A mathematical model of heat shock response to study the competition between protein folding and aggregation

Sushmita Pal[1] and Rati Sharma[1,*]

[1]*Department of Chemistry,*

Indian Institute of Science Education and Research (IISER) Bhopal

Bhopal Bypass Road, Bhauri, Bhopal 462066 INDIA

Proteins, under conditions of cellular stress, typically tend to unfold and form lethal aggregates leading to neurological diseases like Parkinson's and Alzheimer's. A clear understanding of the conditions that favor dis-aggregation and restore the cell to its healthy state after they have been stressed is therefore important in dealing with these diseases. The heat shock response (HSR) mechanism is a signaling network that deals with these undue protein aggregates and aids in the maintenance of homeostasis within a cell. This framework, on its own, is a mathematically well studied mechanism. However, not much is known about how the various intermediate mis-folded protein states of the aggregation process interact with some of the key components of the HSR pathway such as the Heat Shock Protein (HSP), the Heat Shock transcription Factor (HSF) and the HSP-HSF complex. In this article, using kinetic parameters from the literature, we propose and analyze two mathematical models for HSR that also include explicit reactions for the formation of protein aggregates. Deterministic analysis and stochastic simulations of these models show that the folded proteins and the misfolded aggregates exhibit bistability in a certain region of the parameter space. Further, the models also highlight the role of HSF and the HSF-HSP complex in reducing the time lag of response to stress and in re-folding all the mis-folded proteins back to their native state. These models therefore call attention to the significance of studying related pathways such as the HSR and the protein aggregation and re-folding process in conjunction with each other.

* rati@iiserb.ac.in

I. INTRODUCTION

Survival under stressful conditions is one of the primary aspects of a thriving species. In general, cells face various kinds of stresses such as Heat [18], Oxidative stress [39, 43], Endoplasmic Reticulum stress [11] or stresses induced by chemical agents[10]. Therefore, living organisms have evolved various defense mechanisms to respond to such stresses. Heat stress, in particular, leads to protein unfolding and misfolding, eventually causing protein aggregation in the cell. Such aggregation impairs protein function for normal cellular processes [23, 29]. Response to this therefore requires re-folding these proteins to their native state, which is achieved through a mechanism known as the heat shock stress response (HSR) pathway.

In broad terms, the HSR pathway follows the titration feedback mechanism [1]. Under unstressed conditions, Heat Shock Proteins (HSPs), which are small molecular chaperones, exist in the form of a complex with Heat Shock Factors (HSFs). On accumulation of misfolded/unfolded proteins in the cell, the HSPs are titrated away from the HSF-HSP complex and bind to the hydrophobic residues of the unfolded proteins to fold them back. The free HSFs now dimerize, trimerize, get phosphorylated (activated) and through the process of transcription and translation produce more HSPs. These HSPs can now fold back more of the accumulated protein aggregates.

Modeling of HSR has closely followed experimental evidences that revealed pieces of the mechanism right from the beginning. The modeling frameworks have explained data from heat shock experiments in various species such as bacteria [4, 13, 16, 40], yeast [15, 46], worms [3], plants [22, 36, 37] and mammalian cells [27, 28, 30, 35, 38]. The earliest HSR model that fit the data from heat shock experiments in mammalian cells [27], introduced a mathematical framework of the titration feedback loop. Following this, there were several other models which included varying degrees of details into them, such as, transcriptional regulation [30], translation and activation of HSF [28, 35] and later a more simplified model [38].

Although the heat shock experiments have revealed the HSR mechanism, they say noth-

ing about the different mis-folded states of the proteins themselves. In fact, many proteins are known to undergo spontaneous aggregation. This happens during amyloid formation that begins with a nucleation phase and then goes through a series of intermediate elongation steps and self propagates to spontaneously aggregate [12]. Such protein aggregation has been seen, for example, in β-lactoglobulin and its peptides when they spontaneously aggregate to form amyloid fibrils [9]. Spontaneous aggregation also occurs in PHF6, a key hexapeptide of tau protein, the mechanism for which was shown in a recent molecular dynamics simulation study [21]. In addition to such aggregation, proteins are also known to show hysterisis between unfolding and folding. This kind of bistability (the presence of two stable states of proteins) is exhibited in the guanidine hydrochloride induced unfolded and refolded states of Transthyretin [17] and MTase proteins [2]. Such threshold crossing is therefore quite common in protein folding and aggregation.

The above mentioned mathematical models of the HSR explained the various mechanistic details. However, these did not look into the the folded and aggregated states of the proteins themselves. In recent years, there have been a few simplified models which looked at the interplay between folded and aggregated states [26, 31, 34]. These models represent over-simplified threshold cross-over (or bistability) between folded and aggregated states. The first of these three models only included the HSP counts and ignored every other feature of the HSR framework, while the latter models did not even include that. Therefore, it is evident that more work needs to be carried out to explore the bistable phenomenon in conjunction with HSR, where the interaction network is not just restricted to the folded and aggregated states, but also includes specific entities of the HSR pathway, such as, HSF, HSP and the HSF-HSP complex.

In light of this, we have in this work, explored bistability in the competition between folded and aggregated states while still incorporating features of the HSR framework. We included interaction of proteins not just with HSP but also with the HSF-HSP complex in addition to activation of HSF in response to heat shock. We explore these features through two specific models with increasing levels of detail.

Our main findings from these explorations are that the presence of HSF and HSF-

HSP complex is important for conversion of misfolded proteins to theirnative folded state. Further, there is delayed response to heat shock which manifests as a slowly decreasing aggregated steady state which then crosses the threshold of bistability to become completely folded. This feature only results when there is production of more HSP through HSF activation due to increased stress and not otherwise. We develop the mathematical models through a system of ODEs and then carry out stochastic simulations to explore the probability distributions of the folded and aggregated states of the protein. The methods used are discussed in Section II. Section III gives a detailed account of the specific results obtained via the two models, while Section IV provides discussions and implications regarding the larger context of these results.

II. MATHEMATICAL MODEL AND METHODOLOGY

Model Structure

The basic pathway of the heat shock response is as follows. The cell activates its response mechanism when it receives a stress signal. In the case of heat stress, the response is elucidated as increased production of heat shock proteins (HSPs) [6], which are small molecular chaperones that act to fold back mis-folded proteins. The HSPs usually exist in the form of a complex with the Heat Shock Transcription Factor (HSF). Unfolded proteins titrate away the HSPs from this complex. Increased stress causes more proteins to become unfolded thereby increasing the demand for HSPs. This increased production of HSPs is facilitated through activation of free HSFs. The mis-folded proteins also go through several intermediates, both during the aggregation and the re-folding processes. In this study, we develop models that capture these intermediate states along with the HSR mechanism.

Taking into consideration the folding and aggregation process of proteins along with the HSR pathway, we propose and analyze two possible models for protein aggregation in this study. Both these models incorporate HSF and its interaction with HSP along with the different stages of the folding pathway, something that had not been considered before. In Model I, we assume that the total HSPs and HSFs in the system are conserved. Model II

