsp, see Fig. 3c and 3b) and very small (less than 0.035 for hsp : mfp and less than 0.007 for hsp : hsf, see Fig. 3e and 3d, respectively), the results of independent runs of the system are expected not to vary much and the presented outcomes of 1000 stochastic simulations together with the estimated mean should provide a statistically adequate picture of the evolution of the chemical system in time. One might argue that the ratio for hsf_3 : hse is however quite big: it peaks at about 1.1 and stabilises below 0.6 (see Fig. 3a). In this particular case the mean converges to approximately 3 molecules of hsf_3 : hse and the standard deviation is around 1.6, which all in all gives a narrow range of possible values of molecule number and hence this result can be accepted.

We first investigated the number of steady-states of the deterministic model. Since our attempts to analytically solve the algebraic system of steady state equations obtained from the differential ones did not bring any results, we performed some numerical investigations. We randomly chose 10000 sets of initial particle numbers for the continuous model from a wide range of val-

ues, but in such a way that the total amounts of hse, hsf and proteins in the resulting system would always be the same as in the case of the original deterministic model presented in Petre et al. (2009a). For each of these sets we run numerical time-course simulations and waited for the considered system to stabilize. In all these cases the systems converged to exactly the same state as the original model, i.e. no other steady states were found by this method. Additionally, bifurcation analysis performed with the AUTO software (XPPAUT was used as the front-end, Doedel et al. (1997); Ermentrout (2002)) with respect to parameter values did not reveal multistationarity (data not shown). These results suggest that the heat shock response mechanism is rather monostable.

Next, we were interested in investigating the range of behaviour the stochastic model was likely to exhibit. As shown in Section 3, there exists only one stationary limit distribution π given by Eq. (13), which governs the transitions of the Markov jump process when the number of iterations goes to infinity. In particular, we analysed the unimodality of the hsp level by computing some appropriate statistics from the performed 1000 stochastic realisations.

First, we computed the median $m(t)$ of the 1000 stochastic realisations on the time interval $T = \{130000s, \ldots, 150000s\}$. It is depicted in Fig. 4 as the middle black line. The upper and lower black lines are $m(t)$ $+/-$ $\frac{1}{4} \cdot s$ respectively, where s is the range of dynamics the model exhibits in the 1000 realisations on the considered time interval, i.e.

$$s = \max_{t \in T,\, i \in I} \{r_i(t)\} - \min_{t \in T,\, i \in I} \{r_i(t)\},$$

where $I = \{1, \ldots, 1000\}$ and r_i is the i-th realisation. The mean (brown line) basically coincides with the median on the whole time interval.

Next, in order to check whether the realisations r_i, $i = 1, \ldots, 1000$, can be divided into subgroups such that the means of the subgroups would differ significantly from each other, we applied the following procedure. We defined two subsets:

$$S_U = \{r_i \quad : \quad \forall_{t \in T}\, r_i(t) > m(t) - \frac{s}{4} \\ \wedge\, \exists_{t \in T}\, r_i(t) > m(t) + \frac{s}{4}\Big\}$$

and

$$S_L = \{r_i \quad : \quad \forall_{t\in T}\, r_i(t) < m(t) + \frac{s}{4} \\ \wedge\, \exists_{t\in T}\, r_i(t) < m(t) - \frac{s}{4}\Big\}.$$

In our case, there are 253 realisations in S_U and 189 in S_L. The means computed from the realisations of each of these subsets are depicted by red lines in Fig. 4. The means are close to the global mean on the whole considered time interval and since the numbers of elements in the S_U and S_L subsets are rather small, i.e. approximately 1/4 and 1/5 of all the 1000 considered realisations, this result does not indicate any significant split.

