






# Bayesian genome scale modelling identifies thermal determinants of yeast metabolism

Gang Li <sup>1</sup>, Yating Hu<sup>1,7</sup>, Jan Zrimec <sup>1,7</sup>, Hao Luo<sup>1</sup>, Hao Wang <sup>1,2,3</sup>, Aleksej Zelezniak<sup>1,4</sup>, Boyang Ji<sup>1,5</sup> & Jens Nielsen <sup>1,5,6</sup> 

The molecular basis of how temperature affects cell metabolism has been a long-standing question in biology, where the main obstacles are the lack of high-quality data and methods to associate temperature effects on the function of individual proteins as well as to combine them at a systems level. Here we develop and apply a Bayesian modeling approach to resolve the temperature effects in genome scale metabolic models (GEM). The approach minimizes uncertainties in enzymatic thermal parameters and greatly improves the predictive strength of the GEMs. The resulting temperature constrained yeast GEM uncovers enzymes that limit growth at superoptimal temperatures, and squalene epoxidase (ERG1) is predicted to be the most rate limiting. By replacing this single key enzyme with an ortholog from a thermotolerant yeast strain, we obtain a thermotolerant strain that outgrows the wild type, demonstrating the critical role of sterol metabolism in yeast thermosensitivity. Therefore, apart from identifying thermal determinants of cell metabolism and enabling the design of thermotolerant strains, our Bayesian GEM approach facilitates modelling of complex biological systems in the absence of high-quality data and therefore shows promise for becoming a standard tool for genome scale modeling.

<sup>1</sup>Department of Biology and Biological Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden. <sup>2</sup>National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Chalmers University of Technology, SE-41258 Gothenburg, Sweden. <sup>3</sup>Wallenberg Center for Molecular and Translational Medicine, University of Gothenburg, SE-41258 Gothenburg, Sweden. <sup>4</sup>Science for Life Laboratory, Tomtebodavägen 23a, SE-171 65 Stockholm, Sweden. <sup>5</sup>Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark. <sup>6</sup>BioInnovation Institute, Ole Måløes Vej 3, DK2200 Copenhagen N, Denmark. <sup>7</sup>These authors contributed equally: Yating Hu, Jan Zrimec. email: [nielsenj@chalmers.se](mailto:nielsenj@chalmers.se)

Temperature is a key environmental and evolutionary factor that shapes the physiology of living cells. Organisms have successfully adapted to survive in diverse temperature ranges<sup>1–3</sup>, where minor deviations from the optimal temperature by merely a few degrees can dramatically impair cell growth. For instance, the model eukaryotic organism *Saccharomyces cerevisiae* has an optimal growth temperature (OGT) of ~30 °C, whereas a temperature of 42 °C is already lethal to the organism<sup>4,5</sup>. Since cell growth fundamentally requires all cellular components to be functional in the temperature window of cell growth, proteins, an abundant group of biomolecules that carry out the majority of catalytic functions and are also highly sensitive to changes in temperature<sup>5–7</sup>, are considered to have a critical effect on cell physiology in relation to temperature. However, despite all our knowledge of temperature effects at both the cellular and molecular levels, including recent breakthroughs in temperature-dependent protein folding<sup>7–10</sup> and enzyme kinetics<sup>11,12</sup>, the temperature association between proteins and cell physiology is still poorly understood.

Multiple studies have attempted to model the temperature effects on cell growth, though with very few proteome-wide parameters<sup>13</sup>. Examples include the dominant activation barrier and the number of essential proteins to cell growth<sup>14</sup> and the activation energy of the growth process and the free energy change of protein denaturation<sup>15</sup>. These models showed excellent performance when describing the general cell growth rate at various temperatures, however, they could not pinpoint the specific rate-limiting enzymes, nor predict the amount of improvement in growth rate by replacing these enzymes with temperature-insensitive homologs.

To this end, genome-scale metabolic models (GEMs)<sup>16–18</sup>, which are a comprehensive mathematical representation of cellular biochemical reactions<sup>19</sup>, have been used to model the thermosensitivity of metabolism in *Escherichia coli*, for instance by associating metabolic reactions with protein structures<sup>20</sup> or by modeling protein-folding networks<sup>21</sup>. It however remains challenging to model more complex, eukaryotic organisms, such as *S. cerevisiae*, due to their metabolic complexity<sup>16</sup> as well as due to the lack of availability of the required enzymatic data<sup>7,22</sup>, including high-quality protein structures<sup>20,21</sup>. In addition, such GEMs rely on thousands of parameters to describe the temperature effects on protein folding and kinetics<sup>16</sup>, which have to be empirically or computationally estimated<sup>20,21</sup>. This leads to large statistical uncertainties in model parameters and can make the models unreliable, due to inaccurate temperature associations between proteins and cell physiology. Therefore, in order to enable accurate modeling of the temperature dependence of cell metabolism, a key requirement is to develop a modeling approach that resolves the issues with large uncertainties of temperature-related parameters and produces accurate temperature constrained predictions.

Hence, in the present study, we introduce a Bayesian genome-scale modeling approach to model the temperature effects on cellular metabolism in *S. cerevisiae*, the most widely used industrial organism with readily available thermal experimental data<sup>5,23,24</sup> and highly sophisticated GEMs<sup>16,18,25</sup>. We first quantify and reduce the large uncertainties in the parameters describing enzyme thermosensitivity using Bayesian statistical learning<sup>26</sup> to simulate phenotypic data. We show that the resulting models are capable of reproducing various experimental datasets and provide explicit insight into how yeast metabolism is affected by temperature. Our approach identifies the sterol metabolism as a key factor in the yeast thermal adaptation and predicts the flux-controlling enzymes in superoptimal temperature ranges as potential targets for the future design of thermotolerant yeast strains. We then experimentally validate the

predicted most rate-limiting enzyme by replacing it with an ortholog from a known thermotolerant yeast *Kluyveromyces marxianus*. We hereby demonstrate the power of Bayesian genome-scale modeling for studying complex biological systems.

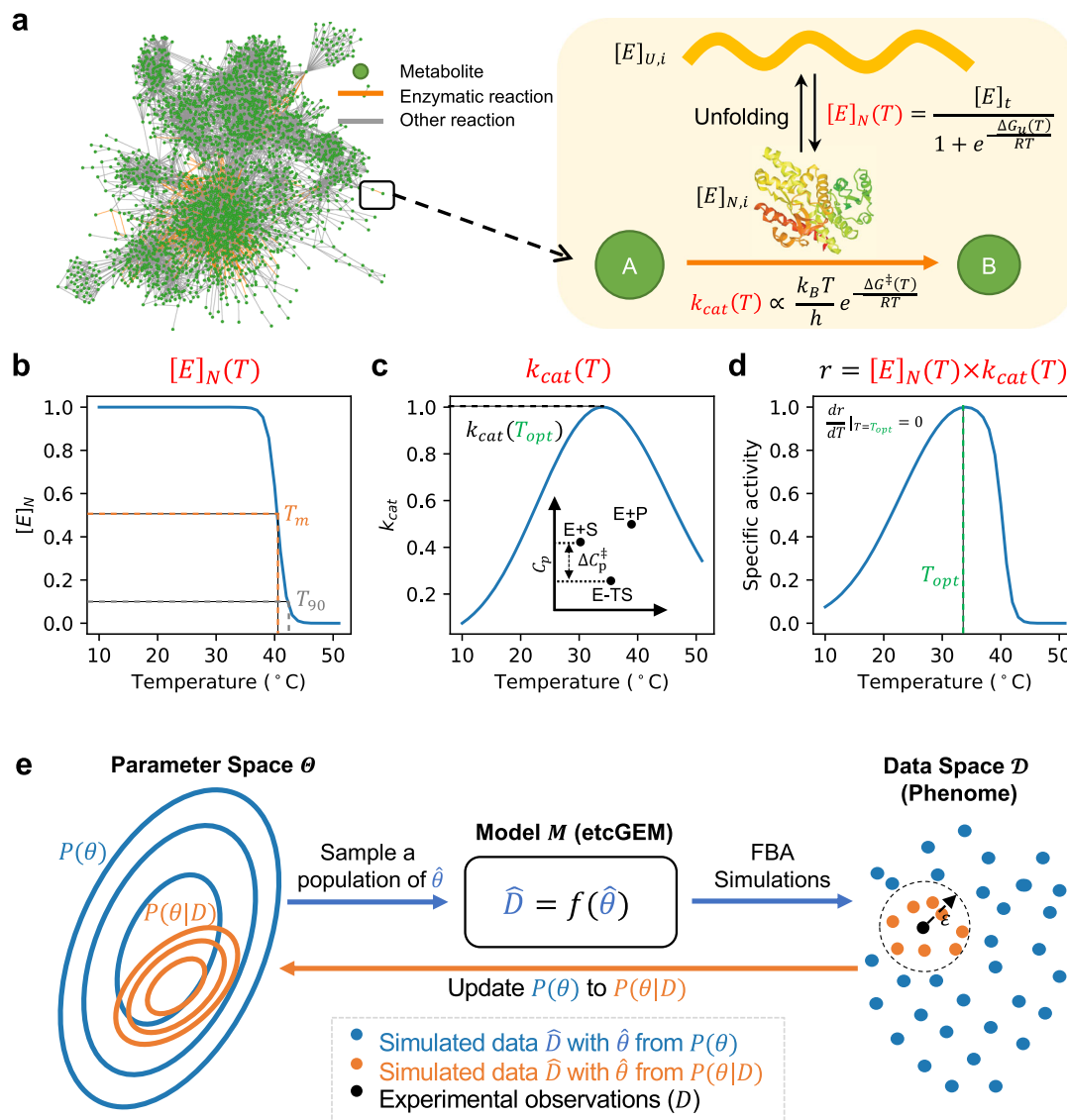
## Results

**Using Bayesian statistical learning to integrate temperature dependence in ecGEMs.** In this study, we developed a novel approach for incorporating temperature dependence into an enzyme-constrained GEM (ecGEM)<sup>16</sup> (Fig. 1) with the resulting model termed enzyme and temperature constrained GEM (etcGEM). The approach combined the following steps: (i) etcGEM construction (Fig. 1a–d), (ii) flux balance analysis (FBA), and (iii) Bayesian statistical learning (Fig. 1e). The ecGEM, which includes, besides the traditional stoichiometric matrix, also enzyme abundances and activities, provided an excellent template to directly integrate the enzyme temperature effects. Firstly, for a given reaction, the flux cannot exceed the capability of the enzyme, which is defined as the product of the functional enzyme concentration  $[E]_N$  and its  $k_{cat}$ . Secondly, the total amount of enzymes that the cell can afford is also limited<sup>27</sup>. Inclusion of temperature constraints into ecGEM was thus achieved by making  $[E]_N$  and  $k_{cat}$  temperature-dependent, and by incorporating the additional cost of enzymes in the denatured state (Fig. 1a, “Methods”). Three thermal parameters were required for each enzyme in the resulting etcGEM, including (i) the melting temperature  $T_m$  (Fig. 1b), (ii) the heat capacity change  $\Delta C_p^\ddagger$  (Fig. 1c), and (iii) the optimal temperature  $T_{opt}$  (Fig. 1d, “Methods”). Moreover, to capture the temperature effects on the energy cost of non-growth associated maintenance (NGAM), a temperature-dependent NGAM expression term was estimated from experimental data and included in the model.