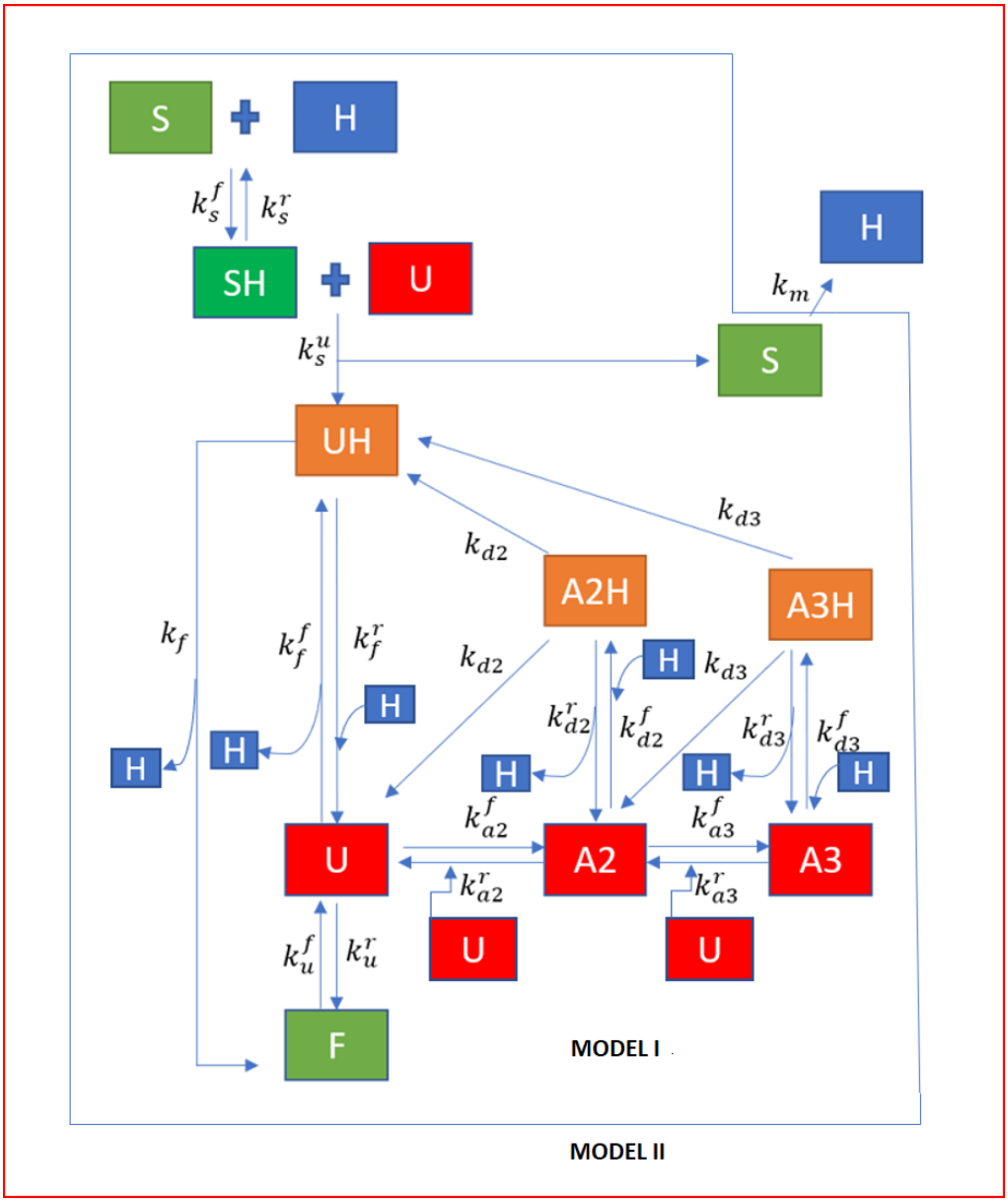

FIG. 1: **Our mathematical model.** The single headed arrows represents irreversible reaction whereas single headed two arrows running opposite to each other represent reversible reactions. U, A2, A3 are unfolded, aggregate with 2 monomers and aggregate with 3 monomers respectively. UH, A2H, A3H are unfolded proteins and aggregates bound to HSP. S, H, F, SH are HSF, HSP, Folded Protein and HSF:HSP complex respectively.

is an extension of Model I where we let the HSPs be dynamically produced. We consider two models because they represent varying degrees of detail in the HSR pathway, as will become clear when we explore the individual models through simulations. These models give a better understanding of the Heat Shock Response and and elucidate how HSPs are

produced when the cell is exposed to Heat Shock. A combined picture of Models I and II is shown in Figure 1 and the reactions specific to individual models are listed in Table I.

The models proposed consider aggregate formation from upto 3 unfolded monomers of protein, taking into consideration that the higher degrees of aggregation show a similar pattern of aggregation. Such a model helps us examine the unique behaviour of two distinct processes i.e, monomer-monomer dimerization and monomer-oligomer dimerization and the effect of their rate constants on bistability. In a real cell, HSPs exist in several different forms such as HSP16, HSP70, HSP90, etc [5, 6]. Similarly, many different kinds of proteins can be folded back by these chaperones when they get mis-folded. However, for the sake of simplicity and as per standard practice, we choose not to differentiate between various kinds of HSPs and proteins in both these models. Therefore, in our models, HSPs, folded proteins (F) and unfolded proteins (U) serve as stand-ins for several possible different molecules of the same class.

Parameter estimation

As listed in Table II, we chose our parameters from literature on mammalian cells. We found that Robinson et al's [33] work on Chaperone activity and protein folding process inside eukaryotic cells provided us with the range of values for several of our parameters. In order to study the catalytic role of chaperone in protein folding, we also assumed self re-folding of unfolded proteins to be several orders of magnitude lower than chaperone assisted refolding [31].

The stress was introduced in the model by specifying the value of the parameter k_u^f, i.e., the rate at which the folded proteins unfold. Previous studies performed on various mammalian cells [19, 27], show the nature of denaturation of proteins as a function of time. The proteins in these experiments were exposed to gradually increasing temperature and the differential scanning calorimetric data was procured for transition or melting temperatures. The curve obtained from the fractional protein denaturation as a function of temperature was approximated to obtain Eq. 1. The parameter k_u^f, according to this equation, increases

	Model I	**Model II**
Reactions	$\mathrm{F} \iff U$ $\mathrm{U} + \mathrm{U} \iff \mathrm{A}_2$ $A_2 + \mathrm{U} \iff \mathrm{A}_3$ $\mathrm{U} + \mathrm{H} \iff U : H$ $A_2 + \mathrm{H} \iff \mathrm{A}_2\mathrm{:H}$ $A_3 + \mathrm{H} \iff \mathrm{A}_3\mathrm{:H}$ $A_2\mathrm{:H} \implies U + U : H$ $A_3\mathrm{:H} \implies \mathrm{A}_2 + \mathrm{U:H}$ $\mathrm{U:H} \implies F + H$ $\mathrm{S:H} \iff S + H$ $\mathrm{S:H} + \mathrm{U} \implies U : H + S$	$\mathrm{F} \iff U$ $\mathrm{U} + \mathrm{U} \iff \mathrm{A}_2$ $A_2 + \mathrm{U} \iff \mathrm{A}_3$ $\mathrm{U} + \mathrm{H} \iff U : H$ $A_2 + \mathrm{H} \iff \mathrm{A}_2\mathrm{:H}$ $A_3 + \mathrm{H} \iff \mathrm{A}_3\mathrm{:H}$ $A_2\mathrm{:H} \implies U + U : H$ $A_3\mathrm{:H} \implies \mathrm{A}_2 + \mathrm{U:H}$ $\mathrm{U:H} \implies F + H$ $\mathrm{S:H} \iff S + H$ $\mathrm{S:H} + \mathrm{U} \implies U : H + S$ $\mathrm{S} \implies H$

TABLE I: The complete list of reactions corresponding to each model.

as a power law with increase in temperature. This equation depicts the rate of denaturation in the temperature range of 37-45 °C.

$$k_u^f = (1 - \frac{0.4}{e^{T-37}}) \times 1.4^{T-37} \times 1.45 \times 10^{-5} \tag{1}$$

In experimental studies of the HSR mechanism, cells are made to undergo hyperthermia, which is defined as the exposure of cells to temperatures above 37°C. In line with these experimental studies, hypothetical heat stress induced in our model was around 40°C and Eq. 1 was used to calculate k_u^f at this temperature. Following the above basic principles, we compiled the complete set of parameters. These are listed along with their respective references in Table II.

