# Supplementary Material of: Data-driven dynamical model indicates that the heat shock response in *Chlamydomonas reinhardtii* is tailored to handle natural temperature variation

Stefano Magni[1,2], Antonella Succurro[3,5], Alexander Skupin[2,4], and Oliver Ebenhöh[1,5,†]

[1]Institute of Quantitative and Theoretical Biology, Heinrich Heine University, Düsseldorf, Germany
[2]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
[3]Botanical Institute, University of Cologne, Cologne, Germany
[4]University California San Diego, La Jolla, USA
[5]Cluster of Excellence on Plant Sciences (CEPLAS), Düsseldorf, Germany
[†]Author for correspondence. E-mail: oliver.ebenhoeh@hhu.de

Journal of the Royal Society Interface, 2018

## Contents

# A Supplementary Material: mathematical description of the model

Our mathematical model is graphically represented by the signalling network schematically depicted in Fig. 1 A.

The model is described by twelve dynamic variables (see Table 1 of the Supplementary Material), each representing the concentration of the corresponding component of the network. In some of the panels of the figures (e.g. Fig. 1 B, the concentrations are expressed as a function of the total amount of the corresponding species (for the quantities where this is conserved), while in some other panels (e.g. Fig. 1 D) the concentrations are expressed in arbitrary units (a.u.), which are used to normalize each panel to a reference value. These reference values are the same across different figures, and are listed in table 3 of the Supplementary Material. They are necessary because, since no data for the absolute values of the concentration of any of the species involved are available, the model variable could be fit only to relative data.

The temporal dynamics of the variables are governed by a set of ordinary differential equations (ODEs), reported in Table 2 of the Supplementary Material. The equations we employ depend on rate expressions (see Table 5 of the Supplementary Material) describing the various regulatory processes ($\omega$), activation and deactivation steps ($\nu$), synthesis rates ($\pi$) and degradation rates ($\eta$). For the majority of these rate laws we assume mass action kinetics. We also employ some additional non-linear terms having a behaviour of the type of Michaelis-Menten kinetics or Hill kinetics. The term following a Hill kinetics is the one describing the regulatory process $\omega_{PS}$ involved in the reaction by which the denatured proteins induce the activation (phosphorylation) of the SK. This allows to have a response following the sigmoidal shape which captures a threshold effect in the activation of this reaction. The action of the phosphorylated SK$^*$, the enzyme phosphorylating HSF, is described by a Michaelis-Menten behaviour, typical of enzymatic kinetics. The effect of the temperature, which denatures (unfolds or misfolds) proteins, is described by means of the Arrhenius law, with an activation energy in the wide range reported in the literature for the activation energies of protein denaturation due to thermal stress (as discussed in the literature [23, 24]). The values of the parameters (rate constants) employed in the model are provided in Table 4 and the parameterization of the model is described in detail in Section B of the Supplementary Material.

| **Variable** | **Representing concentration of** |
|---|---|
| $[P]$ | Proteins |
| $[P^{\#}]$ | Degenerated proteins |
| $[SK]$ | Stress kinases (SK) |
| $[SK^*]$ | Phosphorylated SK |
| $[HSF]$ | Heat shock factor (HSF) |
| $[HSF^*]$ | Phosphorylated HSF |
| $[G]$ | Free gene |
| $[HSF^*G]$ | HSF bound to gene, active |
| $[HSFG]$ | HSF bound to gene, inactive |
| $[mR_F]$ | mRNA for the HSF |
| $[mR_{HP}]$ | mRNA for the HSP |
| $[HP]$ | Heat shock protein (HSP) |

Table 1: **Variables and names used in the model.**

| **ODEs** |
|---|
| $\frac{d[P^{\#}]}{dt} = -\nu_P + \nu'_P$ |
| $\frac{d[SK^*]}{dt} = -\nu_S + \nu'_S$ |
| $\frac{d[HSF]}{dt} = \nu_F + \pi_F + \nu'_{FG} - \nu_{FG} - \nu'_F - \eta_F$ |
| $\frac{d[HSF^*]}{dt} = -\nu_F + \nu'_F + \nu'_{F^*G} - \nu_{F^*G}$ |
| $\frac{d[HSF^*G]}{dt} = \nu_{F^*G} + \nu_{F^*} - \nu'_{F^*} - \nu'_{F^*G}$ |
| $\frac{d[HSFG]}{dt} = \nu'_{F^*} + \nu_{FG} - \nu'_{FG} - \nu_{F^*}$ |
| $\frac{d[mR_F]}{dt} = \pi_{RF} - \eta_{RF} + \pi_{RFbasal}$ |
| $\frac{d[mR_{HP}]}{dt} = \pi_{RP} - \eta_{RP} + \pi_{RPbasal}$ |
| $\frac{d[HP]}{dt} = \pi_{HP} - \eta_{HP}$ |
| **Conserved quantities** |
| $[P] + [P^{\#}]$ |
| $[SK] + [SK^*]$ |
| $[G] + [HSFG] + [HSF^*G]$ |

Table 2: **The ODEs used in the model and the conserved quantities.** Even if the system has twelve variables (listed in Table 1), only nine ODEs are required to model it. In fact, there are three conserved quantities: $[P]+[P^{\#}]$, $[SK]+[SK^*]$ and $[G]+[HSFG]+[HSF^*G]$ are constants. The initial conditions used for the twelve variables are: $[P] = 100000\ \mu$M, $\left[P^{\#}\right] = 1\ \mu$M, $[SK] = 0.1\ \mu$M, $[SK^*] = 0.05\ \mu$M, $[HSF] = 10.5\ \mu$M, $[HSF^*] = 1\ \mu$M, $[G] = 0.0012\ \mu$M, $[HSF^*G] = 0.0002\ \mu$M, $[HSFG] = 0.0008\ \mu$M, $[mR_F] = 0.0036\ \mu$M, $[mR_{HP}] = 0.0036\ \mu$M, $[HP] = 1\ \mu$M. Let us remark that the values of the variables are initiated at the initial conditions above, but before applying any HS we let the system run for a long time, so that it has reached the steady state (as confirmed by inspection of the Jacobian matrix of the system as discussed in Supplementary Material D.2) when we apply any HS. This transient part of each simulation is not shown in the figures. Similarly, the model was run for a long time to reach steady state conditions also during each step of the model fitting stage (including the gradient search).

| **Variable** | **Reference value** |
|---|---|
| Proteins | 100001 $\mu$M |
| SKs | 0.105 $\mu$M |
| HSFs | 300 $\mu$M |
| Genes | 0.0022 $\mu$M |
| mRNAs | 15 $\mu$M |
| HSP | 10000 $\mu$M |

Table 3: **Reference values employed for the normalization of model's variables.** For Proteins, SKs and Genes the variables are rescaled using the corresponding conserved quantity of Table 2.

| **Rate constant** | **Fiducial value** | **Final value** | **Unit of measure** |
|---|---|---|---|
| $k_P$ | 10 | 11.49 | $(\mu\text{M s})^{-1}$ |
| $k'_P$ | 100 | 111.9 | $\text{s}^{-1}$ |
| $k_S$ | 100 | 100.7 | $\text{s}^{-1}$ |
| $k'_S$ | 500 | 533.1 | $\text{s}^{-1}$ |
| $k'_F$ | 1 | 0.9557 | $\text{s}^{-1}$ |
| $k_F$ | 1 | 0.9640 | $\text{s}^{-1}$ |
| $k'_{FG}$ | 0.10 | 0.08371 | $\text{s}^{-1}$ |
| $k_{FG}$ | 0.0050 | 0.005574 | $(\mu\text{M s})^{-1}$ |
| $k_{F^*G}$ | 1 | 0.9131 | $(\mu\text{M s})^{-1}$ |
| $k'_{F^*G}$ | 0.50 | 0.4136 | $\text{s}^{-1}$ |
| $k'_{F^*}$ | 0.010 | 0.01064 | $\text{s}^{-1}$ |
| $k_{F^*}$ | 0.010 | 0.01192 | $\text{s}^{-1}$ |
| $k_{\pi_{RF}}$ | 16 | 18.16 | $\text{s}^{-1}$ |
| $k_{\pi_{RH}}$ | 4.5 | 4.193 | $\text{s}^{-1}$ |
| $k_{\pi_{HP}}$ | 0.5 | 0.5142 | $\text{s}^{-1}$ |
| $k_{\pi_F}$ | 0.02 | 0.02112 | $\text{s}^{-1}$ |
| $d_F$ | 0.001 | 0.0008728 | $\text{s}^{-1}$ |
| $d_{HP}$ | 0.000086 | 0.0009384 | $\text{s}^{-1}$ |
| $d_{RF}$ | 0.0015 | 0.001719 | $\text{s}^{-1}$ |
| $d_{RP}$ | 0.0012 | 0.001017 | $\text{s}^{-1}$ |