Further, a clustering algorithm was applied in order to determine whether some subsets of realisations could be isolated and the computed means would point to potential multimodality. To this aim, we utilised the *Agnes* algorithm (implementation of an agglomerative hierarchical clustering method, Kaufman and Rousseeuw (1990)) with the *manhattan* metric, i.e. the distance between two realisations r_i and r_j is defined as $d(r_i, r_j) = \sum_{t\in T} |r_i(t) - r_j(t)|$, thus the realisations are treated as points in a $|T|$-dimensional space. By applying this metric the characteristics of the realisations on the whole considered time interval are taken into account, hence they are compared in a "global" sense. The obtained dendrogram is presented in Fig. 5. The *agglomerative coefficient* (AC), which measures the clustering structure of the dataset, is 0.82. This indicates that the clustering algorithm did find some rather clear structuring[1]. We isolated two groups of realisations that stand out on the obtained dendrogram. They are marked in Fig. 5 by two rectangles which enclose the dendrogram branches constituting these groups. The two resulting subclusters are at almost the same height in the clustering tree. The means of the stochastic realisations belonging to these two groups at time point $t = 150000s$ are 757 (left subcluster) and 794 (right subcluster). Although the agglomerative coefficient indicates some clustering structure of the realisations, the mean values are very close to each other and agree well with the steady state value of the deterministic model (767).

Finally, as suggested in (Wilkinson, 2006), we investigated the empirical probability mass function by drawing histograms of the realisations at some

[1]AC is a dimensionless quantity, varying between 0 and 1 – AC close to 1 shows that a very clear structure has been found, while value 0 implies that the data consists of only one big cluster, see e.g. Kaufman and Rousseeuw (1990) for details.

time point in the considered time interval T. Figure 6 shows the histograms overlaid with the normal distribution curve with mean and standard deviation computed from all 1000 realisations at time point $t = 150000s$. In the case of Fig. 6a, where the bin width is set to 20, the obtained results indicate that the distribution is unimodal. Changing the bin size to 10 (Fig. 6b) does not change the picture significantly.

Although due to the small particle numbers of some of the reagents the stochastic modelling is more reasonable, the presented results do not reveal any qualitative discrepancy in the dynamics of the two considered models of the heat shock response. The range of behaviour the stochastic model is likely to exhibit, which can be observed based on the performed 1000 simulations, confirms the dynamics of the continuous model. The performed analysis of the stochastic realisations does not reveal any clear signs of multistationarity of the HSR mechanism. Although unimodality of a stationary probability density function does not necessarily imply the uniqueness of the stable steady state of the deterministic approximation (as well as bimodality does not determine the existence of bistability, etc.), usually this is the case and to this extent the stochastic results agree with the deterministic outcomes indicating that there exists only one stable steady state. This shows that the approximation of a discrete system with a continuous model is valid and strengthens the trust in the deterministic description. Additionally, the presented stochastic formulation, together with the performed analysis of its behaviour and comparison to the continuous description, let us gain more insight into the dynamics of the HSR mechanism, especially in respect of the number of steady states, which, as discussed previously, is important from a biological point of view.

5. Conclusions and Further research

In this paper we presented a stochastic model associated with a previously described (Petre et al., 2009a) model of the heat shock response in eukaryotic cells. The stochastic model was viewed as a Markov jump process and the existence and uniqueness of the stationary distribution was shown. Further, the model was compared to the deterministic description of heat shock response (Petre et al., 2009a). The aim with the comparison was to show that in this particular case the approximation of a discrete system with a continuous model is reasonable. This is not true in general, especially when the numbers of metabolites in the considered biochemical system are small.

The presented results indicate that the stochastic and deterministic models provide a qualitatively consistent picture of the dynamics of the heat shock response mechanism. Additionally, the development of the stochastic model and the effort of performing 1000 stochastic simulations enabled gaining some more information about the dynamics of the heat shock response. The outcomes of the analysis of the stochastic realisations lead towards the conclusion that the heat shock response mechanism is a rather monostable system. Moreover, this is in agreement with the results of the analysis performed on the deterministic model. All in all, the presented results strengthen the trust in the deterministic description of the HSR mechanism in eukaryotic cells proposed in Petre et al. (2009a).

