SUPPORTING TEXT

Description of the Full-Order Heat Shock Model

In building the heat-shock model, we made the following assumptions. During transcription initiation, RNA polymerase (RNAP) binds reversibly to the promoter region. We assume that this process, in addition to the subsequent formation of an open complex, achieves rapid equilibrium (1). Transcription initiation is assumed to be a pseudofirst-order reaction with rate K_{tr}. Similarly, translation initiation is assumed to proceed with a pseudofirst-order rate K_{TL}. For most *Escherichia coli* operons, initiation and elongation rates are such that ribosome queuing does not occur (2). We therefore take each transcription and translation initiation reaction to be independent. For exponential growth in cell volume, proceeding as e^{kt}, the dilution of species concentration is directly incorporated in the degradation rate. The degradation of σ^{32} by the protease FtsH in this model is implemented through its interaction with DnaK, and its cochaperone DnaJ. Raising the temperature produces an increase in the cellular levels of unfolded proteins that then titrate DnaK/J away from σ^{32}, allowing it to bind to RNAP (resulting in increased transcription) and stabilizing it in the process (3). Other possible models favor the direct titration of proteases by unfolded proteins as the mechanism underlying σ^{32} stabilization. We do not expect qualitative alterations to the predictions of our model with this alternative scenario.

Based on these assumptions, the mathematical model that we propose to describe the dynamics of the heat shock response uses mass-action first order kinetics to describe both the synthesis of new proteins (σ factors, chaperones, and proteases), and the association/dissociation activity of molecules. This modeling approach produces a set of ordinary differential equations. Upon simulation, those equations exhibited numerical stiffness. Usually, this behavior is due to the interaction of some fast and slow dynamics and is a manifestation of rate constants different by several orders of magnitude. The observed numerical stiffness imposed the necessity of transforming the differential equations that describe the fast states into algebraic constraints through a singular perturbation argument. It is a common practice to assume the binding rates (association and dissociation) between proteins or between proteins and specific DNA promoters are fast compared with the rate of synthesis and degradation of mRNAs and proteins. Therefore, we assumed that the binding dynamics reach their steady state very fast compared with other reactions in the system. We also used mass-

balance equations to relate the total quantity of a species in the system to its free concentration and the concentration of the different compounds where it appears. The resulting model is a set of differential algebraic equations (DAEs), which are of the form:

$$\dot{\mathrm{X}}(\mathrm{t}) = \mathrm{F}(\mathrm{t};\mathrm{X};\mathrm{Y}) \quad [1]$$

$$0 = \mathrm{G}(\mathrm{t};\mathrm{X};\mathrm{Y}) \quad [2]$$

where X is an 11-dimensional vector whose elements are the differential variables, and Y is a 20-dimensional vector whose elements are algebraic variables. This form is known as a semi-explicit DAE, with Eq. 2 being the constraint equation. If we differentiate Eq. 2 with respect to time, we get the following:

$$0 = \mathrm{G_X}(\mathrm{t};\mathrm{X};\mathrm{Y})\dot{\mathrm{X}} + \mathrm{G_Y}(\mathrm{t};\mathrm{X};\mathrm{Y})\dot{\mathrm{Y}} + \mathrm{G_t}(\mathrm{t};\mathrm{X};\mathrm{Y})$$

If $\mathrm{G_Y}(\mathrm{t},\mathrm{X},\mathrm{Y}) = \frac{\partial \mathrm{G}(\mathrm{t},\mathrm{X},\mathrm{Y})}{\partial \mathrm{Y}}$ is nonsingular, the system is an implicit set of ordinary differential equations. Therefore, the DAE system is of index one and is solvable by backward differentiation formulas, as implemented in specialized software packages such as DASSL (4).