Table 4: **Values of the rate constants used in the model.** The values employed in all the simulations shown in this work are those labelled as final values, while those labelled as fiducial values are those employed as a starting point for the optimization procedure described in Section 2.3.

| **Rate law** | **Reaction** |
|---|---|
| $\nu_P = k_P \cdot [P^{\#}] \cdot [HP]$ | $P^{\#} + HP \rightarrow P + HP$ |
| $\nu'_P = k'_P \cdot [P] \cdot \omega_{TP}$ | $P \rightarrow P^{\#}$ |
| $\nu_S = k_S \cdot [SK^*]$ | $SK^* \rightarrow SK$ |
| $\nu'_S = k'_S \cdot [SK] \cdot \omega_{PS}$ | $S + P^{\#} \rightarrow SK^* + P^{\#}$ |
| $\nu'_F = k'_F \cdot [HSF] \cdot \frac{[SK^*]}{1+[SK^*]}$ | $HSF + SK^* \rightarrow HSF^* + SK^*$ |
| $\nu_F = k_F \cdot [HSF^*]$ | $HSF^* \rightarrow HSF$ |
| $\nu'_{HSFG} = k'_{FG} \cdot [HSFG]$ | $HSFG \rightarrow HSF + G$ |
| $\nu_{HSFG} = k_{FG} \cdot [G] \cdot [HSF]$ | $HSF + G \rightarrow HSFG$ |
| $\nu_{HSF^*G} = k_{F^*G} \cdot [G] \cdot [HSF^*]$ | $HSF^* + G \rightarrow HSF^*G$ |
| $\nu'_{F^*G} = k'_{F^*G} \cdot [HSF^*G]$ | $HSF^*G \rightarrow HSF^* + G$ |
| $\nu'_{HSF^*} = k'_{F^*} \cdot [HSF^*G]$ | $HSF^*G \rightarrow HSFG$ |
| $\nu_{HSF^*} = k_{F^*} \cdot [HSFG]$ | $HSFG \rightarrow HSF^*G$ |
| $\pi_{RF} = k_{\pi_{RF}} \cdot [HSF^*G]$ | $HSF^*G : HSF^*G + mRF$ |
| $\pi_{RP} = k_{\pi_{RP}} \cdot [HSF^*G]$ | $HSF^*G : HSF^*G + mRHP$ |
| $\pi_{HP} = k_{\pi_{HP}} \cdot [mR_{HP}]$ | $mR_{HP} : HP + mR_{HP}$ |
| $\pi_F = k_{\pi_F} \cdot [mR_F]$ | $mR_F : HSF + mR_F$ |
| $\eta_F = d_F \cdot [HSF]$ | $HSF \dashrightarrow$ |
| $\eta_{HP} = d_{HP} \cdot [HP]$ | $HP \dashrightarrow$ |
| $\eta_{RF} = d_{RF} \cdot [mR_F]$ | $mR_F \dashrightarrow$ |
| $\eta_{RP} = d_{RP} \cdot [mR_{HP}]$ | $mR_{HP} \dashrightarrow$ |
| $\pi_{RFbasal} = p_{RFbasal}$ | $\rightarrow mR_F$ |
| $\pi_{RPbasal} = p_{RPbasal}$ | $\rightarrow mR_P$ |

Table 5: **Kinetic rate laws used in the model.** The reactions are those represented in the scheme of Fig. 1 A, and the rate laws are mainly based on mass action kinetics, apart from some terms which follow Arrhenius law or have a Michaelis-Menten or Hill kinetics behaviour. The last is $\omega_{PS} = \frac{[P^{\#}]^m}{P_0^m + [P^{\#}]^m}$, where $m = 5$ and $P_0 = 600$ $\mu$M. The term following the Arrhenius law is $\omega_{TP} = A \cdot \exp\left(-\frac{E_a}{R \cdot T}\right)$, with an activation energy of $E_a = 174.440$ kJ mol$^{-1}$, within the wide range reported in the literature for the activation energies of protein denaturation due to thermal stress (as discussed in the literature [23, 24]), and with $A = 9.4318 \times 10^{28}$, $R = 8.3144598$ J mol$^{-1}$ K$^{-1}$ the ideal gas constant and $T$ the temperature expressed in degrees Celsius. Finally, the basal rates are computed as $p_{RFbasal} = d_{RF} \cdot d_F \cdot 0.02125$ $\mu$M$/k_{\pi_F}$ and $p_{RPbasal} = d_{RP} \cdot d_{HP} \cdot 17.5$ $\mu$M$/k_{\pi_{HP}}$. Where not otherwise specified, the values of the parameters mentioned in this caption, as e.g. $m$ and $P_0$, were manually tuned and were not part of the parameter fitting procedure described in section 2.3, which only focused on the rate constants. Section E of the Supplementary Material demonstrates the robustness of the main findings of this work w.r.t. changes in the parameters $m$ and $P_0$.

## B Supplementary Material: extended description of the calibration of the model

In this section we provide more details on the procedure that we have employed to calibrate the model. We describe in detail the objective function used. We then describe the random sampling of the parameter space. We finally explain in detail the gradient search employed to determine the final set of parameter values.

### B.1 Definition of the objective function

To assess how well the model simulations reproduce the behaviour of the corresponding data points from Schmollinger et al. [6], we employ as objective function the root mean square deviation (RMS) of the simulation results w.r.t. the data on mRNA expression for HSF and HSP, for the six controls of the feeding experiments (Fig. 2 A,B). It should be remarked that the data obtained with northern blot analysis by Schmollinger et al. [6] consist of relative concentrations and not of absolute concentrations, as can be seen from the figures therein. Indeed, the data points of each control curve are normalized to the maximum among all the time points and all the curves (controls and feedings) for that particular feeding experiment. Thus, to compute the RMS we normalize the values of the concentrations to the maximum of each curve, to be able to compare these with the experimental data. The model can quantitatively capture the proportional increases or decreases observed in experiments; whether it also captures their absolute values awaits future experimental results. The data employed do not contain absolute (dimension full) measures of the concentrations, but only relative measures (i.e. normalized to a maximum). Thus the concentrations appearing in each figure are normalized to a reference value (see Table 3).

### B.2 Random exploration of the parameter space employing a Monte-Carlo analysis

The fiducial parameter set from which we start, manually tuned by means of our understanding of the mechanism underlying the model, already allows the model to roughly reproduce the qualitative behaviour observed in the experiments. Nevertheless, we decided to explore more systematically the parameter space represented by the twenty rate constants of Table 4, and study how the RMS just defined changes if we move around from the fiducial parameter set.

For this purpose we first performed a Monte Carlo (MC) scan of the parameter space. We did so by assuming a flat prior probability distribution between half and two times of the fiducial value of each parameter (fiducial values are listed in the second column of Table 4). The part of the parameter space which we explored is thus a 20-dimensional hypercube, every point having the same probability of being randomly selected. Then, by randomly extracting a value for each parameter from these distributions, we generated $10^5$ randomized sets of parameters.