To resolve the challenges arising from the uncertainties in the parameter values, we used Bayesian statistical learning<sup>26</sup>, which is a probabilistic framework that has been successfully applied for quantifying and reducing uncertainties in various fields, including deep learning<sup>28</sup>, ordinary differential equations<sup>29</sup>, and biochemical kinetic models<sup>30</sup>. The approach uses experimental observations ( $D$ ) to update *Prior* distributions ( $P(\theta)$ ) of model parameters to *Posterior* ones ( $P(\theta|D)$ ) (Fig. 1e). We refer to the model equipped with  $\theta$  sampled from  $P(\theta)$  or  $P(\theta|D)$  as a *Prior* or *Posterior* etcGEM, respectively. The resulting *Posterior* etcGEMs provided a more reliable platform to study the thermal dependence of cell metabolism, with an inherent benefit that the uncertainty in the interpretation and prediction from the improved *Posterior* etcGEMs could also be quantified.

## Bayesian modeling improves etcGEM performance by reducing parameter statistical uncertainties.

We next applied the above approach to model the temperature dependence of yeast metabolism (Fig. 1). This was done by incorporating temperature effects into the ecYeast7.6<sup>16</sup> model and the resulting model was termed etcYeast7.6. Enzyme  $T_m$  and  $T_{opt}$  parameters were either collected from literature or predicted by machine-learning models (“Methods”). The heat capacity change  $\Delta C_p^\ddagger$  was estimated as  $-6.3$  kJ/mol/K by fitting the macromolecular rate theory to the yeast specific growth rate at various temperatures<sup>31</sup> and then applied for all enzymes. As a result, the etcYeast7.6 model was obtained with an expansion of 2292 temperature-associated parameters for a total of 764 metabolic enzymes (Fig. 1a). The temperature dependence of NGAM was inferred from experimental data (“Methods”, Supplementary Fig. 1). To assess the quality of the models, three datasets were used that included (i) the maximal specific growth rate in aerobic batch cultivations (number of temperature points  $n = 8$ )<sup>4</sup>, (ii) anaerobic batch



**Fig. 1** Using Bayesian statistical learning to integrate temperature dependence in enzyme-constrained GEMs. **a** An illustration of the temperature effects on enzyme-catalyzed reactions and their integration into an etcGEM (see detailed description and equations in “Methods” section). The metabolic network ecYeast7.6<sup>16</sup> is shown. **b** A two-state denaturation model<sup>20,21,58</sup> was used to describe the temperature-dependent unfolding process.  $[E]_N$  is the concentration of the enzyme in the native state;  $T_{opt}$  is the optimal temperature at which the specific activity is maximized;  $T_m$  and  $T_{90}$  are temperatures at which there is a 50 and 90% probability that an enzyme is in the denatured state, respectively. **c** Macromolecular rate theory<sup>31,33</sup> describing the temperature dependence of enzyme turnover number  $k_{cat}$ . Inset shows the heat capacity difference between ground state (E + S) and transition state (E – TS), adapted from Hobbs et al.<sup>31</sup>. **d** Temperature dependence of enzyme-specific activity  $r$ , which is a product of **(b)** and **(c)**. **e** Overview the Bayesian statistical learning approach, where the problem can be formulated as given a generative model ( $M$ ) (enzyme and temperature constrained genome-scale metabolic model, etcGEM in this study) corresponding to a set of parameters  $\theta$  and a set of measurements  $D$  (phenome data), Bayes’ theorem provides a direct way of updating the Prior distribution of parameters  $P(\theta)$  to a Posterior distribution  $P(\theta|D)$ :  $P(\theta|D) = \frac{P(D|\theta) \times P(\theta)}{P(D)}$ .  $P(\theta|D)$  is thereby a less uncertain description of the real  $\theta$ . Since  $P(D|\theta)$  is, in most applications, computationally expensive or even infeasible to obtain, a sequential Monte Carlo based approximate Bayesian computation (SMC-ABC) approach was implemented (“Methods”) to sample a list of parameter sets from the Posterior.

cultivations ( $n = 8$ )<sup>5</sup>, and (iii) fluxes of carbon dioxide ( $\text{CO}_2$ ), ethanol, and glucose in chemostat cultivations ( $n = 6$ )<sup>23</sup> (Supplementary Fig. 2, “Methods”).

We observed that ecYeast7.6 predictions made using the initial parameter values could only accurately reproduce the experimental observations when the temperature was lower than 30  $^{\circ}\text{C}$ , while the method failed at temperatures above 30  $^{\circ}\text{C}$  (Supplementary Fig. 2). This was assumed to occur since at lower temperatures, the temperature dependence of enzyme  $k_{cat}$  values is the major determining factor, while at higher temperatures there are additional factors, such as the protein denaturation and increased energy for maintenance, that affect the cell growth<sup>5</sup>

(Fig. 1b, c). Particularly, the metabolic shift (Supplementary Fig. 2c) happens within about 2 $^{\circ}$  (36–38  $^{\circ}\text{C}$ ) and accurate prediction of this metabolic flux shift may require more precise enzyme parameter values. Moreover, we measured a high level of uncertainty associated with the initial parameter values, as the average standard deviation was estimated as 3.4 and 5.9  $^{\circ}\text{C}$  for enzymes with experimentally measured  $T_m$  and those without experimentally measured  $T_m$ , respectively, and increased up to 13  $^{\circ}\text{C}$  with the  $T_{opt}$  values predicted by machine learning (“Methods”). Another potential source of error was due to assuming the same  $\Delta C_p^\ddagger$  values for all enzymes.

We, therefore, applied the Bayesian statistical learning approach. First, a threefold cross-validation using the above three datasets, showed that there was both overlapping and orthogonal information among the datasets (Supplementary Fig. 3), suggesting that all three datasets should be used for updating the model parameters. Next, the three datasets were combined and split into training (50%) and test (50%) datasets based on the temperature points in each dataset (Supplementary Fig. 4). The training dataset was used to update the *Prior*, which was then tested on the test dataset after each iteration (Supplementary Fig. 4:  $R_{\text{test}}^2$  increased proportionally with the  $R_{\text{train}}^2$ ). After ~80 iterations, the *Posterior* models achieved a median  $R_{\text{train}}^2$  score of 0.90 (5–95% percentile range: [0.89–0.93]) and a median  $R_{\text{test}}^2$  score of 0.84 (5–95% percentile range: [0.71–0.91]), demonstrating the high generalizability of the *Posterior* models obtained from the SMC-ABC approach. Finally, we used all three datasets to update the *Prior* to *Posterior* and sampled 100 *Posterior* etcGEMs, where each model achieved an average  $R^2$  higher than 0.9 on all three datasets (Supplementary Fig. 5) and could therefore accurately describe the observed measurements (Fig. 2a–c and Supplementary Fig. 6). The increased performance on all three datasets clearly demonstrated the need to update the parameter *Prior* distribution to a *Posterior* one.

We next explored which parameters had been updated in the Bayesian approach. The principal component analysis applied to the 21504 parameter sets generated in the approach showed a clear trend of how the *Prior* distributions were gradually updated to distinct *Posterior* distributions, despite the first two components explaining less than 2% of the total data variance (Fig. 2d). Further comparison between *Prior* and *Posterior* distributions revealed that in all three parameter categories, a reduced variance in the updated parameters was more likely than a change in mean values (Fig. 2e, protein-wise comparison shown in Supplementary Fig. 7). Particularly for enzyme  $T_{\text{opt}}$ s, a significant (Šidák adj. one-tailed  $F$ -test  $p$  value < 0.01) reduction in variance was observed with 59% (449/764), whereas a significant (Šidák adj. Welch's  $t$  test  $p$  value < 0.01) change in the mean value was found with merely 26% (200/764). Importantly, we observed that the approach tended to change the enzyme  $T_{\text{opt}}$  rather than its  $T_m$  and  $\Delta C_p^\ddagger$  parameters (Fig. 2e). In addition, a machine learning approach (“Methods”) further revealed that, out of all three parameter types, the largest contribution to the improved *Posterior* etcGEM performance during the Bayesian approach was from enzyme  $T_{\text{opt}}$ s (Fig. 2f). A note, even though the uncertainties of most parameters were reduced (Fig. 2e), there were still big uncertainties in the *Posterior* models (e.g.,  $T_{\text{opt}}$ , 10.9 vs. 7.1 °C;  $T_m$ , 4.9 vs. 4.0 °C;  $\Delta C_p^\ddagger$ , 2.0 vs. 1.8 kJ/mol/K, average standard variance comparison between *Prior* and *Posterior*, Supplementary Fig. 7).

