# Evidence for a short-lived resonance state in enzyme catalysis via rate-equation convolution

Moira L. Steyn-Ross and D. A. Steyn-Ross*
*School of Engineering, University of Waikato, Hamilton, New Zealand*

Erica J. Prentice, Emma J. Walker, and V. L. Arcus
*School of Science, University of Waikato, Hamilton, New Zealand*

(Received 12 November 2022; revised 2 March 2023; accepted 30 May 2023; published 26 June 2023)

At the cellular level, all biological function relies on enzymes to provide catalytic acceleration of essential biochemical processes driving cellular metabolism. The enzyme is presumed to lower the activation energy barrier separating reactants from products, but the precise mechanism remains unresolved. Here we examine the temperature dependence of the enzyme-catalyzed dissociation of *p*-nitrophenyl-$\alpha$-D-glucopyranoside (*p*NPG), a chromogenic analog for maltose, isomaltose, and sucrose disaccharide sugars, into *p*-nitrophenol (*p*NP) and glucose (monosaccharide). The enzymes of interest are the wild type and mutant forms of glucosidase MalL produced by the probiotic bacterium *Bacillus subtilis*. The per-enzyme production rates $k(T)$ for the *p*NPG $\rightarrow$ glucose reaction all show a characteristic temperature profile with an Arrhenius-like (approximately exponential) slow acceleration at low temperatures, rising through a point of inflexion to reach a maximum, then turning over to decline steeply towards zero production at high temperatures. This asymmetric profile is found to be well fitted by convolving an exponential growth function $f(T)$ with a Gaussian temperature distribution $g(T)$ to produce an exponentially modified Gaussian function $h(T)$. To give a physical interpretation of the convolution components, we make the temperature mapping $\Theta \equiv T_{\text{ref}} - T$ where $T_{\text{ref}}$ marks the temperature at which a given mutant becomes fully denatured (unfolded) and therefore inactive, then convert the convolution components to probability density functions which obey the convolution theorem of statistics. Working in $\Theta$ space, we identify $f(\Theta)$ as the density function for an Arrhenius-like transition from ground-state A to metastable-state B, and $g(\Theta)$ as the Gaussian distribution of offset-temperature fluctuations for the metastable state. By mapping the standard thermodynamic relations for temperature and energy fluctuations to the enzyme frame of reference, we are able to derive an expression for the lifetime for the metastable B state. For the 15 enzyme experiments, we obtain a mean value $\langle \Delta t \rangle \gtrsim (29.0 \pm 1.3) \times 10^{-15}$ s, in remarkably good agreement with the ~30-fs estimate for the period of glycosidic bond oscillations extracted from published infrared spectroscopy. We suggest that the metastable B state provides a low-energy target that has the effect of lowering the activation energy barrier by presenting an alternative axis for the reaction coordinate.

DOI: 10.1103/PhysRevE.107.064407

## I. INTRODUCTION

Proteins with enzymatic function provide an essential catalytic pathway to accelerate chemical reactions inside cells. All biological function is linked to catalytic activity [1]. Without enzyme assistance, reaction rates would run orders of magnitude slower, causing fundamental biochemical processes—such as metabolism—to become completely infeasible. This catalytic acceleration arises because of the ability of the enzyme macromolecule to lower $E_a$, the activation energy barrier separating reactants and products. Based on computational studies, the most common explanation for barrier reduction is a preorganization of the enzyme active site as it binds to the substrate molecule [2,3], but this is contentious. In fact, the multiple factors enabling high enzyme catalytic efficiency remain poorly understood [4].

In this paper, we investigate the enzyme catalysis of *p*-nitrophenyl-$\alpha$-D-glucopyranoside (*p*NPG) into its *p*-nitrophenol (*p*NP) and glucose (G) components:

$$p\text{NPG} \xrightarrow{\text{MalL}} p\text{NP} + \text{G}. \tag{1}$$

The specific enzyme of interest is the $\alpha$-glucosidase MalL (EC 3.2.1.10)[1], produced by the bacterium *Bacillus subtilis*. MalL is a large protein consisting of 561 amino acids with a molecular mass of 66 kDa. MalL is classified as oligo-1,4-1,6-$\alpha$-glucosidase (sucrase-maltase-isomaltase) [5], meaning that it can cleave the 1-4 and 1-6 glycosidic bonds in the

*asr@waikato.ac.nz



[1]Enzyme Commission number: classifies the enzyme based on the chemistry it catalyses.

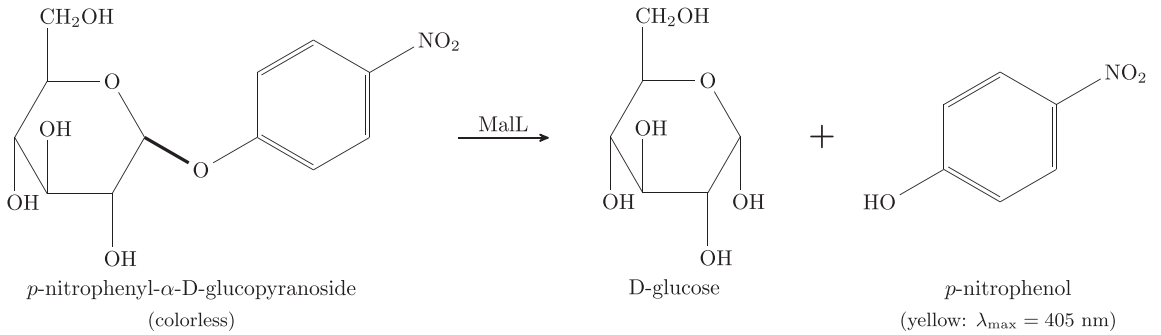



FIG. 1. Chemical structure diagram for reaction (1) showing the MalL-catalyzed decomposition of *p*NPG into its glucose and nitrophenol components. The MalL enzyme acts to cleave the C–O glycosidic linkage (shown in bold).

disaccharide sugars sucrose, maltose, and isomaltose to release free glucose. To quantify enzyme activity, we need to count the number of glucose molecules released per MalL enzyme molecule per unit time. This is conveniently done by replacing the natural disaccharide substrate with the artificial substrate *p*NPG, which is colorless in solution but produces a yellow product (*p*NP) when the *p*NPG glycosidic bond (drawn in bold in Fig. 1) joining the D-glucose residue to the *p*-nitrophenyl moiety is cleaved (hydrolyzed) [6]. This color change enables real-time spectroscopic detection of enzyme activity at the *p*NP absorbance peak of 405 nm.

Enzyme activity is strongly affected by changes in temperature and pH level. It is also affected by engineered mutations of the amino acid residues close to the binding pocket (also called the active site) that captures the substrate molecule. In the present paper, we examine the temperature dependence of MalL activity for three settings of pH for both wild-type (WT) MalL and four single-point mutations, giving a total of 15 distinct rate-versus-temperature experiments.

Because enzymes exist in a liquid water milieu, enzyme function is only possible over the temperature range for which water is liquid, i.e., $0 < T/^{\circ}\mathrm{C} < 100$ (assuming standard atmospheric pressure), but, in fact, most enzymes will begin to denature (lose their 3D structure and hence their functionality) at temperatures well below the boiling point of water. For example, the melting temperature for the MalL maltase enzyme is around 50 °C (e.g., see melting curves in Fig. 4).

As a general rule, the rate of chemical reaction $k(T)$ increases with temperature, showing a roughly exponential growth with absolute temperature,

$$k(T) = Ce^{\lambda T} \quad \text{(simple exponential form)}, \tag{2}$$

where $C$ and $\lambda$ are appropriate fitting constants with units $\mathrm{s}^{-1}$ and $\mathrm{K}^{-1}$, respectively.

In 1899, Arrhenius [7] showed empirically that the temperature dependence of the reaction rate is more accurately modeled by

$$k(T) = Ae^{-\frac{E_a}{RT}} \quad \text{(Arrhenius form)}, \tag{3}$$

where $E_a$ is the activation energy (J/mol), $R$ is the universal gas constant, $A$ is the number of reactant collisions per second, and $e^{-Ea/RT}$ is the probability that a collision produces a reaction. Reaction rate $k$ carries the same units as the pre-exponential factor $A$ ($\mathrm{s}^{-1}$), and we are assuming the reaction is first order.

Eyring's 1935 transition state theory [8] provided an alternative form for the rate equation with Gibbs energy of activation $\Delta G^{\ddagger}$ replacing activation energy $E_a$ in the Eq. (3) exponent,

$$\begin{aligned} k(T) &= \frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger}}{RT}} \\ &= \left(\frac{k_B T}{h} e^{\frac{\Delta S^{\ddagger}}{R}}\right) e^{-\frac{\Delta H^{\ddagger}}{RT}} \\ &= BT\, e^{-\frac{\Delta H^{\ddagger}}{RT}} \quad \text{(Eyring form)}, \end{aligned} \tag{4}$$

where $\Delta G^{\ddagger} \equiv \Delta H^{\ddagger} - T\Delta S^{\ddagger}$; $\Delta H^{\ddagger}$ is the activation enthalpy, $\Delta S^{\ddagger}$ is the activation entropy, and both are assumed to be temperature independent. The Arrhenius (3) and Eyring (4) equations are mathematically very similar, apart from the fact that the bracketed prefactor in (4) scales weakly (linearly) with temperature. The applicability of these models can be tested by plotting [$\ln(k)$ vs $1/T$] (Arrhenius) or [$\ln(k/T)$ vs $1/T$] (Eyring); straight-line graphs are expected with slopes $-\frac{E_a}{R}$ and $-\frac{\Delta H^{\ddagger}}{R}$, respectively, with $E_a \approx \Delta H^{\ddagger} + RT$.

In practical application, it is not unusual to find that the Arrhenius and Eyring models fit the data equally well, and this is certainly the case for the *low-temperature* portion of the enzyme-catalyzed sugar reaction that we report here; see Fig. 2. However, it is clear from Fig. 2 that exponential extrapolations to higher temperatures ($T \gtrsim 310$ K, 37 °C) dramatically overestimate measured reaction rates and fail to predict the emergence of an optimum temperature $T_{\mathrm{opt}} \approx 318$ K that maximizes production rate.

Similar production rate-versus-temperature experiments have been repeated for the WT variant of MalL at three pH levels (6.5, 7.0, 8.0) and for four distinct point mutations of the native WT enzyme (V200 → V200S, V200T, V200A, S536R) for a total of 15 experiments (see Appendix for details). In every case, we find ∼exponential growth at low bath temperatures, emergence of an optimum temperature, then rapid decay towards zero at higher temperatures. Thus, Fig. 2 (WT at pH 6.5) can be taken as representative of the thermal behavior of all of the MalL mutants tested.