Parameter	Value [$(item \times s)^{-1}$ or s^{-1}]	Description	Reference
k_u^f	10^{-4}	Temperature dependent misfolding of folded proteins	Eq. 1
k_u^r	10^{-5}	Self Refolding of misfolded proteins	Robinson et al,1996 [33]
k_{a2}^f	0.02	Rate of aggregation of two U	Zettlmeissel et al,1979 [45]
k_{a2}^r	0.04	Self disaggregation of A_2	Rieger et al,2006 [31]
k_{a3}^f	0.01	Rate of aggregation of A_2 and U	Zettlmeissel at al,1979 [45]
k_{a3}^r	0.01	Self disaggregation of aggregate A_3	Rieger et al,2006 [31]
k_f^f	0.01	Rate of U and H binding	Flynn et al.,1989 [7]
k_f^r	0.1	Rate of disaggregation of U:H	Flynn et al., 1989 [7]; Liberek et al., 1991 [20]
k_{d2}^f	0.1	Rate of entrapment of H with A_2	Robinson et al,1996 [33]
k_{d2}^r	0.1	Rate of disaggregation of A_2:H	Robinson et al,1996 [33]
k_{d3}^f	0.001	Rate of entrapment of H with A_3	Robinson et al,1996[33]
k_{d3}^r	0.1	Rate of disaggregation of A_3:H	Robinson et al,1996 [33]
k_{d2}	3×10^{-3}	Rate constant for degradation of A_2H to U and U:H	Rieger et al,2006 [31]
k_{d3}	3×10^{-3}	Rate constant for degradation of A_3H to A_2 and U:H	Rieger et al,2006 [31]
k_f	10^{-3}	Catalytic refolding of proteins from UH	Rieger et al,2006 [31]
k_s^r	3.56	Rate of dissociation of S:H into S and H	Petre et al, 2011 [28]
k_s^f	9.73	Rate of association of S and H	Petre et al, 2011 [28]
k_s^u	10^{-4}	Rate of S:H binding to U	Mikolaj et al, 2013 [41]; Zuzanna et al, 2009 [47]
k_m	$10^{-4} - 10^{-3}$	Rate of HSP translation	Scheff et al, 2015 [35], Petre et al 2011 [28], Sivery et al 2016 [38]

TABLE II: List of parameters. The values of rate has units of $(item \times s)^{-1}$ for a second order reaction and s^{-1} for a first order reaction.

Deterministic analysis and stochastic simulations

As a first step to defining our models, we wrote down the chemical kinetic rate equations corresponding to the changes in flux of the individual components undergoing the reactions. This system of ordinary differential equations (ODEs) is shown in Equations 2-11. As mentioned in Table II, the kinetic rate parameters were chosen from literature. The descriptions of the various kinetic parameters of the models are also stated in Table II. We then solved this system of equations at steady state and determined whether the folded, unfolded and aggregated proteins exhibited bistability. We further explored this bistability through stochastic simulations.

Exploring the system through simulations is important because reactions taking place inside a living cell are associated with stochasticity due to both population heterogeneity of the various components of the reaction system and the fluctuations inherent in the cellular environment. This inherent stochasticity in the system leads to fluctuations both in the reaction time and in the copy number of individual species undergoing the reactions. Therefore, a need for carrying out stochastic simulations arises in order to account for these fluctuations. The magnitude of these fluctuations (also called noise), has an inverse relationship with system size. This is reflected in the copy number fluctuations in our system.

The system under consideration exhibits bistability. This means that the system fluctuates between two stable states through an unstable state. The switch between the two states occurs when a fluctuation around one stable state pushes the system across the threshold taking it to the other stable state. This switch can be seen in the individual trajectories of the simulations (Figs. 3 through 8).

Therefore, it is of utmost importance to perform stochastic simulations to get a better insight into the bistable behaviour of the system. The stochastic behaviour can be studied by writing down the corresponding rate of change of probabilities of the system being in a certain composition with respect to its components. This equation, known as the Chemical Master Equation (CME), cannot be solved analytically for a system of this level of complexity. Stochastic simulations of such systems using the Gillespie algorithm [8] therefore provides an alternative.

The simulations in this study were run for 1000 replicates to study the noise in the system. The resulting probability distribution of the copy numbers of all the components were plotted to see the effect of chaperones and HSF dynamics in determining the state of the cell. The simulations themselves were performed using the simulation software package, Lattice Microbes [32] which incorporates the Gillespie algorithm. The simulations for both the models were carried out by setting up every reversible reaction as two individual independent reactions. The stoichiometric matrix and dependency matrix were explicitly written to describe the nature of the reactions and the effect on species count after every time step.

III. RESULTS

We have, in this article, presented two models (I and II), which are representations of the protein folding and aggregation pathway with increasing levels of details from the HSR pathway. We analyze the two models to show how incorporating features from the HSR pathway makes for a better model for the interaction of the HSPs with the unfolded proteins.

A. Model I

Model I is represented in Table I and its mathematical equations representing the change in concentrations of the individual species are represented in Equations (2) through (11).

Deterministic analysis

We first carried out a deterministic analysis of Model 1. The conservation of total proteins, total HSP and total HSF are represented mathematically in Equations (2) to (4). The kinetic rate equations of the individual species are shown in Equations (5) through (11).

$$[P]_{tot} = U + F + UH + 2 \times (A_2 + A_2H) + 3 \times (A_3 + A_3H) \tag{2}$$

$$\alpha = UH + A_2H + A3H + H + SH \tag{3}$$

$$\beta = S + SH \tag{4}$$

$$\frac{df}{dt} = k_f \times UH + k_u^r \times U - k_u^f \times F \tag{5}$$

$$\frac{dA_2}{dt} = ka_2^f \times U \times U - ka_2^r \times A_2 + kd_2^r \times A_2H - kd_2^f \times A_2 \times H + ka_3^r \times A_3 - ka_3^f \times U \times A_2 + kd_3 \times A_3H \tag{6}$$

$$\frac{dA_3}{dt} = kd_3^r \times A_3H - kd_3^f \times A_3 \times H + ka_3^f \times U \times A_2 - ka_3^r \times A_3 \tag{7}$$

$$\frac{dUH}{dt} = k_f^f \times U \times H - k_f^r \times UH - k_f \times UH + kd_2 \times A_2H + kd_3 \times A_3H + k_s^u \times SH \times U \tag{8}$$

$$\frac{dA_2H}{dt} = kd_2^f \times A_2 \times H - kd_2^r \times A_2H - kd_2 \times A_2H \tag{9}$$

$$\frac{dA_3H}{dt} = kd_3^f \times A_3 \times H - kd_3^r \times A_3H - kd_3 \times A_3H \tag{10}$$

$$\frac{dSH}{dt} = k_s^f \times S \times H - k_s^r \times SH - k_s^u \times SH \times U \tag{11}$$

We solved the system of ODEs along with the constraints, i.e., Eqs. (2) through (11), under steady state conditions (where the change in copy numbers with respect to time is set to zero) to find regions of parameter space that have 3 or 1 physically relevant solutions for the copy numbers of folded (F), unfolded (U) and aggregated (A_2 and A_3) proteins. The region of space that has 3 steady state solutions shows bistable behavior while that with only one steady state solution shows monostable behavior.

Since the total HSP and HSF are conserved in Model I, the components of the system in the bistable regime can potentially redistribute itself in such a way that there are two stable steady states with an unstable state in the middle. The deterministic analysis was explicitly performed for Model 1. The corresponding parameter scan with the two regions (bistable and monostable) explicitly displayed are plotted in Fig 2. Following this, stochastic

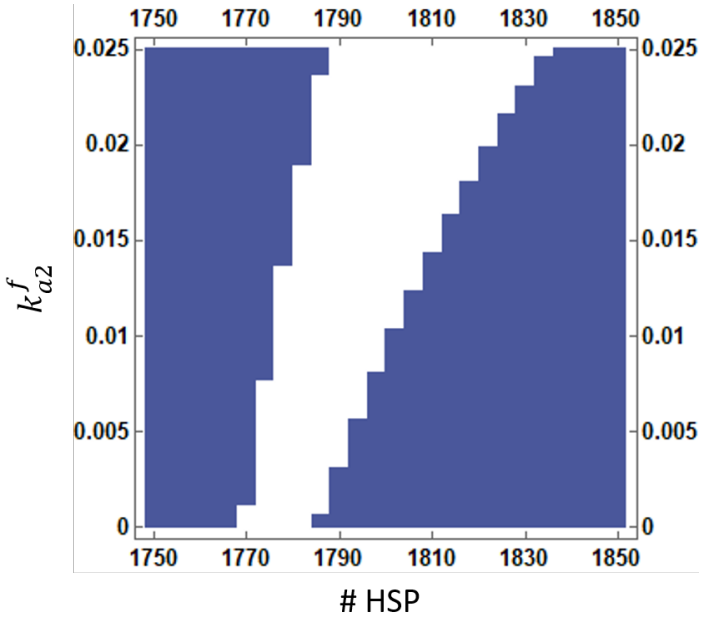

FIG. 2: **Parameter scan to determine the bistable regime.** The white shaded region of the figure shows the bistable regime where the system has of 3 steady states. The system of ODEs were solved with a two parameter scan over a range in parameter space. k_{a2}^f and concentration of HSP were the two parameter chosen for scan. The total protein count, S or SH count is 5000 and 1000 respectively.

simulations were also carried out.