Although it was shown in Section 3 that the Markov jump process has a unique stationary distribution, there is no certainty that it was reached already in the considered time interval $T = \{130000s, \ldots, 150000s\}$. It was chosen based on the results of many stochastic simulations, which suggest that the process stabilises relatively long before the time point $t = 130000s$. Nevertheless, some assessment of the convergence to the stationary distribution in this case would be desired. One of possible approaches is to measure the rate of convergence by the *mixing time* (Sinclair, 1992). For ergodic Markov chains the rate is governed by the second largest eigenvalue in absolute value λ_2, in particular the *spectral gap* $1 - \lambda_2$ is both a necessary and sufficient condition for rapid mixing, see Sinclair (1992) for details. The problem of determining λ_2 of the presented Markov chain underlying the stochastic model of heat shock response is subject of further research.

The rate constant values for the presented stochastic model were obtained from the corresponding values of the deterministic model presented in Petre et al. (2009b), which in turn were fitted to available experimental data. As suggested in Wilkinson (2006), another way of deducing the rate constant values for the stochastic model could utilise methods that are based on Bayesian inference and take advantage of Markov Chain Monte Carlo (MCMC) algorithms such as the Metropolis-Hastings algorithm or the Gibbs Sampler. However, such methods demand high-quality, calibrated, high-resolution time-course measurements for a reasonably large subset of model metabolites (Wilkinson, 2006). Unfortunately, experimental data of such quality are still seldom if ever available and make a challenge for experimental biology.

Acknowledgements

Both deterministic and stochastic models were implemented and run in Copasi, a software application for simulation and analysis of biochemical networks (Hoops et al., 2006). The stochastic simulations were performed using the Gibson and Bruck algorithm (Gibson and Bruck, 1998). The obtained time-course data were analysed and plotted in R, a software environment for statistical computing and graphics (R Development Core Team, 2008).

The authors give special thanks to Anna Gambin from the Faculty of Mathematics, Informatics and Mechanics, University of Warsaw for helpful advice as well as thorough and detailed comments.

Andrzej Mizera would like to express his gratefulness to Jan Westerholm, Mats Aspnäs and Evren Yurtesen from the Department of Information Technologies, Åbo Akademi University for making available their computational resources and providing essential help for performing the stochastic simulations. The author would like to thank Ion Petre from the Computational Biomodelling Laboratory, Åbo Akademi University for invaluable discussions.

The authors were partially supported by Ministry of Science and Higher Education (grant number N N518 426936).

References

Balch, W. E., Morimoto, R. I., Dillin, A., Kelly, J. W., 2008. Adapting proteostasis for disease intervention. Science 319, 916–919.

Chen, Y., Voegli, T., Liu, P., Noble, E., Currie, R., 2007. Heat shock paradox and a new role of heat shock proteins and their receptors as anti-inflammation targets. Inflamm Allergy Drug Targets 6 (2), 91–100.

Doedel, E. J., Champneys, A. R., Fairgrieve, T. F., Kuznetsov, Y. A., Sandstede, B., Wang, X., 1997. AUTO 97, software for continuation and bifurcation problems in ordinary differential equations. Tech. rep., Concordia University, Montreal, Canada.

Donati, Y., Slosman, D., Polla, B., 1990. Oxidative injury and the heat shock response. Biochem Pharmacol 40, 2571–2577.

Ermentrout, B., 2002. Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students. Vol. 14 of Software, Environment and Tools. SIAM.

Gibson, M., Bruck, J., 1998. An efficient algorithm for generating trajectories of stochastic gene regulation reactions. Tech. rep., California Institute of Technology.

Gillespie, D. T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics 22, 403–434.

Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., Kummer, U., 2006. Copasi – a COmplex PAthway SImulator. Bioinformatics 22 (24), 3067–3074.

Jones, C. M., Henry, E. R., Hu, Y., Chan, C. K., Luck, S. D., Bhuyan, A., Roder, H., Hofrichter, J., Eaton, W. A., 1993. Fast events in protein folding initiated by nanosecond laser photolysis. Proc. Natl. Acad. Sci. USA 90, 1186064.

Kampinga, H. K., 1993. Thermotolerance in mammalian cells: protein denaturation and aggregation, and stress proteins. J. Cell Science 104, 11–17.

