

# Robustness Analysis of the Heat Shock Response in *E. coli*

H. El-Samad \*, M. Khammash \*, H. Kurata †, J. Doyle ‡

\* Electrical and Computer Engineering

Iowa State University

Ames, Iowa 50011

† Department of Biochemical Science and Engineering

Kyushu Institute of Technology

Izuka, 820-8502, Japan

‡ Control and Dynamical Systems

California Institute of Technology

Pasadena, CA 91125

## Abstract

The bacterial heat shock response refers to the mechanism by which bacteria react to a sudden increase in the ambient temperature of growth. The consequences of such an unmediated temperature increase at the cellular level is the unfolding, misfolding, or aggregation of cell proteins, which threatens the life of the cell. Cells respond to the heat stress by initiating the production of heat-shock proteins whose function is to refold denatured proteins into their native states. The heat shock response, through the elevated synthesis of molecular chaperones and proteases, enables the repair of protein damage and the degradation of aggregated proteins. In a previous work [1], we have devised a dynamic model for the heat shock response in *E. coli*. In the present paper, we provide a thorough discussion of the dynamical nature of this model. We use sensitivity analysis and simulation tools to illustrate the remarkable efficiency, robustness, and stability of the heat shock response system.

## 1 Introduction

An important environmental challenge that organisms like *E. coli* face during their lifetime is the exposure to higher than normal temperatures. To fight the detrimental effects of high temperatures, *E. coli* has evolved an elaborate heat shock response mechanism. In a previous work [1], we have devised a model that describes this mechanism. The model proposed was able to reproduce the dose-response of the production of *hsps* as a function of temperature, while taking into account all the known biology of the heat shock response. In addition, the model could explain the altered phenotype of

mutants in which genes involved in the pathway have been deleted. We had identified both feedforward and feedback components to the heat shock control mechanism. In this paper, we explore the dynamic structure of the heat shock response model developed in [1]. We show that the feedforward component guarantees a fast response that insures the quick and efficient refolding of proteins, while the feedback structure regulates the transcription of the heat shock proteins and ensures both remarkable stability and robustness to parameter variations and uncertainties. For this purpose, we use extensive simulations and study the sensitivity of the model to different parameters. Finally, we discuss the optimality of the remarkably conserved heat shock response system in terms of efficiency, performance, and robustness.

## 2 Biology of the Heat Shock Response and the Heat Shock Model

In *E. coli*, much of the regulation of the concentrations of proteins occurs at the level of transcription and translation [2]. The enzyme RNA polymerase (RNAP) bound to a regulatory sigma factor recognizes the promoter and then transcribes specific genes into messenger RNA (*mRNA*). The *mRNA* is translated by the ribosome into protein. At physiological temperatures (30°C to 37°C), RNAP is bound to the major sigma factor  $\sigma^{70}$ . The *RNAP* :  $\sigma^{70}$  complex transcribes the genes necessary for growth at normal temperatures. When *E. coli* are exposed to high temperatures, the special heat shock sigma factor,  $\sigma^{32}$  encoded by the *rpoH* gene, is rapidly induced.  $\sigma^{32}$  binds to RNA polymerase and directs the transcription of a small set (approximately 20) of heat shock genes [2]. The heat shock

genes encode for molecular chaperones (GroEL, DnaK, DnaJ, GroES, GrpE, etc.) that are involved in refolding denatured proteins. Another class of heat shock proteins are proteases (Lon, FtsH, etc.) that function to degrade unfolded proteins. In an *rpoH* null mutant that does not make  $\sigma^{32}$ , the heat shock proteins are not induced and the cells are viable only at temperatures below 20°C [3].