Bistability

We varied the total HSP and the kinetic rate constant for aggregate formation, k_{a2}^f, to carry out the two parameter scan in order to determine the number of physically relevant solutions of the ODEs at steady state. We did this while keeping fixed all the other parameters of the system. The choice of k_{a2}^f as one of the parameters to be scanned over for the bistability analysis is arbitrary and a similar parameter scan could be performed by choosing the other parameters as well and the results for bistability would still hold true. However, as we later show using sensitivity analysis, k_{a2}^f is also one of the sensitive parameters in the models, thereby showing a restrictive range for the bistability regime (as shown in Fig. 2).

The parameter scan confirmed the presence of bistability in our model. As mentioned

earlier, bistablility leads to the threshold phenomenon where the system crosses the barrier and in this case goes to the higher aggregated state which could be lethal for the cell. In the parameter scan, shown in Fig. 2, the white region depicts the bistable region where the system has 3 steady state solutions, where, two states (lowest and highest values) represent the stable steady states and the one in the middle, the unstable steady state.

Stochastic Simulations

Once we had the bistable region identified from Fig. 2, we could choose parameters that would show either a single or a bimodal distribution of folded and aggregate proteins. As described in the previous section, we laid down our model to study the role of HSP, HSF and the HSP-HSF complex in the bistability that the system exhibits. In Model I, HSP and HSF are conserved in the system.

The first set of simulations, as shown in Fig. 3, were performed using total HSP counts = 2050, total proteins = 5000 and HSF = 1000, which are typical values [25]. The simulations resulted in a transient state for short time and eventually went to the higher aggregated state. The simulations were run for 1000 replicates in order to get a smooth distribution function. From Fig. 3(B), we could see that some replicates reach higher aggregated state sooner than others. From the probability distribution in Fig. 3(C), we could infer that the probability for the system to be in lower folded protein concentration is higher than that of higher folded protein concentration.

Later, a second set of simulations was performed to see the role of HSP in the system. We hypothesized that the presence of more HSP would help the system to remain healthy and it would spend more time in folded state and after a definite time would cross the barrier and go to the higher aggregated state. Therefore, to test this hypothesis, we increased the HSP count to 2100 in our simulation of the system. This is shown in Fig 4. The simulations tested positive for our hypothesis. Indeed, increasing HSP in the system helped it to be in lower aggregated state. The system stayed in the lower aggregated state for a long duration of time, which is validated from the probability distribution curve where most of the proteins stayed in folded state.

The simulations of Model I rightly represent the scenario that increasing the amount of HSPs increase the concentration of the folded protein compared to the unfolded protein, whereas, some of the proteins remain in the aggregated or unfolded state even at higher concentrations of HSP. However, in a real cell, the aggregated proteins do become folded under increasing concentrations of HSP. We addressed this drawback of Model I in Model II.

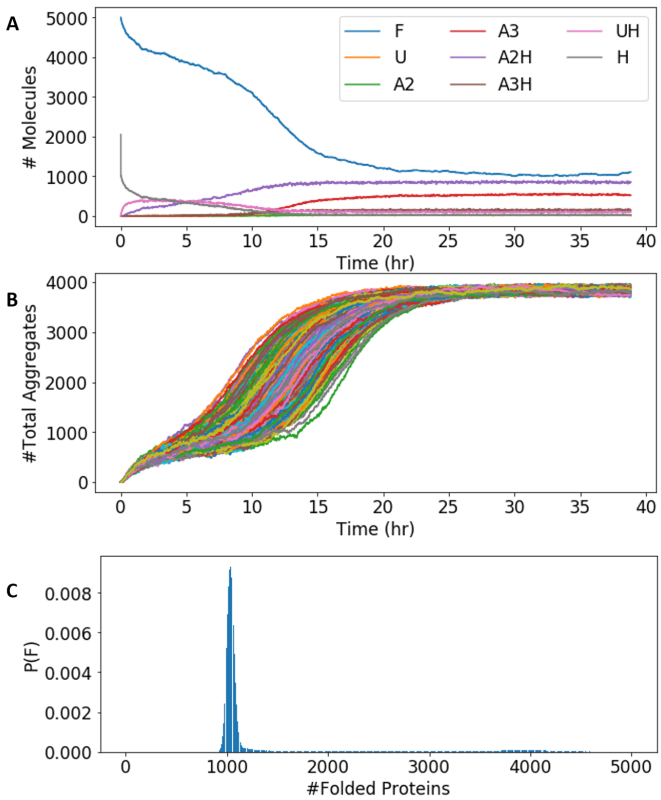

FIG. 3: **Simulations for 1000 replicates of Model I for conserved HSP and HSF.** (A) The presence of bistability for the entities present in the system for one replicate. (B) Depiction of the stochasticity in the system and bistable behaviour for the formation of aggregates. (C) The Probability distribution function for 1000 replicates. Total Protein=5000, Total HSP=2050, Total HSF= 1000

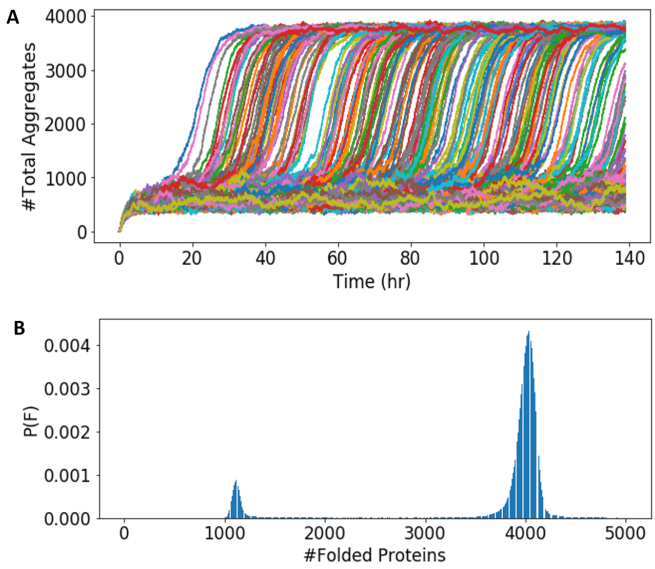

FIG. 4: **Simulations for 1000 replicates of Model I for conserved HSP and HSF.** (A) Depiction of stochasticity in the system and the exhibition of bistability in the formation of aggregates. (B) The Probability distribution function of folded proteins for a collection of 1000 replicates. Total Protein=5000, Total HSP=2100, Total HSF= 1000

B. Model II

The previous model was devised under the assumption that the HSPs are conserved. However, in the real system, the HSPs are produced continuously with increase in stress. This phenomenon was addressed in this model. Therefore, as listed in Table I, Model II has an additional reaction where HSP (H) is created dynamically from HSF (S). This means that the total HSP and the total HSF are no longer conserved which implies that Equations (3) and (4) are no longer valid for Model II. We built on the results of Model I and further explored this dynamical system through stochastic simulations.