For each randomized set of parameters we computed the corresponding value of the RMS. We obtained values of the RMS ranging approximately from 0.13 to 0.70. The fiducial parameter set has a RMS w.r.t. the controls of the feeding experiments of 0.147, which is already remarkably close to the lower edge of the range obtained using random combinations of the parameters.

For visualization we then select the 300 points corresponding to the best (i.e. lowest) values of the RMS (which then range between 0.130 and 0.149). By plotting for these points the value of the RMS versus the value of each parameter (as shown in Fig. 2 C for the 20 parameters), we observe that for the vast majority of the parameters no preferred interval in which the lowest values of the RMS occur more often can be identified. Very low values of the RMS can occur everywhere in the interval used for the random scan, depending on the values assumed by the other parameters.

We subsequently studied the correlation between each couple of parameters, by plotting the RMS of each of these 300 parameters sets (which as we stated previously where selected among the randomly generated $10^5$) as function of each couple of parameters (Fig. 5 of the Supplementary Material). We found that only seldomly there are clear correlations or anti-correlations between the preferred values for the two parameters of the couple. Nevertheless, for the majority of the couples of parameters no such correlation can be identified.

Interestingly, the parameter set which provides

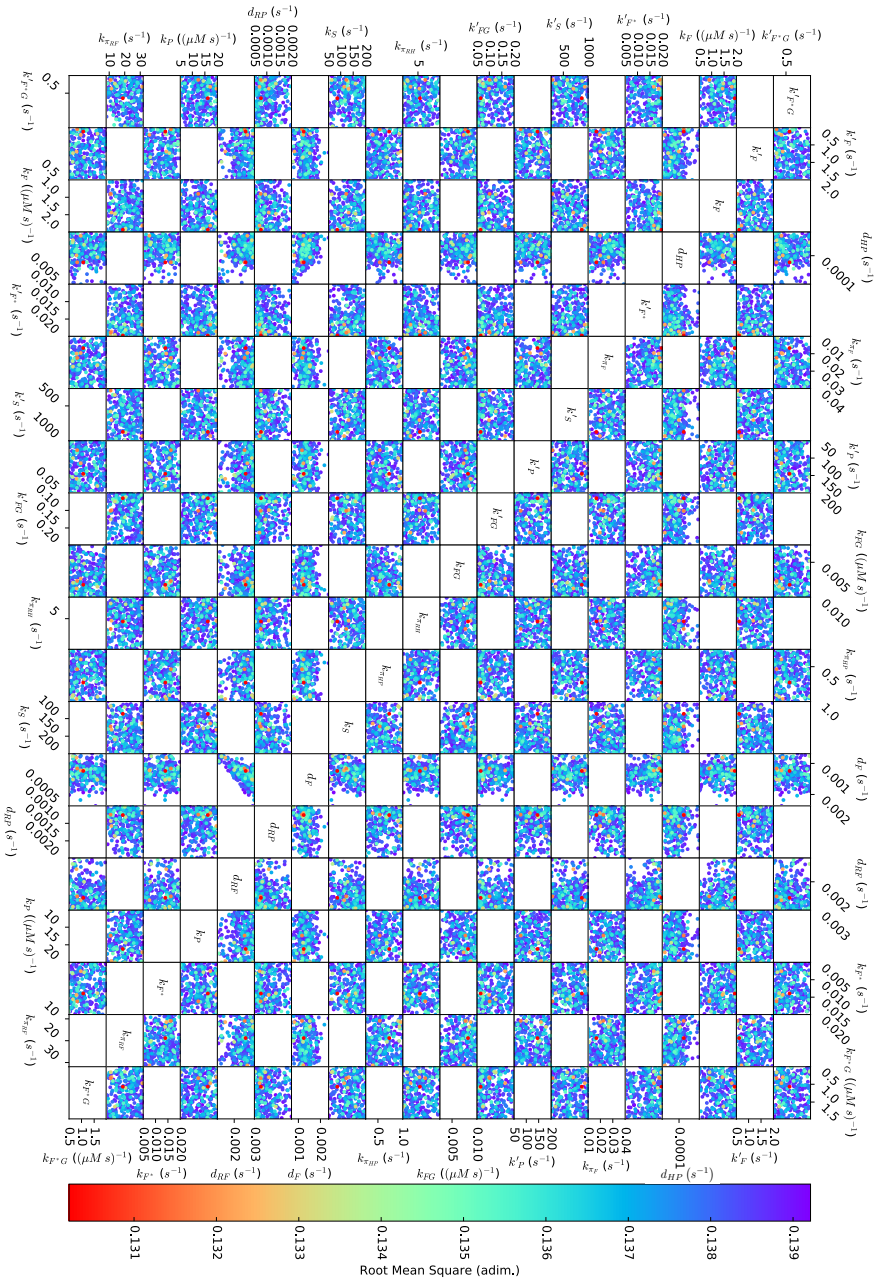


Figure 5: **Projection w.r.t. each couple of parameters of the MC scan of the parameter space.** Colors show the RMS for the same points of Fig. 2 C, to investigate the presence of correlations among them.

the best RMS among this randomly generated sets (i.e. the isolated point at the very bottom of each panel in Fig. 2 C), fails to reproduce key features that the real system has shown in the experiments, when used to simulate other situations, in particular the double heat shock experiment shown in Fig. 3 C. For instance the capability of showing a second HSR of the same amplitude as the first only when the duration between the two consecutive HSs lasts for at least 5h cannot be reproduced. Furthermore, other parameter sets providing a low RMS in the MC approach were exhibiting oscillations in the concentrations of $mR_F$ and $mR_{HP}$, which have not been observed in experiments. In principle, the double heat shock experiment data could also be included in the calibration procedure, but this would require some rather arbitrary assumptions as described in Supplementary Material B.4, and thus we discarded that approach.

These observations lead us to conclude that many very different configurations distributed in the whole parameter space would result in a small RMS, but fail to reproduce key features. Since the fiducial parameter set reproduces all desired features and, moreover, results in an only slightly higher RMS than the lowest obtained by the random sampling approach, we chose a gradient search for a local optimization of the manually tuned fiducial parameter set.

On the other hand, if we consider small variations of one parameter at a time, we observe smooth variations in the RMS values, indicating that the RMS is roughly parabolic for perturbations of each one of the majority of the parameters around the corresponding fiducial value, and that the fiducial value often lies already close to the minimum.

In Fig. 6 we show that the final parameter values (indicated by yellow starts), obtained with the local optimization (gradient search, see next section), lie indeed around the minimum of an RMS landscape which is roughly parabolic w.r.t. individual perturbations of (the great majority of) the parameters. This figure shows how much the RMS decays for small changes in any of the 20 rate constants considered in our fitting procedure and listed in Table 4. We have here considered small variations of one parameter at a time, adding or subtracting up to 80% of its final value, for each parameter. For a parameter with fiducial value e.g. 100 $s^{-1}$, this means considering the range from 20 $s^{-1}$ to 180 $s^{-1}$, which contributes to the asymmetry in the majority of the sub-panels of Fig. 6, where the RMS appears to grow more when lowering the corresponding rate constant. We can also remark how the fiducial parameter values, indicated by a red vertical line, are indeed already close to the values minimizing the RMS for each parameter.

We clearly see that for similar perturbations different parameters produce drastically different changes in the RMS. While over the range of perturbations considered some parameters induce changes in the RMS spanning only the range from 0.14 to 0.16, some other parameters lead to increases of the RMS up to about 0.40. This analysis provides thus an idea of how important each parameter is to the overall quality of the model fit and it might offer some motivation for future experimental work, by suggesting which parameters it would be most useful to measure empirically, where possible.

Thus, after having performed a global random scan of the parameter space to gain a better understanding of how the RMS behaves in it, we decided to determine the final parameter set by employing instead a local optimization method, namely a gradient search starting from the fiducial parameter set. This provides a parameter set allowing a better fit to the data used to compute the RMS, while not moving too far away from our already plausibly behaving fiducial set.