Karmakar, R., Bose, I., 2007. Posititve feedback, stochasticity and genetic competence. Physical Biology 4, 29–37.

Kaufman, L., Rousseeuw, P. J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons.

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

Lepock, J. R., Frey, H. E., Rodahl, A. M., Kruuv, J., 1988. 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.

Lindquist, S., Craig, E. A., 1988. The heat-shock proteins. Annual Review of Genetics 22, 631–677.

Lipniacki, T., Hat, B., Faeder, J. R., Hlavacek, W. S., 2008. Stochastic effects and bistability in T cell receptor signaling. Journal of Theoretical Biology 254 (1), 110–122.

Liu, B., DeFilippo, A. M., Li, Z., 2002. Overcomming immune toerance to cancer by heat shock protein vaccines. Molecular cancer therapeutics 1, 1147–1151.

Lukacs, K. V., Pardo, O. E., Colston, M., Geddes, D. M., Alton, E. W., 2000. Heat shock proteins in cancer therapy. In: Habib (Ed.), Cancer Gene Therapy: Past Achievements and Future Challenges. Kluwer, pp. 363–368.

McAdams, H. H., Arkin, A., 1999. It's a noisy business! Genetic regulation at the nanomolar scale. Trends in Genetics 15 (2), 65–69.

Morimoto, R., 2008. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. Genes Dev 22, 1427–1438.

Norris, J. R., 1998. Markov Chains. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Pahle, J., 2009. Biochemical simulations: stochastic, approximate stochastic and hybrid approaches. Briefings in Bioinformatics 10 (1), 53–64.

Parsell, D., Lindquist, S., 1993. The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. Ann Rev Genetics 27, 437–496.

Peper, A., Grimbergent, C., Spaan, J., Souren, J., van Wijk, R., 1997. A mathematical model of the hsp70 regulation in the cell. Int. J. Hyperthermia 14, 97–124.

Petre, I., Mizera, A., Back, R.-J., 2009a. Computational heuristics for simplifying a biological model. In: Ambos-Spies, K., Löwe, B., Merkle, W. (Eds.), Mathematical Theory and Computational Practice: 5th Conference on Computability in Europe, CiE 2009, Heidelberg, Germany, July 19-24, 2009, Proceedings. Vol. 5635 of Lecture Notes in Computer Science. Springer, pp. 399–408.

Petre, I., Mizera, A., Hyder, C. L., Mikhailov, A., Eriksson, J. E., Sistonen, L., Back, R.-J., 2009b. A new mathematical model for the heat shock response. In: Condon, A., Harel, D., Kok, J. N., Salomaa, A., Winfree, E. (Eds.), Algorithmic Bioprocesses. Natural Computing Series. Springer, pp. 411–425.

Pockley, A., 2003. Heat shock proteins as regulators of the immune response. The Lancet 362 (9382), 469–476.

Powers, M., Workman, P., 2007. Inhibitors of the heat shock response: Biology and pharmacology. FEBS Lett. 581 (19), 3758–3769.

R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL `http://www.R-project.org`

Remondini, D., Bernardini, C., Forni, M., Bersani, F., Castellani, G., Bacci, M., 2006. Induced metastable memory in heat shock response. J. Biol. Physics 32, 49–59.

Resnick, S., 1992. Adventures in Stochastic Processes. Birkhäuser.

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

Sandmann, W., 2008. Discrete-time stochastic modeling and simulation of biochemical networks. Computational Biology and Chemistry 32, 292–297.

Sinclair, A., 1992. Improved bounds for mixing rates of markov chains and multicommodity flow. Combinatorics, Probablility and Computing 1, 351–370.

Srivastava, R., You, L., Summers, J., Yin, J., 2002. Stochastic vs. deterministic modeling of intracellular viral kinetics. Journal of Theoretical Biology 218, 309–321.

Szymańska, Z., Żylicz, M., 2009. Mathematical modeling of heat shock protein synthesis in response to temperature change. Journal of Theoretical Biology 259, 562–569.

