Modeling and Analysis of Bacterial Stress Response: A Control Perspective

Mustafa Khammash, UCSB

Joint work with:
H. El-Samad, UCSB
H. Kurata, Kyushu Institute of Tech.
John Doyle, Caltech
Outline

• The biology of the heat-shock response

• Architecture of the heat-shock system
  – A biologist’s perspective
  – A control theorist’s perspective

• Modeling and Simulation
  – Deterministic
  – Stochastic

• Analysis of the heat-shock system
Loss of Protein Function

Network failure

Cell

Death

Unfolded Proteins

Aggregates

Folded Proteins

Temp cell

Temp environ
Heat-Shock Response

• The *heat-shock response* is a protective cellular response to deal with heat-induced protein damage.

• Involves building and dispatching heat-shock proteins (HSPs)
  
  – **Chaperones:**
    e.g. DnaK, DnaJ, GrpE, GroES, GroEL, …
  
  – **Proteases:**
    e.g. Lon, FtsH, HslVU, Clp, …
Function of Heat-Shock Proteins

I. Protein Folding

Unfolded/partially folded Proteins

DnaK

GroEL/
GroES

II. Protein Degradation

Proteases

Proteins Aggregates

Amino Acids

Folded Proteins
The Heat-Shock System
Heat-Shock Response

• In *E. coli*, induction of the heat-shock response is attributed to an increase in the regulator $\sigma^{32}$

• $\sigma^{32}$ regulates heat-shock protein gene expression

• Upon a temperature upshift from 30$^\circ$C to 42$^\circ$C the levels of $\sigma^{32}$ rapidly increases before reaching a new steady-state level
Heat-Shock Gene Transcription

σ^{32} factor

RNA Polymerase

start

promoter

DNA

hsp2

end

terminator
mRNA Translation

Heat-Shock Proteins

mRNA

ribosomes
Regulation of HSP Synthesis

Achieved through tight regulation of $\sigma^{32}$ at 3 levels:

- **Heat-Induced $\sigma^{32}$ Synthesis**
  - Translational induction is due to temperature melting of $rpoH$ mRNA secondary structure
• Regulation of $\sigma^{32}$ Activity

− Chaperones inhibit $\sigma^{32}$ activity by sequestering $\sigma^{32}$ away from RNAP. Unfolded proteins increase $\sigma^{32}$ activity.
- Regulation of $\sigma^{32}$ Degradation
  - $\sigma^{32}$ is transiently stabilized upon temperature upshift

but...

$\sigma^{32}$ is protected from degradation when bound to RNAP
Regulation of Heat-Shock Response

Heat → rpoH gene

Transcription & Translation

DnaK
GroEL
GroES
Chaperones

HslVU
FtsH
Proteases

Translation

σ 32

Heat

Degradation

Activity

Transcription
Heat treatment leads to the degradation of proteins. The unfolded proteins aggregate and are degraded by DnaK, RNAP, rpoH, FtsH, and Lon enzymes. The sigma factor (σ) is involved in the transcription of mRNA by RNAP. The rpoH protein is activated by heat, promoting the degradation of other proteins.
A Control System decomposition

A control theorist’s view. How do these two views relate?
\[ \dot{x} = f(x, Temp) \]

\[ x_k = \text{concentration} \]

\[ \dot{x}_1 = f_1(x, T) \]
\[ \dot{x}_2 = f_2(x, T) \]
\[ \vdots \]
\[ \dot{x}_n = f_n(x, T) \]

\[ 0 = g_1(x, T) \]
\[ 0 = g_2(x, T) \]
\[ \vdots \]

Developing the math model:
Start with the “modularity” of concentrations of each chemical component plus conservation constraints.
Mechanisms Modeled

- $\sigma^{32}$ synthesis (temperature dependent)
- Binding of $\sigma^{32}$ to RNAP
- Binding of $\sigma^{32}$:RNAP complex to gene promoters
- Transcription and translation of chaperones (represented by DnaK)
- Transcription and translation of FtsH
- Transcription and translation of other proteases (represented by HslVU)
• Competition among DnaK and RNAP for $\sigma^{32}$
• Degradation of $\sigma^{32}$ when unprotected by RNAP
• Protein denaturing (temperature dependent)
• Binding of DnaK to unfolded proteins
• Protein folding via DnaK
Mathematical Model