There are two mechanisms by which  $\sigma^{32}$  levels are increased when the temperature is raised. First, the translation rate of the *rpoH* mRNA increases immediately, resulting in a fast 10-fold increase in the concentration of  $\sigma^{32}$  [4]. Second, under non-stress conditions,  $\sigma^{32}$  is recognized and sequestered by the *hsp* chaperone *DnaK*. The concentration of the  $\sigma^{32} : DnaK$  complex depends on the amount of *DnaK* that is bound to unfolded proteins. Raising the temperature produces an increase in the cellular levels of unfolded proteins which titrate *DnaK* away from  $\sigma^{32}$ , resulting in more  $\sigma^{32}$  that is capable of binding to RNA polymerase and initiating the transcription of the heat shock genes [5]. The accumulation of high levels of heat shock proteins leads to the down regulation of the response. During this phase, the abundant chaperones efficiently refold most of the denatured proteins thereby decreasing the pool of unfolded protein, freeing up *DnaK* to sequester  $\sigma^{32}$  from RNA polymerase. In addition, heat shock proteases such as *FtsH* degrade  $\sigma^{32}$ . The result is a decrease in the concentration of  $\sigma^{32}$  to a new steady state concentration that is dictated by the balance between the temperature-dependent translation of the *rpoH* mRNA and the level of  $\sigma^{32}$  activity modulated by the *hsp* chaperones and proteases acting in a negative feedback fashion. The above relationships are illustrated in Figure 1. To put those relationships in a more mathematical context, we have devised a model that describes the transcription, translation, and protein folding dynamics by differential equations, and the binding dynamics by algebraic equations. The result was a *Differential-Algebraic* set of equations of the form:

$$\begin{aligned}\dot{X}(t) &= F(t; X; Y) \\ 0 &= G(t; X; Y) \\ X(t = t_0) &= X_0 \\ Y(t = t_0) &= Y_0\end{aligned}$$

where  $\mathbf{X}$  is a 11-dimensional vector whose elements are the differential variables and  $\mathbf{Y}$  is a 20-dimensional vector whose elements are the algebraic variables. This form is known as a *semi-explicit* DAE and will be explored in the next section. The full model equations are detailed and fully explained in [1].

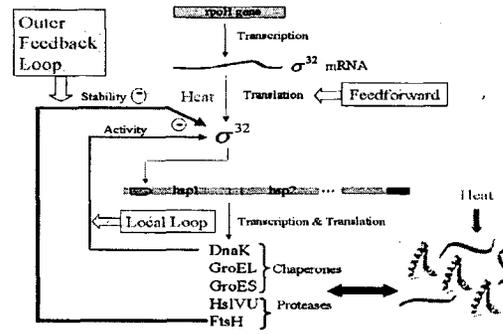


Figure 1: The heat shock response in *E. coli*

### 3 Differential-Algebraic Equations and Sensitivity Equations

In the most general form, a nonlinear DAE is given by:

$$F(t, \dot{X}, X, \theta) = 0$$

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

A special form arises in classical singular perturbation problems, where after simplifying the model to get a first order approximation of its behavior, the DAE takes the form:

$$\begin{aligned}\dot{X}(t) &= F(t; X; Y; \theta) \\ 0 &= G(t; X; Y; \theta) \\ X(t = t_0) &= X_0(\theta) \\ Y(t = t_0) &= Y_0(\theta)\end{aligned}\tag{1}$$

Here  $X$  is an  $n$ -dimensional vector of differential variables,  $Y$  is an  $m$ -dimensional vector of algebraic variables, and  $\theta$  is a  $q$ -dimensional vector of time-independent parameters whose nominal values are given. This form is known as a *semi-explicit* DAE, with (1) being the constraint equation. As mentioned before, and explained in [1], the heat shock response model takes this form after considering the fast and slow dynamics associated with synthesis and binding dynamics respectively. If we differentiate (1) with respect to time, we get the following set of equations:

$$\begin{aligned}\dot{X} &= F(t; X; Y; \theta) \\ G_X(t; X; Y; \theta)\dot{X} + G_Y(t; X; Y; \theta)\dot{Y} \\ &+ G_t(t; X; Y; \theta) = 0\end{aligned}$$

If  $G_Y(t, X, Y, \theta) = \frac{\partial G(t, X, Y, \theta)}{\partial Y}$  is nonsingular, the system is an implicit ODE. Therefore, the DAE system is of index one and is solvable [6]. For the heat shock

model derived in [1], the  $G_Y$  matrix is nonsingular, therefore the system is solvable through standard algorithms involving the *Backward Differentiation Formulas* as implemented in specialized software packages such as *DASSL* [6].