### B.3 Local optimization using a gradient search to fit the parameters' values to the data

As discussed above, the fiducial set of parameters allows already to obtain a value of the RMS which is close to the lower value of the RMS obtained considering the random parameter sets. In fact we explored the parameter space with the MC analysis and we concluded that the RMS can be improved only marginally w.r.t. the value corresponding to this set. Having shown that no region of the parameter space is preferred by the RMS, we opted for a local optimization procedure, thus performing a gradient search starting from the point represented by the fiducial set of parameters, employing the steepest descent method.

This means that we start from a point $\vec{x}_0$ in the parameter space represented by the fiducial values for the 20 rate constants employed in our model (see

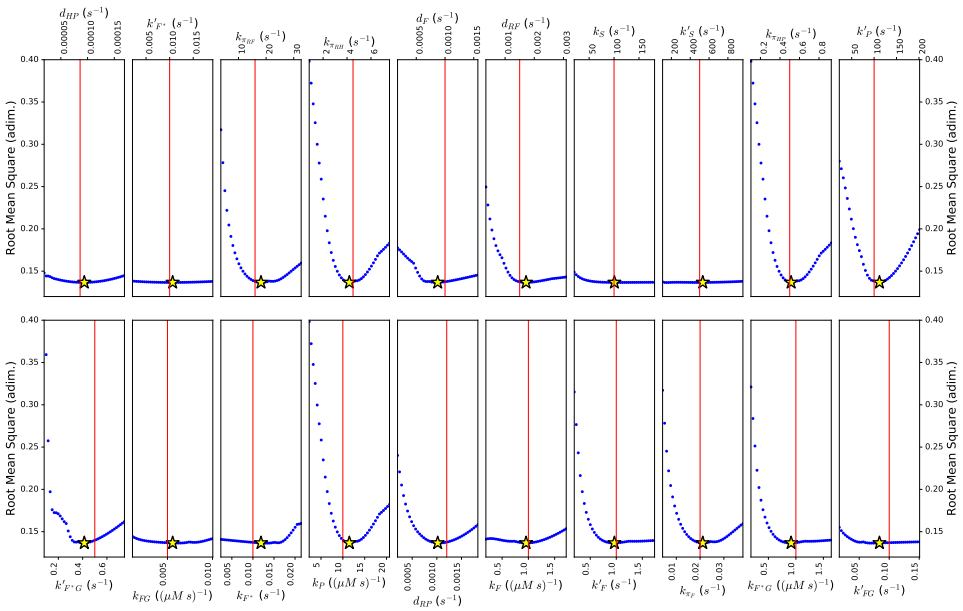


Figure 6: **Sensitivity analysis: effect on the RMS of small perturbations in each parameter.** Each sub-panel shows the changes in the RMS occurring for small perturbations of one parameter at a time. These perturbations are performed around the final parameter values (indicated in the figure by yellow stars) obtained with the local optimization. The vertical red line in each sub-panel indicates the fiducial value of the corresponding parameter, i.e. the starting point of the two alternative calibration approaches used, the MC and the gradient search finally employed. For each parameter separately, perturbations are performed by adding or subtracting up to 80% of its final value to itself.

Table 4). We compute the corresponding value of the root mean square $RMS\left(\vec{x}_0\right)$. We then compute numerically the gradient of the RMS at that point $\vec{\nabla} RMS\left(\vec{x}_0\right)$ (by approximating partial derivatives using the symmetric difference quotient).

We then proceed along the direction opposite to the gradient towards a new point $\vec{x}_{n+1} = \vec{x}_n - \gamma \vec{\nabla} RMS\left(\vec{x}_n\right)$ in the parameter space which provides a smaller value of the RMS. We do so iteratively until a termination criterion described below is satisfied, and label the iteration number by $n$. At each iteration, we need to decide which is the length of the step $\gamma$ that we want to use in the direction opposite to the gradient. To do so, we implement a line search with the aim to loosely minimize the function $f\left(\gamma\right) \doteq RMS\left(\vec{x}_n - \gamma \vec{\nabla} RMS\left(\vec{x}_n\right)\right)$ w.r.t. $\gamma$, i.e. along the direction opposite to the gradient. This means finding the value of $\gamma$ which minimizes the function $f\left(\gamma\right)$. We do so numerically employing a modification of the bisection rule based on the Golden ratio to save computation time.

Since the orders of magnitude of the parameters are very different, we expect the isosurfaces of the function $RMS\left(\vec{x}\right)$ to be far away from being spherical. This would lead to a very slow convergence of the method, because the gradient at each step would hardly point roughly toward the minimum. We thus employed also a preconditioning of the function that we want to minimize, i.e. the RMS. This means that we applied the numerical procedure of minimization not to the actual $RMS\left(\vec{x}\right)$, but to a function $RMS'\left(\vec{x}'\right)$ which we obtain by transforming $RMS\left(\vec{x}\right)$ via a rescaling of all the parameters using their fiducial values. In this way all the parameters are of order of magnitude 1, which is more suitable for the application of the described numeric algorithm. Once the minimization has been performed, we applied the inverse transformation to re-obtain the parameters in their original form.

The termination criterion we employed forces the algorithm to stop when the average RMS decrease over the last ten iterations is lower then a threshold value. This criterion is of course somewhat arbitrary, but we have selected it by empirically verifying that it allows a better fit to the control data (a lower RMS), avoiding an over-fitting which would lead to model behaviour too far away from the behaviour expected w.r.t. other situations as e.g. the double HS. The algorithm stopped after 31 iterations (Fig. 2 D) and returned the set of parameters listed in the second column of Table 4 as final values. The corresponding value of the RMS w.r.t. the controls of the feeding experiments is 0.137. We have employed this set of parameters to perform all the simulations shown in this work (apart from the calibration runs).

### B.4 Exploring the possibility of an alternative RMS

As a final test we defined, similarly to what we have done above, a RMS distance between the model simulations and the experimental data from the double HS experiment of Schroda et al. [28]. Defining such an RMS is much more arbitrary than defining the RMS w.r.t. the controls of the feeding experiments, because of the previously mentioned hypothesis on the proportionality between the enzyme concentration and its activity. We do not combine the two RMSs, as this would require to attribute to the two a weight which would be highly arbitrary.

For these reasons we only compute this second RMS a posteriori as a check, for the best (i.e. here those with $RMS < 0.149$ w.r.t. the data of the feeding experiments) among the random parameters sets selected above, and show the distribution of the values of the two different RMS in Fig. 7 of the Supplementary Material. This shows that minimizing both at the same time cannot be obtained, and one needs to find a trade-off between the two. The values of the RMS w.r.t. double HS run between 0.11 and 0.55 (not all visible in the figure, which is magnified). We can a posteriori compute this RMS for the fiducial parameter set, finding 0.170, and for the final parameter set, obtaining 0.168.

Finally, we verified that in the gradient search performed in the previous section, if we employ the sum of the two RMS instead of the RMS w.r.t. to feedings only, we can improve the fit to the double HS data, but at the cost of having non-realistic behaviours in the model consisting of big oscillations in the concentrations after onset of HS (and of introducing an arbitrary weight between the two RMSs). We eventually decided to employ for the calibration only the RMS w.r.t. the feeding data.

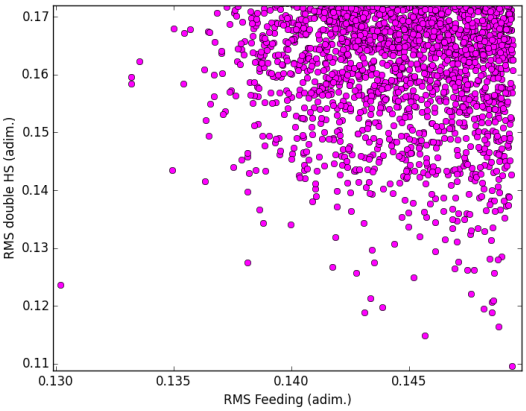


Figure 7: **RMS w.r.t. the double HS data of Fig. 3 C versus RMS w.r.t. the controls of the feeding experiments of Fig. 2 (A-B).** The figure shows a magnification of the region where the points with lowest values of the RMSs lie.