Because a given enzyme loses catalytic function when it melts (unfolds), it is plausible to assume that the downturn in production is a direct result of enzyme denaturation as

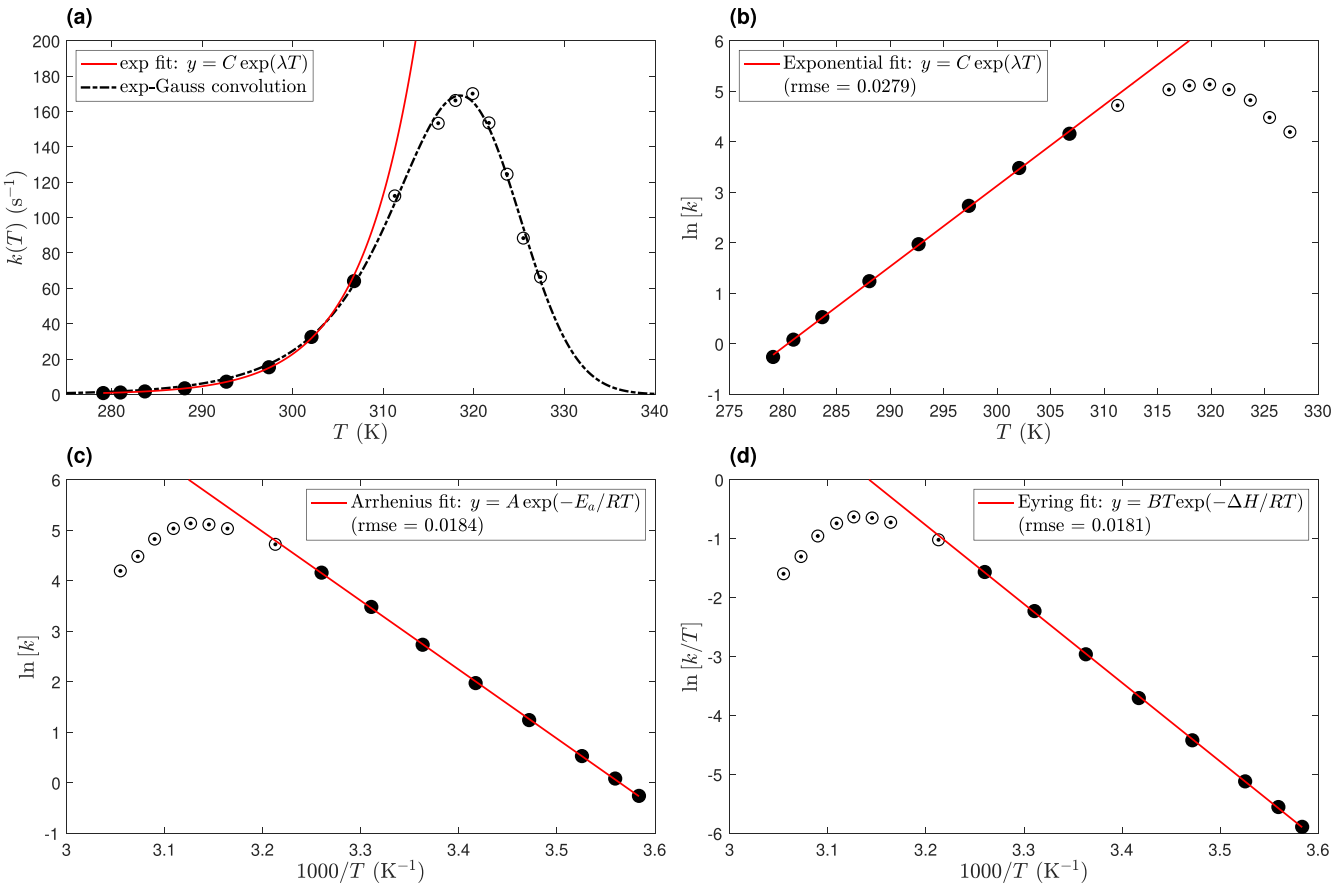



FIG. 2. Rate of glycosidic bond cleavage, $k(T)$, for the wild-type (WT) MalL enzyme-catalyzed hydrolysis reaction *p*NPG $\rightarrow$ (nitrophenol + glucose) at pH 6.5 as a function of bath temperature $T$. Production rate increases ∼exponentially with temperature for $T \rightarrow 310$ K [filled black circles: see red curve-fits in (a), (b)], but falls below the exponential growth projection for $T \gtrsim 310$ K (circled dots). Peak production occurs at $T_{\rm opt} \approx 318$ K, followed by a monotonic decline towards zero production at higher temperatures. The early growth (filled circles) is equally well described by (c) an Arrhenius fit [Eq. (3)] or (d) an Eyring fit [Eq. (4)]; the rmse values give the respective root-mean-square curve-fitting errors for the low-temperature measurements. The black-dashed curve in (a) demonstrates that a $k(T)$ fit across the *full* temperature range can be realized by convolving early exponential growth with late Gaussian decay: see Eq. (6d). Low-temperature curve fits: (a), (b) exponential: [$C = 3.40 \times 10^{-20}$ s$^{-1}$, $\lambda = 0.160$ K$^{-1}$]; (c) Arrhenius: [$A = 1.34 \times 10^{21}$ s$^{-1}$, $E_a = 113.5$ kJ/mol]; (d) Eyring: [$B = 1.69 \times 10^{18}$ s$^{-1}$, $\Delta H = 111.1$ kJ/mol].

the bath temperature approaches, then exceeds, the melting temperature $T_{\rm m}$. This explanation cannot be sufficient, however, since protein unfolding is a slow process that runs on timescales several orders of magnitude slower than the fast kinetics associated with the cleavage reaction [9,10]. This difference in timescales means that it is possible to obtain meaningful production rates at temperatures above $T_{\rm m}$ (as illustrated later in Fig. 4). Consequently, the emergence of an optimum temperature has been described as strange (i.e., not explained by denaturation), suggesting the existence of hidden conformational states [10].

Arcus *et al.* [11,12] have developed a macromolecular rate theory (MMRT) to account for the emergence of an optimum temperature in the absence of denaturation; this theory has been used in more recent work by Bunzel *et al.* [13]. In this picture, the deviations from Arrhenius kinetics at higher temperatures are the result of a significant difference in heat capacity between the enzyme-substrate complex and the enzyme-transition-state complex. This follows from the dramatic increase in binding affinity when proceeding along the reaction coordinate from the enzyme-substrate to enzyme-transition state. A change in heat capacity implies that both $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ will be strongly temperature dependent, resulting in the observed curvature in the Arrhenius plot (Fig. 2). In the present paper, we consider a complementary explanation that analyzes enzyme rate curves to extract probability distributions, associating a very short-lived species with single-particle energy fluctuations. We note that while MMRT accounts for the deviations from Arrhenius kinetics due to $\Delta C_p^{\ddagger} < 0$, the same enzyme kinetics data may also be used to extract the lifetime of a metastable state close to the transition state for the chemical reaction, as we set out below.

As illustrated by the black-dashed curve in Fig. 2(a), and the dotted red curves of Fig. 4, we can achieve a good fit to the $k(T)$ rate measurements across the *full* temperature range by convolving a (truncated) exponential growth curve (representing low-temperature acceleration) with a Gaussian temperature profile (at higher temperatures) to give the characteristic peaked distribution whose rates fall off asymmetrically either side of the production peak at $T_{\rm opt}$. We find

that this exponential-Gaussian convolution approach works equally well across all 15 mutant experiments, and thus recommends itself as a universal fitting function for enzyme performance with respect to temperature. This fitting success raises the immediate question: What is the significance of the location and width of the Gaussian profile?

We will argue that the convolution fit, when expressed as a probability density function, represents the probability summation of two independent processes: an Arrhenius-like transition from a ground-state A to a metastable-state B, and Gaussian-distributed single-particle temperature fluctuations of a short-lived resonance state. This convolutional probability summation requires the introduction of an offset temperature scale $\Theta = (T_{\rm ref} - T)$, where $T_{\rm ref}$ is the temperature at which the enzyme has become inactive, marking an absolute zero for enzyme activity.

Assuming that the standard thermodynamic relations for temperature and energy fluctuations can be mapped to this enzyme-relative $\Theta$ frame of reference, the width of the Gaussian profile allows us to calculate the lifetime of the resonance state via application of Heisenberg's energy-time uncertainty principle. We obtain resonance lifetimes which are in good agreement with the putative duration of a single oscillation of the glycosidic bond, as computed from wave-number absorbtion signatures reported in the infrared spectroscopy literature.

The paper is structured as follows. In Sec. II A, we define the exponentially modified Gaussian (EMG) probability distribution as the convolution product of a (truncated) exponential growth density function with a Gaussian distribution. This is the EMG form used for fitting to the $k(T)$ enzyme rate curves, but needs to be mapped back to the standard exponentially decaying EMG form to make use of the fundamental convolution theorem of statistics for addition of independent random variables.

In Sec. II B, we compare four source Gaussians—retrieved from $k(T)$ convolutional fitting—against MalL melting curves obtained from differential scanning colorimetry (DSC) measurements. In each case, the retrieved Gaussian distribution is significantly broader than the DSC signal, indicating that the high-temperature roll-off in enzyme production rate is *not* primarily driven by protein denaturation.

Section II C invokes the convolution theorem of statistics to motivate the mapping of the three probability distributions [$f(\cdot) =$ Arrhenius-like acceleration; $g(\cdot) =$ Gaussian distribution; $h(\cdot) =$ resultant EMG convolution] from temperature $T$ space to an enzyme-based offset-temperature $\Theta$-space.

We present an energy cartoon and enzyme-recycling scheme in Sec. II D. We highlight the vast difference in timescales between the rate of enzyme turnover (of order milliseconds) versus duration of the putative metastable B state that comes into existence immediately prior to bond cleavage (tens of femtoseconds). The extreme rarity and brevity of the B state means it can be treated as a single particle whose internal energy is uncoupled from the thermal temperature of the reaction chamber. This leads naturally in Sec. II E to the notion of single-particle fluctuations in temperature and internal energy, as measured in the enzyme frame of reference. These fluctuations are coupled via the effective single-particle heat capacity, allowing us to derive an expression for the lifetime of the B-state resonance in Sec. II F.

Finally, in Sec. III we show how the Gaussian fluctuation, when expressed as a standard normal probability curve, serves to unify all 15 MalL enzyme experiments. We discuss the implications of our theoretical findings and suggest future work.