In order to study the sensitivity of the model to parametric uncertainty characterized by  $\theta$ , 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  $\theta$  we get:

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

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  $X_\theta$  with respect to  $t$ , we obtain the following set of differential equations:

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

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:

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

$$\begin{aligned} 0 &= G(t; X; Y; \theta) \\ 0 &= \frac{\partial G}{\partial X}(t; X; Y; \theta) X_\theta(t, \theta) \\ &+ \frac{\partial G}{\partial Y}(t; X; Y; \theta) Y_\theta(t, \theta) \\ &+ \frac{\partial G}{\partial \theta}(t; X; Y; \theta) \\ X(t = t_0) &= X_0(\theta) \\ Y(t = t_0) &= Y_0(\theta) \\ X_\theta(t_0, \theta) &= 0 \end{aligned}$$

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 the cases where the change in the 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. In what follows, we will use sensitivity analysis to study the robustness of the feedback architecture in the heat shock response system.

#### 4 The Feedback Architecture of the Heat Shock Response and Sensitivity Analysis Results

The main objective of the heat shock regulatory system is to refold denatured proteins upon exposure of the cell to higher than normal temperatures. There are different potential designs that could help achieve this task. However, any evolutionary advantageous design would have, in addition to satisfactory performance, some robustness features, so as to keep the outputs of interest constant despite potential fluctuations or uncertainties in the system parameters. In *E. coli* and a large number of other organisms, the task of refolding proteins upon heat shock is implemented through the up-regulation of cellular chaperones. However, these heat shock proteins occupy precious space in the cell and use its resources. Therefore, the cell must maintain a fine balance between the protective effect of these *hsps* and the metabolic burden of overexpressing them. In other terms, the number of chaperones should be large enough to achieve proper refolding of denatured

proteins and prohibit their aggregation, but must be kept reasonably low as not to cause a metabolic stress to the cell. Furthermore, the number of chaperones should be robustly constant despite small fluctuations in the number of the  $\sigma^{32}$  molecules, and uncertainties in the parameters involved in the synthesis of the *hsps*, such as the binding between  $\sigma^{32}$  and the gene promoter, or the transcription and translation rates.

In what follows, we will argue that the heat shock response has a simple design architecture that achieves efficiency, robustness, and performance. We will show that this architecture consists of three components: a *feedforward term*, an *inner loop* and an *outer loop*. These components act in concert to achieve different performance objectives. The task of achieving a fast and efficient response to the heat stress is implemented through a *feedforward* control term for  $\sigma^{32}$ . The biological implementation of this feedforward component is through the partial melting of the secondary structure of the *mRNA* upon exposure to high temperature, which enhances ribosome entry and translational initiation. This feedforward term drives the transition to the high temperature operation mode and assures reliable folding of proteins, therefore enhancing performance. The important task of maintaining a low metabolic burden on the cell and assuring robustness has necessitated the evolution of an *inner local loop* which acts through the regulation of the activity of  $\sigma^{32}$ . By sequestering  $\sigma^{32}$ , the chaperones modulate the number of the free  $\sigma^{32}$ , therefore limiting their activity in the transcription of unneeded heat shock proteins. This sequestration of free  $\sigma^{32}$  also buffers the effects of parameter variations and uncertainties, therefore enhancing robustness and maintaining the output signal *DnaK* within a narrow range, despite possibly fluctuating inputs and parameters. At the same time, an outer loop which acts through the degradation of  $\sigma^{32}$  by proteases, assures the stability of the whole system through hindering the build-up of  $\sigma^{32}$  resulting from continuous production. A block diagram depicting the full dynamics of the heat shock response is shown in Figure 2, where for simplicity, the folding dynamics are depicted as a disturbance to the free *DnaK* pool. Figure 3 is a highly simplified diagram that shows the basic architecture of the system as described above.