Protein Synthesis

\[
\begin{align*}
\frac{d[mRNA(DnaK)]}{dt} &= K_{tr1}.[\sigma^{32} : RNAP : ph] - \alpha_{mRNA}.[mRNA(DnaK)] \\
\frac{d[DnaK_i]}{dt} &= K_{TL}.[mRNA(DnaK)] - \alpha_{prot}.[DnaK_i] \\
\frac{d[mRNA(FtsH)]}{dt} &= K_{tr2}.[\sigma^{32} : RNAP : ph] - \alpha_{mRNA}.[mRNA(FtsH)] \\
\frac{d[FtsH_i]}{dt} &= K_{TL}.[mRNA(FtsH)] - \alpha_{prot}.[FtsH_i] \\
\frac{d[mRNA(protease)]}{dt} &= K_{tr3}.[\sigma^{32} : RNAP : ph] - \alpha_{mRNA}.[mRNA(protease)] \\
\frac{d[protease_i]}{dt} &= K_{TL}.[mRNA(protease)] - \alpha_{prot}.[protease_i] \\
\frac{d[mRNA(HslVU)]}{dt} &= K_{tr4}.[\sigma^{32} : RNAP : ph] - \alpha_{mRNA}.[mRNA(HslVU)] \\
\frac{d[HslVU_i]}{dt} &= K_{TL}.[mRNA(HslVU)] - \alpha_{prot}.[HslVU_i] \\
\frac{d[mRNA(\sigma^{32})]}{dt} &= K_{tr5}.[\sigma^{70} : RNAP : pg] - \alpha_{mRNA}.[mRNA(\sigma^{32})] \\
\frac{d[\sigma^{32}]}{dt} &= \eta(T) \cdot K_{TL} \cdot [mRNA(\sigma^{32})] - r\alpha_{prot}.[\sigma^{32}] - \alpha_{FtsH}.[\sigma^{32} : DnaK : FtsH] \\
&- \alpha_{protease}(T).[\sigma^{32} : DnaK : protease] - \alpha_{HslVU}(T).[\sigma^{32} : HslVU] \\
\frac{d[P_{folded}]}{dt} &= K_{fold}.[P_{unfolded} : DnaK] - K(T).[P_{folded}]
\end{align*}
\]
Binding Equations

\[
\begin{align*}
[\sigma^{70}: \text{RNAP}] &= K_1.[\sigma^{70}_j].[\text{RNAP}_j] \\
[\sigma^{32}: \text{RNAP}] &= K_2.[\sigma^{32}_j].[\text{RNAP}_j] \\
[\text{RNAP} : D] &= K_3.[\text{RNAP}_j].[D_j] \\
[\sigma^{32}: \text{DnaK} : \text{FtsH}] &= K_4.[\sigma^{32}: \text{DnaK}].[\text{FtsH}_j] \\
[\sigma^{32}: \text{DnaK}] &= K_5.[\sigma^{32}_j].[\text{DnaK}_j] \\
[\sigma^{32}: \text{DnaK} : \text{protease}] &= K_6.[\sigma^{32}: \text{DnaK}].[\text{protease}_j] \\
[\sigma^{32}: \text{HslVU}] &= K_7.[\sigma^{32}_j].[\text{HslVU}_j] \\
[\text{Punfolded} : \text{DnaK}] &= K_8.[\text{Punfolded}].[\text{DnaK}_j] \\
[\sigma^{32}: \text{RNAP} : \text{ph}] &= K_9.[\sigma^{32}: \text{RNAP}].([\text{ph}_i] - [\sigma^{32}: \text{RNAP} : \text{ph}]) \\
[\sigma^{70}: \text{RNAP} : pg] &= K_{10}.[\sigma^{70}: \text{RNAP}].([pg_i] - [\sigma^{70}: \text{RNAP} : pg]) \\
[\sigma^{70}: \text{RNAP} : D] &= K_{11}.[\sigma^{70}: \text{RNAP}].[D_i] \\
[\sigma^{32}: \text{RNAP} : D] &= K_{12}.[\sigma^{32}: \text{RNAP}].[D_i]
\end{align*}
\]