## II. THEORY

### A. Convolution fitting

The exp-Gauss convolution fit to the full $k(T)$ rate curve shown in Fig. 2(a) is obtained in two steps. First, we convolve $f(T)$, an exponential growth curve truncated at selected maximum temperature $T_{\rm ref}$, with $g(T)$, a Gaussian distribution centered at temperature $\mu_T$ with variance $\sigma_T^2$; this generates the full convolution function $h'(T)$. In the second step, we shrink the domain of the convolution curve by displacing the $h'(T)$ curve to the left by temperature offset $T_{\rm ref}$ to form the reduced fitting function $h(T)$. We write the full convolution as

$$\begin{aligned} h'(T) &= f(T) \otimes g(T) \\ &\equiv \int_0^\infty f(\tau)\, g(T-\tau)\, d\tau, \end{aligned} \tag{5}$$

where $f(T)$, $g(T)$, $h'(T)$ are probability density functions (i.e., each has unit area) defined by

$$f(T) = \lambda\, e^{-\lambda(T_{\rm ref}-T)}\, u(T_{\rm ref} - T), \tag{6a}$$

$$g(T) = \frac{1}{\sqrt{2\pi}\,\sigma_T}\, e^{-(T-\mu_T)^2/2\sigma_T^2}, \tag{6b}$$

$$\begin{aligned} h'(T) &= h(T - T_{\rm ref}), \\ h(T) &= \lambda \exp\left(\lambda(T-\mu_T) + \frac{(\lambda\,\sigma_T)^2}{2}\right) \\ &\quad \times \frac{1}{2}\,\mathrm{erfc}\left[\frac{1}{\sqrt{2}}\left(\frac{(T-\mu_T)}{\sigma_T} + \lambda\,\sigma_T\right)\right]. \end{aligned} \tag{6c}$$

The growth function $f(T)$ can be interpreted as Arrhenius-like transition probability density. The unit-step $u(\cdot)$ in Eq. (6a) truncates the exponential growth curve at an appropriate maximum temperature $T_{\rm ref}$,

$$u(T_{\rm ref} - T) = \begin{cases} 1, & T \leqslant T_{\rm ref} \\ 0, & T > T_{\rm ref.} \end{cases} \tag{7}$$

Here, erfc$(\cdot)$ in Eq. (6c) is the complementary error function. The three density functions $f(\cdot)$, $g(\cdot)$, $h(\cdot)$ are plotted in Fig. 3(b). The reduced-domain temperature convolution of $f(T)$ (truncated exponential growth, in dot-dashed black), with $g(T)$ (Gaussian, blue), produces $h(T)$, an optimized variant of the EMG (dotted red) that we designed to provide a good fit to $k(T)$ enzyme production rate curves with their characteristic gradual rise to peak, followed by faster decay. Note that this slow-rise-fast-decay profile is the *mirror image* of that generated by the standard EMG definition commonly used for curve fitting in chromotography, psychophysiology, cell and molecular biology (see Golubev [14] and references therein) as illustrated in Fig. 3(a).

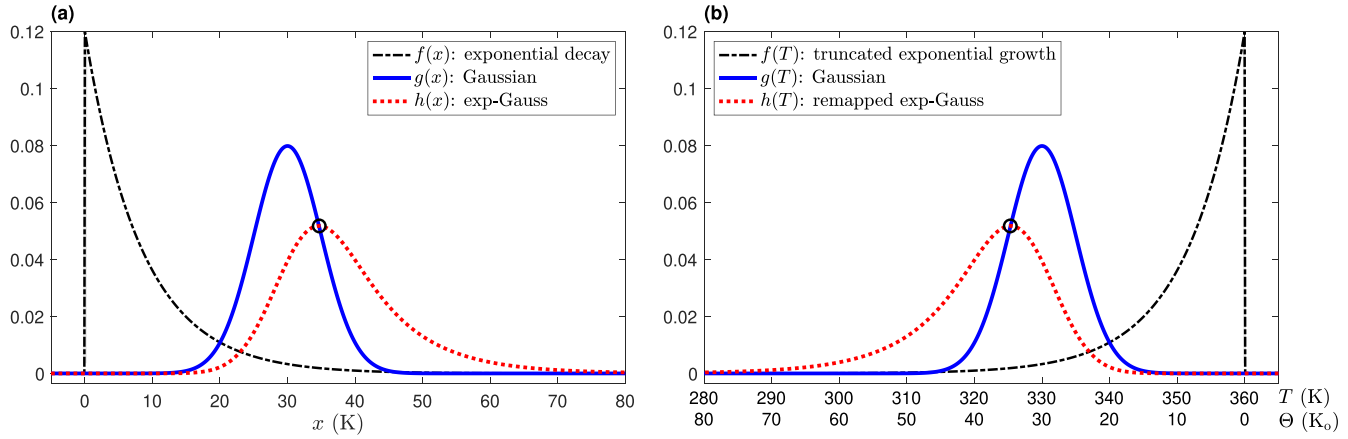



FIG. 3. Illustrative probability density functions for $f(\cdot)$ and $g(\cdot)$, and their convolution $h(\cdot) = f(\cdot) \otimes g(\cdot)$ plotted as a function of $x$ (left panel) and $T$ (right panel). (a) Standard exponentially modified Gaussian with $x$ taken to represent absolute temperature; model settings: $\lambda = 0.12$ K$^{-1}$ (exponential decay), $(\mu_x, \sigma_x) = (30, 5)$ K (Gaussian). Expected mean values: E$[f(x), g(x), h(x)] = [8.33,\ 30,\ 38.33]$ K; variances: var$[f(x), g(x), h(x)] = [69.44,\ 25,\ 94.44]$ K$^2$. (b) Remapped EMG for representative enzyme-rate model $\lambda = 0.12$ K$^{-1}$, $T_{\text{ref}} = 360$ K (truncated exponential growth); $(\mu_T, \sigma_T) = (330, 5)$ K (source Gaussian). The $g(\cdot)$ and $h(\cdot)$ curves intersect at the convolution peak (circled point). The subsidiary $\Theta$ scale in (b) is given by $\Theta = T_{\text{ref}} - T$, with $\Theta = 0$ K$_o$ corresponding to the temperature $T = T_{\text{ref}}$ at which the enzyme-substrate complex becomes fully denatured.

The linear mapping from our customized enzyme form (6) back to EMG standard form (9) is given by

$$\left.\begin{aligned} T &= -x + T_{\text{ref}} \\ \mu_T &= -\mu_x + T_{\text{ref}} \\ \sigma_T &= \sigma_x. \end{aligned}\right\} \tag{8}$$

With these substitutions, Eqs. (6) reduce to

$$f(x) = \lambda\, e^{-\lambda x}\, u(x), \tag{9a}$$

$$g(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x}\, e^{-(x-\mu_x)^2/2\sigma_x^2}, \tag{9b}$$

$$h(x) = \lambda \exp\left(-\lambda(x-\mu_x) + \frac{(\lambda\,\sigma_x)^2}{2}\right) \times \frac{1}{2}\,\text{erfc}\left[\frac{1}{\sqrt{2}}\left(\frac{-(x-\mu_x)}{\sigma_x} + \lambda\,\sigma_x\right)\right], \tag{9c}$$

with the $u(\cdot)$ gating function in Eq. (9a) forcing the exponential to zero for $x < 0$,

$$u(x) = \begin{cases} 1, & x \geqslant 0 \\ 0, & x < 0. \end{cases} \tag{10}$$

With mapping (8), the sign reversal of $x$ flips the exponential growth (approximating Arrhenius acceleration at low temperatures) of (6a) to an exponential decay in (9a), and the $+T_{\text{ref}}$ offset shifts the exponential truncation point from $T = T_{\text{ref}}$ to $x = 0$. The mirror equivalence of the two EMG forms is demonstrated in Fig. 3. Note that the value of $T_{\text{ref}}$ will be specific to a given enzyme experiment, since it depends on both MalL mutant type and selected pH level.

Although we use Eqs. (6) for $k(T)$ curve fitting, we need to return to standard EMG form (9) for the analysis of fluctuation statistics and subsequent determination of the lifetime of the proposed metastable resonance state. This is because the $h(T)$ fitting function has been constructed to provide a good match over the reduced temperature range ($\sim$280 K to $T_{\text{ref}}$) (rather than the full convolution domain which extends out to $2 \times T_{\text{ref}}$); consequently, the convolution theorem for addition of random variables does not apply to $h(T)$ when plotted in physical temperature space. Instead, we define a remapped temperature scale $\Theta = T_{\text{ref}} - T$, thus the $x$ scale for the left panel of Fig. 3 becomes the left-running $\Theta$ scale shown in the right panel; we revisit this temperature rescaling in Sec. II C.

For the purposes of initial curve fitting, it is sufficient to set $T_{\text{ref}}$ to a value well above $T_{\text{opt}}$, the temperature of maximum production (e.g., $T_{\text{ref}} = T_{\text{opt}} + 40$ K). But once the mean $\mu_T$ and variance $\sigma_T^2$ of the source Gaussian of (6b) have been determined, we can refine $T_{\text{ref}}$ to represent the reference temperature at which the enzyme has fully unfolded. This more precise value then allows us to estimate the lifetime of the proposed enzyme-substrate resonance state.

Close inspection of Fig. 2 shows that Arrhenius and Eyring models provide a more accurate low-temperature fit (lower rms fitting error) than the simple exponential model. This raises the obvious question: Why not use (appropriately truncated) Arrhenius (3) or Eyring (4) forms—instead of simple exponential (2)—for the convolution? We have tested both alternatives against simple exponential for all 15 enzyme experiments. We find that while Arrhenius and Eyring retrievals for the source Gaussian parameters $(\mu_T, \sigma_T)$ are essentially identical (rounding to 4 sig figs in $\mu_T$; 3 sig figs in $\sigma_T$), on average the exponential convolution *underestimates* the Arrhenius/Eyring value for Gaussian mean $\mu_T$ by 0.2 K but *overestimates* the standard deviation $\sigma_T$ by 0.15 K. See Table I. These small offset errors do not have any material impact on our conclusions.

Although Arrhenius and Eyring convolutions give better fits at low temperatures, they both share a major disadvantage: unlike the simple exponential, there is *no closed form* for their convolution with a Gaussian. This means that the fitting must be done by iterative numerical convolution integrations; consequently, the fitting algorithm runs about two orders of magnitude slower than does fitting against the EMG closed form.