The importance and function of the feedforward term was investigated by comparing the protein folding regimes in the presence and absence of this term. The result is shown in Figure 4. It could be clearly seen that in the absence of the feedforward term, protein folding is impaired due to the absence of fast  $\sigma^{32}$  production. This results in insufficient chaperone production and consequently, impaired and slightly delayed protein folding. Indeed, as could be seen in Figure 4, in the

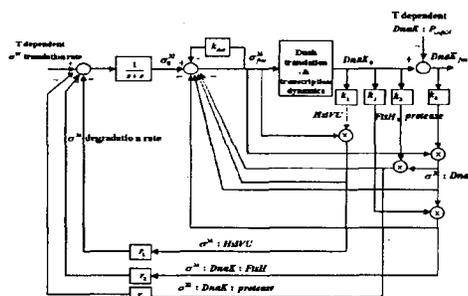


Figure 2: Control diagram of the heat shock response in *E. coli*

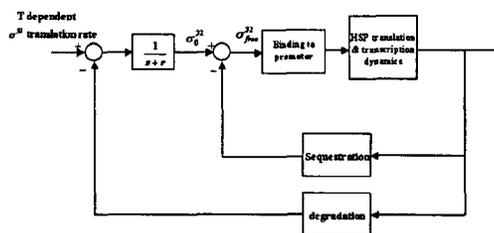
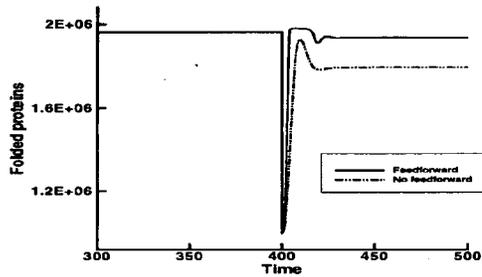


Figure 3: Simplified control diagram of the heat shock response in *E. coli*

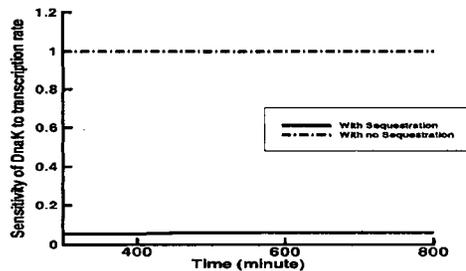
absence of the feedforward term, the level of unfolded proteins at steady-state after heat shock is around 11% of the total protein content of the cell, while it is less than 3% in the presence of the feedforward term.

In order to explore the advantages of the feedback architecture implemented through the inner and outer loop in the heat shock response system, we performed numerical sensitivity analysis. Motivated by the previous arguments, we put special emphasis on the sensitivity of the total chaperone level to the binding of  $\sigma^{32}$  to *RNAP* and gene promoter, in addition to the transcription and translation parameters of the heat shock proteins. In this analysis, we considered the full model, as well as a modified model where binding of  $\sigma^{32}$  to *DnaK* was inhibited, hence disabling the inner loop. In order to solve numerically for the sensitivity equations, we used DASPK3.0, an algorithm developed by Petzold *et. al* for the sensitivity analysis of large-scale differential algebraic systems [7]. We plotted the normalized sensitivity solutions  $\frac{\partial u}{\partial \theta_i} = \frac{u}{\theta_i} \frac{\partial \theta_i}{\partial u}$  where  $u$  is the output of interest and  $\theta_i$  is the parameter of interest. This normalized sensitivity function can be interpreted as the approximate percentage change in the state variable  $u$  for a 1% increase in the parameter  $\theta_i$ .