Stochastic Simulations

As per the response mechanism, unfolded proteins tend to aggregate with an increase in the cell's exposure to heat stress. The HSPs are chaperones that help fold back these unfolded proteins. The cells therefore, translate HSP as a response to their fight against such a stress. Fig. 5 shows the threshold phenomenon inside the cell under stress. Under stress,

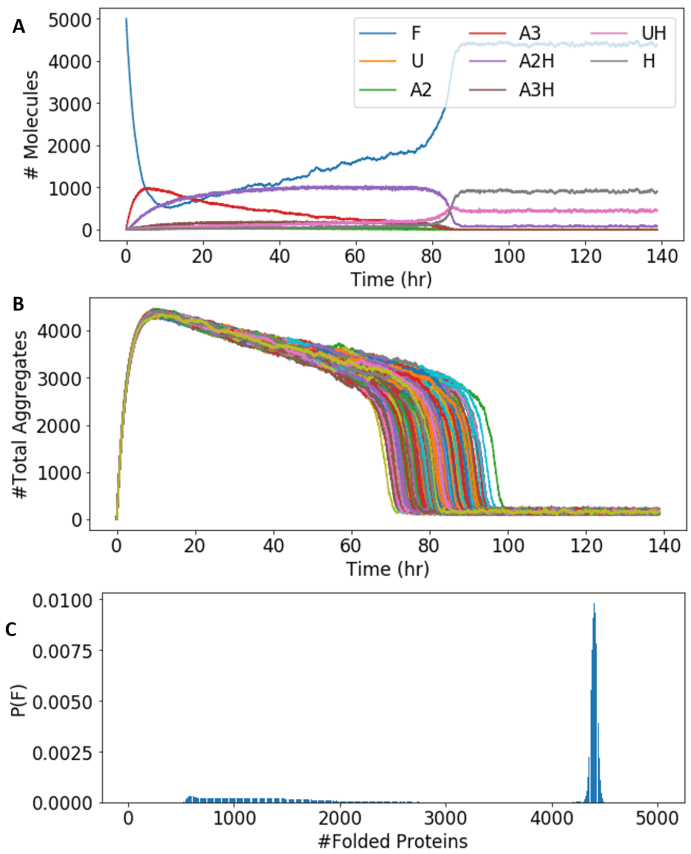

FIG. 5: **Simulations for 1000 replicates of the Heat Shock Response Model (Model II).** (A)The presence of bistability for the entities present in the system for one replicate (B) Depiction of stochasticity in the system and the exhibition of bistability in the dissolution of aggregates. (C) The Probability distribution function for 1000 replicates. Total Protein=5000, Total initial basal HSP=100,Total initial HSF= 1500

there is rapid misfolding and aggregate formation, but the presence of basal level HSP as an initial response of the cell tends to gradually decrease the aggregates and after a certain time due to the Heat Shock Response, the system crosses the threshold barrier and goes to the state of lower aggregates. The threshold barrier is crossed as the system reaches a certain level of HSP concentration. In this model, the HSPs are produced from the HSFs. In the real scenario, it is this HSF in its activated form that acts as transcription factors and translates HSP. The initial HSFs in the system are consumed to produce more HSP which helps in protein folding. Due to the gradual decrease in aggregate concentration, there is a wider distribution of folded proteins in the lower state i.e, around 1000 folded proteins (Fig. 5(C)).

Description	Sensitivity Coefficient	Value
Temperature induced mis-folding of folded proteins	$\frac{\partial \ln f}{\partial \ln k_u^f}$	-1
Forward aggregation decreases the complex U:H	$\frac{\partial \ln uh}{\partial \ln k_{a3}^f}$	-0.550
Smaller aggregates tend to decrease as chaperones bind to higher aggregates	$\frac{\partial \ln a_2}{\partial \ln k_{d3}^f}$	-0.557
Folded proteins increases as chaperones bind to the smaller aggregates	$\frac{\partial \ln f}{\partial \ln k_{d2}^f}$	0.506
Chaperones assisted refolding enhances its binding to aggregates	$\frac{\partial \ln a_{2h}}{\partial \ln k_f^f}$	0.769
Folded proteins increases as S:H binds to unfolded proteins	$\frac{\partial \ln f}{\partial \ln k_s^u}$	0.158
Folded proteins increases as the rate of S and H binding increases	$\frac{\partial \ln f}{\partial \ln k_s^f}$	1.18
U:H complex increases as the rate of S and H binding increases	$\frac{\partial \ln uh}{\partial \ln k_s^f}$	1.18
S:H complex increases as the rate of S and H binding increases	$\frac{\partial \ln sh}{\partial \ln k_s^f}$	7.46

TABLE III: Sensitivity coefficients describing the role of parameters in the counts for various species. The listed values are the ones that significantly alters the processes inside the cell and are most sensitive.

In the simulations depicted in Fig. 5, we assumed that there are no HSP-HSF complex at the beginning of the simulation. But, the system could exist in the form where all the HSPs and HSFs are either segregated or in the form of a complex. Therefore, we next carried out simulations where the system in its initial state had all the HSPs in the form of the HSP-HSF complex. Since it had higher HSP concentration initially, the system immediately went to the lower aggregate state as shown in Fig 6(A). However, it is to be noted that eventually simulations depicted in both Figs. 5 and 6 had nearly equal amounts of HSP (counts 1000) in the final state. Further, Fig. 6(B) shows the probability distribution for the folded proteins in the system. The probability of finding approximately 4500 proteins in the folded state is higher, thereby depicting the healthy condition of the cell.

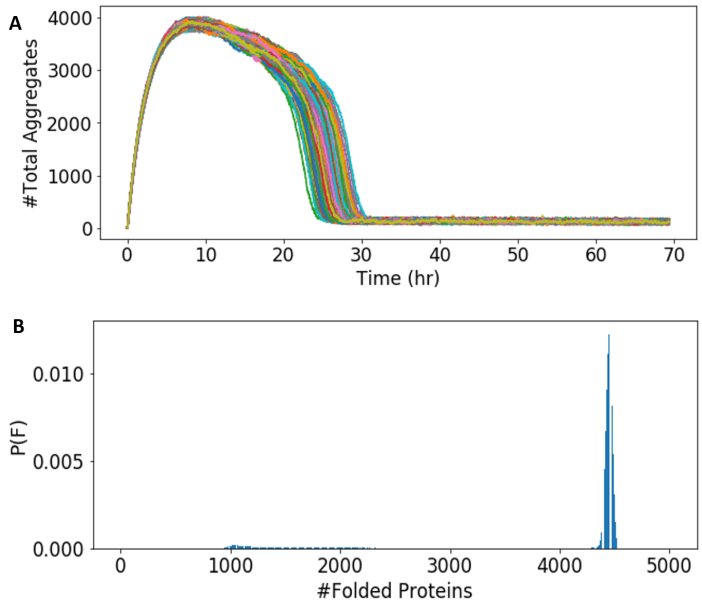

FIG. 6: **Simulations of Model II with HSF in the form of HSF-HSP complex initially.** (A)1000 replicates for the system where all HSF and HSP are in the bound state.(B) Probability distribution curve for Total Protein count=5000, Free HSF=0, Initial HSF:HSP=1500, Initial Basal level HSP=100

We next studied the role of the basal level of HSP in the system. The amount of the basal level of HSP present decides the response time for the cell. To study this, we simulated our system with (HSP count=100) and without basal level HSP, shown in Figs. 7(A) and 7(B) respectively. It was evident that the cell tried to be in slowly decreasing aggregated steady

state for longer when we had no HSP initially. This indicated a delayed response in this scenario.

C. Sensitivity Analysis

To find the relationship between the species and the parameters used in our model, we estimated the sensitivity coefficients for every species and parameter [42]. Carrying out sensitivity analysis gives a quantitative estimate to questions such as which parameters are of utmost importance and which ones do not affect the model. It also gives a quantitative measure of how changes in different parameters of the model can potentially alter the processes inside the cell. Since we were only interested in studying the steady state kinetics of the bistable system, we calculated the sensitivity coefficient using Eq. 12.

$$\frac{\partial c}{\partial k_j} = -J^{-1}F_j \tag{12}$$

where J is the Jacobian Matrix and F_j is the jth column of the matrix $F = \frac{\partial f_i}{\partial k_j}$. Here, f_i denotes the ith function for the corresponding steady state differential equation with $1 \leq i \leq 8$, c denotes the concentration of the species and k_j denotes the jth parameter with $1 \leq j \leq 18$.