Tijms, H. C., 2003. A First Course in Stochastic Models. John Wiley & Sons.

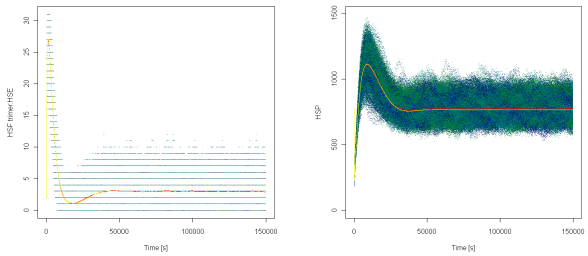
Voellmy, R., Boellmann, F., 2007. Chaperone regulation of the heat shock protein response. Adv Exp Med Biol 594, 89–99.

Wilkinson, D. J., 2006. Stochastic Modelling for Systems Biology. Mathematical & Computational Biology. Chapman & Hall/CRC.

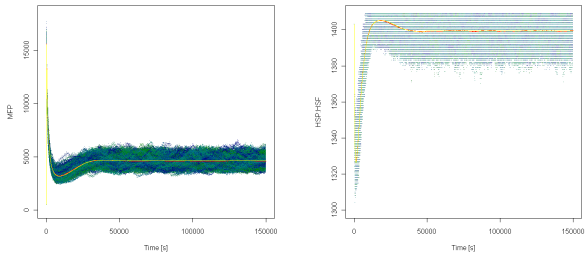
Workman, P., de Billy, E., 2007. Putting the heat on cancer. Nature Medicine 13 (12), 1415–1417.

$$
\begin{aligned}
2\,\mathrm{hsf} &\rightarrow \mathrm{hsf}_2 && (R_1)\\
\mathrm{hsf}_2 &\rightarrow 2\,\mathrm{hsf} && (R_2)\\
\mathrm{hsf} + \mathrm{hsf}_2 &\rightarrow \mathrm{hsf}_3 && (R_3)\\
\mathrm{hsf}_3 &\rightarrow \mathrm{hsf} + \mathrm{hsf}_2 && (R_4)\\
\mathrm{hsf}_3 + \mathrm{hse} &\rightarrow \mathrm{hsf}_3 : \mathrm{hse} && (R_5)\\
\mathrm{hsf}_3 : \mathrm{hse} &\rightarrow \mathrm{hsf}_3 + \mathrm{hse} && (R_6)\\
\mathrm{hsf}_3 : \mathrm{hse} &\rightarrow \mathrm{hsf}_3 : \mathrm{hse} + \mathrm{hsp} && (R_7)\\
\mathrm{hsp} + \mathrm{hsf} &\rightarrow \mathrm{hsp} : \mathrm{hsf} && (R_8)\\
\mathrm{hsp} : \mathrm{hsf} &\rightarrow \mathrm{hsp} + \mathrm{hsf} && (R_9)\\
\mathrm{hsp} + \mathrm{hsf}_2 &\rightarrow \mathrm{hsp} : \mathrm{hsf} + \mathrm{hsf} && (R_{10})\\
\mathrm{hsp} + \mathrm{hsf}_3 &\rightarrow \mathrm{hsp} : \mathrm{hsf} + 2\,\mathrm{hsf} && (R_{11})\\
\mathrm{hsp} + \mathrm{hsf}_3 : \mathrm{hse} &\rightarrow \mathrm{hsp} : \mathrm{hsf} + \mathrm{hse} + 2\,\mathrm{hsf} && (R_{12})\\
\mathrm{hsp} &\rightarrow && (R_{13})\\
\mathrm{prot} &\rightarrow \mathrm{mfp} && (R_{14})\\
\mathrm{hsp} + \mathrm{mfp} &\rightarrow \mathrm{hsp} : \mathrm{mfp} && (R_{15})\\
\mathrm{hsp} : \mathrm{mfp} &\rightarrow \mathrm{hsp} + \mathrm{mfp} && (R_{16})\\
\mathrm{hsp} : \mathrm{mfp} &\rightarrow \mathrm{hsp} + \mathrm{prot} && (R_{17})
\end{aligned}
$$