Mass Balance Equations

\[
\begin{align*}
[\text{RNAP}_i] &= [\text{RNAP}_j] + [\sigma^{70}: \text{RNAP}] + [\sigma^{32}: \text{RNAP}] + [\text{RNAP} : D] + [\sigma^{70}: \text{RNAP} : D] \\
&+ [\sigma^{32}: \text{RNAP} : D] + [\sigma^{70}: \text{RNAP} : pg] + [\sigma^{32}: \text{RNAP} : \text{ph}] \\
[\sigma^{32}_i] &= [\sigma^{32}_j] + [\sigma^{32}: \text{DnaK} : \text{protease}] + [\sigma^{32}: \text{RNAP}] + [\sigma^{32}: \text{RNAP} : D] \\
&+ [\sigma^{32}: \text{DnaK} : \text{FtsH}] + [\sigma^{32}: \text{DnaK}] + [\sigma^{32}: \text{RNAP} : \text{ph}] + [\sigma^{32}: \text{HslVU}] \\
[\text{DnaK}_i] &= [\text{DnaK}_j] + [\sigma^{32}: \text{DnaK} : \text{FtsH}] + [\sigma^{32}: \text{DnaK}] + [\text{Punfolded} : \text{DnaK}] \\
&+ [\sigma^{32}: \text{DnaK} : \text{protease}] \\
[\sigma^{70}_i] &= [\sigma^{70}_j] + [\sigma^{70}: \text{RNAP}] + [\sigma^{70}: \text{RNAP} : D] + [\sigma^{70}: \text{RNAP} : pg] \\
[\text{FtsH}_i] &= [\text{FtsH}_j] + [\sigma^{32}: \text{DnaK} : \text{FtsH}] \\
[\text{HslVU}_i] &= [\text{HslVU}_j] + [\sigma^{32}: \text{HslVU}] \\
[\text{protease}_i] &= [\text{protease}_j] + [\sigma^{32}: \text{DnaK} : \text{protease}] \\
[\text{Protein}_i] &= [\text{Punfolded}] + [\text{Punfolded} : \text{DnaK}] + [\text{Pfolded}]
\end{align*}
\]
Full Model Simulations
FtsH Null Mutant Simulations

**Graphs:**
- **Left Graph:**
  - Title: SIGMA32 TOTAL
  - X-axis: TIME
  - Y-axis: 0 to 900
  - Lines:
    - Green: Mutant
    - Red: Wild Type
  - Annotations:
    - 30° to 42°

- **Right Graph:**
  - Title: DNAK TOTAL
  - X-axis: TIME
  - Y-axis: 0 to 45000
  - Lines:
    - Green: Mutant
    - Red: Wild Type
  - Annotations:
    - 30° to 42°

**Legend:**
- Green: FtsH Null mutant
- Red: Wild Type
Stochastic vs. Deterministic

- Total number of sigma-32 molecules per cell is very small (~30 per cell)
- Number of free sigma-32 molecules per cell is even smaller
- Does it make sense to treat these quantities as concentrations?
- Stochastic models need to be considered
Stochastic Simulation Algorithm

• N species. Number of molecules of the reactants is: $X_1, \ldots, X_N$

• M reaction channels: $R_1, \ldots, R_M$.

• Model is discrete and stochastic

• At any time $t$, the reaction is governed by reaction probability density function $P(\tau, \mu)$
$P_X(\tau, \mu) \ d\tau$ is the probability that, given the state $X=(X_1, \ldots, X_n)$ at time $t$, the next reaction will occur in the infinitesimal time $(t+\tau, t+\tau+d\tau)$, and will be an $R_\mu$ reaction.
Free $\sigma_{32}$

Stochastic Simulation Algorithm

- Sigma32 free (1st sample run)
- Sigma32 free (2nd sample run)
- Sigma32 free (mean of 36 runs)
- Deterministic Sigma32 free
Fast and Frequent reactions

Fast, rare and important reaction

Slow reactions

Multiple time scales

Multiple concentration scales

σ_{free}: 0-1 molecules/cell
σ_{total}: 30-100 molecules/cell
DnaK total: 10000 molecules/cell

Heat

σ70
rpoH
RNAP
σ mRNA

σ
RNAP

DnaK

σ
FtsH

DnaK

Lon

Folded

Unfolded

Aggregate

Degradation
Analysis of the HS System

• What are the advantages of the various control strategies?
  – Disturbance feedforward
  – Sequestration loop
  – Degradation loop
  – Fast synthesis and degradation of sigma-32