TABLE I. Extracted parameters $(\mu_T, \sigma_T)$ (in K) for source Gaussians obtained via (A) exponential-Gaussian and (B) Arrhenius-Gaussian convolution for each of the 15 MalL enzyme experiments. (Note that at the displayed resolution (4 sig figs for $\mu_T$, 3 sig figs for $\sigma_T$), the Eyring-Gaussian convolution results are indistinguishable from Arrhenius-Gaussian, so the Eyring retrievals are not shown here.)

| | | | (A) exp Gauss | | (B) Arrh Gauss | | $\Delta\mu_T$ | $\Delta\sigma_T$ |
|---|---|---|---|---|---|---|---|---|
| Expt. | Mutant | pH | $\mu_T$ | $\sigma_T$ | $\mu_T$ | $\sigma_T$ | $\mu_B - \mu_A$ | $\sigma_B - \sigma_A$ |
| 1 | WT | 6.5 | 322.9 | 5.42 | 323.1 | 5.33 | 0.2 | −0.09 |
| 2 | | 7 | 321.0 | 5.05 | 321.2 | 4.97 | 0.2 | −0.08 |
| 3 | | 8 | 317.6 | 6.66 | 317.7 | 6.61 | 0.1 | −0.05 |
| 4 | V200S | 6.5 | 331.1 | 8.28 | 331.4 | 7.84 | 0.3 | −0.44 |
| 5 | | 7 | 328.7 | 6.58 | 329.0 | 6.32 | 0.3 | −0.26 |
| 6 | | 8 | 324.4 | 8.93 | 324.7 | 8.77 | 0.3 | −0.16 |
| 7 | V200T | 6.5 | 327.9 | 5.83 | 328.1 | 5.58 | 0.2 | −0.25 |
| 8 | | 7 | 327.1 | 6.01 | 327.3 | 5.78 | 0.2 | −0.23 |
| 9 | | 8 | 324.4 | 8.02 | 324.7 | 7.87 | 0.3 | −0.15 |
| 10 | S536R | 6.5 | 323.2 | 5.81 | 323.3 | 5.73 | 0.1 | −0.08 |
| 11 | | 7 | 322.1 | 4.87 | 322.2 | 4.80 | 0.1 | −0.07 |
| 12 | | 8 | 317.9 | 5.63 | 318.0 | 5.55 | 0.1 | −0.08 |
| 13 | V200A | 6.5 | 326.1 | 7.22 | 326.3 | 7.14 | 0.2 | −0.08 |
| 14 | | 7 | 327.7 | 6.08 | 327.8 | 5.98 | 0.1 | −0.10 |
| 15 | | 8 | 324.2 | 5.82 | 324.4 | 5.74 | 0.2 | −0.08 |
| | | | | | | Mean discrepancy: | 0.19 | −0.147 |

### B. Comparison of source Gaussian with enzyme melting curves

The denaturation characteristics of an enzyme can be quantified using DSC. The difference in the amount of energy required to increase the temperature of a sample (buffer-enzyme) versus a reference (buffer-buffer) is measured as a function of temperature as it is increased linearly at ∼1 °C/min. The differential power delivered to the enzyme sample typically shows a well-defined endothermic peak at $T_{\mathrm{m}}$, the so-called melting temperature, marking the point at which 50% of the protein is presumed to have unfolded to its denatured (inactive) state. The time integral of the differential power curve gives the enthalpy change for the (folded) → (unfolded) melting transition.

The panels of Fig. 4 show the curve fits, expressed as probability densities, for four enzyme mutants at pH 6.5. The dotted red curves are the Eq. (6) EMG fits to the measured data, and the blue and dot-dashed black curves show, respectively, the Gaussian and exponential density components of the convolution. These are superimposed on the DSC melt curves (obtained in separate thermostatics experiments) for the selected MalL mutant. We note that in each case, the melt curve (filled gray) is much narrower than the production curve (dotted red), with melt onset consistently occurring beyond the point of inflexion of the production curve. This means that production starts to fall away from the (low-temperature) accelerating Arrhenius trend at a temperature well below melt onset.

Further, the data points show that useful production is still possible beyond the peak of the DSC curve. This feature is not so surprising when we consider that production kinetics are orders of magnitude faster than unfolding kinetics, and that production is measured over a time span of a few seconds (∼10 s at low temperatures, dropping to ∼3 s at the highest temperatures), while the melting curve is obtained over a period of 10–15 min.

It is clear that the turnover in the production curve is not primarily a function of temperature-driven protein unfolding. Instead, we will argue that the turnover characteristics are, in fact, determined by the source Gaussian revealed via the convolution fitting. The variance of the Gaussian provides a measure of the single-particle (enzyme-substrate complex) temperature fluctuations from which we can deduce the heat capacity of the particle, the internal energy fluctuations (relative to $T_{\mathrm{ref}}$, the temperature at which the protein is completely unfolded), and hence estimate the lifetime of the metastable complex. Thus, we interpret the emergence of the Gaussian component as providing evidence of the existence of a short-lived metastable state.

### C. Interpretation of convolution components

As shown in Fig. 4, the $k(T)$ MalL-catalyzed *p*NPG → glucose production curves—expressed as probability density functions—are well-fitted by convolving a (truncated) exponential growth with a broad Gaussian distribution whose peak appears in the vicinity of the MalL melting temperature. The fact that the production density function $h(\cdot)$ [Eq. (6c)] can be decomposed into two convolutional components $f(\cdot)$ and $g(\cdot)$ suggests that the MalL enzyme function can be modeled as the sum of two independent random processes: $f(T)$ is the Arrhenius-like probability density for an A → B barrier crossing; $g(T)$ describes Gaussian-distributed temperature fluctuations (mean $\mu_T$, variance $\sigma_T^2$) for the metastable state.

To make this probabilistic interpretation more precise, we invoke the *fundamental convolution theorem of statistics*

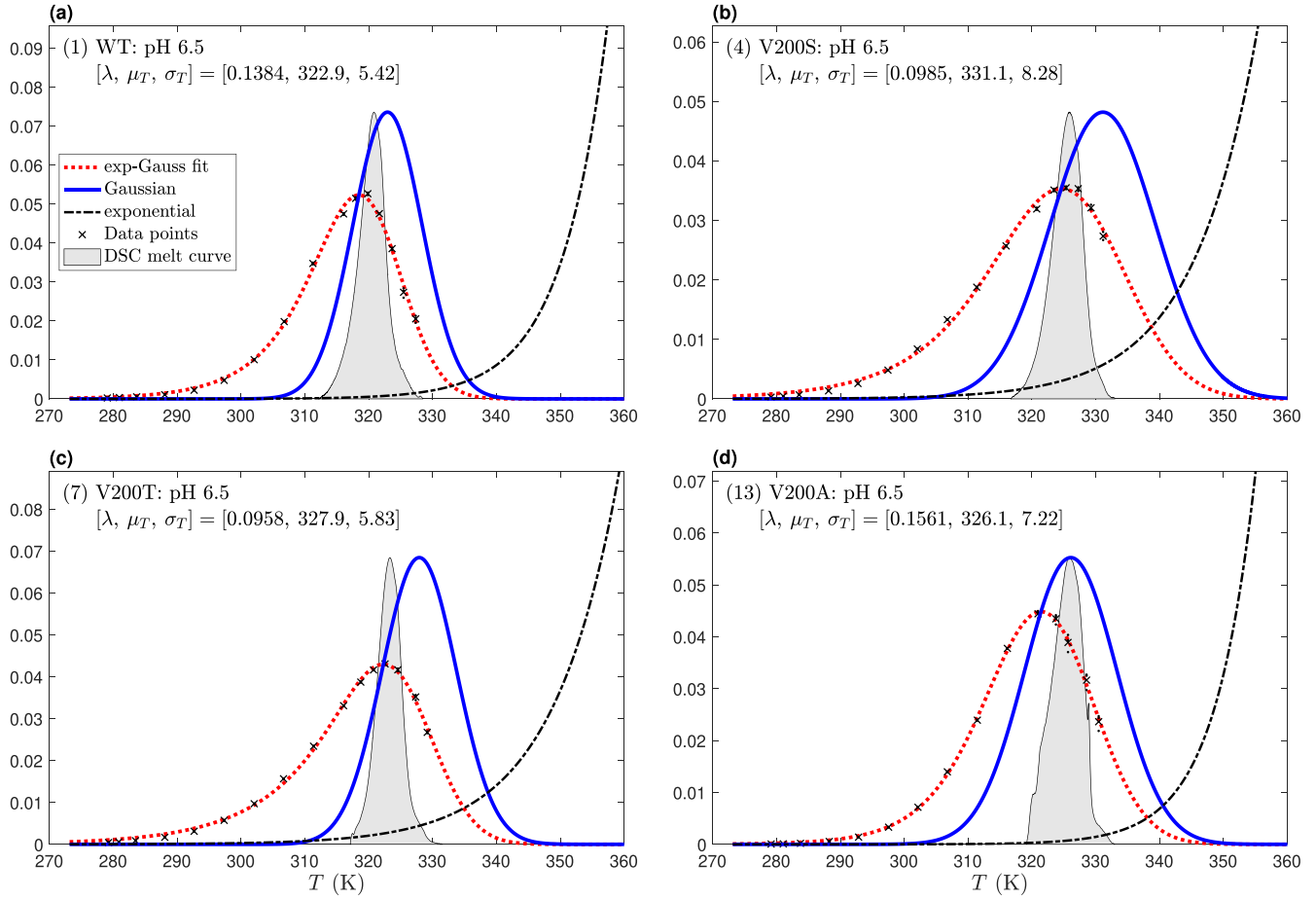



FIG. 4. Exponential (dot-dashed black), Gaussian (solid blue), and EMG (dotted red) probability density curve fits superimposed on DSC (differential scanning colorimetry) melting curves for four enzyme experiments. To aid visual comparison, each DSC plot (filled gray) has been scaled to the same height as the source Gaussian. We note that at half height, the Gaussians are two to three times *broader* than the melt profiles, indicating that protein unfolding is not the primary driver for the temperature dependance of enzyme dynamics.

[15,16]. In Fig. 3(a), let $X_1$ be a random variable with exponentially decaying probability density $f(x)$, and $X_2$ be a second random variable with Gaussian density function $g(x)$. Provided that $X_1$ and $X_2$ are independent, then the probability density function for the sum $X_3 = X_1 + X_2$ is given by the EMG convolution product $h(x) = f(x) \otimes g(x)$. This implies that the mean of the EMG distribution equals the sum of the means of the exponential and Gaussian distributions, and similarly the variance of the EMG distribution equals the sum of the variances of the two component distributions:

$$\langle X_3 \rangle = \langle X_1 \rangle + \langle X_2 \rangle, \tag{11a}$$

$$\mathrm{var}[X_3] = \mathrm{var}[X_1] + \mathrm{var}[X_2] \tag{11b}$$

[see caption of Fig. 3(a) for numerical confirmation]. However, it is not possible to make similar statements for the enzyme temperature distributions of Fig. 3(b). That is, if $T_1$ is the random variable described by the $f(T)$ truncated exponential growth distribution, $T_2$ the random variable for the $g(T)$ Gaussian distribution, and $T_3$ the random variable for (reduced) convolution product $h(T)$, then it is clearly *not* the case that $T_3 = T_1 + T_2$, and therefore $\langle T_3 \rangle \neq \langle T_1 \rangle + \langle T_2 \rangle$. (This is despite the fact that the *shapes* of the distributions are preserved so the variances add: $\mathrm{var}[T_3] = \mathrm{var}[T_1] + \mathrm{var}[T_2]$.)