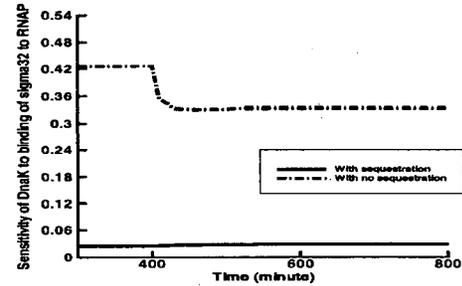
**Figure 4:** Level of folded proteins for the heat shock response system. Heat shock occurs at  $t=400$  min.

Figure 5 shows the sensitivity of the total *DnaK* level to the transcription rate in the full model along with a plot for the same sensitivity in the case where the inner loop was eliminated. In the no sequestration of  $\sigma^{32}$  by *DnaK* case, the sensitivity values increase tremendously, supporting the idea that the local loop is efficiently increasing the overall system robustness. A similar conclusion could be drawn from Figure 6 which depicts the sensitivity of *DnaK* to the binding constant between  $\sigma^{32}$  and *RNAP*. Again the plotted sensitivities are for the full model and the model where there is no sequestration of  $\sigma^{32}$  by *DnaK*.



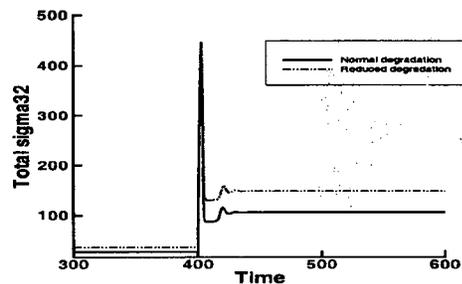
**Figure 5:** Normalized sensitivity of *DnaK* to the transcription rate. Heat shock occurs at  $t=400$ min.

In order to investigate the role of the inner feedback loop in improving robustness of the *DnaK* level to variations in the total number of  $\sigma^{32}$ , we attempted to alter the level of  $\sigma^{32}$  by changing the degradation constants. In particular, we decreased the degradation constant of  $\sigma^{32}$  by *FtsH* by one third of its value. Obviously, the level of the total  $\sigma^{32}$  increased by almost one third of its value at low temperature and almost 40% of its value at high temperature (Figure 7). However, as expected, the change in level of *DnaK* was highly attenuated (only 1.18% change at low temperature and 1.35% at



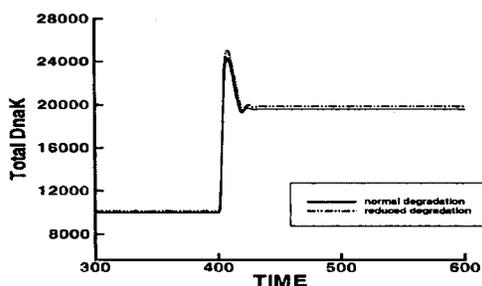
**Figure 6:** Normalized sensitivity of *DnaK* to the binding between  $\sigma^{32}$  and *RNAP*.

high temperature) (Figure 8).

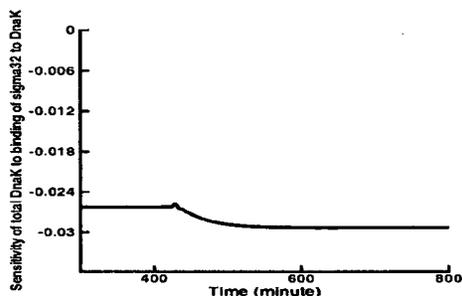


**Figure 7:** Total level of  $\sigma^{32}$  for different degradation constants of  $\sigma^{32}$  by *FtsH*. (—) Normal degradation constant, (---) Degradation constant reduced by one third of its normal value. Heat shock occurs at  $t=400$ min.