The stationary sensitivity coefficient represents the change of stationary species concentration as a result of a differential change in parameters. After close introspection of all the sensitivity coefficients, the least sensitive parameter was k^r_{a2} whose sensitivity coefficient value was way less than 0.1 for all the species in our system. To test this, we changed the value of k^r_{a2} with an increment of 10% and tested the steady state values. Since, this was the least sensitive parameter, the values of none of the species changed significantly. The sensitivity coefficient for self refolding turned out to be negligible in our analysis. This further supported our assumption that the rate constants (k^r_u) for self folding of mis-folded proteins were significantly smaller than the catalytic refolding. For e.g, when we compared the sensitivity coefficients for mis-folding and self refolding, we observed that the parameter k^f_u was more sensitive than k^r_u. The 10 fold increase in the parameter k^f_u changed the protein counts of folded proteins from 1800 to 150 whereas, a 10 fold increase in k^r_u changed the

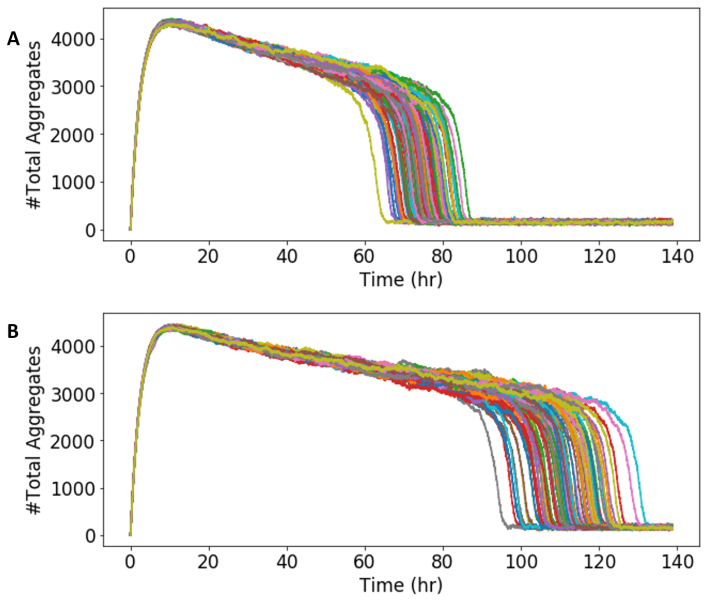

FIG. 7: **Significance of basal HSP.** Simulations for 100 replicates (A) with basal HSP (counts = 100) and (B) without basal HSP (counts = 0) in the system. In both these simulations, total Protein=5000 and Initial HSF=1500.

protein counts from 1800 to 1900 only. Further, the parameter k_u^f is temperature dependent (Eq. 1), thereby experiencing the direct consequence of heat stress.

The proper cellular functioning can only be performed by the cells that contain folded proteins. All the response that is mounted in the system drives the reactions towards folded state of the protein. Also, we assume that initially all the proteins are in their healthy state i.e, the folded state. Therefore, the folded proteins tend to be affected the most during the Heat Shock Response where almost all the parameters tend to affect its existence.

The parameter k_{a2}^f is a sensitive parameter which means that the aggregation process is indeed an important step in our model. We also performed our parameter scan taking k_{a2}^f as one of the parameters to get our bistable regime. The other parameters which showed a similar trend in terms of sensitivity were k_{a3}^f, k_{a3}^r, k_{d2}^f and k_{d2}^r.

In our simulations, we also studied the role of HSF and the HSF-HSP complex. The rate constants for these reactions directly affect the U:H complex and Folded proteins. Therefore,

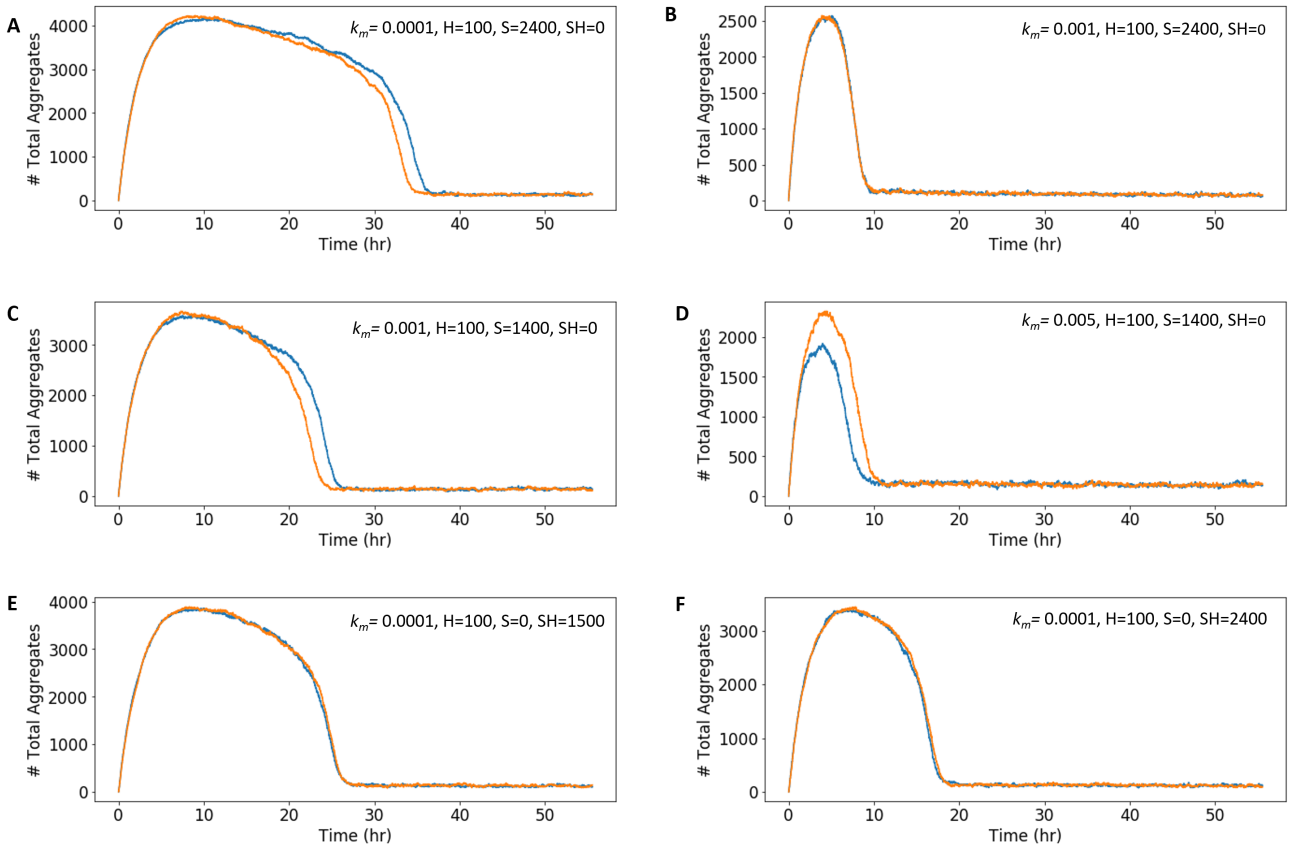

FIG. 8: **Role of HSF in Model II.** Each figure depicts the role of molecular counts of HSPs and HSFs and the rate of production of HSP on the response time. The stochastic simulations were carried out for 2 replicates. The value of k_m is in the units s^{-1}. We compare Fig. (A) with (B), Fig. (C) with (D) and Fig. (E) with (F).

the reactions involving these components are significant. The kinetic rate constant, k_s^f, in fact, is a very sensitive parameter and corresponds to the production of the complex S:H which in turn produces UH and F. Therefore, any small increment in this parameter could lead to more folded proteins in the system. The S:H complex when bound to unfolded protein activates the response mechanism leading to a stable state consisting of higher concentration of folded proteins.

The complete set of sensitivity coefficients are listed in Table S1 in the supplementary information.

DISCUSSION

We have, in this work, developed and carried out Monte Carlo simulations of two specific model systems that tie the HSR pathway with the various intermediate states of the protein folding and aggregation mechanism. In coming up with these models, we have tapped into two basic assumptions. One is that the total protein copy number is conserved due to the associated energy cost in increasing gene expression [44]. Also, proteins lost as irreversible aggregates due to very high stresses are replenished by translation which ensures protein homeostasis [24]. We have incorporated this phenomenon in our models and simulations in the form of Eq. 1. The second assumption is that chaperone assisted refolding kinetics is much faster than free refolding. This comes from many well studied systems and models which show that chaperones inhibit aggregation by binding to the exposed hydrophobic amino acid residues. Subsequently, these chaperone bound proteins go through a series of folding intermediates which have lowered energy barriers to get to the native state [14] .