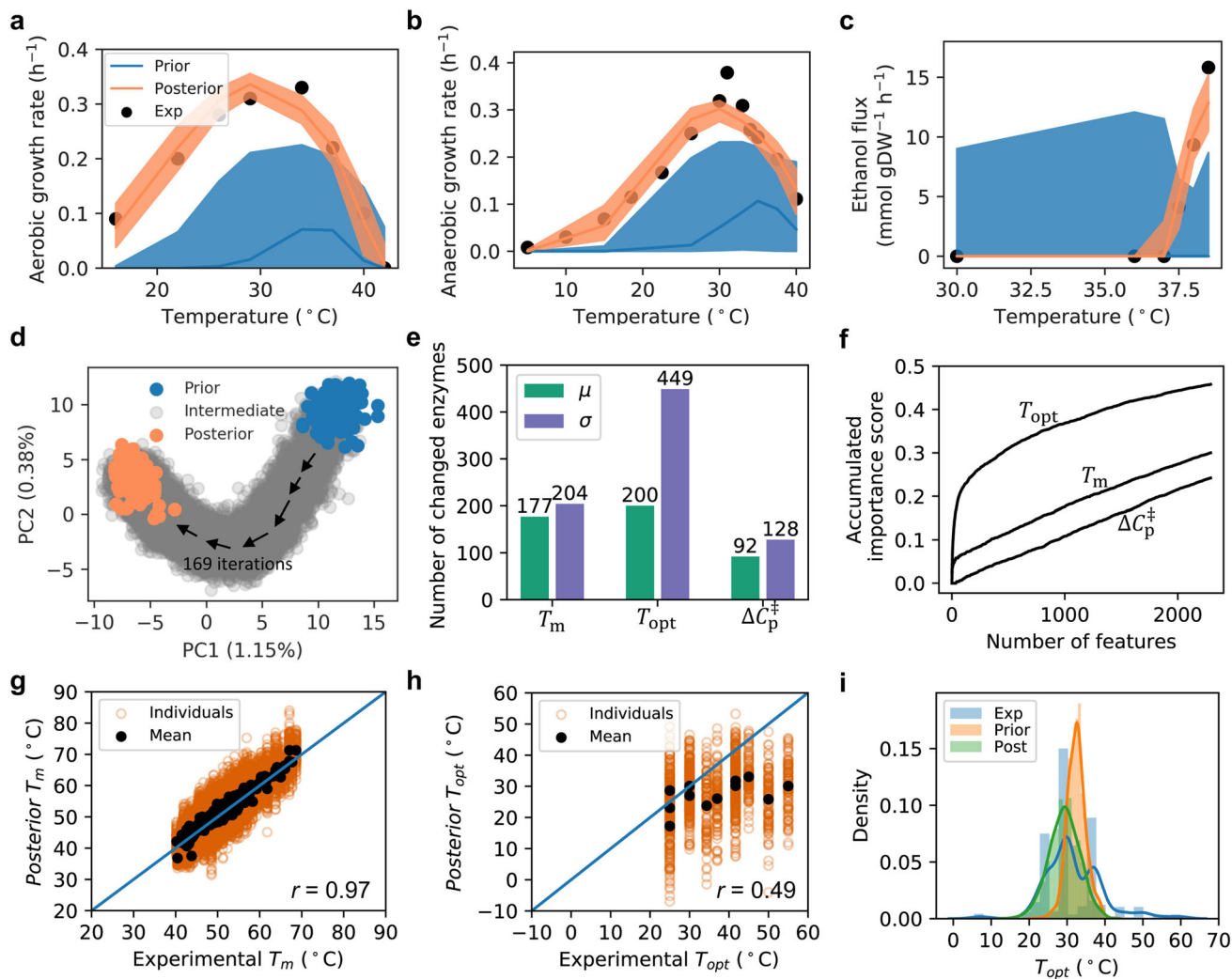
To evaluate the *Posterior* parameter sets, the parity plot comparing the  $T_m$  in the *Posterior* models and experimental values used in the *Prior* showed that the *Posterior* mean values were strongly correlated with the experimental ones (Pearson's  $r = 0.97$ ,  $p$  value <  $1e-32$ ) (Fig. 2g). This indicated that the parameter values returned by the SMC-ABC approach, and with which the model fit the experiment data well, were not very different from the experimentally measured values. We also compared the  $T_{\text{opt}}$  values in the *Posterior* models to the experimental data collected from BRENDA, where 14 enzymes with known  $T_{\text{opt}}$  in BRENDA could be mapped to the etcYeast model based on their UniProt IDs. A weak correlation (Pearson's  $r = 0.49$ ,  $p$  value = 0.075) was found between the *Posterior* mean values and the experimental data (Fig. 2h). Further comparison of the distribution of *Posterior* mean values of all enzymes and 662 enzymes  $T_{\text{opt}}$  records for *S. cerevisiae* in BRENDA showed that the *Posterior*  $T_{\text{opt}}$  values were more similarly distributed to the

experimental  $T_{\text{opt}}$  values than were the original *Prior* values, showing that the SMC-ABC approach indeed improved the estimation of  $T_{\text{opt}}$  values (Fig. 2i).

**The yeast growth rate is explained by temperature effects on its enzymes.** With the *Posterior* etcGEMs capable of describing various experimental observations (Fig. 2a–c), we analyzed how the temperature effects on each of the three processes—NGAM,  $k_{\text{cat}}$  and the protein denaturation process—contribute to whole-cell growth (Fig. 3a). We observed that, at temperatures below 29 °C, the temperature-dependent  $k_{\text{cat}}$  was the only factor that affected the cell growth rate. In the range between 29 and 35 °C, both  $k_{\text{cat}}$  and NGAM determined the growth rate. The contribution of enzyme denaturation to the temperature dependence of cell growth, however, was observed only at temperatures higher than 35 °C, with the denaturing effect becoming the dominant effect at ~40 °C and lack of cell growth at 42 °C. Therefore, in contrast to previous reports indicating that an over tenfold increase in NGAM cost with the temperature change from 30 to 33 °C was the major limiting factor to cell growth<sup>5,32</sup>, our modeling approach showed that the increased NGAM has a merely moderate effect on growth rate (Fig. 3a).

Interestingly, the temperature dependence of enzyme  $k_{\text{cat}}$  alone could explain the temperature dependence of cell growth below 35 °C, including the decline in cell growth right after the optimal growth point defined by OGT. According to the macromolecular rate theory<sup>31,33</sup>,  $k_{\text{cat}}$  degeneration at temperatures above the optimal point can be attributed to the negative values of  $\Delta C_p^\ddagger$  for enzyme catalysis. This can explain the negative curvature of enzyme-specific activities in the absence of the denaturation process<sup>31,33,34</sup>. Given that experimentally measured enzyme melting temperatures ( $T_m$ ) are on average 20 °C higher than enzyme  $T_{\text{opt}}$ s collected from BRENDA<sup>35</sup> (Fig. 3b), protein denaturation alone seems to be insufficient to explain the thermal mechanism underlying enzyme  $T_{\text{opt}}$ s. In addition, all posterior  $T_{\text{opt}}$ s showed a similar distribution as experimental  $T_{\text{opt}}$ s, even though the etcGEM had never seen those experimental  $T_{\text{opt}}$ s (Fig. 2i), which supported our use of the macromolecular rate theory in the model. In the *Posterior* models, the degeneration of enzyme-specific activities clearly depended on the  $k_{\text{cat}}$  degeneration instead of protein denaturation (Fig. 3c). This indicates that  $k_{\text{cat}}$  degeneration, in addition to protein denaturation, plays an important role in the temperature dependence of yeast cell growth.

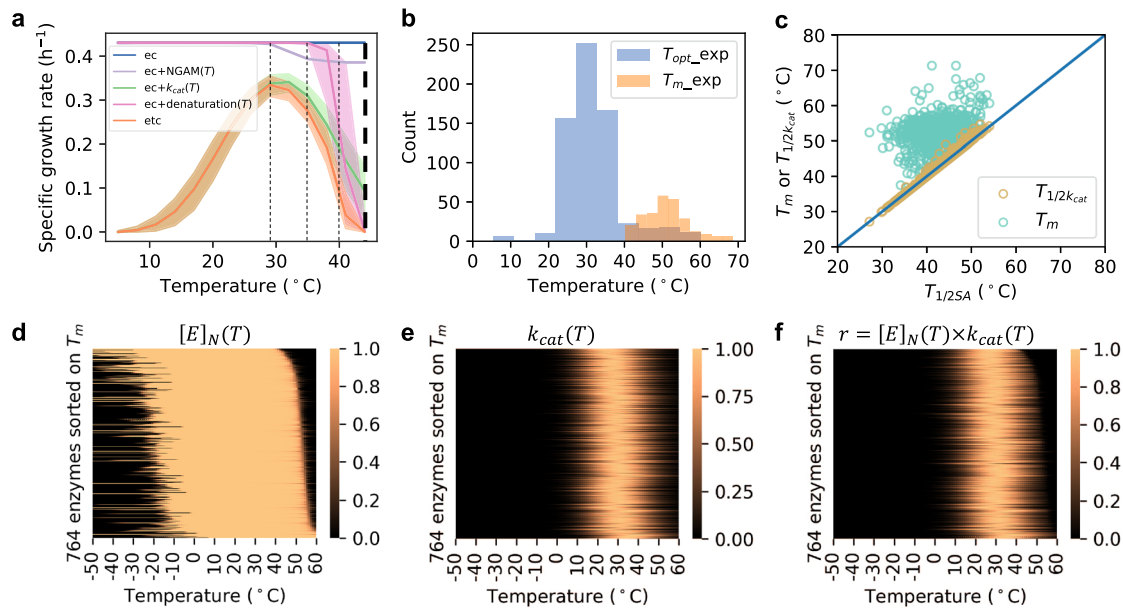
We further observed that, even though the model contained only 764 enzymes from a total of ~6700 proteins<sup>36</sup>, protein denaturation alone could still explain the termination of cell growth at 42 °C (Fig. 3a). However, in the *Posterior* etcGEMs, only 9 enzymes (1%) with a mean melting temperature below 42 °C were present (ERG1, ATP1, ALA1, KRS1, SER1, HEM1, PDB1, ADH1, and TRP3) (Supplementary Fig. 8), of which three (ATP1, HEM1, and PDB1) are located in the mitochondria<sup>37</sup>. The other enzymes remained in the native state even at temperatures several degrees higher than 42 °C (Fig. 3d), though they were enzymatically active only in the temperature window of cell growth between 10 and 42 °C (Fig. 3f), due to the low  $k_{\text{cat}}$  values beyond this temperature range (Fig. 3e). Nearly all yeast enzymes were found to be cold-denatured only at temperatures lower than 0 °C, which is beyond the temperature window for cell growth (Fig. 3d), although there have been studies showing that some yeast proteins have a cold-induced denaturation above 0 °C<sup>38</sup>. This may be due to the fact that the experimental data used to improve the model performance were mainly obtained in the supra-optimal temperature range, with few data available from the suboptimal range.



**Fig. 2 Bayesian modeling improves etcGEM performance by reducing parameter statistical uncertainties.** **a, b** Simulated **(a)** aerobic and **(b)** anaerobic growth rates in batch cultivations at various temperatures with *Prior* and *Posterior* etcGEMs. **c** Simulated ethanol secretion flux in a chemostat at various temperatures. In **a–c**, lines indicate median values and shaded areas indicate regions between the 5th and 95th percentiles ( $n = 128$  for *Prior* and  $n = 100$  for *Posterior* models). **d** Principal component analysis (PCA) 21,504 parameter sets ( $\hat{\theta}$ ) sampled in the Bayesian approach. Each parameter in the set  $\hat{\theta}$  was standardized by subtracting the mean and then be divided by the standard deviation before PCA.  $\hat{\theta}$  of 128 *Prior* and 100 *Posterior* etcGEMs are highlighted in blue and orange, respectively. All other  $\hat{\theta}$  were termed as “intermediate” and marked in gray. **e** The number of enzymes, out of all 764, with a significantly changed mean (Šidák adj. Welch’s  $t$  test  $p$  value  $< 0.01$ , two-sided) and variance (Šidák adj. one-tailed  $F$ -test  $p$  value  $< 0.01$ ) in  $T_m$ ,  $T_{opt}$ , and  $\Delta C_p^\ddagger$  between *Prior* and *Posterior*. Parameters from 128 *Prior* and 100 *Posterior* etcGEMs were used for statistical tests. **f** A random forest model was used to score the importance of all 2292 parameters during the Bayesian approach (“Methods”). The plot shows the accumulated importance score for each of the three-parameter categories. **g** Parity plot comparing the experimental  $T_m$  values and ones in the 100 *Posterior* models. **h** Parity plot comparing the experimental  $T_{opt}$  values from BRENDA and ones in the 100 *Posterior* models. In **g, h**,  $r$  denotes the Pearson’s correlation coefficient between experimental values and *Posterior* mean values. **i** Comparison among distributions of the mean of *Prior*  $T_{opt}$  s, mean of *Posterior*  $T_{opt}$  s (Post) and experimental  $T_{opt}$  s (Exp) from BRENDA. The bimodal distribution of  $T_{opt}$  values is due to the common values in BRENDA occurring at room temperature 25 °C (111 times), the optimal growth temperature of yeast 30 °C (244 times), and 37 °C (142 times). Source data are provided as a Source Data file.