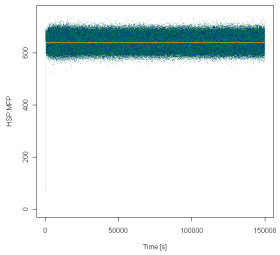
Table 1: The simplified model for the eukaryotic heat shock response



(a) hsf_3 : hse (b) hsp

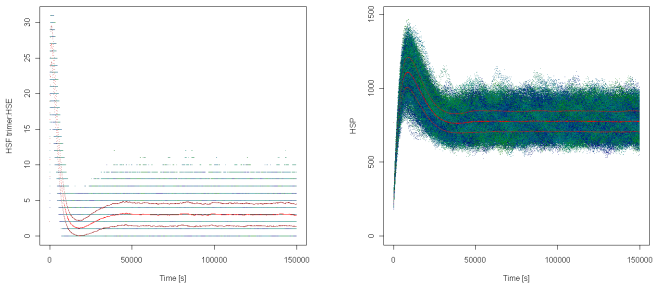


(c) mfp (d) hsp : hsf



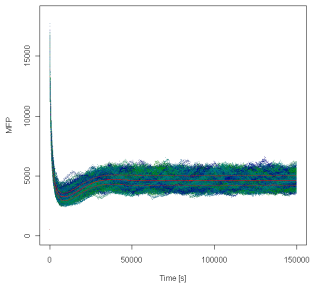
(e) hsp : mfp

Figure 1: Results of 1000 independent discrete stochastic simulation runs. The trajectories of individual realisations are plotted with blue and green points (each run with separate shade). The red points show the average taken over all runs and the yellow line is the outcome of the continuous deterministic simulation.

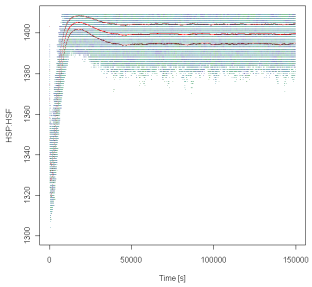


(a) hsf_3 : hse

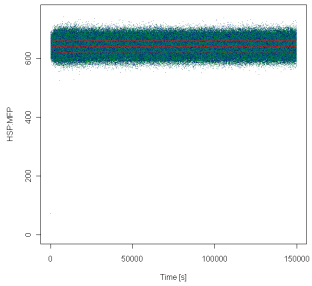
(b) hsp



(c) mfp

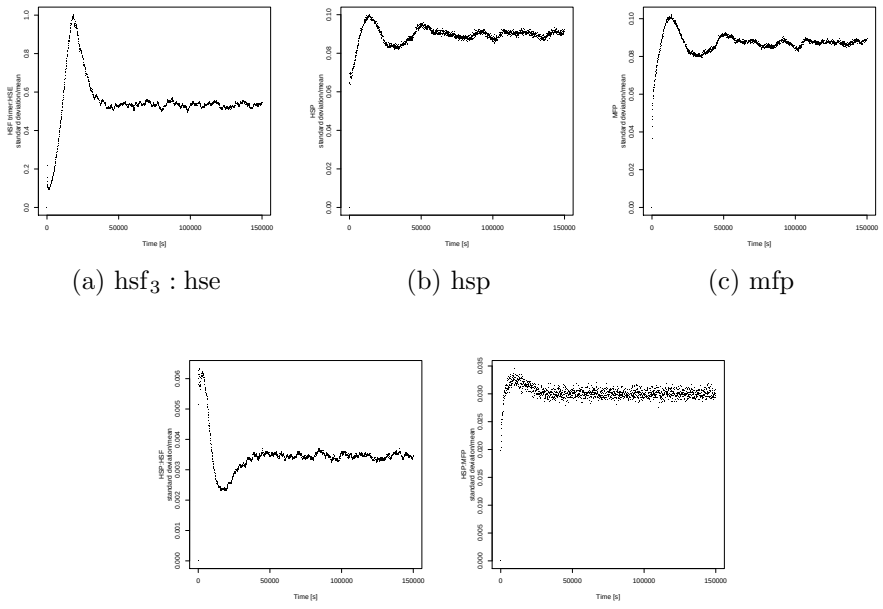


(d) hsp : hsf



(e) hsp : mfp

Figure 2: The mean taken over the outcome of 1000 independent stochastic simulations of the system (red points) and the mean +/− standard deviation (upper/lower brown points).