Following these features, we studied two specific models. Model I simply incorporated the HSF-HSP interaction and the related reactions while keeping the total HSP constant. Model II, on the other hand, in addition to the reactions of Model I also incorporated the fact that HSP is produced continuously with corresponding increase in stress. Model II is therefore closer to reality and this is reflected in the dynamical changes in the copy number of the folded and aggregated states of the protein. In the simulations of Model I (Figs. 3 and 4), due to reaction kinetic fluctuations, some of the proteins switched to the high aggregated state early. And since the total HSP was held constant, these high aggregated states once reached were long lived and highly stable. In Model II as well, stress did cause the folded proteins to become aggregated. But because of dynamical HSP concentration, as shown in the trajectories of Figs. 5A and 5C, all the proteins in the high aggregated state eventually re-folded. This also came about via the spontaneous transition characteristic of a bistable system, but as can be seen from Fig. 5C, there was a clear slowly decreasing aggregated state which was missing in Model I. Both, Models I and II, give a bimodal distribution as shown in Figs. 4B and 5C respectively. However, because of the slowly decreasing aggregated state, Model II gives a much broader distribution for the low folded (highly aggregated) state. The main advantage for the system to be explained by the latter

model is that almost all the unfolded and aggregated proteins fold back to their native state. This is only possible because of the well-studied role of HSF in translating new HSPs [1].

The concentration of HSPs in the system is what determines the fate of the cell. Therefore, with a varying HSP concentration, we could observe a change in the response time of the cell. We define response time as the time at which the cell toggles between the two steady states in the bistable regime. For Model I, we calculated the average time for 1000 replicates to switch from one state to another. With HSP count=2050, it took about 10 hours to respond and switch into a higher aggregated state whereas, an increase in the HSP count by 50, doubled the response time for folded to aggregated state. Hence, these response times were consistent to our theory i.e, the HSP concentration affects the amount of time the cell stays in one particular state.

Finally, as shown in Figs. 6 and 7, we also extended the features of Model II to study the role of free versus HSP bound HSF and the role of basal HSP respectively. It is clear that initial high concentration of HSP either in the form of the HSF-HSP complex (Fig. 6) or in the form of free basal HSP (Fig. 7) leads to faster refolding dynamics. This is because the aggregated proteins then do not have to wait as long for the synthesis of new HSPs. Therefore, the re-folding process can start earlier. However, as can be seen, both from Figs. 6 and 7, there is some time lag in the response to be established and the aggregated state of the proteins are stable for a long time. The response time can be reduced by changing the number of HSFs, HSPs and the kinetic rate for the production of HSP from HSF. This could be observed in the simulation studies shown in Fig. 8. Initially, as per the parameter range listed in Table II, we changed the rate constant, k_m, for the production of HSPs. Since, k_m is an important parameter for the cell to mount a response, its value [28, 35, 38] plays a key role in setting the response time. The wide range of k_m could be a reflection of the different forms of HSPs that are part of the HSR mechanism. We therefore increased the value of k_m by an order of magnitude and observed a significant change in the response time between the simulation trajectories shown in Figs. 8(A) and 8(B). On further increasing the value of k_m from $0.001s^{-1}$ to $0.005s^{-1}$, we could clearly see a decrease in the response time by several hours, as shown in Figs. 8(C) and 8(D), even though the number of HFs (S) in the system is reduced to 1400. As mentioned before, the state of the cell can also change depending

on the nature of molecules present inside it. For example, HSPs and HSFs could either be present in the free form or as a complex. We simulated a case where all the HSFs are in the form of the HSF:HSP complex. By changing the amount of HSF:HSP complex, there is a change in the HSP molecules which affects the response time inside the cell. The presence of 900 (1500 to 2400) more HSPs inside the cell, even at a low $k_m = 0.0001$, reduced the response time by several hours. This comparison is shown in Figs. 8(E) and 8(F). Such an effect of increased HSF that we predict here can be tested experimentally by overexpressing HSF in the cell. Further, this kind of time lag between stress and response is a general phenomenon observed in several model organisms. For example, in yeast, delayed proteomic response is observed under severe heat stress leading to delayed HSR [24]. Similar delayed response was also observed in *C. elegans* under 34 deg Celsius high heat stress for 1 hour [3].

In conclusion, the study of the two models revealed novel features regarding the specific roles of HSF, the HSF-HSP complex and the synthesis of HSP under stress in tipping the balance between folded and aggregated states. This would not have been possible without connecting the HSR mechanism to the protein aggregation framework. Therefore, future exploration of stress response mechanisms should be directed towards connecting related pathways. This might make the system more complicated, adding several more reactions and parameters, but will eventually give better insights into the larger inter-connectivity of signaling pathways.

ACKNOWLEDGMENTS

Sushmita Pal is supported by the National Fellowship INSPIRE (Innovation in Science Pursuit of Inspired Research) funded by the Department of Science and Technology (DST), Government of India. This work has also been funded by the Science and Engineering Research Board (SERB) Startup Research Grant (Ref. No. SRG/2019/000726) awarded by

DST, India.

---

[1] Julius Anckar and Lea Sistonen. Regulation of HSF1 Function in the Heat Stress Response: Implications in Aging and Disease. *Annual Review of Biochemistry*, 80:1089–1115, 2011.

[2] Dominique T. Capraro and Patricia A. Jennings. Untangling the Influence of a Protein Knot on Folding. *Biophysical Journal*, 110(5):1044–1051, mar 2016.

[3] Erin K Dahlstrom and Erel Levine. Dynamics of Heat Shock Detection and Response in the Intestine of Caenorhabditis elegans. *bioRxiv*, 2019.

[4] H El-Samad, H Kurata, J C Doyle, C A Gross, and M Khammash. Surviving heat shock: Control strategies for robustness and performance. *Proceedings of the National Academy of Sciences of the United States of America*, 102(8):2736–2741, 2005.

[5] Martin E Feder and Gretchen E Hofmann. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annual review of physiology*, 61(1):243–282, 1999.

[6] Andrija Finka, Vishal Sood, Manfredo Quadroni, Paolo De Los Rios, and Pierre Goloubinoff. Quantitative proteomics of heat-treated human cells show an across-the-board mild depletion of housekeeping proteins to massively accumulate few HSPs. *Cell Stress and Chaperones*, 20(4):605–620, jul 2015.

[7] Gregory C Flynn, Thomas G Chappell, and James E Rothman. Peptide Binding and Release by Proteins Implicated as Catalysts of Protein Assembly. 45593(June), 1989.

[8] Daniel T Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. 81(25):2340–2361, 1977.

[9] Daizo Hamada, Toshiki Tanaka, Gian Gaetano Tartaglia, Amol Pawar, Michele Vendruscolo, Mei Kawamura, Atsuo Tamura, Naoki Tanaka, and Christopher M. Dobson. Competition between Folding, Native-State Dimerisation and Amyloid Aggregation in β-Lactoglobulin. *Journal of Molecular Biology*, 386(3):878–890, feb 2009.

[10] Hiroyuki Hamada, Tsutomu Arakawa, and Kentaro Shiraki. Effect of additives on protein aggregation. *Current pharmaceutical biotechnology*, 10(4):400–407, 2009.

[11] Norfadilah Hamdan, Paraskevi Kritsiligkou, and Chris M Grant. ER stress causes widespread protein aggregation and prion formation. 2017.

[12] Mathias Jucker and Lary C Walker. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature*, 501:45–51, 2013.

[13] Hye Won Kang. A multiscale approximation in a heat shock response model of E. coli. *BMC Systems Biology*, 6(1):143, nov 2012.

[14] Yujin E. Kim, Mark S. Hipp, Andreas Bracher, Manajit Hayer-Hartl, and F. Ulrich Hartl. Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annual Review of Biochemistry*, 82(1):323–355, jun 2013.