Model Equations

Transcription and translation equations

$$\begin{aligned}
fracd[mRNA(DnaK)]dt &= K_{tr_1}.[\sigma^{32}: RNAP : ph] - \alpha_{mRNA}.[mRNA(DnaK)] \\
\frac{d[DnaK_t]}{dt} &= K_{TL}.[mRNA(DnaK)] - \alpha_{prot}.[DnaK_t] \\
\frac{d[mRNA(FtsH)]}{dt} &= K_{tr_2}.[\sigma^{32}: RNAP : ph - \alpha_{mRNA}.[mRNA(FtsH)] \\
\frac{d[FtsH_t]}{dt} &= K_{TL}.[mRNA(FtsH)] - \alpha_{prot}.[FtsH_t] \\
\frac{d[mRNA(protease)]}{dt} &= K_{tr_3}.[\sigma^{32}: RNAP : ph] - \alpha_{mRNA}.[mRNA(protease)] \\
\frac{d[protease_t]}{dt} &= K_{TL}.[mRNA(protease)] - \alpha_{prot}.[protease_t] \\
\frac{d[mRNA(HslVU)]}{dt} &= K_{tr_4}.[\sigma^{32}: RNAP : ph] - \alpha_{mRNA}.[mRNA(HslVU)] \\
\frac{d[HslVU_t]}{dt} &= K_{TL}.[mRNA(HslVU)] - \alpha_{prot}.[HslVU_t] \\
\frac{d[mRNA(\sigma^{32})]}{dt} &= K_{tr_5}.[\sigma^{70}: RNAP : pg] - \alpha_{mRNA}.[mRNA(\sigma^{32})] \\
\frac{d[\sigma_t^{32}]}{dt} &= K_{TL}.\eta(T).[mRNA(\sigma^{32})] - \alpha_{prot}.[\sigma_f^{32}] - \alpha_{FtsH}.[\sigma^{32}: DnaK : FtsH] \\
&- \alpha_{protease}(T).[\sigma^{32}:DnaK:protease] - \alpha_{HslVU}(T).[\sigma^{32} : HslVU] \\
\frac{d[P_{folded}]}{dt} &= K_{fold}.[P_{unfolded} : DnaK] - K(T).[P_{folded}]
\end{aligned}$$

Algebraic binding equations

$$
\begin{aligned}
[\sigma^{70}\text{: } RNAP] &= K_1.[\sigma^{70}_f].[RNAP_f] \\
[\sigma^{32}\text{: } RNAP] &= K_2.[\sigma^{32}_f].[RNAP_f] \\
[RNAP : D] &= K_3.[RNAP_f].[D_t] \\
[\sigma^{32}\text{: } DnaK : FtsH] &= K_4.[\sigma^{32}\text{: } DnaK].[FtsH_f] \\
[\sigma^{32}\text{: } DnaK] &= K_5.[\sigma^{32}_f].[DnaK_f] \\
[\sigma^{32} : DnaK : protease] &= K_6.[\sigma^{32} : DnaK].[protease_f] \\
[\sigma^{32} : HslVU] &= K_7.[\sigma^{32}_f].[HslVU_f] \\
[P_{unfolded} : DnaK] &= K_8.[P_{unfolded}].[DnaK_f] \\
[\sigma^{32}\text{: } RNAP : ph] &= K_9.[\sigma^{32}\text{: } RNAP]([ph_t] - [\sigma^{32}\text{: } RNAP : ph]) \\
[\sigma^{70}\text{: } RNAP : pg] &= K_{10}.[\sigma^{70}\text{: } RNAP]([pg_t] - [\sigma^{70}\text{: } RNAP\text{: } pg]) \\
[\sigma^{70}\text{: } RNAP : D] &= K_{11}.[\sigma^{70}\text{: } RNAP].[D_t] \\
[\sigma^{32}\text{: } RNAP : D] &= K_{12}.[\sigma^{32}\text{: } RNAP].[D_t]
\end{aligned}
$$