**Metabolic shifts are explained by temperature-induced proteome constraints.** Published reports show that at temperatures above 37 °C in chemostat cultures with a dilution rate of 0.1 h<sup>-1</sup>, yeast shifts its metabolism from a completely respiratory one to a partly fermentative one, which is also accompanied by a large increase in glycolytic flux<sup>23</sup>. Since our updated *Posterior* etcGEMs are able to simulate this metabolic shift (Fig. 2c and Supplementary Fig. 6), we used them to further explore the mechanisms behind the observed process. We observed that the shift occurs due to a proteome constraint, meaning that the total protein level in the cell reaches an upper bound (Fig. 4a). The proteome

constraint occurs due to the decrease in enzyme specific activities with increasing temperature (Fig. 3f) and since the maximal protein amount in the cell is limited<sup>27</sup>. As a result, the cell has to synthesize more enzymes to maintain cell growth at the given growth rate (Fig. 4a) until the enzyme amount hits the upper bound. Doubling this protein constraint can delay the shift point from 37 to 38 °C (Fig. 4b and Supplementary Fig. 9). These findings are consistent with earlier studies showing that the activation of the Crabtree effect in chemostat cultures at 30 °C is due to a proteome constraint<sup>16,39</sup>. When the temperature increases above 36 °C, ATP production by glycolysis is



**Fig. 3** Yeast growth rate is explained by temperature effects on its enzymes. **a** Illustration of how the temperature dependence of different processes combines to affect the growth rate. Fig. legend: ec—predictions with the enzyme constrained model; ec+NGAM( $T$ )—incorporates the temperature effects on nongrowth associated maintenance into the ec model (Supplementary Fig. 1); ec+k<sub>cat</sub>( $T$ )—incorporates the temperature effects on enzyme  $k_{cat}$  values into the ec model; ec+denaturation( $T$ )—incorporates the temperature effects on enzyme denaturation into the ec model; etc—enzyme and temperature constrained model that includes the temperature effects on NGAM,  $k_{cat}$  and enzyme denaturation into ec model. The growth rate at each temperature point was simulated with all 100 *Posterior* etcGEMs. Lines indicate median values and shaded areas indicate regions between the 5th and 95th percentiles ( $n = 100$ ). **b** Comparison between distributions of experimentally measured enzyme  $T_{opt}$  values ( $n = 662$ ) from BRENDA<sup>35</sup> and enzyme  $T_m$  values ( $n = 265$ ) from Leuenerger et al.<sup>7</sup> in *S. cerevisiae*. **c** Comparison between  $T_{1/2SA}$ , the temperature at which the specific activity is 50% of its maximum, and  $T_{1/2k_{cat}}$ , the temperature at which the  $k_{cat}$  value is 50% of its maximum, and  $T_m$  in the *Posterior* models. **d** Probability of 764 enzymes in the native state. From top to bottom, the enzymes showed increased  $s$ . Each pixel represents one probability value of an enzyme at a specific temperature. **e** Normalized  $k_{cat}$  values of 764 enzymes at different temperatures. Each pixel represents one normalized value of an enzyme at a specific temperature. **f** Normalized specific activities of 764 enzymes at different temperatures. The values in (**f**) are products of (**d**) and (**e**). In **d–f**, the same ordering of enzymes is shown. Source data are provided as a Source Data file.

dramatically increased, while ATP production by the mitochondria decreases (Fig. 4a). Even though the respiratory pathway produces more ATP per glucose amount, the fermentative pathway produces more ATP per protein mass and therefore becomes more energetically efficient when the cell reaches a proteome constraint<sup>39</sup>. In addition, three key mitochondrial enzymes (ATP1, HEM1, and PDB1) (Supplementary Fig. 8) were found to be unstable, which makes the respiratory pathway even more resource-inefficient for ATP production. When increasing the  $T_m$  of these three enzymes to 50 °C, the same shift could still be observed while the ATP production shift from mitochondria to glycolysis was reduced (Fig. 4c and Supplementary Fig. 9).

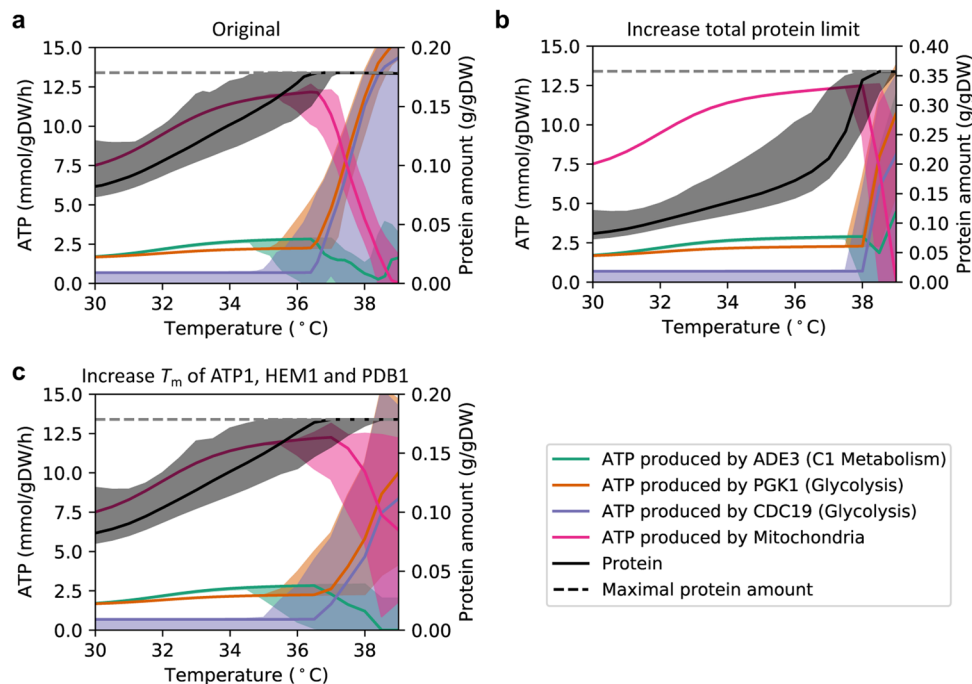
**etcGEM uncovers growth rate-limiting enzymes.** To investigate which enzymes limit the cell growth at superoptimal temperatures, the flux sensitivity coefficient of each enzyme was calculated (“Methods”). Among all the enzymes in the model, the squalene epoxidase ERG1 displayed order of magnitude higher median flux sensitivity coefficient than other enzymes, indicating that it is the most flux-controlling enzyme at 40 °C (Fig. 5a) and above (Supplementary Fig. 10). Furthermore, the removal of the temperature constraint on ERG1 increased the simulated specific growth rate from 0.09 to 0.14 h<sup>-1</sup> (Fig. 5b). We, therefore, evaluated the impact of replacing the wild-type *ERG1* gene with *ERG1* from the thermotolerant yeast *Kluyveromyces marxianus* (KmERG1, “Methods”). At first, at the lethal temperature of 42 °C, only a small improvement in growth rate (from 0.01 to 0.06 h<sup>-1</sup>) was predicted and no significant growth difference was detected between the wildtype and the strain with kmERG1

(Supplementary Fig. 11). However, already after 2 passages of adaptation at 40 °C, the strain with KmERG1 indeed showed significantly better growth than the wild type (Fig. 5c).

The reduced growth rate at 42 °C is likely caused by an impaired function of several different enzymes, and rescuing a single enzyme is insufficient to improve the growth rate. Therefore, in order to characterize the set of growth rate-limiting enzymes at 42 °C, we gradually removed the temperature constraints on enzymes (set  $k_{cat}$  and denaturation temperature independent) in the order of decreasing flux sensitivity coefficient values in each of the *Posterior* etcGEMs. Interestingly, in the case of recovering the cell growth rate to 0.2 h<sup>-1</sup>, we found an agreement among all *Posterior* etcGEMs that ten enzymes are required to be fully functional at 42 °C (Fig. 5d). Since each model predicted a different subset of such enzymes, an ensemble approach was used to count the number of models (votes) in which an enzyme is predicted to be one of ten such enzymes (Fig. 5e). In total, 82 enzymes were predicted by at least 1 *Posterior* etcGEM, and only 24 (out of 82) enzymes were each predicted by more than 10% of the *Posterior* etcGEMs (Fig. 5e, inset). Among these 24 enzymes, 12 enzymes were engaged with Glycolysis and 3 enzymes were involved in sterol biosynthesis: ERG1, and HMG1,2 catalyzing the flux-controlling steps in sterol biosynthesis<sup>40</sup>. The remaining enzymes were mainly involved in DNA or protein synthesis related pathways.

## Discussion

Here, we present a Bayesian genome-scale modeling approach to resolve the temperature dependence of cellular metabolism. Using