This apparent failure of the convolution probability addition in physical temperature space $T$ arises because, as discussed earlier in Sec. II A, fitting function $h(T)$ has been constructed to have a reduced convolution domain. This difficulty is resolved by recognizing that the rightward exponential growth function $f(T)$ can be mapped to a leftward exponential decay $f(T_{\mathrm{ref}} - T) \equiv f(\Theta)$ relative to the denaturation reference temperature $T_{\mathrm{ref}}$. The (reflected and offset) $\Theta$ space is marked in the subsidiary abscissa scale of Fig. 3(b). With this revised temperature mapping, random variables $[\Theta_1,\ \Theta_2,\ \Theta_3]$, respectively associated with probability distributions $[f(\Theta),\ g(\Theta),\ h(\Theta)]$, can be summed via $\Theta_3 = \Theta_1 + \Theta_2$, with statistics,

$$\langle \Theta_3 \rangle = \langle \Theta_1 \rangle + \langle \Theta_2 \rangle, \tag{12a}$$

$$\mathrm{var}[\Theta_3] = \mathrm{var}[\Theta_1] + \mathrm{var}[\Theta_2]. \tag{12b}$$

Thus, $\Theta$ provides an offset-temperature measure for enzyme catalysis, with denaturation point $T_{\mathrm{ref}}$ defining the absolute zero of enzyme activity. We now revisit and refine the probabilistic interpretation for MalL enzyme catalysis (as outlined above in the first paragraph of this section) by replacing physical temperature $T$ (in kelvin, K) with offset temperature $\Theta = (T_{\mathrm{ref}} - T)$, measured in units of offset-kelvin, $\mathrm{K_o}$, the

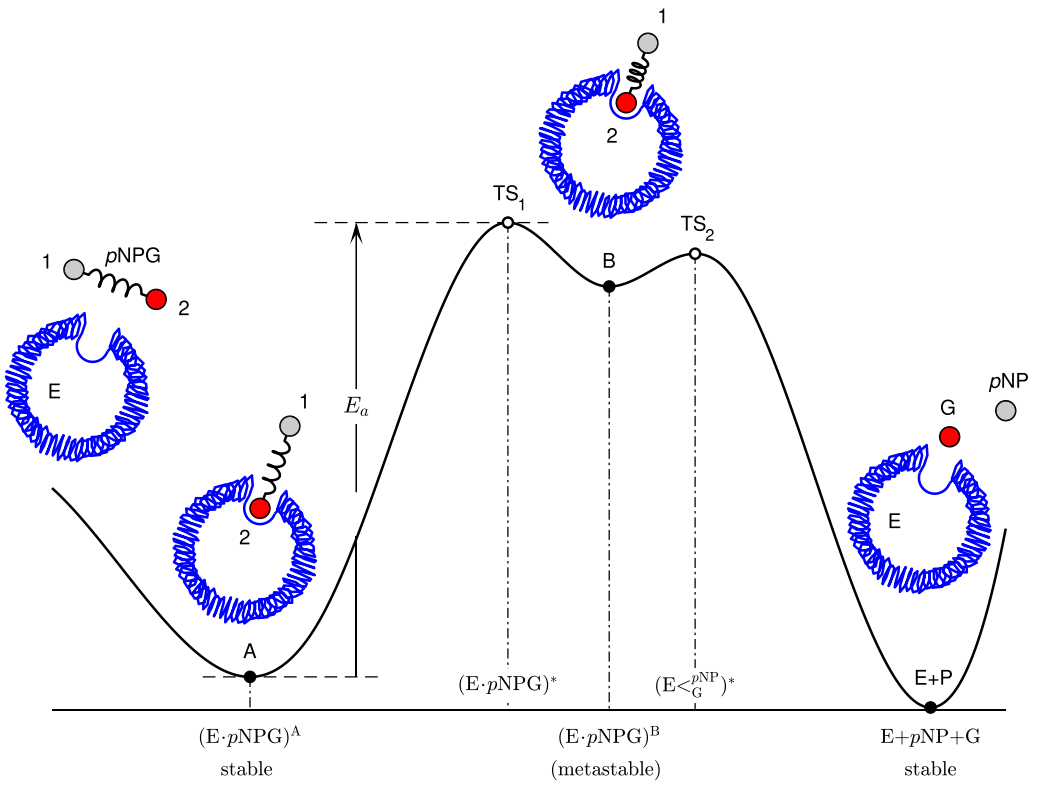



FIG. 5. Potential energy and enzyme cartoons for MalL enzyme-catalyzed (*p*NPG → glucose) sugar reaction as detailed in enzyme-flow sequence (13). Coiled spring represents glycosidic bond between *p*NP (1) and glucose (2) residues. The *p*NPG substrate bonds to MalL via lodgement of glucose residue within enzyme pocket to form bound-state A. Favorable conformational changes lead to the rare metastable-state B (tightly coiled spring), followed promptly by cleavage of glycosidic bond (spring disappears) and release of free glucose and *p*NP products. Key: $E_a$ = activation energy; E = MalL enzyme; *p*NPG = *p*-nitrophenyl-$\alpha$-D-glucopyranoside; *p*NP = *p*-nitrophenol; G = glucose; TS = transition state; E+P = enzyme + products; • = stable or metastable states (valley minima); ○ = unstable transition states (energy maxima). The implicit reaction coordinate runs from left to right. Cleavage of the glycosidic bond occurs during the (B → $TS_2$) step.

temperature displacement below the fully unfolded reference $T_{\rm ref}$, and write

$$f(\Theta_1) = \text{distribution of offset temperatures } \Theta_1 \text{ for A} \to \text{B transition across activation energy barrier,}$$
$$g(\Theta_2) = \text{distribution of } \Theta_2 \text{ offset-temperature fluctuations for short-lived metastable state,}$$
$$h(\Theta_3) = \text{distribution of } \Theta_3 = (\Theta_1 + \Theta_2) \text{ for A} \to \text{B transition } (\Theta_1) \text{ and metastable fluctuations } (\Theta_2).$$

Since $h(\Theta_3)$ is the probability density function computed from the $k(\Theta_3)$ enzyme turnover rate, we infer that the $h(\Theta_3)$ distribution gives the probability (per unit $K_o$) for MalL-catalyzed hydrolysis of substrate *p*NPG to release free glucose. We will argue below in Sec. II E that the $g(\Theta_2)$ Gaussian density corresponds to the distribution of single-particle temperature fluctuations from which we can estimate a lifetime for the metastable B state.

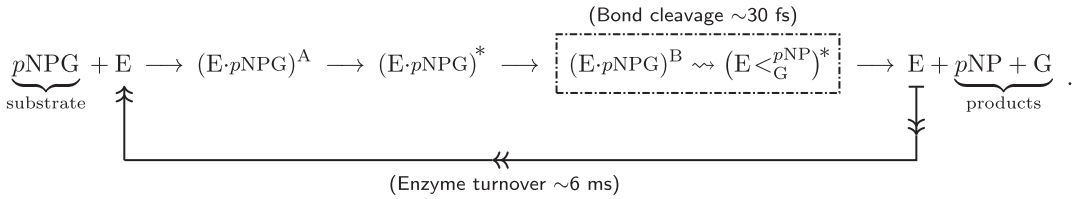
### D. Energy diagram for enzyme-catalyzed sugar cleavage

Figure 5 cartoons a potential energy diagram for the MalL-catalyzed cleavage of synthetic substrate *p*NPG (*p*-nitrophenyl-$\alpha$-D-glucopyranoside) into its constituent glucose and *p*-nitrophenol components. The energy diagram is a simplified representation of the full chain of enzyme-related events shown in sequence (13).

The substrate *p*NPG spontaneously bonds with the MalL enzyme (E) to form an enzyme-substrate complex (E·*p*NPG) that can exist in one of three forms: $(\text{E}\cdot p\text{NPG})^{\rm A}$ = stable ground-state A, $(\text{E}\cdot p\text{NPG})^{*}$ = unstable transition-state $TS_1$, and $(\text{E}\cdot p\text{NPG})^{\rm B}$ = metastable short-lived state B.

The transition from A to B requires sufficient energy to overcome the $E_a$ potential barrier to reach the $TS_1$ turning point representing formation of the $(\text{E}\cdot p\text{NPG})^{*}$ unstable complex; this then spontaneously decays to metastable-state B. We envisage that B is a short-lived state which terminates with the scission of the *p*NPG glycosidic bond to form the unstable $TS_2$ product complex $(\text{E}<_{\rm G}^{p\rm NP})^{*}$: see dot-dashed box

in enzyme-flow sequence (13):




(13)

The unbonded nitrophenol ($p$NP) and glucose (G) components are then released from the enzyme, completing the catalysis reaction and freeing the enzyme molecule to begin another reaction cycle. Note that flow sequence (13) shows only forward arrows since we are measuring initial reaction rates from a system with reactant only, so the reverse reaction is expected to be very slow and therefore negligible.

Figure 2(a) shows that at optimum temperature ∼318 K, each WT MalL enzyme molecule captures and cleaves about 170 $p$NPG molecules per second, implying an enzyme turnover period of $1/(170\ \mathrm{s}^{-1}) \approx 6$ ms. In contrast, the time required for bond cleavage (fracture of the glycosidic bridge that binds the glucose and nitrophenyl residues) is many orders of magnitude smaller. We can estimate the timescale for bond fracture by noting that infrared spectroscopy indicates stretching vibration resonances for the glycosidic bond lie within the frequency range (980–1160) $\mathrm{cm}^{-1} = (2.94$–$3.48) \times 10^{13}$ Hz (see Fig. 2(c) of Ref. [17]), corresponding to a bond oscillation period in the range $(28.8$–$34.0) \times 10^{-15}$ s. That is, in round numbers, the bond period is of order ∼30 fs. We assume that this bond oscillation period is also the time required to cleave the glycosidic bond.

This means that, within each 6-ms enzyme turnover cycle, the 30-fs bond cleavage event serves as a near instantaneous reset of the catalysis production engine, forcing a recycling through the states $\mathrm{A} \to \mathrm{B} \xrightarrow{\text{reset}} \mathrm{A}$.

We acknowledge that this two-state description is an oversimplification since it neglects both (a) the time required for the $(\mathrm{E}<_{\mathrm{G}}^{p\mathrm{NP}})^*$ activated complex to release the glucose and nitrophenol molecules at the end of the cycle and (b) the time required for the enzyme to bond to the $p$NPG substrate at the start of the cycle. However, the experiment is set up so the substrate is available in excess (concentration of $p$NPG is $\sim 10^5$ times greater than that of enzyme), so the latter neglect is not unreasonable.

This vast difference between the turnover and resetting timescales (6 ms : 30 fs $= 2 \times 10^{11} : 1$) implies a correspondingly vast difference in species abundance at any instant of time, i.e., for each enzyme complex in the B state, there will be $\sim 10^{11}$ enzyme-complex molecules in the A state [1]. Thus the B state is both extremely rare and extremely short-lived, when compared with the dominant A state. The fact that each B species has a femtosecond-scale lifetime means that there is no opportunity for it to come into thermodynamic equilibrium with the bath thermostat. This has significant implications for single-particle temperature fluctuations, as we now discuss.