• Tools
  – Simulations
  – Sensitivity analysis
  – Robustness analysis
Feedforward Control

Heat

FF sensor

Computer

FB sensor

DnaK

σ mRNA

RNAP

Actuator

DnaK

FtsH

Lon

Plant

folded proteins

unfolded proteins

aggregate

degradation
Wild type
Without feedforward

Total $\sigma_{32}$

Folded Protein
Sequestration Feedback

- Plant
- Computer
- Sensor
- Actuator
- FF sensor
- FB sensor
- RNAP
- DNAK
- σ mRNA
- Heat
- Plant
- tor
- RNAP
- DnaK
- FtsH


Effect of Sequestration on Sensitivity

Sensitivity of DnaK to model parameters

- No feedback loop
- Wild type

30° 42°
% change in transcription rate

Without sequestration

With sequestration

% change in DNAK

% change in transcription rate
Degradation Feedback

Heat

σ mRNA

FF sensor

σ RNAP

σ DNAK

σ ftsH

RNAP

DnaK

FtsH

Plant
Wild type
Constitutive degradation
(Without regulated degradation)

Total $\sigma_{32}$

Folded Protein
Role of the degradation loop in attenuating stochastic fluctuations

Wild type
Constitutive degradation

DNAK
Linearized Heat Shock Models

Regulated Degradation
Heat Shock

\[
\begin{bmatrix}
S_r \\
D_r
\end{bmatrix} = \begin{bmatrix}
-\theta & -(\alpha_1 + \alpha_2) \\
K & -\gamma_1 - \gamma_2
\end{bmatrix}
\begin{bmatrix}
S_r \\
D_r
\end{bmatrix}
\]

Constitutive Degradation
Heat Shock

\[
\begin{bmatrix}
S_r \\
D_r
\end{bmatrix} = \begin{bmatrix}
-\theta & -\alpha_1 \\
K & -\gamma_1
\end{bmatrix}
\begin{bmatrix}
S_r \\
D_r
\end{bmatrix}
\]
**H₂ norm (variance) Computations**

Fact: Consider the state and output equations given by

$$\dot{X} = \begin{bmatrix} -\theta & -\alpha \\ K & -\gamma \end{bmatrix} X + B\eta(t) = AX + B\eta(t)$$

$$Y = CX$$

where $\eta(t)$=Gaussian noise. The $H₂$ norm of the system (variance due to this noise) is given by

$$\|H\|₂^2 = tr(CPC^T)$$

Where $P$ is the solution to the equation:

$$AP + PA^T = -BB^T$$
Comparison of the Two Models

\[
\|H^{cd}\|^2 = \frac{\theta_1^2 + (\alpha_1 + K)K + \theta_1 \gamma_1}{2(\theta_1 + \gamma_1) + (\alpha_1 K + \theta_1 \gamma_1)}
\]

\[
\|H^{rd}\|^2 = \frac{\theta_1^2 + (\alpha_1 + \alpha_2 + K)K + \theta_1 (\gamma_1 + \gamma_2)}{2(\theta_1 + \gamma_1 + \gamma_2) + ((\alpha_1 + \alpha_2)K + \theta_1 (\gamma_1 + \gamma_2))}
\]

\[
\|H^{cd}\|^2 > \|H^{rd}\|^2
\]

The variance due to additive noise is larger in the constitutive case.
SSA Results and Comparison of the Two Modes of Regulation
• Histogram showing the distribution of chaperones after 2400 realizations
• X-axis number of chaperones at time 1000 (min)
• Y-axis frequency of the number of chaperone occurring in the corresponding bin
Wild type
Degradation of free $\sigma_{32}$

Total $\sigma_{32}$

Folded Protein
Wild type
Without “amplifier”

Total $\sigma_{32}$

Folded Protein
σ32 transcription

σ32 Translation

σ32 degradation

Servo

ftsh

dnaK

Feedforward

Feedback

Disturbance

Flux modules

Amplifier
Total $\sigma^{32}$

Free $\sigma^{32}$

DnaK

Wild type
No feedforward
Low $\sigma^{32}$ flux
Constitutive $\sigma^{32}$ degradation
No DnaK interaction