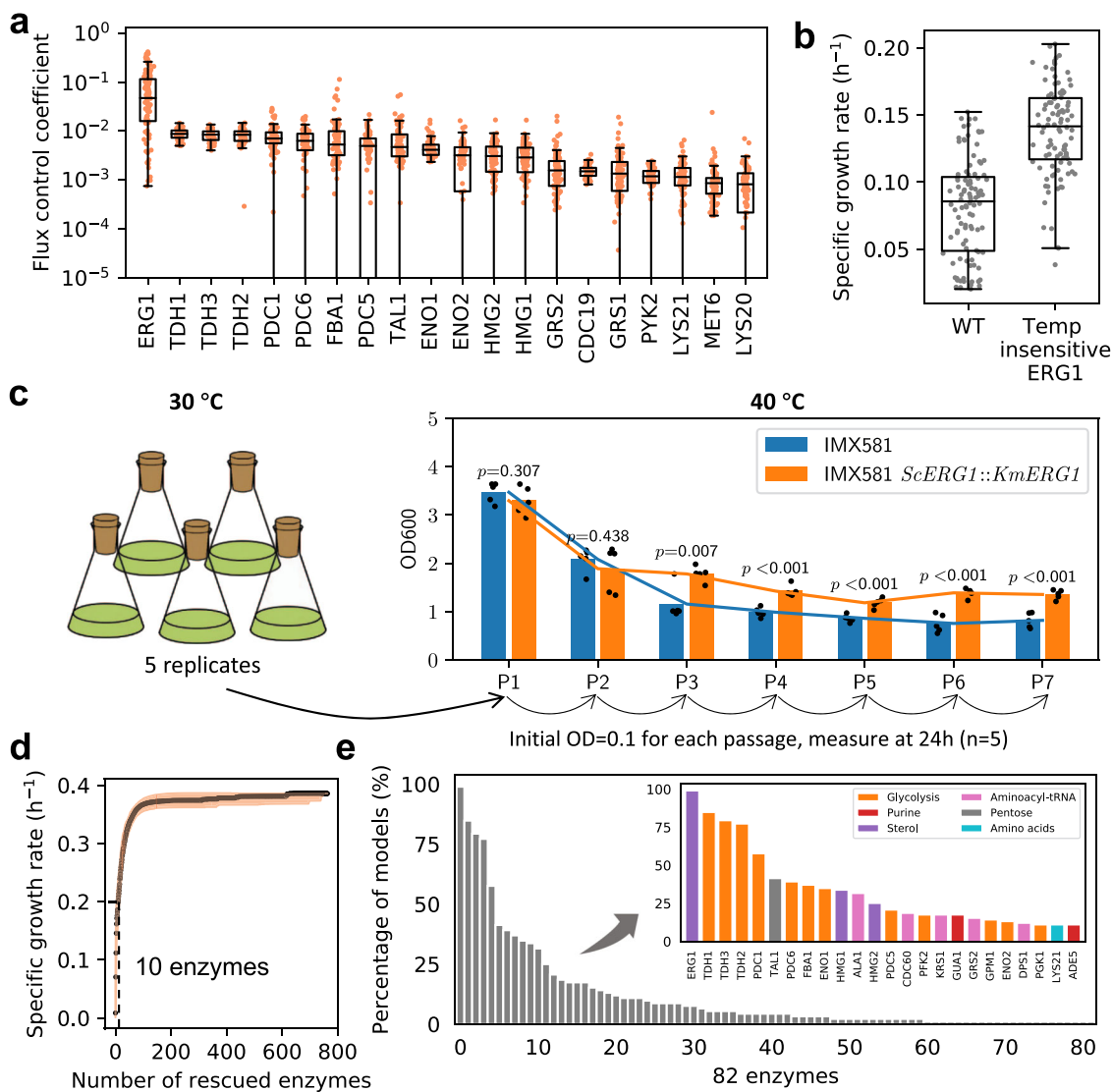
**Fig. 4 Metabolic shifts are explained by temperature-induced proteome constraints.** The ATP production in the cytoplasm and the total protein amount required at different temperatures were simulated using *Posterior* etcGEMs with chemostat culture settings with a dilution rate of  $0.1 \text{ h}^{-1}$  (“Methods”). Simulation results from **a** the original *Posterior* models, **b** the original *Posterior* models with doubled total protein constraint, **c** the original *Posterior* models with increased  $T_m$  of ATP1, HEM1, and PDB1 (to  $50^\circ\text{C}$ ) and are shown. Lines indicate median values and shaded areas indicate the region between the 5th and 95th percentile ( $n = 100$ ). Source data are provided as a Source Data file.

an ecGEM<sup>16</sup> as a template, we modeled the temperature effects on each individual enzyme by including temperature-dependent terms for the independent processes of denaturation as well as catalysis (Fig. 1a). Due to the high level of uncertainty and low accuracy associated with the initial thermal parameter values (Supplementary Fig. 7), which were a result of experimentally measured noise or variability arising from machine learning or theoretical predictions, the model predictions initially could not correctly recapitulate experimental observations (Fig. 2a–c and Supplementary Fig. 2). We, therefore, used Bayesian statistical learning that enabled updating our *Prior* guess of the highly uncertain thermal parameters to a more accurate *Posterior* estimation of these parameters according to observed phenotypic data (Fig. 1e). The resulting *Posterior* etcGEMs accurately describe the experimental observations (Fig. 2a–c) and thus provide a more reliable platform to study the thermal dependence of yeast metabolism.

Previous studies modeling the temperature dependence of enzyme activities have relied mainly on protein denaturation and the Arrhenius equation, where protein denaturation explained the negative curvature for the temperature dependence of enzyme activity<sup>20,21</sup>. However, with the increasing amount of evidence showing that protein denaturation alone is insufficient to explain the decrease in enzyme specific activity above  $T_{\text{opt}}$ , macro-molecular molecular rate theory<sup>31,34</sup> has become a promising alternative. It was successfully applied to many enzymes<sup>31,33,34</sup>, including its use in explaining the evolution of enzyme catalysis<sup>34</sup>. According to the theory, a negative heat-capacity change ( $\Delta C_p^\ddagger$ ) exists between the transition state and the ground state in the enzyme catalytic process (Fig. 1c), which leads to a negative curvature for the temperature dependence of enzyme activity in the absence of denaturation<sup>31</sup>. We found that with this theory, the temperature dependence of  $k_{\text{cat}}$  acts as a major contributor to the cell growth rate at all temperatures, which can especially

explain the decline in cell growth rate right after the OGT (Fig. 3a). Yeast enzymes only maintain high  $k_{\text{cat}}$ s in the temperature window of cell growth (Fig. 3e), which means that the metabolism becomes inefficient at superoptimal temperatures due to the general decrease in enzyme turnover without denaturations (Fig. 3d–f).

Using the Bayesian genome-scale modeling approach to quantitatively depict the temperature effects on yeast metabolism led to insights into the long-standing discussion on the roles of different cellular factors in cellular fitness under heat stresses<sup>4,5,7,23,41</sup>. For instance, protein denaturation has been suspected as one of the main causes of the decline in cell growth beyond the OGT point. However, recent high throughput measurements of melting temperatures ( $T_m$ ) for 707 *S. cerevisiae* proteins revealed a  $T_m$  distribution with a mean value of  $52^\circ\text{C}$  and a minimum of  $40^\circ\text{C}$ <sup>7</sup>, which suggests that protein denaturation alone might not be sufficient to explain the decline of yeast cell growth between  $30^\circ\text{C}$  (OGT) and  $42^\circ\text{C}$  (lethal temperature point). An alternative explanation is provided by the evidence of a significant increase of NGAM observed with yeast cells grown in anaerobic chemostat cultivations at high temperatures ( $33\text{--}40^\circ\text{C}$ ) compared to ones grown at low temperatures ( $5\text{--}31^\circ\text{C}$ )<sup>5</sup>, which suggests an imbalance in cellular energy allocation in the superoptimal temperature range. Quantitative assessment using our modeling approach revealed that impaired cell growth is caused by a combination of decreased  $k_{\text{cat}}$  values, increased NGAM costs, and protein denaturation (Fig. 3). Furthermore, between  $30$  and  $35^\circ\text{C}$ , the combined decrease in  $k_{\text{cat}}$  and increase in NGAM explains the decline in cell growth, whereas, with temperatures above  $35^\circ\text{C}$ , protein denaturation becomes the dominant factor, causing cell death at  $42^\circ\text{C}$ . However, in accordance with published findings that cellular proteomes have a broad distribution of protein stability with only proteins at the tail of the distribution being problematic<sup>42</sup>, using



**Fig. 5 etcGEM uncovers growth rate-limiting enzymes.** **a** Twenty enzymes with the highest flux sensitivity coefficients at 40 °C ( $n = 100$  for each enzyme). **b** Predicted maximal specific growth rate of wild-type yeast and the one without any temperature constraints (fully functional) on ERG1 enzyme at 40 °C ( $n = 100$  for each strain). In **a**, **b**, centerline, median; box limits, upper and lower quartiles; whiskers, 1.5 $\times$  interquartile range; Each dot represents the prediction from one *Posterior* etcGEMs. **c** The effect of KmERG1 expression on thermotolerance in *S. cerevisiae*. The strains were passaged for 7 times (24 h for each passage) to obtain stable growth at 40 °C. Optical densities (600 nm) are shown at 24 h. Data are represented as mean values of five replicates in the bar chart. Dots represent the values of five replicates.  $p$  Values denote Welch's  $t$  test (two-sided). The exact  $p$  values for P1-P7 are 0.307, 0.438, 0.0066, 0.00020, 0.00046, 0.00015, 0.00018, respectively. **d** Simulated maximum specific growth rate by removing the temperature constraints of most rate-limiting enzymes at each step in each *Posterior* etcGEM at 42 °C. Lines indicate median values and shaded areas indicate the region between the 5th and 95th percentiles ( $n = 100$ ). **e** The percentage of *Posterior* etcGEMs ( $n = 100$ ) predicts an enzyme to be in the minimal enzyme set required to be fully functional at 42 °C in order to achieve a maximal specific growth rate of 0.2  $\text{h}^{-1}$ . Inset shows the names and pathways of genes predicted by more than *Posterior* etcGEMs 10% they are involved in. Source data are provided as a Source Data file.

our approach we identified only  $\sim 1\%$  unstable enzymes denatured at the lethal point ( $T_m$  lower than 42 °C, Fig. 3d).

We identified two interesting metabolic pathways involved in yeast thermotolerance: sterol metabolism and mitochondrial energy metabolism. With sterol metabolism (Fig. 5d), it is known that high sterol levels help yeast cells survive under heat stress<sup>43</sup> and changes of the sterol composition of the yeast membrane from ergosterol to fecosterol<sup>44</sup> can significantly increase yeast thermotolerance. However, yeast was found to downregulate its whole ergosterol biosynthesis at both transcription and translation levels when increasing the temperature from 30 to 36 °C (Supplementary Fig. 12). Our modeling approach identified three problematic enzymes (Fig. 5d: HMG1,2 and ERG1) in the sterol

metabolism, which are also flux-controlling enzymes in the sterol biosynthesis pathway<sup>45</sup>. We experimentally confirmed that the replacement of ERG1 with its ortholog in the thermotolerant yeast *K. marxianus* can significantly improve the cell growth at 40 °C (Fig. 5c). Further simulations by downregulating the genes involved in ergosterol pathways (except ERG1) showed that when these enzymes were down-regulated by a percentage higher than  $\sim 40\%$ , a temperature-insensitive ERG1 can rescue the cell growth, while it failed when they were downregulated to less than 40% (Supplementary Fig. 13). We thereby hypothesize that, since those three enzymes are problematic at superoptimal temperatures, there is no need for the cell to maintain high expression and translation levels of other enzymes in the same pathway. Instead,



it has to downregulate its whole ergosterol biosynthesis to save resources and increase fitness.

With mitochondria, previous studies have indicated that the mitochondrial genome plays an important role in yeast thermal adaptation<sup>46–48</sup>. We found that out of the nine unstable enzymes identified with the *Posterior* etcGEMs (with a  $T_m$  lower than 42 °C, Supplementary Fig. 8), three (ATP1, HEM1, and PDB1) belonged to the mitochondrial energy metabolism. Simulation of chemostat data (Fig. 4) revealed that at superoptimal temperatures, yeast prefers to produce ATP via the glycolysis metabolism instead of the mitochondrial energy metabolism in the mitochondria. This can be explained by the limited total protein content in the cell and resource-inefficient mitochondrial energy metabolism. A previous study also found that a mitochondrial matrix protein Mge1p, which is essential for mitochondrial functions, loses its functions at temperatures higher than 37 °C due to denaturation and dimer dissociation<sup>49</sup>. Another study found that the mitochondrial inner membrane is severely affected at 38 °C and may impair the function of mitochondria<sup>50</sup>. Furthermore, mitochondria only exist in eukaryotes and almost all of them have evolved to have an OGT below 40 °C<sup>3</sup>. All these findings indicate that mitochondria are not evolved to be functional at very high temperatures. Since mitochondrial energy metabolism is not essential for yeast cell growth, as there are alternative energy pathways (Fig. 4), this also explains why we could not successfully predict mitochondrial enzymes to be engineering targets for the recovery of cell growth at 42 °C (Fig. 5d), despite the existence of three unstable enzymes in the mitochondrial energy metabolism.