(a) hsf_3 : hse (b) hsp (c) mfp

(d) hsp : hsf (e) hsp : mfp

Figure 3: The ratios of the standard deviation to the sample mean at each considered time point.

$$\begin{bmatrix}
0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
-2 & 2 & -1 & 1 & 0 & 0 & 0 & -1 & 1 & 1 & 2 & 2 & 0 & 0 & 0 & 0 & 0 \\
1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & -1 & -1 & -1 & -1 & 0 & -1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1
\end{bmatrix}$$

Table 2: The stoichiometry matrix of the heat shock response model. The columns correspond to reactions (R_1)–(R_{17}) and the rows to metabolites in the order: hse, hsf, hsf_2, hsf_3, hsf_3 : hse, hsp, hsp : hsf, hsp : mfp, mfp, prot.

Param.	Reaction	Value	Unit
k_1^+	(R_1)	6.98	$\frac{V}{\#\cdot s}$
k_1^-	(R_2)	0.19	s^{-1}
k_2^+	(R_3)	1.07	$\frac{V}{\#\cdot s}$
k_2^-	(R_4)	10^{-9}	s^{-1}
k_3^+	(R_5)	0.17	$\frac{V}{\#\cdot s}$
k_3^-	(R_6)	$1.21 \cdot 10^{-6}$	s^{-1}
k_4	(R_7)	$8.3 \cdot 10^{-3}$	s^{-1}
k_5^+	(R_8)	9.74	$\frac{V}{\#\cdot s}$
k_5^-	(R_9)	3.56	s^{-1}
k_6	(R_{10})	2.33	$\frac{V}{\#\cdot s}$
k_7	(R_{11})	$4.31 \cdot 10^{-5}$	$\frac{V}{\#\cdot s}$
k_8	(R_{12})	$2.73 \cdot 10^{-7}$	$\frac{V}{\#\cdot s}$
k_9	(R_{13})	$3.2 \cdot 10^{-5}$	s^{-1}
k_{10}	(R_{14})	$\varphi(42) = 7.77 \cdot 10^{-5}$	s^{-1}
k_{11}^+	(R_{15})	$3.32 \cdot 10^{-3}$	$\frac{V}{\#\cdot s}$
k_{11}^-	(R_{16})	4.44	s^{-1}
k_{12}	(R_{17})	13.94	s^{-1}

Metabolite	Init. no.
hsf	0
hsf_2	0
hsf_3	0
hse	29
hsf_3 : hse	2
hsp	766
hsp : hsf	1403
mfp	517
hsp : mfp	71
prot	$1.15 \cdot 10^8$

Table 3: The numerical values of the parameters and the initial numbers of molecules in the stochastic model. The numerical quantities are obtained by adopting the corresponding values in Petre et al. (2009a): the initial numbers of molecules are truncated to natural numbers, the value of the rate constant k_1^+ is twice the value of the corresponding deterministic rate constant. # denotes the number of molecules, V is the cell volume and s - second.

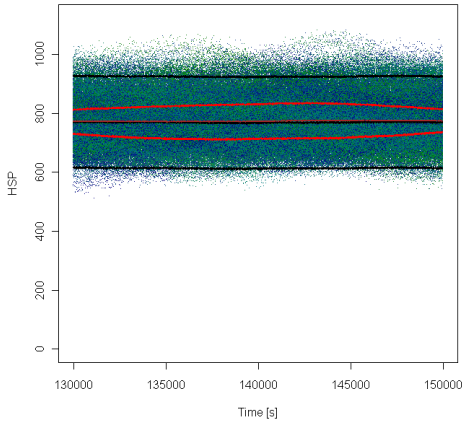


Figure 4: The median of the 1000 realisations on the time interval $T = \{130000s, \ldots, 150000s\}$ (middle black line). The upper and lower black lines are the median $\pm 1/4$ of the range of dynamics the model exhibits in the 1000 realisations on the considered time interval. The mean of all the realisations, of the subset S_U and S_L plotted with brown, upper red and lower red lines, respectively.