Unfolded Proteins

30°  42°

30°  42°
Model Reduction: General Principles for Biological Systems

- Modularity: Well defined boundaries for subsystems.
- Time Scale Separation: Interplay of fast and slow dynamics.
- Concentration Scale Separation: Regulator molecules present in low copy number and actuator molecules present in high copy numbers.
- Robustness: Insensitivity to parameters that can be neglected in the dynamics of the system.
The Reduced Heat Shock Model

We have obtained a simple 3 state nonlinear model of the heat shock system:

\[
\begin{align*}
\frac{dD_t}{dt} &= K_d \frac{S_t}{1 + \frac{K_d D_t}{1+K_s U_f}} - \alpha_d D_t \\
\frac{dS_t}{dt} &= \eta(T) - \alpha_0 S_t - \alpha_s S_t \frac{K_d D_t}{1+K_s U_f} \\
\frac{dU_f}{dt} &= K(T)[P_t - U_f] - K_{Tot} D_t
\end{align*}
\]

\(D_t\) is the concentration of Chaperones

\(S_t\) is the concentration of \(\sigma^{32}\)

\(U_f\) is the concentration of unfolded proteins
\[
\frac{dS_t}{dt} = \eta(T) - \alpha_0 S_t - \alpha_s K_s S_f F_f D_f
\]

\[
S_f = \frac{S_t}{1 + K_s D_f}
\]

\[
F_f = \alpha D_t
\]

\[
D_f = \frac{D_t}{1 + K_u U_f}
\]

\[
\frac{dD_t}{dt} = K_d S_f - \alpha_d D_t
\]

\[
\frac{dU_f}{dt} = K(T) P_t - [K(T) + K_{fold}] D_f
\]
\[
\frac{dS_t}{dt} = \eta(T) - \alpha_0 S_t - \alpha_s K_s S_f F_f + D_f
\]

\[
S_f = \frac{S_t}{1 + K_s D_f}
\]

\[
F_f = \alpha D_t
\]

\[
D_f = \frac{D_t}{1 + K_u U_f}
\]

\[
\frac{dD_t}{dt} = K_d S_f - \alpha_d D_t
\]

\[
\frac{dU_f}{dt} = K(T) P_t - [K(T) + K_{fold}] D_f
\]
Known Biology vs. Phenotype

A dynamic model is to be viewed as a tool for capturing biological knowledge about known components and their frequently complicated interactions.

Question: Are known components and their interactions consistent with experimental observation?

Two scenarios:

- **Yes.** A dynamic model capturing known players and their interactions can robustly reproduce the data.

- **No.** A proof is generated showing that with the existing components it is impossible to reproduce the data.
Case Study: Origin of Adaptation Phase

A hypothesis…
If induction phase is caused by increased sigma32 translation, then surely the adaptation phase is caused by a **shut off in translation**.

Several experiments to search for hypothesized shut-off mechanism

But…

- Dynamic simulations of the model based on known components always exhibit an adaptation phase. **No need for a shut off mechanism!**
- Feedback through the FtsH loop is responsible for such adaptation
- Translation of sigma32 does *not* shut off at higher temperatures.
  
Missing Players?

Question: Is a model without the FtsH feedback loop consistent with the data?

\[
\frac{dS_t}{dt} = \eta(T) - \alpha_0 S_t - \alpha_s K_s S_f D_f \\
\frac{dD_t}{dt} = K_d S_f - \alpha_d D_t \\
\frac{dU_f}{dt} = K(T)P_t - [K(T) + K_{fold}]D_t \\
D_f = f_1(D_t, U_f, F_f) \\
S_f = f_2(D_t, U_f, F_f)
\]
Model Invalidation

Given:

- Proposed state-space model with constraints on parameters
  \[
  \dot{x} = f(x, p),
  \]
  \[
  P = \{ p \in \mathbb{R}^m : q_{k_1}(p) \leq 0, k_1 = 1, \ldots K_1 \}
  \]

- Measurement data
  \[
  Y_0 = \{ x \in \mathbb{R}^n : q_0(x) \leq 0, r_0(x) = 0 \}
  \]
  \[
  Y_{T_f} = \{ x \in \mathbb{R}^n : q_{T_f}(x) \leq 0, r_{T_f}(x) = 0 \}
  \]
  \[
  D = \{ x \in \mathbb{R}^n : q_{k_1}(x) \leq 0, k_1 = K_1 + 1, \ldots K_2 \}.
  \]