As a note, the assumption of two-state denaturation and macromolecular theory to describe the temperature dependence of enzyme activity may be oversimplified for some enzymes. The experimental data used to update the *Prior* guesses of thermal parameters was mainly from the superoptimal temperature range (Fig. 2a–c), whereas there are still big uncertainties in the current *Posterior* (Supplementary Fig. 7). These all indicate that the etcYeast7.6 can still be enhanced in the future by (i) improving the formulation of enzyme temperature-dependence and (ii) more experimental data for model validation and uncertainty reduction.

In conclusion, we demonstrate the usefulness of a Bayesian genome-scale modeling approach for reconciling temperature dependence of yeast metabolism. Describing the link between temperature and cell physiology is of industrial importance, e.g., for finding optimized production of biochemicals<sup>24,51–53</sup>, but also in medicine, e.g., to understand the effects of temperature on human metabolism<sup>54–56</sup>. Furthermore, based on its success here, we foresee that our method can be integrated into genome-scale modeling approaches in general. This approach can also become a staple of GEM modeling in order to resolve uncertainties present in the data, which can be important as GEMs have become a widely used platform for integration of various biological data, such as transcriptomics and proteomics data that are associated with large uncertainties<sup>57</sup>.

## Methods

**The temperature-dependent enzyme-constrained genome-scale metabolic model.** The central concepts of an enzyme constrained model<sup>16</sup> are as follows: (1) the flux through each reaction cannot exceed the capacity of its catalytic enzyme:  $v_i \leq k_{cat,i} \cdot [E]_i$ , where  $[E]_i$  is the concentration of enzyme  $i$ ; (2) the total enzyme amount is constrained by the experimental measurement:  $\sum [E]_i \leq [E]_t$ . Once the temperature-dependent denaturation and  $k_{cat}$  were considered,  $[E]_i$  in the first constraint should be  $[E]_{N,i}$ , which is the concentration of individual active enzymes.  $[E]_i$  in the second constraint should be  $[E]_{t,i} = [E]_{N,i} + [E]_{U,i}$ , which is the total concentration of enzymes in both active and denatured forms (Fig. 1a). In addition, to capture the increased expenditure for maintenance under increased heat stress, a temperature-dependent non-growth-associated ATP maintenance term can be assumed from experimental measurements. In summary, the updated constraints

in etcGEM are

$$\begin{cases} Sv = 0 \\ 0 \leq v_i \leq k_{cat,i}(T) \cdot [E]_{N,i}(T) \\ \sum ([E]_{N,i}(T) + [E]_{U,i}(T)) \leq [E]_t \\ \text{NGAM}(T) = f(T) \end{cases} \quad (1)$$

The effect of temperature on  $k_{cat}$  values can be described with an expanded Arrhenius equation (macromolecular rate theory), by including a nonzero heat-capacity change ( $\Delta C_p^\ddagger$ ) between the transition state and the ground state of the enzyme catalytic process<sup>31,33</sup>:

$$k_{cat}(T) \propto \frac{k_B T}{h} e^{-\frac{\Delta G^\ddagger(T)}{RT}}, \quad (2)$$

in which  $k_B$  is the Boltzmann constant,  $h$  is Planck's constant,  $R$  is the universal gas constant and  $\Delta G^\ddagger(T)$  is the free energy difference between the ground state and the transition state. The latter can be expanded as

$$\Delta G^\ddagger(T) = \Delta H_{T_0}^\ddagger + \Delta C_p^\ddagger(T - T_0) - T \left( \Delta S_{T_0}^\ddagger + \Delta C_p^\ddagger \ln \left( \frac{T}{T_0} \right) \right), \quad (3)$$

where  $\Delta H_{T_0}^\ddagger$ ,  $\Delta S_{T_0}^\ddagger$ , and  $\Delta C_p^\ddagger$  are the differences in enthalpy, entropy, and heat capacity change between the transition and ground states, respectively, and  $T_0$  is the reference temperature. This theory has been successfully applied to study the temperature dependence of enzyme activity<sup>31,33</sup> and evolution<sup>34</sup>.

Since there is not enough detailed information regarding the heat-induced denaturation process of yeast proteins, a simple two-state model denaturation was assumed as in many other studies<sup>20,21,58</sup>. In such a model, a protein molecule could be either in a native state ( $N$ ) or a denatured state ( $U$ ), and an equilibrium state was assumed:  $N \leftrightarrow U$ . Thereby

$$[E]_{N,i} = \frac{1}{1 + e^{-\frac{\Delta G_{u,i}(T)}{RT}}} [E]_{t,i}, \quad (4)$$

in which  $[E]_{t,i} = [E]_{N,i} + [E]_{U,i}$ , where  $[E]_{t,i}$  is the concentration of enzyme  $i$  and  $\Delta G_{u,i}(T)$  is the free energy difference between the denatured state and the native state and can be expressed as

$$\Delta G_{u,i}(T) = \Delta H_{u,i}(T) - T \Delta S_{u,i}(T), \quad (5)$$

where  $\Delta H_{u,i}(T)$  and  $\Delta S_{u,i}(T)$  are the enthalpy and entropy changes between the denatured and native states at temperature  $T$ . It has been found that convergence temperatures  $T_H^*$  (373.5 K) and  $T_S^*$  (385 K) exist for  $\Delta H_{u,i}$  and  $\Delta S_{u,i}$ , respectively<sup>41,59,60</sup>. At such temperatures, the  $\Delta H_{u,i}$  and  $\Delta S_{u,i}$  converge to a common value of  $\Delta H^*$  and  $\Delta S^*$ . Thereby,

$$\Delta G_{u,i}(T) = \Delta H^* + \Delta C_{p,u,i} \left( T - T_H^* \right) - T \Delta S^* - T \Delta C_{p,u,i} \log \left( \frac{T}{T_S^*} \right), \quad (6)$$

in which  $\Delta C_{p,u,i}$  is the difference in heat-capacity change between the denatured and native states.

In summary, the values of  $\Delta G^\ddagger(T)$  and  $\Delta G_{u,i}(T)$  need to be determined in order to model the temperature dependence of enzyme activities, and they can be associated with six unknown parameters:  $\Delta H_{T_0}^\ddagger$ ,  $\Delta S_{T_0}^\ddagger$ , and  $\Delta C_p^\ddagger$  for  $\Delta G^\ddagger(T)$ , and  $\Delta H^*$ ,  $\Delta S^*$ , and  $\Delta C_{p,u,i}$  for  $\Delta G_{u,i}(T)$ .

**Computation of thermal parameters.** Since it is difficult to directly measure those six thermal parameters ( $\Delta H_{T_0}^\ddagger$ ,  $\Delta S_{T_0}^\ddagger$ ,  $\Delta C_p^\ddagger$ ,  $\Delta H^*$ ,  $\Delta S^*$ , and  $\Delta C_{p,u,i}$ ) for each enzyme, indirect measurements have to be used to approximate the larger set of thermal parameters. As there are six free variables in the system, six different equations are required to solve for those parameters.

1. At the protein melting temperature  $T_m$ :
 
$$\Delta G_{u,i}(T_m) = 0. \quad (7)$$
2. At the enzyme optimal temperature  $T_{opt}$ , the enzyme activity is maximized:

$$\frac{dr}{dT} \Big|_{T=T_{opt}} = 0. \quad (8)$$

in which  $r = k_{cat}[E]_{N,i}$

3.  $k_{cat}$  at the enzyme optimal temperature  $T_{opt}$  is known:

$$k_{cat}(T_{opt}) = \frac{k_B T_{opt}}{h} e^{-\frac{\Delta G^\ddagger(T_{opt})}{RT_{opt}}}. \quad (9)$$

4.  $\Delta C_p^\ddagger$  value can be approximate from the temperature dependence of cell growth rate<sup>31</sup>.
5. We found that there is a very strong linear correlation ( $r^2 = 0.998$ , Pearson's correlation) between  $\Delta H^*$  and  $\Delta S^*$  of 116 proteins from Sawle et al.<sup>41</sup> (Supplementary Fig. 14)

$$\Delta H^* = 299.58 \Delta S^* + 20008 \text{ J/mol} \quad (10)$$

6. For some enzymes,  $T_{90}$ , where a 90% possibility exists that an enzyme molecule is in the denatured state, is experimentally measured:

$$\Delta G_u(T_{90}) = -RT_{90} \ln 9. \quad (11)$$

As a result, the six thermal parameters  $\Delta H_{T_0}^\ddagger$ ,  $\Delta S_{T_0}^\ddagger$ ,  $\Delta C_p^\ddagger$ ,  $\Delta H^*$ ,  $\Delta S^*$ , and  $\Delta C_{p,u}$  can be obtained by solving the above equations.

In the case of lacking  $T_{90}$  or failed to obtain a positive  $\Delta C_{p,u}$  protein sequence length was used to estimate  $\Delta H^*$  and  $\Delta S^*$ <sup>41</sup> as below:

$$\Delta H^* = (4.0N + 143) \times 1000. \quad (12)$$

$$\Delta S^* = 13.27N + 448. \quad (13)$$

**Sequential Monte Carlo-based approximate Bayesian computation.** Approximate Bayesian computation<sup>61</sup> was applied to infer parameter sets from *Posterior* distributions. Given an observed dataset  $D$  and a model specified by  $\theta$  sampled from the *Prior* distribution  $P(\theta)$ , if the distance between simulated data  $\hat{D}$  and observed  $D$  is less than a given threshold  $\epsilon$ , then this  $\hat{\theta}$  is accepted as the one sampled from  $P(\rho(D, \hat{D}) < \epsilon)$ .  $P(\rho(D, \hat{D}) < \epsilon)$  is often used to approximate the *Posterior*  $P(\theta|D)$  when  $\epsilon$  is sufficiently small. In the case of high-dimensional parameter space and/or when the  $P(\theta)$  is very different from  $P(\theta|D)$ , the acceptance rate would be very low and thus this approach becomes computationally expensive to generate a population of  $\theta$  from  $P(\rho(D, \hat{D}) < \epsilon)$ . In this work, a sequential Monte Carlo approach was designed (Supplementary Table 1) to generate a population of  $\theta$  sampled from  $P(\rho(D, \hat{D}) < \epsilon)$ . This approach was validated on several toy models with known parameter values and was found to outperform existing other population-based methods when the number of parameters is far greater than the number of samples for fitting (Supplementary Note 1, Supplementary Figs. 16–20).