[15] Joanna Krakowiak, Xu Zheng, Nikit Patel, Zoë A Feder, Jayamani Anandhakumar, Kendra Valerius, David S Gross, Ahmad S Khalil, and David Pincus. Hsf1 and Hsp70 constitute a two-component feedback loop that regulates the yeast heat shock response. *eLife*, 7:e31668, feb 2018.

[16] Hiroyuki Kurata, Hana El-Samad, Rei Iwasaki, Hisao Ohtake, John C Doyle, Irina Grigorova, Carol A Gross, and Mustafa Khammash. Module-Based Analysis of Robustness Tradeoffs in the Heat Shock Response System. *PLoS computational biology*, 2(7):e59, 2006.

[17] Zhihong Lai, Jennifer McCulloch, Hilal A. Lashuel, and Jeffery W. Kelly. Guanidine hydrochloride-induced denaturation and refolding of transthyretin exhibits a marked hysteresis: Equilibria with high kinetic barriers. *Biochemistry*, 36(33):10230–10239, aug 1997.

[18] James R Lepock, Harold E Frey, and Kenneth P Ritchie. Protein Denaturation in Intact Hepatocytes and Isolated Cellular Organelles During Heat Shock. 122(6):1267–1276, 1993.

[19] James R Lepock, Harold E Frey, A Michael Rodahl, and Jack Kruuv. Thermal Analysis of CHL V79 Cells Using Differential Scanning Calorimetry : Implications for Hyperthermic Cell Killing and the Heat Shock Response. 24:14–24, 1988.

[20] Krzysztof Liberek, Jaroslaw Marszalek, Debbie Angt, and Costa Georgopoulostt. Escherichia coli DnaJ and GrpE heat shock proteins jointly stimulate ATPase activity of DnaK. 88(April):2874–2878, 1991.

[21] Hongli Liu, Haiyang Zhong, Xuewei Liu, Shuangyan Zhou, Shuoyan Tan, Huanxiang Liu, and Xiaojun Yao. Disclosing the Mechanism of Spontaneous Aggregation and Template-Induced Misfolding of the Key Hexapeptide (PHF6) of Tau Protein Based on Molecular Dynamics Simulation. *ACS Chemical Neuroscience*, 10(12):4810–4823, dec 2019.

[22] Stefano Magni, Antonella Succurro, Alexander Skupin, and Oliver Ebenhöh. Data-driven dynamical model indicates that the heat shock response in chlamydomonas reinhardtii is tailored

to handle natural temperature variation. *Journal of the Royal Society Interface*, 15:20170965, 2018.

[23] Richard I. Morimoto. The heat shock response: Systems biology of proteotoxic stress in aging and disease. *Cold Spring Harbor Symposia on Quantitative Biology*, 76:91–99, jan 2011.

[24] Moritz Muhlhofer, Evi Berchtold, Chris G Stratil, Martin Haslbeck, Ralf Zimmer, and Johannes Buchner. The Heat Shock Response in Yeast Maintains Protein Homeostasis by Chaperoning and Replenishing Proteins. *Cell Reports*, 29:4593–4607, 2019.

[25] Nagarjuna Nagaraj, Jacek R Wisniewski, Tamar Geiger, Juergen Cox, Martin Kircher, Janet Kelso, Svante Pääbo, and Matthias Mann. Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular Systems Biology*, 7:548, 2011.

[26] Sebastian Pechmann and Michele Vendruscolo. Derivation of a solubility condition for proteins from an analysis of the competition between folding and aggregation. *Molecular BioSystems*, 6(12):2490–2497, 2010.

[27] A Peper, C A Grimbergent, J A E Spaant, J E M Sourens, and R V A N Wijks. A mathematical model of the hsp70 regulation in the cell. 14(1):97–124, 1998.

[28] Ion Petre, Andrzej Mizera, Claire L Hyder, and Annika Meinander. A simple mass-action model for the eukaryotic heat shock response and its mathematical validation. pages 595–612, 2011.

[29] Klaus Richter, Martin Haslbeck, and Johannes Buchner. The Heat Shock Response: Life on the Verge of Death. *Molecular Cell*, 40:253–266, 2010.

[30] Theodore R Rieger, Richard I Morimoto, and Vassily Hatzimanikatis. Mathematical Modeling of the Eukaryotic Heat-Shock Response : Dynamics of the hsp70 Promoter. *Biophysical Journal*, 88(3):1646–1658, 2005.

[31] Theodore R Rieger, Richard I Morimoto, and Vassily Hatzimanikatis. Bistability Explains Threshold Phenomena in Protein Aggregation both In Vitro and In Vivo. *Biophysical Journal*, 90(3):886–895, feb 2006.

[32] Elijah Roberts, John E Stone, and Zaida Luthey-schulten. Lattice Microbes : High-Performance Stochastic Simulation Method for the Reaction-Diffusion Master Equation. pages 1–11, 2012.

[33] Anne Skaja Robinson and Douglas A Lauffenburger. Model for ER Chaperone Dynamics and Secretory Protein Interactions. 42(5):1443–1453, 1996.

[34] Conner I Sandefur and Santiago Schnell. A Model of Threshold Behavior Reveals Rescue Mechanisms of Bystander Proteins in Conformational Diseases. *Biophysical Journal*, 100:1864–1873, 2011.

[35] Jeremy D Scheff, Jonathan D Stallings, Jaques Reifman, and Vineet Rakesh. Article Mathematical Modeling of the Heat-Shock Response in HeLa Cells. *Biophysj*, 109(2):182–193, 2015.

[36] Stefan Schmollinger, Miriam Schulz-Raffelt, Daniela Strenkert, Daniel Veyel, Olivier Vallon, and Michael Schroda. Dissecting the Heat Stress Response in Chlamydomonas by Pharmaceutical and RNAi Approaches Reveals Conserved and Novel Aspects. *Molecular Plant*, 6:1795–1813, 2013.

[37] Michael Schroda, Dagmar Blo, È Cker, and Christoph F Beck. The HSP70A promoter as a tool for the improved expression of transgenes in Chlamydomonas. *The Plant Journal*, 21(2):121–131, 2000.

[38] Aude Sivéry, Emmanuel Courtade, and Quentin Thommen. A minimal titration model of the mammalian dynamical heat shock response. *Physical biology*, 13(6):066008, 2016.

[39] Thomas C Squier. Oxidative stress and protein aggregation during biological aging. 36:1539–1550, 2001.

[40] R. Srivastava, M. S. Peterson, and W. E. Bentley. Stochastic kinetic analysis of theEscherichia coli stress circuit using sigma 32-targeted antisense. *Biotechnology and Bioengineering*, 75(1):120–129, oct 2001.

[41] Zuzanna Szymańska and Maciej Zylicz. Mathematical modeling of heat shock protein synthesis in response to temperature change. *Journal of theoretical biology*, 259(3):562–569, 2009.

[42] Tamas Turanyi. Sensitivity analysis of complex kinetic systems. Tools and applications. *Journal of Mathematical Chemistry*, 5:203–248, 1990.

[43] Luiz R C Vasconcellos, Fabianno F Dutra, Mariana S Siqueira, Heitor A Paula-neto, Jennifer Dahan, and Ellen Kiarely. Protein aggregation as a cellular response to oxidative stress induced by heme and iron. 2016.

[44] Andreas Wagner. Energy costs constrain the evolution of gene expression. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 308B(3):322–324, may 2007.

[45] Gerd Zettlmeissl, Rainer Rudolph, and Rainer Jaenicke. Lactic Dehydrogenase. Noncovalent Aggregation vs. Reactivation. 1. Physical Properties and Kinetics. 18(25):5567–5571, 1979.

[46] Xu Zheng, Joanna Krakowiak, Nikit Patel, Ali Beyzavi, Jideofor Ezike, Ahmad S. Khalil, and David Pincus. Dynamic control of Hsf1 during heat shock by a chaperone switch and phosphorylation. *eLife*, 5:1–26, 2016.

[47] Maciej Zylicz. Mathematical modeling of heat shock protein synthesis in response to temperature change. 259:562–569, 2009.