Provide (if possible) a proof that the measurements are inconsistent with the model.
Model Invalidation using SOStools

\[ \dot{x} = f(x, p), \]
\[ P = \{ p \in \mathbb{R}^m : q_{k_1}(p) \leq 0, k_1 = 1, \ldots K_1 \} \]

\[ Y_0 = \{ x \in \mathbb{R}^n : q_0(x) \leq 0, r_0(x) = 0 \} \]
\[ Y_{T_f} = \{ x \in \mathbb{R}^n : q_{T_f}(x) \leq 0, r_{T_f}(x) = 0 \} \]
\[ D = \{ x \in \mathbb{R}^n : q_{k_1}(x) \leq 0, k_1 = K_1 + 1, \ldots K_2 \}. \]

**Theorem 1** Assume that there is a function \( B : \mathbb{R}^{n+m+1} \rightarrow \mathbb{R} \) such that

\[ B(x_{T_f}, p, T_f) - B(x_0, p, 0) > 0, \quad \forall x_{T_f} \in Y_{T_f}, x_0 \in Y_0, p \in P \]
\[ \frac{\partial B}{\partial x}(x, p, t)f(x, p, t) + \frac{\partial B}{\partial t}(x, p, t) \leq 0, \quad \forall t \in [0, T_f], x \in D, p \in P. \]

Then the measurements \( \{Y_0, Y_{T_f}\} \) is inconsistent with the system
\[ \dot{x} = f(x, p), \quad p \in P. \]
A barrier function $B(x,p,t)$ was constructed using SOStools.

A model w/o FtsH cannot possibly produce measured data.
Is the WT Response optimal?

• Observation: A hypothetical heat-shock response system maybe devised to achieve
  – a very small number of unfolded proteins
  – minimal complexity (no feedback necessary)
• E.g. over expressing chaperones & eliminating feedback
• However… chaperone over-expression
  – involves a high metabolic cost
  – is toxic to the cell
• The existing design appears to achieve a tradeoff: acceptable folding, with minimal number of chaperones
• How optimal is the WT design?
• Cost function:

\[ J_\alpha(\theta) = \int_{t_0}^{t_1} [\text{chaperones}]^2 dt + \alpha \int_{t_0}^{t_1} [P_{un}]^2 dt \]

• For a given \( \alpha > 0 \), solve:

\[ \min_{\theta} J_\alpha(\theta) \]

s.t. \( \dot{x} = F(x, y, \theta) \)

\( 0 = G(x, y, \theta) \)

• The optimal solution yields a point in \( \mathbb{R}^2 \)

\[ p^{opt}(\alpha) := \left( \int_{t_0}^{t_1} [\text{chaperones}^{opt}]^2 dt, \int_{t_0}^{t_1} [P_{un}^{opt}]^2 dt \right) \]

• The curve \( \{ p^{opt}(\alpha) : \alpha > 0 \} \) gives the set of Pareto optimal designs
Pareto Optimal Design of the Heat Shock System

\[ \int_{t_0}^{t_1} [P_{un}]^2 \, dt \]

\[ \int_{t_0}^{t_1} [chaperones]^2 \, dt \]

Wild type heat shock

Pareto Optimal curve

Various nonoptimal values of parameters
New Tools

• Robustness
  – Is the system robust to large variations of parameters?
  – Robustness as a measure of plausibility

• Model validation
  – Given:
    • A collection of experimental data
    • A system model with a given structure
  – Does there exist a choice of model parameters that is consistent with the data?
Control Theory in Biological Systems

• Feedback regulation mechanisms are ubiquitous

• Bring out the dynamic nature of biochemical interactions
  – Explain interactions in the context of regulation

• Identify functional biological modules

• Many similarities with engineering systems
• New understanding of regulation in biology
  – Notions such as robustness, adaptation, amplification, isolation, and nonlinearity are required for a deeper understanding of biological function
  – Many similarities with engineering systems
• Are there general design principles?
Acknowledgements

• Carol Gross, UCSF

• Tau-Mu Yi, Caltech