Mass balance equations

$$
\begin{aligned}
[RNAP_t] &= [RNAP_f] + [\sigma^{70}\text{: } RNAP] + [\sigma^{32}\text{: } RNAP] + [RNAP : D] + [\sigma^{70}\text{: } RNAP : D] \\
&+ [\sigma^{32}\text{: } RNAP : D] + [\sigma^{70}\text{: } RNAP : pg] + [\sigma^{32}\text{: } RNAP : ph] \\
[\sigma^{70}_t] &= [\sigma^{70}_f] + [\sigma^{70}\text{: } RNAP] + [\sigma^{70}\text{: } RNAP : D] + [\sigma^{70}\text{: } RNAP : pg] \\
[\sigma^{32}_t] &= [\sigma^{32}_f] + [\sigma^{32}\text{: } DnaK : protease] + [\sigma^{32}\text{: } RNAP] + [\sigma^{32}\text{: } RNAP : D] \\
&+ [\sigma^{32}\text{: } DnaK : FtsH] + [\sigma^{32}\text{: } DnaK] + [\sigma^{32}\text{: } RNAP\text{:}ph] + [\sigma^{32} : HslVU] \\
[DnaK_t] &= [DnaK_f] + [\sigma^{32}\text{: } DnaK : FtsH] + [\sigma^{32}\text{: } DnaK] + P_{unfolded} : DnaK] \\
&+ [\sigma^{32} : DnaK : protease] \\
[FtsH_t] &= [FtsH_f] + [\sigma^{32}\text{: } DnaK : FtsH] \\
[HslVU_t] &= [HslVU_f] + [\sigma^{32}\text{: } HslVU] \\
[protease_t] &= [protease_f] + [\sigma^{32}\text{: } DnaK : protease] \\
[Protein_t] &= [P_{unfolded}] + [P_{unfolded} : DnaK] + [P_{folded}]
\end{aligned}
$$

Parameter values

The rate parameters used in the model equations were determined by using various sources. The binding and degradation constants were taken from the literature of the heat-shock response. The synthesis rates for different proteins, σ factors, and chaperones were tuned to produce biologically plausible numbers of these quantities in the cell in the wild-type heat shock. Those parameters are given in Table 1. Initial values for the state variables are given in Table 2, and other constants used are given in Table 3.

Sensitivity Analysis

To study the sensitivity of the model to parametric uncertainty characterized by θ, we need to find the derivative of the solution $\begin{bmatrix} X \\ Y \end{bmatrix}$ with respect to each parameter. Assuming that this solution exists and is uniquely defined on the time interval of interest, we can write:

$$\begin{aligned} X(t,\theta) &= X_0 + \int_{t_0}^{t} F(s;X;Y;\theta)ds \\ 0 &= G(t;X;Y;\theta) \end{aligned}$$

Differentiating with respect to θ, we get:

$$\begin{aligned} X_\theta(t,\theta) &= \int_{t_0}^{t} [\frac{\partial F}{\partial X}(s;X;Y;\theta)X_\theta(s,\theta) + \frac{\partial F}{\partial Y}(s;X;Y;\theta)Y_\theta(s,\theta) + \frac{\partial F}{\partial \theta}(s;X;Y;\theta)]ds \\ 0 &= \frac{\partial G}{\partial X}(t;X;Y;\theta)X_\theta(t,\theta) + \frac{\partial G}{\partial Y}(t;X;Y;\theta)Y_\theta(t,\theta) + \frac{\partial G}{\partial \theta}(t;X;Y;\theta) \end{aligned}$$

Here, again $X_\theta(t,\theta) = \frac{\partial X(t,\theta)}{\partial \theta}$ and $Y_\theta(t,\theta) = \frac{\partial Y(t,\theta)}{\partial \theta}$.