### E. Molecular-scale temperature fluctuations

As pointed out by Landau and Lifshitz [18, pp. 8, 341], a thermodynamic quantity, such as temperature, will vary with time, fluctuating about its mean value. The relative size of these fluctuations will decrease rapidly as the size of the system (i.e., the number of particles) increases, so in the macroscopic limit the fluctuations become vanishingly small. However, we are interested here in the opposite case. How large do temperature fluctuations become when we shrink the system size to a single particle?

Consider an aqueous reaction chamber containing $N$ enzyme-substrate molecules, maintained at (mean) macroscopic temperature $T$ by a thermostat. We imagine inserting a sufficiently small thermometer that is able to sense $\tau$, the local temperature of $n$ of these molecules. Following Mazo [19], the mean value of $\tau$ is $\langle\tau\rangle = T$, and its variance is

$$\mathrm{var}[\tau] = \sigma_\tau^2 = \langle\tau^2\rangle - T^2 = \frac{k_B T^2}{n\, c_v} \approx \frac{k_B T^2}{n\, c_p}, \tag{14}$$

where $c_v$ is the per-molecule heat capacity at constant volume. Since water is essentially incompressible, we can replace $c_v$ with $c_p$, the molecular heat capacity at constant pressure. Equation (14) is a well-established result [18–21]. These fluctuations in local temperature will be Gaussian distributed about mean temperature $T$ with probability density

$$p(\tau) = \frac{1}{\sqrt{2\pi}\,\sigma_\tau}\, e^{-(\tau-T)^2/2\sigma_\tau^2}. \tag{15}$$

The reaction chamber used for our MalL enzyme experiments has a volume of 0.5 mL. A typical value for MalL concentration is $3.1 \times 10^{-8}$ mol/L, so the chamber contains $\sim 10^{13}$ enzyme molecules. In the macroscopic limit, $n \to N \sim 10^{13}$, the fluctuation variance in (14) becomes vanishingly small, and the density in (15) tends to a delta-function spike at $\tau = T$. At the opposite extreme, consider the single-molecule limit $n \to 1$ in (14). This limit gives the desired expression for maximal fluctuation variance in temperature:

$$\mathrm{var}[\tau] = \sigma_\tau^2 = \frac{k_B \langle\tau\rangle^2}{c_p} = \frac{k_B T^2}{c_p}. \tag{16}$$

Given the single-particle heat capacity $c_p$ of (16), we may also define Gaussian-distributed fluctuations in internal energy $U$ about mean value $\langle U\rangle$ with variance [18,20,22]:

$$\mathrm{var}[U] = \sigma_U^2 = k_B \langle\tau\rangle^2\, c_p = k_B T^2\, c_p. \tag{17}$$

(We will use this result later in Sec. II F to estimate the lifetime of metastable-state B.)

We assume that these Gaussian-distributed single-particle fluctuation statistics (15)–(17) describe the thermal characteristics of the extremely rare and fleeting B-state species $(\mathrm{E}{\cdot}p\mathrm{NPG})^{\mathrm{B}}$, which comes into existence one bond oscillation prior to the cleavage of the glycosidic bond to form $(\mathrm{E} <_{\mathrm{G}}^{p\mathrm{NP}})^{*}$ and thence product (nitrophenol and glucose). We further assume that the single-particle Gaussian (15) corresponds to the source Gaussian density function (6b) retrieved via $k(T)$ convolution fitting and illustrated in Fig. 4.

As described in Sec. II A and Eq. (6a), the A → B Arrhenius transition $[(\mathrm{E}{\cdot}p\mathrm{NPG})^{\mathrm{A}} \rightarrow (\mathrm{E}{\cdot}p\mathrm{NPG})^{\mathrm{B}}]$ is well approximated by a growing exponential function that is truncated at a suitably selected maximum temperature $T_{\mathrm{ref}}$. This is the temperature at which—within the ∼10-s assay timeframe—the MalL protein has become completely unfolded, and so marks the thermal point at which all enzyme activity ceases. This thermal limit varies with the MalL mutant type and pH setting.

By mapping from physical temperature $T$ to offset-temperature $\Theta = T_{\mathrm{ref}} - T$ [via (8) with $x \equiv \Theta$], we can rewrite Eqs. (9) in terms of independent random variables $\Theta_1$ (exponential decay) and $\Theta_2$ (Gaussian), and their sum $\Theta_1 + \Theta_2 = \Theta_3$ (EMG),

$$f(\Theta_1) = \lambda\, e^{-\lambda\Theta_1}\, u(\Theta_1), \tag{18a}$$

$$g(\Theta_2) = \frac{1}{\sqrt{2\pi\ \mathrm{var}[\Theta_2]}}\ \exp\left(-\frac{(\Theta_2 - \langle\Theta_2\rangle)^2}{2\ \mathrm{var}[\Theta_2]}\right), \tag{18b}$$

$$h(\Theta_3) = \lambda \exp\left(-\lambda(\Theta_3 - \langle\Theta_3\rangle) + \frac{\lambda^2\ \mathrm{var}[\Theta_3]}{2}\right) \times \frac{1}{2}\ \mathrm{erfc}\left[\frac{1}{\sqrt{2\ \mathrm{var}[\Theta_3]}}\big(-(\Theta_3 - \langle\Theta_3\rangle) + \lambda\ \mathrm{var}[\Theta_3]\big)\right], \tag{18c}$$

where the variance of the $g(\cdot)$ Gaussian in (18b) is given by

$$\mathrm{var}[\Theta_2] \ =\ \frac{k_B\,\langle\Theta_2\rangle^2}{c_p'}, \tag{19}$$

obtained from (16) with the mappings $\langle\tau\rangle \rightarrow \langle\Theta_2\rangle$ and $c_p \rightarrow c_p'$. Here, $c_p' = dU'/d\Theta_2$ is the single-particle heat capacity measured in what we refer to as the *enzyme frame of reference*, and $U'$ is the corresponding single-particle internal energy measured in this frame, i.e., as measured relative to unfolding temperature $T_{\mathrm{ref}}$ using $\Theta$ offset-temperature coordinates. Applying the same mappings to (17), we obtain an expression for the variance for the fluctuations in internal energy, as measured in the enzyme frame:

$$\mathrm{var}[U] = k_B\,\langle\tau\rangle^2 c_p \ \rightarrow\ \mathrm{var}[U'] = k_B\,\langle\Theta_2\rangle^2 c_p'. \tag{20}$$

Note that we have assumed that the fluctuation expressions (16) and (17) in $T$ space can be rewritten as fluctuation statements (19) and (20) relevant to offset-temperature $\Theta$ space.

### F. Lifetime estimation for metastable-state B

Located in a local energy minimum of Fig. 5, state B—representing enzyme-substrate configuration $(\mathrm{E}{\cdot}p\mathrm{NPG})^{\mathrm{B}}$—is locally stable, and so can be pictured as a well-defined single-particle quantum state which obeys Heisenberg's energy-time uncertainty principle,

$$\Delta E\ \Delta t\ \gtrsim\ \frac{\hbar}{2},$$

with $\hbar \equiv h/2\pi$; we take $\Delta E$ to be the full width at half maximum of the Gaussian distribution for the $U'$ internal energy fluctuations:

$$\Delta E\ =\ 2\sqrt{2\ln 2}\,\sigma_{U'}\ \approx\ 2.355\,\sigma_{U'}, \quad \text{with } (\sigma_{U'})^2 \equiv \mathrm{var}[U'].$$

By eliminating heat capacity $c_p'$ between fluctuation equations (19) and (20), we can express internal energy variance in terms of displaced-temperature variance,

$$\mathrm{var}[U'] = \frac{(k_B\langle\Theta_2\rangle^2)^2}{\mathrm{var}[\Theta_2]} \Rightarrow \sigma_{U'} = \frac{k_B\langle\Theta_2\rangle^2}{\sigma_{\Theta_2}} = \frac{k_B(T_{\mathrm{ref}} - \mu_T)^2}{\sigma_T},$$

where $\mu_T$ and $\sigma_T$ are the mean and standard deviations of the retrieved $g(T)$ source Gaussian expressed in temperature space. $T_{\mathrm{ref}}$ is the temperature at which the enzyme has become fully unfolded, and we locate this point at temperature $T_{\mathrm{ref}} = \mu_T + z\sigma_T$, i.e., at a point $z$ standard deviations above the Gaussian mean temperature $\mu_T$. A sensible value would be $z = 3$, but larger values could be considered. With this substitution for $T_{\mathrm{ref}}$, the $U'$ standard deviation simplifies to

$$\sigma_{U'}\ =\ k_B\, z^2\, \sigma_T\ \Rightarrow\ \Delta E\ \approx\ 2.355\, k_B\, z^2\, \sigma_T,$$

and, hence, from the uncertainty principle, we can deduce a lower bound for lifetime (in s) for the B-state resonance:

$$\Delta t\ \gtrsim\ \frac{\hbar}{2\,\Delta E} \approx \left(\frac{\hbar}{(2)(2.355)k_B}\right)\frac{1}{z^2\,\sigma_T} \approx \frac{(1.62\times 10^{-12}\ \mathrm{s\,K})}{z^2\,\sigma_T}. \tag{21}$$

For the WT MalL experiment shown in Fig. 4(a), the retrieved source Gaussian has standard deviation $\sigma_T = 5.42$ K. Setting $z = 3$ in Eq. (21) gives a B-state lifetime $\Delta t \gtrsim 33.2 \times 10^{-15}$ s, remarkably close to the 30 fs glycosidic bond oscillation period estimated earlier from infrared spectroscopy (see Sec. II D).

Table I summarizes the statistics for the source Gaussian profiles retrieved via convolutional fitting for each of the 15 enzyme experiments (five MalL mutants × three pH settings). Examining the $\sigma_T$ column for the exp-gauss (EMG) retrievals, the range of standard deviations is (4.87–8.93) K, leading to Eq. (21) lifetime predictions (for $z = 3$) in the range $\Delta t \sim$ (20.2–37.0) fs, with mean and standard errors:

$$\langle\Delta t\rangle_{\mathrm{EMG}}\ \gtrsim\ (29.0 \pm 1.3)\,\mathrm{fs}.$$

The Arrhenius-Gaussian (and Eyring-Gaussian) retrievals (column 7 of Table I) give $\sigma_T$ values that are consistently *smaller* (by ∼2%) than the EMG values, leading to B-state lifetime estimates that are slightly *larger*,

$$\langle\Delta t\rangle_{\mathrm{ArrhG}}\ \gtrsim\ (29.6 \pm 1.3)\,\mathrm{fs},$$

but both are in good agreement with the ∼30-fs spectral estimate for the resonance period of the glycosidic bond.