## C Supplementary Material: double heat shock characterization

We present here some additional simulations which illustrate extensively the features of the response to a double HS, and next we explain how we have extended our model to simulate the double HS experiments performed in Schroda et al. [28].

### C.1 Simulating the double HS experiments to investigate their main features

Fig. 8 shows four groups of panels illustrating how the HSR dynamics changes when we simulate a generic double HS experiment, providing to the system two HSs separated by 30 min (A-F), 2 h (G-L), 3 h 30 min (M-R) or 5 h (S-X). We provide a sudden variation of the temperature between 25°C and 42°C. In each simulation, the first HS lasts 120 minutes and starts at $t = 40$ minutes, while the second HS lasts until the end of the simulation.

As we can see from Fig. 8, when the second HS occurs only 30 min after the first, the system shows almost no response to the second. This is due to the fact that during the first HS, thanks to gene activation (panel D) and subsequent production of mRNA for the HSF (panel E), the quantity of HSF available to the system increases (panel C). When the second HS occurs, 30 min after the end of the first, the HSF available to the cell is already enough and no activation of the SK takes place (panel B). When the second HS occurs a lot of HSP is still available in the system (panel F) and thus the level of $P^{\#}$ does not increase (panel A). When the second HS occurs 2 h after the first, there is a small HSR during the second, visible at the level of SK (panel H) and of mRNAs (panel K). This because even if a lot of HSP is still available (panel L), the HSF concentration is very low (panel I) and then a moderate HSR is necessary to allow the system to quickly refold unfolded proteins. The HSR corresponding to a second HS occurring at 3 h 30 min after the first is similar, but enhanced. When the second HS occurs 5 h or more after the first, the concentration of HSP is approaching the level it had before the HS (panel X), all the other quantities are back to the original values, and an almost full HSR takes place during the second HS (panels T-W).

It is very interesting to remark that, while the concentrations of all the species after the end of the first HS go back to the values that they had before the first HS quite fast, the HSP does not (panel X). This allows to have only a much smaller amount of unfolded protein $P^{\#}$ during the second HS w.r.t. the amount during the first HS. This can be seen in panels A, G, M and S, where the concentration of degenerated proteins $P^{\#}$ increases by a considerable amount during the first HS, while considerably less during the second even when this is occurring many hours after the first. The behaviour we observe in our simulations likely indicates that the production of HSF which follows a first HSR and the accumulation and slow degradation of HSP has the role of preparing the system for a subsequent occurrence of the same stressing situation (HS) already encountered in the past, thus representing a transient molecular memory. This observed behaviour is consistent with the claim of Schroda et al. [28] that *Chlamydomonas reinhardtii* needs around 5 h after a first HS to recover and exhibit a HSR during which the increase in $HSP$ concentration has the same amount as during the first HSR.

### C.2 Extending the model to simulate the experiments on the double HS

To perform the simulation reproducing the data of Schroda et al. [28] (Fig. 3 C of the main text), we extended our model including few new variables and equations. We included the production of ARS enzyme by adding two new variables and four new reactions. These are listed in Table 6, together with the corresponding ODEs and parameters values. The two variables represent the concentrations of the mRNA coding for the ARS enzyme (due to fusion of the ARS gene on the HSP70A promoter), and that of the ARS enzyme. The four new reactions that we added describe the production of mRNA coding for the ARS enzyme from G*HSF, the translation of ARS mRNA into ARS enzyme (i.e. production of ARS), a very slow degradation of the ARS enzyme and degradation of ARS mRNA. We selected the values of the free parameters in order to match the observations. The comparison in Fig. 3 C requires the simplifying, but reasonable, additional assumption that the activity of the ARS enzyme is roughly directly proportional to its concentration. Despite these simplifications, the qualitative agreement between simulation and data is remarkable.

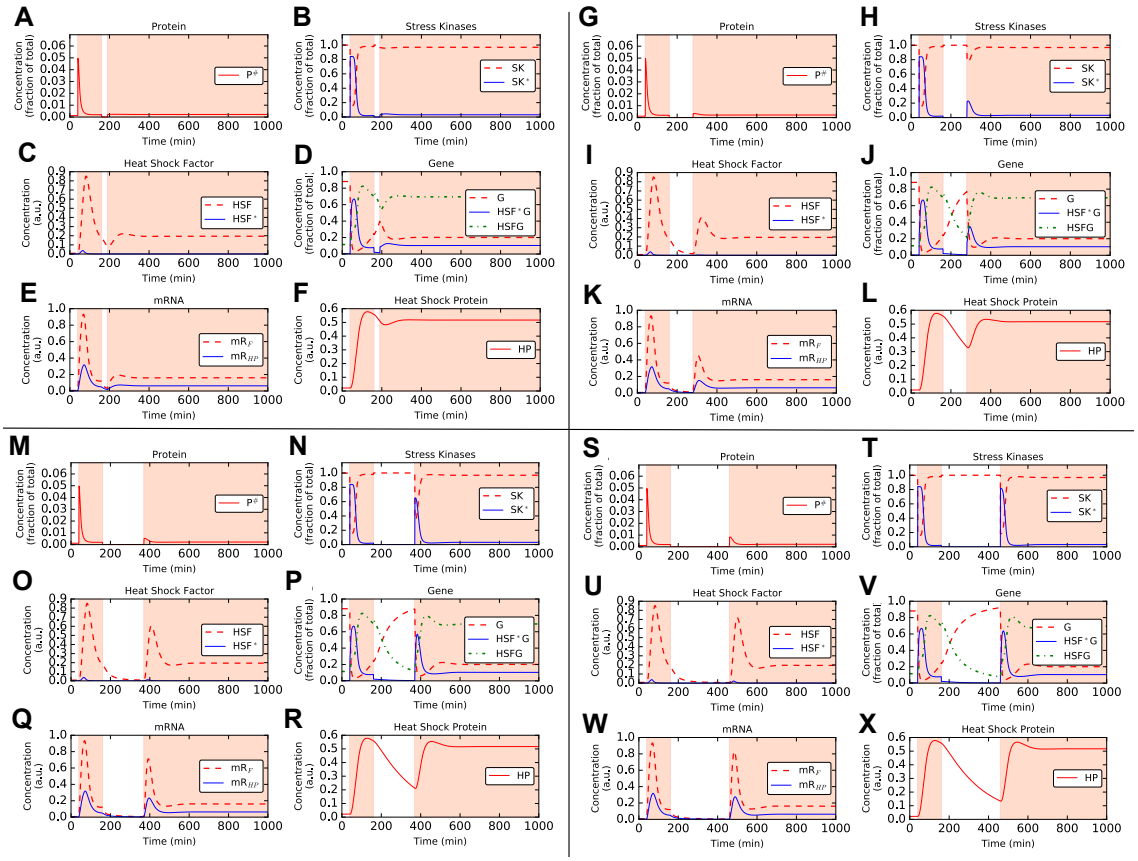


Figure 8: **Simulation of a generic double heat shock experiment.** The second heat shock is provided respectively after 30 min **(A-F)**, 2 h **(G-L)**, 3.5 h **(M-R)** and 5 h **(S-X)**. We can appreciate how the dynamics changes at the level of each species. Particularly relevant is the fact that a full response to the second HS is possible only after about 5 h, as clearly shown by e.g. the SK*, HSF* and mRNAs curves.