Now, differentiating X_θ with respect to t, we obtain the following set of differential equations:

$$\dot{X}_\theta(t,\theta) = \frac{\partial F}{\partial X}(t;X;Y;\theta)X_\theta(t,\theta) + \frac{\partial F}{\partial Y}(t;X;Y;\theta)Y_\theta(t,\theta) + \frac{\partial F}{\partial \theta}(t;X;Y;\theta).$$

These equations, along with the algebraic constraint involving G produce an additional set of $(n+m)\times q$ sensitivity equations, which together with the original system yield:

$$\begin{aligned} \dot{X}(t) &= F(t;X;Y;\theta) \\ \dot{X}_\theta(t,\theta) &= \frac{\partial F}{\partial X}(t;X;Y;\theta)X_\theta(t,\theta) + \frac{\partial F}{\partial Y}(t;X;Y;\theta)Y_\theta(t,\theta) + \frac{\partial F}{\partial \theta}(t;X;Y;\theta) \\ 0 &= G(t;X;Y;\theta) \\ 0 &= \frac{\partial G}{\partial X}(t;X;Y;\theta)X_\theta(t,\theta) + \frac{\partial G}{\partial Y}(t;X;Y;\theta)Y_\theta(t,\theta) + \frac{\partial G}{\partial \theta}(t;X;Y;\theta) \end{aligned}$$

$$X(t=t_0) = X_0(\theta)$$

$$Y(t=t_0) = Y_0(\theta)$$

$$X_\theta(t_0,\theta) = 0$$

These equations form a new $(n+m)\times(q+1)$ DAE, whose solution would give the sensitivity to parameter variations along the trajectory of the system. It is worth noting here that this type of sensitivity analysis is referred to as small sensitivity and, as the name indicates, is applied to cases where the change in

parameters is small. The solution of the perturbed system is then approximated by a Taylor series expansion of the form:

$$\begin{aligned} X(t,\theta+\Delta\theta) &= X(t,\theta)+\Delta\theta\frac{\partial X(t,\theta)}{\partial\theta} \\ &+ \text{higher-order terms.} \end{aligned}$$

For large parameter variations, one should adopt what is referred to as sensitivity in the large, which is mainly a measure of the error between the nominal and perturbed system trajectories.

Model Validation

The level, activity, and stability of σ^{32} are the key regulatory elements of the heat-shock response. Upon exposure to heat, the time profile of the level of σ^{32} shows a fast but transient increase due to increased translation and stabilization of the otherwise very unstable σ^{32} (Fig. 1). The transient stabilization of σ^{32} is a result of the large number of unfolded proteins that titrate σ^{32} away from DnaK. The increased level of σ^{32} leads to the synthesis of large numbers of molecular chaperones and proteases that in turn act as a negative feedback on the level of σ^{32}. This negative feedback leads to the decrease in the level of σ^{32} until it reaches a steady state 3 to 4-fold larger than the value at a low temperature.

To reproduce this behavior and validate our model, we adopted the following strategy. We complemented the parameters picked from the literature by fitting the unavailable parameters to reproduce the steady-state levels of σ^{32} and chaperones at low temperature. The parameter sets that reproduced both the typical transient response and steady-state for chaperones and σ^{32} at high temperature were kept and the other discarded. To discriminate among the remaining sets of parameters, we simulated various heat-shock mutants. The set of parameters that was finally kept corresponded to values of the unavailable parameters that reproduced the steady state, the specific shape of the transient response, and the various heat-shock mutant data without any additional tuning. The results are depicted in Fig. 1, where the simulations are shown for the time response of the σ^{32} cellular level (temperature increases at t = 400min).

The result of this simulation agrees qualitatively with experimental results. For the set of parameters used, the agreement of the model predictions with the data was also quantitative. The levels and rate of synthesis of the chaperones under normal temperature and heat-shock conditions also agree with the literature data (12) (see Fig. 2).

Heat shock was also thoroughly investigated for FtsH-null mutants (12). In this mutant, loss of FtsH function causes marked stabilization of σ^{32} at low temperature ($\approx$ 20-fold of the wild type) and high temperature ($\approx$ 35-fold of the wild type), whereas induction of heat-shock protein synthesis occurred almost normally in these cells. As shown in Fig. 1, our model provides good

agreement with this accumulation of σ^{32} and the qualitative dynamics for the relative DnaK synthesis (Fig. 2).