In addition to studying the robustness of the heat shock response system, sensitivity analysis helps in the identification of the parameters that determine the behavior of the system. This is of crucial importance in the modelling of biological systems, which frequently exhibit complex behavior and involve a large number of parameters. Insight into the parameters that a certain system is most sensitive to is of great value in the design of guided experiments. For example, in our system, it might be obvious that the level of chaperones should be sensitive to their synthesis rate as a result of the direct relationship between these two quantities. However, it may come as a surprise, uncovered by the sensitivity analysis, that the level of *DnaK* is almost equally sensitive to small perturbations in the binding constant between  $\sigma^{32}$  and *DnaK* than to small perturbations in the binding constant between  $\sigma^{32}$  and *RNAP* (see figure 9), although the level of *DnaK* is affected by these two parameters in opposite directions.



**Figure 8:** Total level of *DnaK* for different degradation constants of  $\sigma^{32}$  by *FtsH*. (-) Normal degradation constant, (-.-) Degradation constant reduced by one third of its normal value. Heat shock occurs at  $t=400$ min.



**Figure 9:** Normalized sensitivity of the total *DnaK* level to the binding constant between  $\sigma^{32}$  and *DnaK*. Heat shock occurs at  $t=400$ min.

The categorization of the heat shock response system into multiple loops that achieve different functions is a perfect example of the emerging theme of *functional modularity* in molecular cell biology. Survival of living systems necessitates that critical parameters of these functional modules are robust. We have proven that the block consisting of the transcription and translation of the heat shock proteins along with the feedback loop that regulates this function forms a module that is robust by nature. This module interacts with the rest of the system to achieve a larger function, which is the folding of proteins. These design principles that result in robust modular systems are used extensively in manmade machines.

## 5 Conclusion

In this paper, we discussed the design principles underlying the dynamics of the feedback structure in

a model of the heat shock response. We identified three components to this feedback structure and explored in detail the function of these components. We showed that robustness is an inherent property of the system and that a modular approach could be used to understand the functionality of the whole system. The heat shock response offers a valuable example that illustrates the necessity of the tools borrowed from control engineering in the understanding of molecular networks.

## Acknowledgments

The authors would like to thank Carol Gross at the University of California at San Francisco for useful discussions and valuable insight into the heat shock response. The authors would also like to acknowledge support from the ISU SPRIGS Grant.

## References

- [1] H.Kurata, H. El-Samad, T.-M. Yi, M.Khammash, J.C. Doyle, "Feedback Regulation of the Heat Shock Response in *E. Coli*", *Proceedings of the 40th IEEE Conference on Decision and Control*, pp. 837-842, 2001.
- [2] C.A. Gross, "Function and Regulation of the Heat Shock Proteins" in *Escherichia Coli and Salmonella: Cellular and Molecular Biology*, (eds F.C. Neidhart), ASM press, Washington D.C., pp. 1384-1394, 1996.
- [3] Y.N. Zhou, N. Kusukawa, J.W. Erickson, C.A. Gross, & T. Yura "Isolation and Characterization of *Escherichia Coli* Mutants that Lack the Heat Shock Sigma Factor  $\sigma^{32}$ ", *J. Bacteriol.*, vol. 170, pp. 3640-3649, 1988.
- [4] D.B. Strauss, W.A. Walter, & C.A. Gross, "The Activity of  $\sigma^{32}$  is Reduced Under Conditions of Excess Heat Shock Protein Production in *Escherichia Coli*", *Genes Dev.*, vol. 3, pp. 2003-2010, 1989.
- [5] F. Arsene, T. Tomoyasu, & B. Bukau, "The Heat Shock Response of *Escherichia Coli*", *International Journal of Food Microbiology*, vol. 55, pp. 3-9, 2000.
- [6] K. Brenan, L. Campbell, L. Petzold, *Numerical Solution of Initial Value Problems in Differential-Algebraic Equations*, Elsevier Science Publishing Co, N.Y., 1989.
- [7] S. li, L. Petzold, "Software and Algorithms for Sensitivity Analysis of Large-Scale Differential Algebraic Systems", *J. Comput. and Appl. Math.*, Vol. 125, No. 1-2, pp. 131-146, 2000.