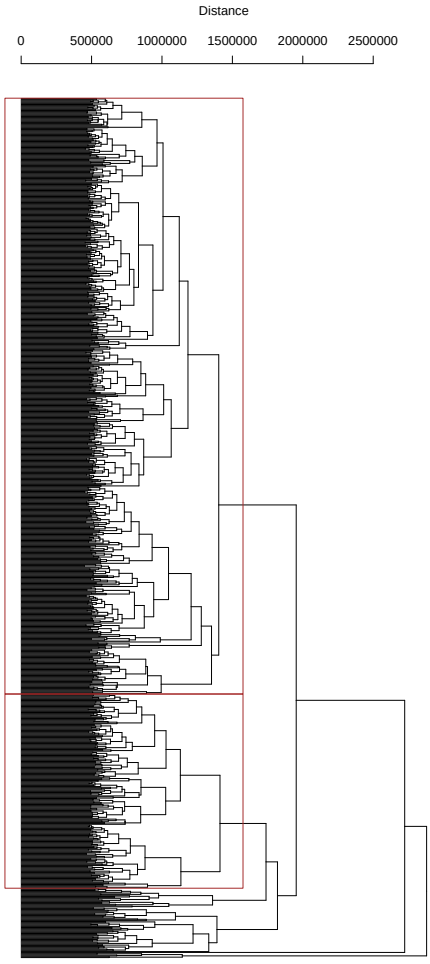
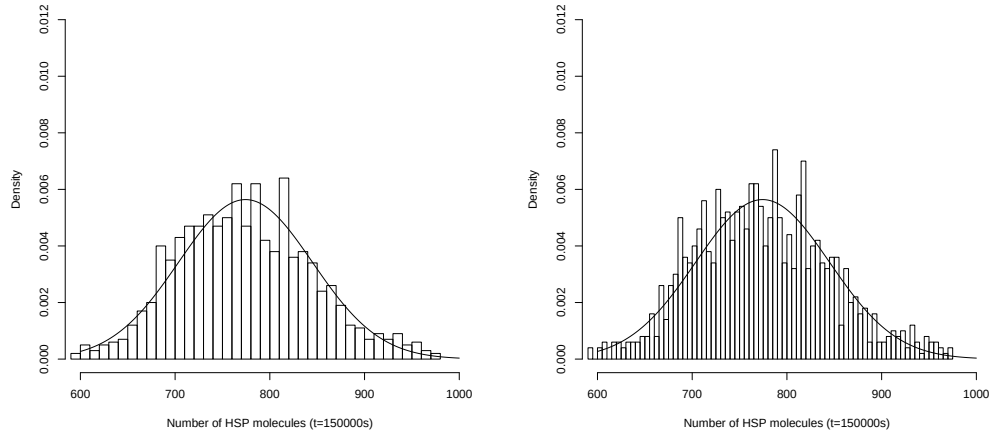


Figure 5: The clustering tree (dendrogram) obtained with the *agnes* clustering algorithm with the *average* method and the *manhattan* metric applied to the 1000 stochastic realisations considered on the time interval 130000 – 150000 seconds. The leaves of the clustering tree are the original realisations. Two branches come together at the distance between the two clusters being merged. The agglomerative coefficient equals 0.82. The rectangles distinguish two subclusters discussed in Section 4.



(a) (b)

Figure 6: Histograms overlaid with the normal distribution curve with mean and standard deviation computed at time point $t = 150000s$ from all 1000 realisations: (a) bin width set to 20, (b) bin width set to 10.