Additionally, the model was tested for the case where σ^{32} was overproduced at $30°C$. This case was investigated by Straus *et. al* (13) who observed that when σ^{32} was overproduced at normal growth temperature, the level of σ^{32} increased and remained elevated 60 min.. It was also observed that the synthesis of DnaK was only transiently induced, reaching a peak at $\approx$ 10 min. after σ^{32} overproduction and then declining. This behavior for the σ^{32} level and rate of synthesis of chaperones could be reproduced by our model (data not shown).

Description of the Reduced-Order Heat-Shock Model

Although the detailed modeling of the heat-shock response involves nonlinear equations with a level of complexity that cannot be analytically tractable, we found that the basic functional modules in this response and their qualitative behavior can be fully captured by a simplified model whose components are exposed in what follows. For ease of notation, we will denote the chaperones by D, the protease by F, the σ^{32} by S, and the unfolded proteins by U_f. Furthermore, we refer to the total quantity of X by X_t and the free quantity of X by X_f. As in the full heat-shock model, we will assume that D is produced at a rate $\mathrm{K_d}$ and is degraded at rate α_d. D can reversibly bind to the σ-factor S or to the unfolded protein $\mathrm{U_f}$. In turn, the level of free σ, S_f, influences the rate of transcription of protein D. S itself is produced at a temperature-dependent rate $\eta(\mathrm{T})$ and degraded through a regulated mechanism involving D and F. The proteins are folded through the interaction with D and unfolded with a temperature-dependent rate $K(T)$. The chemical reactions describing this network are naturally divided into two categories, fast and slow. The fast reactions are assumed to be in equilibrium with respect to the slow reactions. In terms of differential equations, those reactions translate to the following:

$$\begin{aligned}
\frac{dS_t}{dt} &= \eta(T) - \alpha_0 S_t - \alpha_s S : D : F \\
\frac{dD_t}{dt} &= K_d S_f - \alpha_p D_t \\
\frac{dU_f}{dt} &= K(T) P_{folded} - K_{fold} U : D
\end{aligned} \qquad [3]$$

where $U : D$ is the complex formed by the binding of U_f to D, and $S : D : f$ is the complex formed by the binding of S to F mediated by the binding to D. The total concentration of protein D, unfolded proteins U_f, F, and S are constant, so that we can write the mass balance equations

$$\begin{aligned}
S_t &= S_f + S : D + S : D : F \\
D_t &= D_f + U : D + S : D + S : D : F \\
F_t &= F_f + S : D : F,
\end{aligned} \qquad [4]$$

where P_t is the total number of proteins in the cell, considered here to be constant. We can now apply some simplifying assumptions. We eliminate *U:D* and *S:D:F* from Eq. 3 by utilizing the fact that the binding reactions are fast compared with expression and degradation and write algebraic expressions:

$$\begin{aligned} S:D &= K_s.S_f.D_f \\ U:D &= K_u.U_f.D_f \\ S:D:F &= K_f F_f S:D. \end{aligned} \quad [5]$$

Using the relationships in Eqs. 4 and 5, we get the following relationships:

$$S_f = \frac{S_t}{1 + K_s D_f (1 + K_f F_f)}$$

If $S_t \ll D_t$ (as in the case of the heat-shock response), the expression for D_f simplifies to:

$$D_f \simeq \frac{D_t}{1 + K_u U_f}$$

and consequently:

$$S_f \simeq \frac{S_t}{1 + \frac{K_s D_t 1 + K_f F_f}{1 + K_u U_f}}.$$

Under these assumptions, Eqs. 3 become

$$\begin{aligned} \frac{dS_t}{dt} &= \eta(T) - \alpha_0.S_t - \alpha_s.\frac{\alpha K_s K_f D_t^2}{\Gamma - \kappa D_t + K_s D_t (1 + \alpha K_f D_t)} S_t \\ \frac{dD_t}{dt} &= K_d.\frac{\Gamma - \kappa D_t}{\Gamma - \kappa D_t + K_s D_t (1 + \alpha K_f D_t)} S_t - \alpha_d D_t, \end{aligned} \quad [6]$$

where we have defined $\Gamma \doteq 1 + K_u P_t$, $\kappa \doteq \frac{K_u (K_T + K_f)}{K_T}$, and assumed that $F_f = \alpha D_t$. This assumption is justified by the fact that FtsH is part of the σ^{32} regulon, and that is much more abundant in the cell than σ^{32}. P_t is the total number of proteins in the cell (assumed to be constant), K_s, K_f, and K_u are binding rate constants, and α_0, α_d, and α_s are degradation constants.

Under the above-mentioned assumptions, the simplified model reproduces qualitatively the dynamics of the heat-shock response system. The equations involved in this model have the benefit of being amenable to analytical analysis, an unfeasible task for the full model with its layered complexity.

It is worth noting here that even though the description of the reduced-order model seems to be based on empirical interactions, the same description could be exactly derived from the full model through various algebraic manipulations and approximations *. Furthermore, the expressions of the reduced-order model can be exactly cast into a modular decomposition shown in Fig. 3.

*The model reduction procedure involves details and algebraic manipulations that are omitted here.

1. DeHaseth, P. L., Zupancic, M. L. & Record, M. T. (1998) *J. Bacteriol.* **180**, 3019-3025.

2. Yarchuk, O., Jacques, N., Guillerez, J. & Dryfus, M. (1992) *J. Mol. Biol.* **226**, 581-596.

3. Morita, M., Kanemori, M., Yanagi, H. & Yura, T. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 5860-5865.

4. Petzold, L. R. (1983) in *Scientific Computing*, ed. Stepleman R. S. (North-Holland, Amesterdam), pp. 65-68.

5. Record, M. T., Reznikoff, W. S., Craig, M. L., Mcquade, K. L. & Schelax, P. J. (1996) in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, eds. Neidhart, F. C., Curtis, R. I., Ingraham, J. L., Lin, C. C., Low, K. B., Magasnik, B., Reznikoff, W. S., Riley, M., Schaechter, M. & Umbarger, H. E. (Ame. Soc. Microb., Washington, DC), pp. 792-821.

6. Blaszczak, A., Zylicz, M., Georgopoulos, C. & Liberek, K. (1995) *EMBO J.* **14**, 5085-5093.

7. Dehaseth, P. L., Lohman, T. M., Burgess, R. B. & Record, M. T. (1978) *Biochemistry* **17**, 1612-1622.

8. Schmidt, D., Baici, A., Gehring, H. & Christen, P. (1994) *Science* **263**, 971-973.

9. Morita, M., Kanemori, M., Yanagi, H. & Yura, T. (1999) *J. Bacteriol.* **181**, 401-410.

10. Bremer, H. & Dennis, P. (1996) in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, eds. Neidhart, F. C., Curtis, R. I., Ingraham, J. L., Lin, C. C., Low, K. B., Magasnik, B., Reznikoff, W.S., Riley, M., Schaechter, M. & Umbarger, H.E. (Am. Soc. Microb., Washington, DC), pp. 1553-1569.

11. Kanemori, M., Nishihara, K., Yanagi, H. & Yura, T. (1997) *J. Bacteriol.* **179**, 7219-7225.

12. Tatsuta, T., Tomoyasu, T., Bukau, B., Kitagawa, M., Mori, H., Karata, K. & Ogura, T. (1998) *Mol. Microb.* **30**, 583-593.

13. Straus, D. B., Walter, D. B. & Gross, C. A. (1989) *Genes Dev.* **3**, 2003-2010.