**Melting temperatures.** Among the 764 enzymes included in eYeast7.6<sup>16</sup>, the  $T_m$  (melting temperature) and  $T_{90}$  (the temperature at which 90% of the protein is in the denatured state) for 266 yeast proteins have been reported previously<sup>7</sup>. For enzymes lacking an experimentally measured  $T_m$ , a melting temperature of 51.9 °C (the average of existing  $T_m$  s of 707 yeast proteins) was assumed. In the original paper<sup>7</sup>, the 95% confidence interval was reported for peptides measured in the experiments and the average standard error was estimated at 3.4 °C. This same value was used as the uncertainty measure for the experimentally determined  $T_m$  s, since the standard error for protein  $T_m$  was not available. The  $T_m$  of the 266 enzymes was then described with a normal distribution  $N(T_{m,i}, 3.4)$ , in which  $T_{m,i}$  is the experimentally measured melting temperature of protein  $i$ . For enzymes that use the mean  $T_m$  of 707 proteins<sup>7</sup> as  $T_m$  estimation, the corresponding uncertainty is described as the standard deviation of the 707  $T_m$  s, equaling 5.9 °C. Thereby, a normal distribution  $N(51.9, 5.9)$  was used.

**Enzyme optimal temperature.**  $T_{opt}$  values of all enzymes in this study were calculated using a previously developed machine-learning method Tome v1.0 (<https://github.com/EngqvistLab/Tome>)<sup>22</sup>, which predicts enzyme  $T_{opt}$  based on primary sequences. This model has a coefficient of determination ( $R^2$  score) of 0.5 on the test dataset. Root-mean-squared error (RMSE) of the prediction was then estimated with

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - f_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (y_i - f_i)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\text{MSE}}{\delta_{DB}^2}. \quad (14)$$

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{(1 - R^2) \delta_{DB}^2} = \sqrt{(1 - 0.5) \times 337} = 13.0 \text{ } ^\circ\text{C}. \quad (15)$$

where  $f_i$  is the predicted value and  $y_i$  is the observed true value of enzyme  $i$ . Then each one of these predicted  $T_{opt}$  s was described with a normal distribution  $N(T_{opt,i}, 13.0)$ .

**Heat capacity change.**  $\Delta C_p^\ddagger$  value was approximated by assuming temperature dependence of yeast cell growth rate as  $-6.3 \text{ kJ/mol/K}$  for all enzymes<sup>31</sup>. Given that  $\Delta C_p^\ddagger$  should be in the general negative for most enzymes<sup>33</sup>, a standard variance of 2.0 was selected from testing a wide range of values because it covers a broad range of  $\Delta C_p^\ddagger$  and with a very low possibility of getting a positive value (Supplementary Fig. 15). A normal distribution of  $N(-6.3, 2.0)$  was subsequently used to describe the  $\Delta C_p^\ddagger$  of all enzymes.

**Non-growth associated ATP maintenance.** To capture the increased expenditure for maintenance under increased heat stress, an empirical equation (Supplementary Fig. 1) was constructed to estimate the NGAM at different temperatures:

$$\text{NGAM}(T) = 0.740 + \frac{5.893}{1 + e^{31.920 - (T - 273.15)}} + 6.12 \times 10^{-6} \times (T - 273.15 - 16.72)^4, \quad (16)$$

based on the experimental data<sup>5</sup>. Since the experimental data only covers the temperature range of between 5 and 40 °C, any NGAM for temperatures lower than 5 °C was set to the value at 5 °C and for those higher than 40 °C was set to the value at 40 °C. The Eq. (16) was used for the anaerobic growth data as well as for aerobic growth since no experimental data were available for this condition.

**FBA simulations with etcYeast7.6.** At a given temperature, first the  $k_{cat}$  values and  $\frac{[E]_{N,J}}{[E]_{N,J} + [E]_{D,J}}$  were calculated and integrated into the enzyme-constrained model and then the NGAM at this temperature was calculated and included in the model. For batch growth simulations, unlimited substrates were used, the same as described in ref. 16. The enzyme saturation factor  $\sigma$  of 0.5 was used<sup>39</sup>. For the simulation of anaerobic growth, in addition to the above changes, the uptake of oxygen was blocked and fatty acids and sterols were supplied into the medium as described in 16. The growth associated with ATP maintenance (GAM) was estimated from experimental data<sup>5</sup> as 70.17 mmol ATP/gdw. Other parameters were unchanged. For the simulation of fluxes at aerobic chemostat conditions, with the same model settings as aerobic batch conditions, the simulation was carried out by first fixing the growth rate to a given dilution rate (0.1 h<sup>-1</sup>) and minimizing the glucose uptake rate. Then the glucose uptake rate was fixed to the simulated value multiplied by a factor of 1.001 (for simulation purposes). Finally, the total enzyme usage was minimized (same as used in ref. 16). To get the flux sensitivity coefficient of an enzyme at a given temperature, the  $k_{cat}$  of all reactions associated with this enzyme were perturbed by a factor of  $(1 + \delta)$ . Then the maximal growth rates were simulated before ( $u$ ) and after ( $u_p$ ) perturbation. Finally, the flux sensitivity coefficient of enzyme  $i$  was calculated as  $\frac{u_p - u}{\delta}$ , where  $\mu$  and  $u_p$  are maximal specific growth rate before and after perturbation.  $\delta$  of 10 was used in this study.

**The distance function used in SMC-ABC approach.** The observed data used in this study was the maximal specific growth rate in aerobic<sup>4</sup> and anaerobic<sup>5</sup> batch cultivations at different temperatures, and glucose, carbon dioxide, and ethanol flux values at different temperatures measured in chemostat cultivations with a dilution rate of 0.1 h<sup>-1</sup><sup>23</sup>. The distance function was designed as follows: first, the coefficient of determination ( $R^2$ ) between simulated and experimental data was calculated for each of the above conditions. Then the average  $R^2$  across these three conditions multiplied by  $-1$  was used to represent the distance  $\rho(D, \hat{D})$ .  $\epsilon$  of  $-0.9$  was used in the SMC-ABC simulation.

**Statistical tests for comparison between  $P(\theta)$  and  $P(\theta|D)$ .** The significance test for the difference in mean values between *Prior* and *Posterior* was carried out by Welch's  $t$  test<sup>62</sup>. The significance test for reduced variance was carried out by the one-tailed  $F$ -test.  $p$  values were adjusted with the correction<sup>63</sup> using a family-wise error rate of 0.01. The significance cutoff was set to 0.01 (Fig. 2e).

**Machine learning applied to score the importance of parameters.** Totally, 2292 parameters of 21504 parameter sets were used as the input feature matrix and the average  $R^2$  scores obtained with the Bayesian approach were used as target labels. The dataset was split into train (80%), validation (10%), and test (10%) datasets. A random forest regressor with 1000 estimators was used. The train and validation datasets were used to optimize the hyperparameter. The obtained model could explain in total 23% the variance in the test dataset. The feature importance scores were extracted directly from the obtained model.

**Genetic manipulation for the experimental validation of ERG1.** The background strain we used in this study was IMX581 derived from CEN.PK113-5D, which contains an integrated Cas9 expression cassette controlled by TEFp promoter<sup>64</sup>. All the genetic manipulations were conducted based on the CRISPR/cas9 system. The codon-optimized kmERG1 were ordered from GenScript (Supplementary Data 1), and the PrimerSTAR HS polymerase was utilized for gene amplification through PCR. Based on strain IMX581, the codon-optimized gene ERG1 from *K. marxianus* (KmERG1) was integrated to replace the native ERG1 (ScERG1) using CRISPR/cas9, yielding HL01. All the design and construction of the plasmid follows the previously described method<sup>64</sup>. The gRNA cassette for target gene scERG1 was obtained using the single-stranded oligos gRNA-ERG1-F/gRNA-ERG1-R, followed by assembling with the linearized backbone plasmid pMEL10, the single gRNA plasmid was constructed by Gibson assembly. The repair fragment containing kmERG1 with round 60 bp overlap was amplified by primers kmEGR1-scERG1up-F/kmEGR1-scERG1dn-R using codon-optimized kmERG1 as a template. Then the repair fragment and single gRNA plasmid were co-transformed into IMX58. All the strains and primers used in this study were listed in Supplementary Tables 2 and 3.

**Strain cultivation under different temperatures.** The thermotolerance was tested and compared between *S. cerevisiae* IMX581 and HL01. Five single colonies of each strain were selected and precultured in YPD media at 30 °C, and cells were then transferred to flasks in 20 mL YPD media to reach 0.1 initial OD600 cultured at 40 ± 0.5 °C, 200 rpm. After that, the cells were transferred into fresh YPD media every 24 h with 0.1 initial OD600 and cultivated at 40 ± 0.5 °C, 200 rpm.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this article is available as a Supplementary Information file. The datasets and plant materials generated and analyzed during the current study are available from the corresponding author upon request. The data for reproducing the figures in both the main and supplementary files are provided as a Zenodo repository (<https://zenodo.org/record/3996543#.X0J1BNP7S3I>). Source data are provided with this paper.

### Code availability

All simulations of genome-scale models have carried out with Cobrapy<sup>64</sup> with Gurobi (Gurobi Optimization, LLC) solver. All code is available on Github (<https://github.com/SysBioChalmers/BayesianGEM>).