| **Variable** | **Concentration of** |
|---|---|
| $[mR_{ARS}]$ | mRNA for ARS enzyme |
| $[ARS]$ | ARS enzyme |
| **ODE** | |
| $\frac{d[mR_{ARS}]}{dt} = \pi_{RARS} - \eta_{RARS}$ | |
| $\frac{d[ARS]}{dt} = \pi_{ARS} - \eta_{ARS}$ | |
| **Rate law** | **Reaction** |
| $\pi_{RARS} = k_{\pi_{RARS}} \cdot [HSF^*G]$ | $HSF^*G : HSF^*G + mR_{ARS}$ |
| $\pi_{ARS} = k_{\pi_{ARS}} \cdot [mR_{ARS}]$ | $mR_{ARS} : mR_{ARS} + ARS$ |
| $\eta_{RARS} = d_{RARS} \cdot [mR_{ARS}]$ | $mR_{ARS}$ --→ |
| $\eta_{ARS} = d_{ARS} \cdot [ARS]$ | $ARS$ --→ |
| **Rate constant** | **Value** |
| $k_{\pi_{RARS}}$ | 36 $s^{-1}$ |
| $k_{\pi_{ARS}}$ | 0.000125 $s^{-1}$ |
| $d_{RARS}$ | 0.0012 $s^{-1}$ |
| $d_{ARS}$ | 0.0000125 $s^{-1}$ |

Table 6: **Extension of the model to include ARS production.** Variables, ODEs, rate laws and rate constants used to extend the model in order to simulate the double HS experiments which measure the expression of the ARS enzyme put under the control of the HSP70A promoter, performed in Schroda et al. [28]. The parameters values above are manually selected in order to reproduce the observations.

# D Supplementary Material: steady state characterization and further applications of the model

One of the main advantages of a mathematical model, inferred and calibrated from experiments, is that such a model often can be used to simulate situations which are difficult to test with experiments and to compute quantities which are difficult to measure.

Thus, in this section we employ our model to simulate systematically further situations that are interesting while difficult to test or not yet tested experimentally.

With our model we thus show the ability of the system to adapt to temperatures higher than usual during heat shocks longer than three hours by shifting to a new steady state. We subsequently study how the steady state concentrations depend on the temperature at which the steady state is reached. Finally we systematically investigate how the accumulation of HSPs depends on the combination of temperature and duration of the heat shock.

## D.1 Prolonged heat shock

We first investigate which response the model predicts upon exposure to a prolonged HS, and how the system acclimates to persistently high temperatures. Experimentally, the systems-wide response to long-term HS was investigated in Hemme et al. [35], where cells adapted to 25°C were exposed to 42°C for a period of 24 h, followed by 8 h at 25°C (recovery phase).

The simulation results for this scenario, where the temperature increase was simulated at time $t = 0$, are shown in Fig. 9 (A-F). Two distinct phases of the response can be distinguished. The first, early HS phase, lasting approximately 3 h after applying the heat shock, represents the initial heat shock response, in which the internal variables respond similarly to the normal heat shock simulations described above (see in particular Fig. 1 (B-G), and also the control curves in Fig. 2 (A,B)). During the late HS response, lasting from 3 to 24 h during heat shock, the variables approximate a new stationary state. This state is characterized in particular by increased concentrations of mRNAs (panel E), and consequently of HSF and HSP, when compared to the corresponding concentrations before the onset of the HS. These elevated levels suggest an acclimation of *Chlamydomonas reinhardtii* to continuous HS conditions, which allows to efficiently avoid the accumulation of misfolded proteins. After reverting the conditions to normal temperatures (25°C), a recovery phase can be observed, in which the variables relax to the original stationary state over a period of several hours. These results are consistent with the observations of Hemme et al. [35] focusing on HSP production (see in particular Fig. 8 therein).

## D.2 Stationary behaviour

The observations that the system adopts a new stable stationary state when continuously exposed to elevated temperatures raises the question how this long-term response depends on the rate of protein denaturing. The steady state concentrations depend on temperature only via the speed of the reaction $P \to P^{\#}$, which itself depends on the temperature, and on other quantities, as described by the corresponding equation in Table 5 of the Supplementary Material based on the Arrhenius law. We now study how the steady state concentrations depend on the temperature at which the steady state is reached.

First we have intuitively verified that the system reaches a steady state by naively integrating it over very long times. We then do it by applying the rigorous procedure which consists in looking for a root of the system represented by the ODEs of Table 2 of the Supplementary Material, i.e. to find the concentrations which correspond to a steady state. It's best to have a good initial guess as a starting condition for the search, and the integration over a very long time performed above provides that. Then, we repeat this for different values of the temperature, to see how the steady state concentrations change as a function of temperature. The result is plotted in Fig. 9 (G-L). Let us remark that on the horizontal axes of Fig. 9 (G-L) we have the temperature. Each point in any panel corresponds to the concentration of the corresponding species at steady state, when the steady state is reached at that particular temperature. We can see that the concentrations at steady state are different for different temperatures, with for instance higher values of the concentrations of mRNAs and HSP corresponding to higher values of the temperature.

On the one hand it is remarkable that, for values of the temperature not too high, the concentration

of unfolded proteins at steady state is kept very low by the response of all the other species, in particular well below one percent of the total amount of proteins. On the other hand, when the temperature increases considerably the HSR is no longer able to efficiently counteract the accumulation of degenerated proteins which accumulates at concentrations high enough to kill the cell. This accumulation is evident in Fig. 9 G.

Finally, we have also verified that, for each of the values of the temperature that we have considered, the model exhibits a realistic stationary behaviour, i.e. the associated steady state (which is a non-equilibrium one) is stable. To do so, firstly we computed numerically the Jacobian of the vector field associated to the ODEs system summarized in Table 2 of the Supplementary Material. Next, we evaluated the Jacobian of the system at the steady state under investigation. Then we computed the eigenvalues of the Jacobian to determine the stability of the steady state. We repeated the procedure for the steady state associated to each of the values of the temperature considered. For every steady state considered, all the nine eigenvalues result to have negative real part, which implies that each of the steady states considered is a stable steady state.

### D.3 Relationship between HS temperature and HS duration in the production of HSP

The questions which we want to address in this section are the following: "is *Chlamydomonas reinhardtii* more stressed for a short HS with high temperature or a long one but with a lower temperature? What is the relationship between the two?". A related question is "does the production of HSP occur only under very intense HS conditions (high temperature, long duration) or does it occur also for very small temperature increases or very short HSs?".

Moreover, since HSs induce the expression of HSP, exposing plant cells to controlled temperature increases may also represent a mean to drive the accumulation of cells proteins (HSPs, but not only) inside these cells. This might be interesting for instance in view of enriching plants in any protein of interest (by engineering the HSPs genes and use their HS-activable promoter to induce the expression of other genes of interest). The question that naturally arises is then: "which is the HS set-up (duration, temperature) that maximizes the accumulation of HSPs?".

To answer these questions we performed a systematic study of how much HSP is produced under different combinations of HS durations and HS temperatures. For such a study, we provided to the system a sharp increase in temperature starting from 20°C. Since here the goal is mainly to study under which conditions *Chlamydomonas reinhardtii* is more stressed, the response is closed only when HSP is produced and can act to refold unfolded/misfolded proteins and the main interest in eliciting an HSR may be to induce the accumulation of HSP (or other proteins), we perform this study at the level of HSP production.

Thus we simulate the response to the different combinations of HS temperature and HS duration, and we plot the value of $[HP]$ computed right at the end of the HS period as contours in the plane representing HS duration versus HS temperature, and a color map is used to make the figure visually clearer. It is worth to point out that this time point provides a value of $[HP]$ that is not necessarily the highest one that can be obtained with an heat shock of that temperature and duration, in fact $[HP]$ grows under HS, reaches a maximum, decreases a bit and settle to a new steady state value until HS is kept on. For short HSs, as seen previously, even if the increase in temperature is sharp, and the activation of SK follows, there is a certain inertia in the response at the level of mRNA production, and an additional delay in the HP synthesis. As a consequence if the concentration of HP is read out at the end of a short HS, it is possible that the obtained value is lower than the value one would obtain with the same HS, but waiting some more time.