We have defined $T_{\mathrm{ref}}$ to be the temperature at which the MalL protein has become fully unfolded and therefore enzymatically inert. Because we reference $T_{\mathrm{ref}}$ to the $g(\cdot)$

TABLE II. Dependence of B-state lifetime $\Delta t$ on definition for the zero-activity reference temperature $T_{\text{ref}}$. $\Pr(Z > z)$ gives the probability that the enzyme is still functioning at temperatures $T > T_{\text{ref}}$. Here, $\sigma_T = 6.41$ K. Lifetime is expressed in femtoseconds, fs $= 10^{-15}$ s.

| $z$ | $T_{\text{ref}}$ | $\Pr(Z > z)$ | $\Delta t$ (fs) |
|---|---|---|---|
| 0 | $\mu_T$ | 0.5 | $\infty$ |
| 1 | $\mu_T + \sigma_T$ | 0.16 | 252 |
| 2 | $\mu_T + 2\sigma_T$ | 0.023 | 63.2 |
| 3 | $\mu_T + 3\sigma_T$ | 0.0013 | 28.1 |
| 4 | $\mu_T + 4\sigma_T$ | $31.7 \times 10^{-6}$ | 15.8 |
| 5 | $\mu_T + 5\sigma_T$ | $0.29 \times 10^{-6}$ | 10.1 |

fluctuation Gaussian via $T_{\text{ref}} = \mu_T + z\sigma_T$, the transition to inactivity is not abrupt, but rather is graded, following the upper wing of the Gaussian curve towards asymptotic zero. Setting $z = 3$, as we have done here, implies a tail probability $\Pr(Z > 3) = 0.13\%$ that the enzyme remains active at temperatures $T > T_{\text{ref}}$. The single-tail probability implications for other choices for $z$ are illustrated in Table II using the standard-normal curve identity:

$$\Pr(Z > z) \;=\; \tfrac{1}{2}\,\text{erfc}[z/\sqrt{2}].$$

Also shown are the corresponding Eq. (21) lifetime predictions for the 15-experiment mean value $\langle \sigma_T \rangle_{\text{EMG}} = 6.41$ K. It is clear from Table II that $z = 3$ is the *minimum* viable choice for setting the value of $T_{\text{ref}}$, the nominal zero-activity reference temperature. Choosing larger values for the $z$ offset leads to smaller estimates for B-state lifetime; these scale as the inverse-square of $z$: $\Delta t \propto 1/z^2$.

By invoking single-particle fluctuation theory, we are making the implicit assumption that the B-state transient exists in some degree of local thermal equilibrium on reactive timescales. Recent work on biological energy transport (e.g., Refs. [23–25]) shows that direct ballistic (nondiffusive) thermal transport can occur on picosecond timescales over nanometer distances. Thus, if the lifetime of a reactive intermediate falls within the range of several bond oscillations, then our assumption of local thermal equilibrium might become questionable.

Generalizing from Rubtsova and Burin [24] and Elenewski *et al.* [25], let us assume that ballistic thermal transport can occur along the protein backbone of the MalL macromolecule at speeds of order $v \sim 1\,\text{nm/ps} = 10\,\text{Å/ps}$. Based on crystallographic data (PDB 4M56), the WT MalL protein molecule has a volume (estimated via convex hull, `convhull` in MATLAB) of $V \approx 107 \times 10^3\,\text{Å}^3$, containing $N_{\text{atm}} \sim 4600$ atoms and $N_{\text{res}} = 561$ amino-acid residues. A localized thermal impulse could propagate ballistically until it encounters a scattering center, then more slowly via thermal diffusive processes thereafter. Assume the volume-averaged distance between residues is a measure of the thermal mean-free path, $d = (V/N_{\text{res}})^{\frac{1}{3}} \approx 5.8\,\text{Å}$. This implies a ballistic transit time of $\tau = d/v \approx 0.6\,\text{ps}$, about 20 times larger than the glycosidic bond oscillation period of 30 fs.

If the B-state lifetime is as small as $\Delta t \sim 30\,\text{fs}$, then there is no opportunity for ballistic (or diffusive) thermal transfer from the *p*NPG substrate to the bulk protein, suggesting that the single-particle fluctuation profile retrieved from the fitted Gaussian is describing thermal motion that is local to the C–O glycosidic linkage. That is, the fluctuating particle is the glycosidic linkage itself, rather than the cluster of nearest-neighbor residues that define the reaction pocket, or some larger residue domain. If, however, the lifetime were an order of magnitude larger, then the assumption of a truly localized thermal equilibrium would be less robust because of stronger coupling to the pocket and its residue surrounds. Nevertheless, this enlarged fluctuation domain could still be thermally uncoupled from the bath temperature given the fleeting nature of the B-state resonance.

## III. DISCUSSION

We have put forward a simplified two-state mechanism to describe the MalL-catalyzed sugar reaction that splits *p*NPG (*p*-nitrophenyl-$\alpha$-D-glucopyranoside) into its nitrophenol and glucose components by cleaving the glycosidic bond between them. As shown earlier in flow sequence (13), we propose that the reaction proceeds via a slow–fast process in which ground-state A transforms to metastable-state B on a millisecond timescale, whereupon state B promptly dissociates into products on the femtoseconds timescale of a single bond oscillation, allowing the enzyme molecule to reset to its ground state,

$$(\text{E}{\cdot}p\text{NPG})^{\text{A}} \xrightarrow{\sim 6\ \text{ms}} \boxed{(\text{E}{\cdot}p\text{NPG})^{\text{B}} \overset{\sim 30\ \text{fs}}{\rightsquigarrow} \left(\text{E}{<}^{p\text{NP}}_{\text{G}}\right)^{*}} \longrightarrow \text{E} + p\text{NP} + \text{G}$$

to repeat the catalytic cycle. We have noted previously that this idealized two-state picture ignores the time required for the release of products at the end of the cycle, and also the time it takes for the free enzyme E to bond to a nearby substrate molecule via $\text{E} + p\text{NPG} \rightarrow (\text{E}{\cdot}p\text{NPG})^{\text{A}}$ at the start of cycle; effectively, these steps have been subsumed into the slow $\text{A} \rightarrow \text{B}$ Arrhenius process.

The motivation for developing this two-state description comes from the fact that the experimental $k(T)$ rate curves for MalL-catalyzed cleavage of *p*NPG are well fitted by convolving a low-temperature Arrhenius-like exponential growth, appropriately bounded, with a high-temperature Gaussian distribution, giving an EMG function of temperature that exhibits slow acceleration to a peak, followed by rapid deceleration towards zero production at higher temperatures. This loss of enzyme function at higher temperatures is unsurprising given that the probability of protein unfolding (denaturation) rapidly increases as the melting point of the enzyme is approached and then exceeded. However, for each of the four MalL mutants shown in Fig. 4, the left-hand point of inflection for the $h(T)$ production curve occurs well before the onset of melting. In addition, we observed that the melt curves are considerably narrower than the $g(T)$ source Gaussians obtained from convolution fitting. These two observations suggest that protein denaturation is *not* the primary driver for the high-temperature rolloff in enzyme productivity, nor does it prescribe the width and location of the source Gaussian.

Each productivity experiment involves measuring the per-enzyme turnover rate ($\text{s}^{-1}$) at $n \sim 16$ temperature points (three technical replicates at each point) over the operational

range of the enzyme (∼5 to 55 °C). By applying an EMG convolution fit to these point measurements, then normalizing to unit area, we constructed $h(T)$, a probability density over temperature which gives the probability, per unit kelvin, of enzyme turnover at temperature $T$. In this way, the discrete $k(T_i)$ [with $i = 1 \ldots n$] point measurements have been mapped to a continuous density $h(T)$ with $T$ now being a continuous random variable. Similarly, the exponential and Gaussian source components of $h(T)$ are density functions over their respective random variable domains, which we distinguish with subscripts, writing $f(T_1)$, $g(T_2)$, $h(T_3)$.

To develop an interpretation for the convolution components, we invoked the fundamental convolution theorem of statistics for the addition of independent probability distributions, but this theorem could not be applied in temperature space because—by design—the fitting function $h(T)$ has a reduced convolution domain. By making the change of variable to offset-temperature $\Theta \equiv (T_{\text{ref}} - T)$, where $T_{\text{ref}}$ is the truncation temperature for the $f(T)$ exponential, we obtained new density functions $f(\Theta_1)$, $g(\Theta_2)$, $h(\Theta_3)$ whose random variables can be summed,

$$\Theta_3 = \Theta_1 + \Theta_2 \;\Rightarrow\; \langle\Theta_3\rangle = \langle\Theta_1\rangle + \langle\Theta_2\rangle, \quad \text{and}$$
$$\text{var}[\Theta_3] = \text{var}[\Theta_1] + \text{var}[\Theta_2],$$

where $\Theta = 0$ K$_\text{o}$ corresponds to $T = T_{\text{ref}}$, the temperature at which the protein becomes fully unfolded, and so represents an absolute zero of enzyme activity. The enzyme pool becomes more potent (less denatured) as $\Theta$ increases (i.e., as $T$ decreases below $T_{\text{ref}}$).

In this enzyme frame of reference, we proposed that exponential decay $f(\Theta_1)$ is the probability, per unit offset-kelvin, of an Arrhenius-like A → B transition across the activation energy barrier. Gaussian $g(\Theta_2)$ describes B-state single-particle fluctuations about mean value $\langle\Theta_2\rangle$ with variance var$[\Theta_2]$. This state is presumed to be both metastable and short-lived. The convolution of these two densities then gives the probability, per unit K$_\text{o}$, of enzyme turnover.

Assuming that the thermodynamic concepts of heat capacity $c_p$ and internal energy $U$ can be mapped to corresponding quantities $c_p'$ and $U'$ in $\Theta$ space, we reformulated the classical expressions for single-particle fluctuations in temperature (16) and energy (17) as offset-temperature analogs (19) and (20), respectively, and hence were able to derive estimates for the lower bound of the B-state lifetime (21). For the 15 mutant experiments reported in Table I, these minimum lifetimes range from $\Delta t = 20.2$ fs (V200S at pH 8) to 37.0 fs (S536R at pH 7), with 15-experiment mean $\langle\Delta t\rangle \gtrsim (29.0 \pm 1.3)$ fs, in remarkably good agreement with the ∼30-fs estimate for the period of glycosidic bond oscillations as computed from published IR spectroscopy [17,26].

This concordance between $\Theta$-space theory and IR spectroscopic measurement suggests that the two-state model for MalL enzyme catalysis has some predictive utility. Within the enzyme frame of reference, the B-state Gaussian fluctuation acts as an organizing center for catalysis by providing a metastable low-energy target that is unavailable to the naked (i.e., nonenzyme assisted) *p*NPG → glucose reaction. The metastable target provides an alternative axis for the reaction coordinate by presenting a much reduced activation energy barrier.