Received: 1 April 2020; Accepted: 25 November 2020;

Published online: 08 January 2021

### References

- Boussau, B., Blanquart, S., Necsulea, A., Lartillot, N. & Gouy, M. Parallel adaptations to high temperatures in the Archaeaneon. *Nature* **456**, 942–945 (2008).
- Hickey, D. A. & Singer, G. A. C. Genomic and proteomic adaptations to growth at high temperature. *Genome Biol.* **5**, 117 (2004).
- Engqvist, M. K. M. Correlating enzyme annotations with a large set of microbial growth temperatures reveals metabolic adaptations to growth at diverse temperatures. *BMC Microbiol.* **18**, 177 (2018).
- Caspeta, L. & Nielsen, J. Thermotolerant yeast strains adapted by laboratory evolution show trade-off at ancestral temperatures and preadaptation to other stresses. *MBio* **6**, e00431 (2015).
- Zakharov, M., Yang, X., Reuss, M. & Pörtner, H. O. Metabolic efficiency in yeast *Saccharomyces cerevisiae* in relation to temperature dependent growth and biomass yield. *J. Therm. Biol.* **52**, 117–129 (2015).
- Fersht, A. R. & Daggett, V. Protein folding and unfolding at atomic resolution. *Cell* **108**, 573–582 (2002).
- Leuenerger, P. et al. Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **355**, eaai7825 (2017).
- Guo, M., Xu, Y. & Gruebele, M. Temperature dependence of protein folding kinetics in living cells. *Proc. Natl Acad. Sci. USA* **109**, 17863–17867 (2012).
- Rocklin, G. J. et al. Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357**, 168–175 (2017).
- Mateus, A. et al. Thermal proteome profiling in bacteria: probing protein state. *Mol. Syst. Biol.* **14**, e8242 (2018).
- Arcus, V. L. et al. On the temperature dependence of enzyme-catalyzed rates. *Biochemistry* **55**, 1681–1688 (2016).
- DeLong, J. P. et al. The combined effects of reactant kinetics and enzyme stability explain the temperature dependence of metabolic rates. *Ecol. Evol.* **7**, 3940–3950 (2017).
- Grimaud, G. M., Mairet, F., Sciandra, A. & Bernard, O. Modeling the temperature effect on the specific growth rate of phytoplankton: a review. *Rev. Environ. Sci. Bio/Technol.* **16**, 625–645 (2017).
- Dill, K. A., Ghosh, K. & Schmitz, J. D. Physical limits of cells and proteomes. *Proc. Natl Acad. Sci. USA* **108**, 17876–17882 (2011).
- Villadsen, J., Nielsen, J. & Lidén, G. *Bioreaction Engineering Principles*. (Springer Science & Business Media, 2011).
- Sánchez, B. J. et al. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol. Syst. Biol.* **13**, 935 (2017).
- Förster, J., Famili, I., Fu, P., Palsson, B. Ø. & Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244–253 (2003).
- Chen, Y., Li, G. & Nielsen, J. Genome-scale metabolic modeling from yeast to human cell models of complex diseases: latest advances and challenges. *Methods Mol. Biol.* **2049**, 329–345 (2019).
- Price, N. D., Reed, J. L. & Palsson, B. Ø. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* **2**, 886–897 (2004).
- Chang, R. L. et al. Structural systems biology evaluation of metabolic thermotolerance in *Escherichia coli*. *Science* **340**, 1220–1223 (2013).
- Chen, K. et al. Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation. *Proc. Natl Acad. Sci. USA* **114**, 11548–11553 (2017).
- Li, G., Rabe, K. S., Nielsen, J. & Engqvist, M. K. M. Machine learning applied to predicting microorganism growth temperatures and enzyme catalytic optima. *ACS Synth. Biol.* **8**, 1411–1420 (2019).
- Postmus, J. et al. Quantitative analysis of the high temperature-induced glycolytic flux increase in *Saccharomyces cerevisiae* reveals dominant metabolic regulation. *J. Biol. Chem.* **283**, 23524–23532 (2008).
- Mohd Azhar, S. H. et al. Yeasts in sustainable bioethanol production: a review. *Biochem. Biophys. Rep.* **10**, 52–61 (2017).
- Lu, H. et al. A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat. Commun.* **10**, 3586 (2019).
- Yau, C. & Campbell, K. Bayesian statistical learning for big data biology. *Biophys. Rev.* **11**, 95–102 (2019).
- Lahtvee, P.-J. et al. Absolute quantification of protein and mRNA abundances demonstrate variability in gene-specific translation efficiency in yeast. *Cell Syst.* **4**, 495–504.e5 (2017).
- Kingma, D. P. & Welling, M. Auto-encoding variational bayes. Preprint at <https://arxiv.org/abs/1312.6114> (2013).
- Girolami, M. Bayesian inference for differential equations. *Theor. Comput. Sci.* **408**, 4–16 (2008).
- Miskovic, L., Béal, J., Moret, M. & Hatzimanikatis, V. Uncertainty reduction in biochemical kinetic models: enforcing desired model properties. *PLoS Comput. Biol.* **15**, e1007242 (2019).
- Hobbs, J. K. et al. Change in heat capacity for enzyme catalysis determines temperature dependence of enzyme catalyzed rates. *ACS Chem. Biol.* **12**, 868 (2017).
- Lahtvee, P.-J., Kumar, R., Hallström, B. M. & Nielsen, J. Adaptation to different types of stress converge on mitochondrial metabolism. *Mol. Biol. Cell* **27**, 2505–2514 (2016).
- van der Kamp, M. W. et al. Dynamical origins of heat capacity changes in enzyme-catalyzed reactions. *Nat. Commun.* **9**, 1177 (2018).
- Nguyen, V. et al. Evolutionary drivers of thermoadaptation in enzyme catalysis. *Science* **355**, 289–294 (2017).
- Jeske, L., Placzek, S., Schomburg, I., Chang, A. & Schomburg, D. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.* **47**, D542–D549 (2019).
- Li, G., Ji, B. & Nielsen, J. The pan-genome of *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **19**, foz064 (2019).
- Malina, C., Larsson, C. & Nielsen, J. Yeast mitochondria: an overview of mitochondrial biology and the potential of mitochondrial systems biology. *FEMS Yeast Res.* **18**, foy040 (2018).
- Pastore, A. et al. Unbiased cold denaturation: low- and high-temperature unfolding of yeast frataxin under physiological conditions. *J. Am. Chem. Soc.* **129**, 5374–5375 (2007).
- Nilsson, A. & Nielsen, J. Metabolic trade-offs in yeast are caused by F1F0-ATP synthase. *Sci. Rep.* **6**, 22264 (2016).
- Friesen, J. A. & Rodwell, V. W. The 3-hydroxy-3-methylglutaryl coenzyme-A (HMG-CoA) reductases. *Genome Biol.* **5**, 248 (2004).
- Sawle, L. & Ghosh, K. How do thermophilic proteins and proteomes withstand high temperature? *Biophys. J.* **101**, 217–227 (2011).
- Ghosh, K. & Dill, K. Cellular proteomes have broad distributions of protein stability. *Biophys. J.* **99**, 3996–4002 (2010).
- Swan, T. M. & Watson, K. Stress tolerance in a yeast sterol auxotroph: role of ergosterol, heat shock proteins and trehalose. *FEMS Microbiol. Lett.* **169**, 191–197 (1998).
- Caspeta, L. et al. Altered sterol composition renders yeast thermotolerant. *Science* **346**, 75–78 (2014).
- Ma, B.-X., Ke, X., Tang, X.-L., Zheng, R.-C. & Zheng, Y.-G. Rate-limiting steps in the *Saccharomyces cerevisiae* ergosterol pathway: towards improved ergosta-5,7-dien-3 $\beta$ -ol accumulation by metabolic engineering. *World J. Microbiol. Biotechnol.* **34**, 55 (2018).
- Baker, E. P. et al. Mitochondrial DNA and temperature tolerance in lager yeasts. *Sci. Adv.* **5**, eaav1869 (2019).
- Wolters, J. F. et al. Mitochondrial recombination reveals mito-mito epistasis in yeast. *Genetics* **209**, 307–319 (2018).
- Paliwal, S., Fiumera, A. C. & Fiumera, H. L. Mitochondrial-nuclear epistasis contributes to phenotypic variation and coadaptation in natural isolates of *Saccharomyces cerevisiae*. *Genetics* **198**, 1251–1265 (2014).
- Moro, F. & Muga, A. Thermal adaptation of the yeast mitochondrial Hsp70 system is regulated by the reversible unfolding of its nucleotide exchange factor. *J. Mol. Biol.* **358**, 1367–1377 (2006).
- Postmus, J. et al. Dynamic regulation of mitochondrial respiratory chain efficiency in *Saccharomyces cerevisiae*. *Microbiology* **157**, 3500–3511 (2011).

51. Ou, M. S., Ingram, L. O. & Shanmugam, K. T. L. (+)-Lactic acid production from non-food carbohydrates by thermotolerant *Bacillus coagulans*. *J. Ind. Microbiol. Biotechnol.* **38**, 599–605 (2011).
52. Matsushita, K. et al. Genomic analyses of thermotolerant microorganisms used for high-temperature fermentations. *Biosci. Biotechnol. Biochem.* **80**, 655–668 (2016).
53. Arora, R., Behera, S. & Kumar, S. Bioprospecting thermophilic/thermotolerant microbes for production of lignocellulosic ethanol: a future perspective. *Renew. Sustain. Energy Rev.* **51**, 699–717 (2015).
54. Repasky, E. A., Evans, S. S. & Dewhirst, M. W. Temperature Matters! And why it should matter to tumor immunologists. *Cancer Immunol. Res.* **1**, 210–216 (2013).
55. Protsiv, M., Ley, C., Lankester, J., Hastie, T. & Parsonnet, J. Decreasing human body temperature in the United States since the industrial revolution. *Elife* **9** (2020).
56. Baracos, V. E., Whitmore, W. T. & Gale, R. The metabolic cost of fever. *Can. J. Physiol. Pharmacol.* **65**, 1248–1254 (1987).
57. Sánchez, B. J. & Nielsen, J. Genome scale models of yeast: towards standardized evaluation and consistent omic integration. *Integr. Biol.* **7**, 846–858 (2015).
58. Kumar, S. & Nussinov, R. How do thermophilic proteins deal with heat? *Cell. Mol. Life Sci.* **58**, 1216–1233 (2001).
59. Murphy, K. P. & Gill, S. J. Solid model compounds and the thermodynamics of protein unfolding. *J. Mol. Biol.* **222**, 699–709 (1991).
60. Robertson, A. D. & Murphy, K. P. Protein structure and the energetics of protein stability. *Chem. Rev.* **97**, 1251–1268 (1997).
61. Sunnåker, M. et al. Approximate Bayesian computation. *PLoS Comput. Biol.* **9**, e1002803 (2013).
62. Welch, B. L. The generalization of ‘Student’s’ problem when several different population variances are involved. *Biometrika* **34**, 28 (1947).
63. Šidák, Z. Rectangular confidence regions for the means of multivariate normal distributions. *J. Am. Stat. Assoc.* **62**, 626–633 (1967).
64. Mans, R. et al. CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **15** (2015).

## Acknowledgements

The authors would like to thank Tyler W. Doughty, Benjamin J. Sánchez, Avlant Nielsen, and Ibrahim Elseman for the helpful discussions. G.L. and J.N. have received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie program, project PAcMEN (grant agreement No. 722287). J.N. also acknowledges funding from the Novo Nordisk Foundation (grant No. NNF10CC1016517), the Knut and Alice Wallenberg Foundation. J.Z. and A.Z. are supported by SciLifeLab funding. The computations were performed on resources at Chalmers Centre for Computational Science and Engineering (C3SE) provided by the Swedish National Infrastructure for Computing (SNIC).

## Author contributions

G.L. and J.N. conceptualized the project; G.L. designed and performed all computations; G.L. and H.L. designed and performed validation of the S.M.C.-A.B.C. approach; G.L., Y.H., J.Z., and J.N. designed the experiments. G.L., J.Z., B.J., H.W., A.Z., and J.N. interpreted the results. Y.H. performed the experimental validations; G.L., J.Z., H.W., and Y.H. wrote the initial draft paper. All authors carried out revisions on the initial draft and wrote the final version.

## Funding

Open Access funding provided by Chalmers University of Technology.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-20338-2>.

**Correspondence** and requests for materials should be addressed to J.N.

**Peer review information** *Nature Communications* thanks Roger Chang, Gertien Smits and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021