Fig. 9 M shows the concentration of HSP at the end of the HS, as a function of HS duration and HS temperature. Firstly, looking at how $[HP]$ changes for any fixed temperature, we can appreciate the same features that we already noticed in Fig. 1 (B-G). There is no response at the level of $[HP]$ for duration of less then about 10 min, then HP rapidly goes up for longer HS, and the maximum HP concentration occurs around 80 to 100 min. For longer HSs $[HP]$ is somewhat lower, and does not change any more when increasing further the duration of the HS (new steady state reached, i.e. acclimation occurred). Second, considering increasing temper-

atures, we can see that even a small increase of few degrees in temperature results in an increase of $[HP]$. We can conceptually divide the plot in four regions (from left to right). No matter what temperature, for short HS (i.e. shorter than about 10 min) there is not enough time to lead to a significant increase in $[HP]$. For durations between approximately 10 min to 80 min and temperature above an increasing value, the contours are almost vertical. For high enough temperature, the HSR is activated and, no matter how much temperature is increased, the dynamics of the response is the same (this comes from the behavior described by the Arrhenius law by which temperature enters the model by unfolding proteins). In the region of the maximum of $[HP]$, the contour lines are almost horizontal, telling us that the maximum reached by $[HP]$ strongly depends, and increases with, the temperature of the HS. Finally, on the right side of the plot we see that the contour lines are now horizontal and $[HP]$ does not change with the duration any more: the system is acclimated (and thus has reached a new steady state). The values of $[HP]$ are somewhat smaller than those corresponding to the region of the maxima.

### D.4 Discussion of additional results

We have shown in Section D.1 that the system can acclimate to higher temperatures during heat shocks longer than three hours, by shifting to a new steady state. Two distinct phases are clearly visible: an early HS lasting for about the first 3 h, and a late HS in which the system shows acclimation (a new steady state is reached). The recovery phase is characterized by a recovery of the conditions pre-HS that occurs over several hours.

We have studied in Section D.2 the variation of the steady state concentrations w.r.t. changes in the temperature. The concentrations of HP, as well as of the mRNAs, increase with increasing temperature, but for not too high temperatures the concentration of unfolded proteins $\left[P^{\#}\right]$ is kept very close to zero, in particular well below 1% of the total amount of proteins $[P] + \left[P^{\#}\right]$. For temperatures too high the HSR cannot prevent the accumulation of degenerated proteins and the cell dies.

We finally used the model to systematically investigate how the accumulation of HSPs depends on the combination of temperature and duration of the heat shock, in Section D.3. Short (smaller than 10 min) HSs do not provide enough time for a significant response at the level of $[HP]$, the maximum of $[HP]$ for any given temperature is obtained at around 80 to 100 min, after that a somewhat smaller $[HP]$ is reached and maintained, and for long enough HS a higher temperature provides higher $[HP]$.

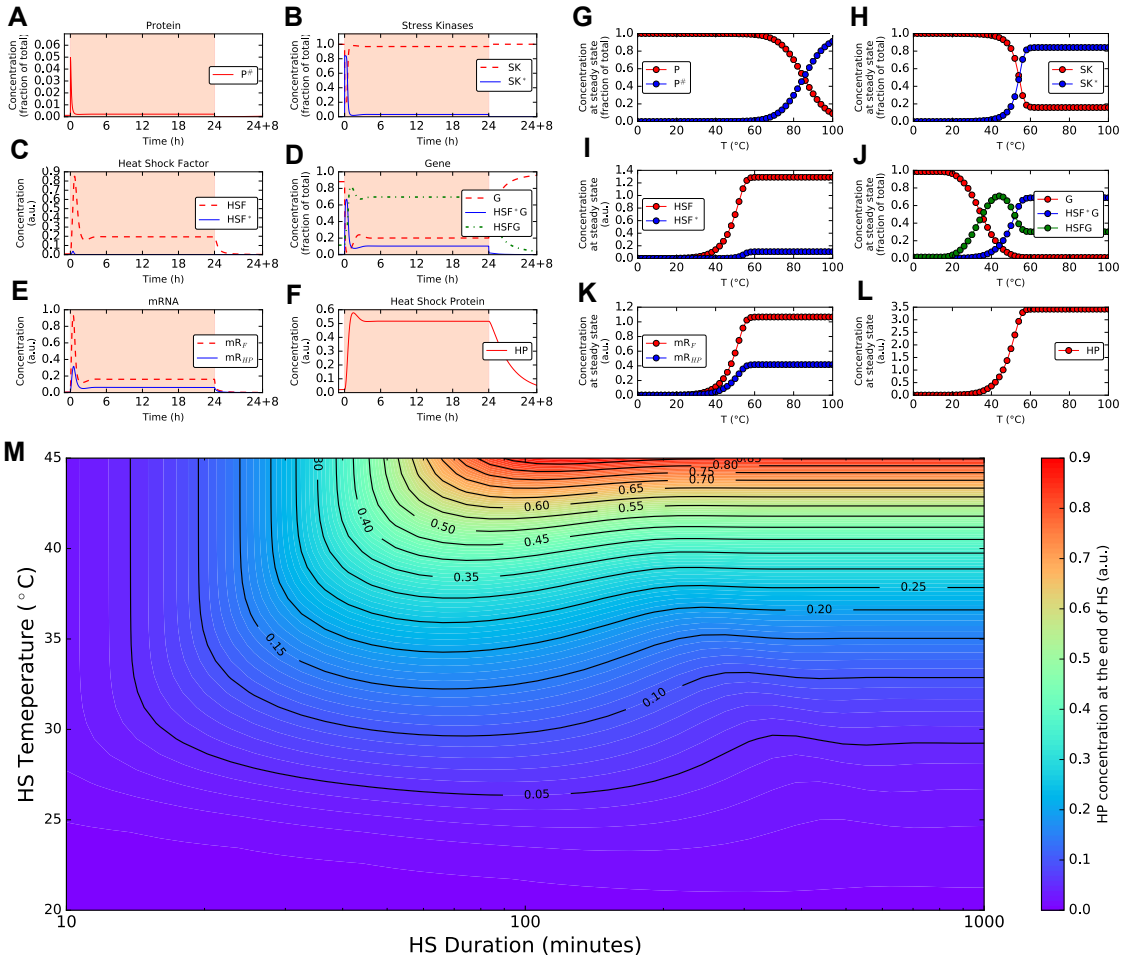


Figure 9: **The organism acclimates to long term HS by reaching a new steady state, but only up to reasonable temperatures. (A-F)** Simulation of the HSR under long-term HS and recovery, provided by shifting the temperature from 25°C to 42°C at $t = 0$ and back to 25°C after 24 h. Two distinct phases are clearly visible: an early HS lasting for about the first 3 h, and a late HS in which the system shows acclimation (a new steady state is reached). After reverting the conditions to normal temperature (25°C), a recovery phase can be observed, in which the variables relax to the original stationary state over a period of several hours. **(G-L)** These simulations illustrate how different steady state concentrations are reached for different temperatures. Each point represents the value of the corresponding concentration at the steady state reached for that particular temperature. The concentrations of HP, as well as that of the mRNAs, increase with increasing temperature. The concentration of unfolded proteins $\left[P^{\#}\right]$ is kept very close to zero for low values of the temperature. When the temperature increases considerably the HSR is no more able to efficiently counteract the accumulation of degenerated proteins which accumulates at concentrations high enough to kill the cell. This accumulation is evident in panel G. Already at about $T \approx$40°C the concentration of accumulated $P^{\#}$ at steady state starts to grow considerably, while above $T \approx$60°C it starts to become overwelming. **(M)** Systematic study of the HSP production as a function of different HS temperatures and HS durations. Short (smaller than 10 min) HSs do not provide enough time for a significant response at the level of $[HP]$, the maximum of $[HP]$ for any given temperature is obtained at around 80 to 100 min, after that a somewhat smaller $[HP]$ is reached and maintained, and for long enough HS a higher temperature provides higher $[HP]$. From this plot one can understand the combined effect of duration and temperature of the HS on the expression of $HSP$.

## E Supplementary Material: Robustness of the main results against changes in Hill kinetics parameters

The non-linearity introduced by the Hill kinetics term $\omega_{PS}$ by which the stress kinases $SK$ get activated (Table 5) is crucial to achieve the current model behavior, which would be impossible to achieve with e.g. a linear term. This non-linearity is a critical component necessary for the model to provide the highly sensitive initiation of the heat shock response. On the other hand, the values of the two parameters $m$ and $P_0$ are not that crucial, and thus were not part of the parameter estimation procedure, which involved only the 20 rate constants. Their values were manually set before the fitting procedure, which was then performed with these two parameters fixed at the nominal values of $m = 5$ and $P_0 = 600\ \mu M$.

To investigate how relevant the values of these two parameters are w.r.t the results which we obtain, and to assess the robustness of our results against changes in $m$ and $P_0$, we perform here additional simulations to show what is the impact of perturbations in these two parameters. We consider the effects on the result of the calibration procedure (Fig. 2 A and Fig. 2 B) on the one hand, and on the main result of this paper, i.e. how the maximum concentration of unfolded protein $P^{\#}$ during HS depends on how fast the heat shock was applied (Fig. 4 I), on the other hand. These figures are reproposed in Fig. 10 A to simplify the comparison. The value in the green circle in the third panel corresponds to the value of the peak in Fig. 1 B, because the HS of that figure exhibits an increase in temperature which occurs within less then 1 minute. We thus consider the effects produced by variations in the two parameters $m$ and $P_0$ on the simulations reported in panel A. We first consider effects of variations only of $m$, decreasing it, panel B, or increasing it, panel C. We then consider the effects of variations only of $P_0$, decreasing it, panel D, or increasing it, panel E. We finally consider the effects of simultaneous variations of both $m$ and $P_0$, aiming at maintaining a low RMS, panel F.

We clearly see that in all cases the RMS is higher (i.e. worse) than using the original values $m = 5$ and $P_0 = 600\ \mu M$, due to the fact that these values were kept constant during the fitting of the 20 rate constants to the data. As we can see, the qualitative behavior of the maximum concentration of unfolded protein $P^{\#}$ (right column) in any case does not change significantly, and the switch between the two regimes always occurs in the range between around 1 $min$ and 100 $min$ for the duration of the HS onset. Thus our main conclusion, i.e. that the accumulation of unfolded proteins dramatically differs for abrupt (e.g. $< 1\ min$), or gradual (e.g. $> 100\ min$) temperature increases, is robust. One might notice that the value of the maximum concentration of unfolded protein $P^{\#}$ for short times (values in the circles) changes across half an order of magnitude. While this might seem a lot at a first glance, it should be considered that changes in only one of the two parameters $m$ and $P_0$ lead often to a dramatic increase in the RMS (by far worse than the RMS value corresponding to the final or fiducial parameter set). A low RMS can be achieved for instance by simultaneous changes of both parameters, panel F. In this case we see that the maximal value of $[P^{\#}]$ is $0.035\ \mu M$, close to the original one which was $0.05\ \mu M$.

To summarize, both parameters could have been included in the fitting procedure, which would have tended to minimize the RMS, but we considered this not necessary. In fact, the drastic difference in accumulated unfolded proteins between sharp and gradual HSs is very robust and occurs for all the variations considered, despite changes in the absolute values of the accumulated unfolded protein. All the parameter changes considered here are anyway discouraged by the calibration data, as the considerably increased RMS obtained in all cases proves.

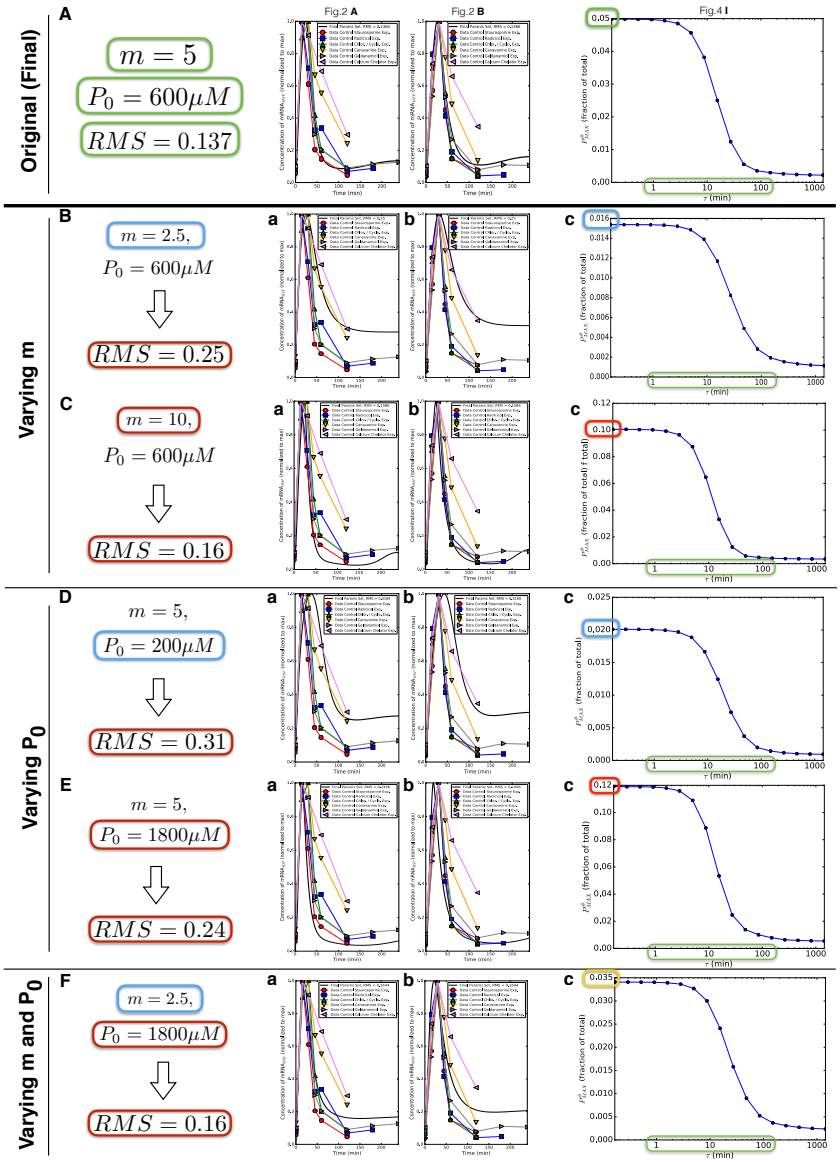


Figure 10: **Robustness of the results against changes in Hill kinetics parameters. (A)** Calibration data and corresponding model simulations reproposed from Fig. 2 A and Fig. 2 B, and main result on the maximum concentration of unfolded proteins accumulated at steady state reproposed from Fig. 4 I. **(B,F)** Effects on the simulations of panel A produced by changes in the two parameters $m$ and $P_0$ appearing in the Hill kinetics term $\omega_{PS}$ by which the stress kinease $SK$ gets activated (Table 5). These two parameters where not part of the model fitting procedure but rather manually selected, this figure thus describes the robustness of the result shown in Fig. 4 I against changes in $m$ and $P_0$. **(B,C)** Effects of variations only of $m$, decreasing it **(B)** or increasing it **(C)**. **(D,E)** Effects of variations only of $P_0$, decreasing it **(D)** or increasing it **(E)**. **(F)** Effects of simultaneous variation of both $m$ and $P_0$, aiming at maintaining a low RMS. The drastic difference in accumulated unfolded proteins between sharp and gradual HSs is very robust and occurs for all the variations considered, despite changes in the absolute values of the accumulated unfolded protein. All the parameter changes considered here are anyway discouraged by the calibration data, as the considerably increased RMS obtained in all cases proves.