To illustrate the notion that—in the enzyme frame of reference—the distribution of B-state fluctuations act as a catalytic organizing center, in Fig. 6 we have plotted the 15 enzyme-rate experiments from three different perspectives: Figures 6(a) and 6(b) show measured $k(T)$ data points and EMG curve-fits in the laboratory frame of reference; Fig. 6(c) shows the same rate curves, but now expressed in $\Theta$ space where the $\Theta = 0$ K$_\text{o}$ origin corresponds to full-denaturation temperature $T = T_{\text{ref}}$; Fig. 6(d) converts the curves of Fig. 6(c) into probability densities, showing the set of source exponentials $f(\cdot)$, the source Gaussian $g(\cdot)$ (heavy black curve), and the EMG convolution resultants $h(\cdot)$. Here the abscissa has been expressed in $Z$-normal coordinates in which $Z_\Theta = (\Theta - \langle\Theta\rangle)/\sigma_\Theta$. In this space, all source Gaussians map to the standard normal, and the variation between enzyme probability densities lies entirely in the exponential and EMG components. As expected from EMG theory, the trailing edge of the Gaussian intersects each $h(\cdot)$ productivity curve at the peak of enzyme production.

The present paper assumes that the $\Delta t$ timescale for bond cleavage is paired to the stretching period of the glycosidic linkage. From the energy-time uncertainty principle, the $\Delta t \gtrsim$ 30 fs estimate gives a lower bound for the lifetime of the short-lived bond-breaking intermediate, but this (fast) timescale does not determine the overall catalytic rate. In our picture, the rate-determining step is the (slow) A → B Arrhenius transition from the enzyme-substrate complex (ground state) to the activated B-state intermediate, with this slow transition running on a millisecond timescale. A more complicated enzymatic reaction might require a sequence of several bond cleavages represented by distinct shallow energy minima (intermediates) along the reaction coordinate, each characterized by a distinct lifetime. Given sufficient thermal resolution in the measurements, it may be possible to discern these events as fine structure details in the $k(T)$ rate curve.

The analysis presented here has been entirely theoretical and based solely on the MalL–*p*NPG hydrolysis reaction, so it would be useful to learn if the convolution-fitting rationale is applicable to other enzyme-catalyzed reactions. But, moving beyond simple rate-curve fitting, one would like to verify (or invalidate) experimentally our claim that the source Gaussian derived from EMG fitting captures the single-particle temperature fluctuations of the short-lived metastable B-state enzyme-substrate complex. One obvious test would be to use an isotopically labeled substrate whose period of bond vibration is changed significantly. The resulting alteration in the zero point energy for the bond should manifest as a change in the width of the source Gaussian for the temperature fluctuations. Further, it might be possible to monitor the growth in kinetic energy fluctuations for individual enzyme–complex molecules, tagged with gold nanoparticles, as the complex transitions from its ground-state A towards metastable-state B using molecular imaging techniques such as fluorescence quenching [27].

Our assumption of a Gaussian fluctuation profile implies harmonic energy fluctuations about the local minimum of the B-state energy well. This is the simplest case. If there were a pair of closely spaced energy minima (corresponding to a pair

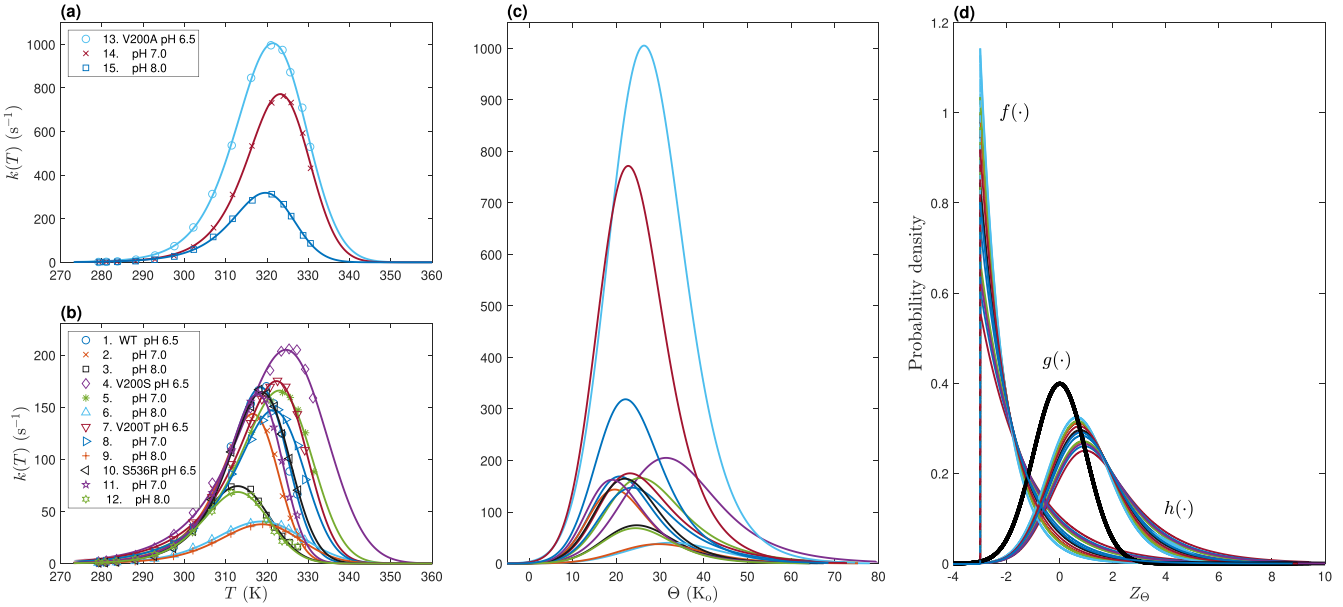



FIG. 6. Enzyme rate curves and probability density functions for all 15 MalL experiments listed in Table I. (a), (b) Measured production rates (discrete symbols) and $k(T)$ EMG convolution fits (smooth curves) for each experiment. V200A mutant productivities (experiments 13–15) are ∼5× higher than those of the other mutants, so are displayed in separate panel (a) to improve visibility. (c) Productivity curves expressed in the enzyme displaced-temperature frame of reference using $\Theta = (T_{\text{ref}} - T)$ coordinates. (d) Probability densities for the exponential $f(\cdot)$ and Gaussian $g(\cdot)$ source components, and the resultant exponentially modified Gaussian $h(\cdot)$, expressed in $Z_\Theta$ coordinates where $Z_\Theta = (\Theta - \langle\Theta\rangle)/\sigma_\Theta$. In this view, all source Gaussians collapse to the standard normal distribution.

of intermediate metastable states), then one might expect to see some non-Gaussian structure (e.g., broadening, multiple peaks, asymmetry) in the fluctuation spectrum with a shape dependent on the degree of mode coupling. In principle, one could form a curve fit to the $k(T)$ kinetics data by convolving an exponential (or Arrhenius) growth function against any suitable non-Gaussian target function, but this would only be feasible if the kinetics measurements had (i) sufficiently fine temperature sampling and (ii) adequate signal-to-noise quality to allow these second-order features to be resolved.

## APPENDIX: MEASUREMENT OF MALL ENZYME TURNOVER RATES

### 1. Protein expression and purification

MalL enzyme WT and variants (V200S, V200T, S536R, V200A) were produced by standard biochemistry protocols that yield a quantifiable solution of isolated protein in a defined buffer system. Specifically, expression systems of *Bacillus subtilis* MalL and the single amino-acid variants were set up in *E. coli* BL21 DE3 cells, genetically encoded within a pPROEX plasmid containing an N-terminal hexa-histidine tag. These cell lines were used for protein production via induction with 0.75 mM isopropyl $\beta$-D-1-thiogalactopyranoside of Luria broth cultures in exponential phase, followed by growth overnight at 18 °C.

Purification from the crude cell protein was carried out in two steps via immobilized metal affinity chromatography and size exclusion chromatography (IMAC and SEC, respectively) at pH 7.0. Initially, cell pellets were lysed via sonication on ice. Target protein was initially isolated by IMAC over a 50-mL imidazole gradient (25 mM to 0.5 M), utilizing the hexa-histidine tag to isolate target protein from cellular proteins. Further purification by size was carried out by SEC in 20-mM HEPES buffer. Enzymes were dialyzed into 40-mM $NaPO_4$ buffer with 150-mM NaCl (at the specified assay pH of 6.5, 7.0, or 8.0).

### 2. MalL temperature assays

Enzyme assays were specifically designed to measure the catalytic capacity of MalL over a wide temperature range, with the heating protocols and timing optimized to eliminate effects of enzyme denaturation prior to or during the measured assay. The KinetAsyst Stopped-Flow System (TgK Scientific, UK) with a connected circulating water bath for temperature control was used to characterize the catalytic rate of MalL via cleavage of saturating concentrations of the enzyme substrate *p*-nitrophenyl-$\alpha$-D-glucopyranoside (*p*NPG). In this system, the reaction is measured in real time by the absorbance at 405 nm of the formed reaction product, *p*-nitrophenol. Reactions were completed in triplicate with five 0.2-s dummy shots to clear the system between each reaction. Each reaction was carried out for 45 s with fresh enzyme (stored on ice), rapidly heated (0.2 seconds) prior to the assay commencing (through mixing of enzyme and substrate solutions). Temperature values reported are those from the thermostat control monitoring the reaction chamber. Stability of the enzyme master mix, stored on ice over the course of the experiment (∼5 h), was confirmed by a repeated midrange

temperature assay at the end of the experimental time period.

### 3. Turnover-rate calculation

Linear regression of (at most) the first 10 s of the reaction was carried out using Kinetic Studio (TgK Scientific, UK) to eliminate any denaturation effects over time. Observed catalytic rates $k(T)$ [$s^{-1}$] were determined using the molar extinction coefficient ($\epsilon$) specific to the given assay pH to convert from measured absorbance to *p*-nitrophenol concentration: $\epsilon = [4061, 7413, 13\,249]$ L $mol^{-1}$ $cm^{-1}$ for pH values [6.5, 7.0, 8.0], respectively.


[1] V. L. Schramm, Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes, Annu. Rev. Biochem. **80**, 703 (2011).

[2] G. Jindal and A. Warshel, Misunderstanding the preorganization concept can lead to confusions about the origin of enzyme catalysis, Proteins **85**, 2157 (2017).

[3] M. Garcia-Viloca, J. Gao, M. Karplus, and D. G. Truhlar, How enzymes work: Analysis by modern rate theory and computer simulations, Science **303**, 186 (2004).

[4] P. K. Agarwal, A biophysical perspective on enzyme catalysis, Biochemistry **58**, 438 (2019).

[5] S. Schönert, T. Buder, and M. K. Dahl, Identification and enzymatic characterization of the maltose-inducible $\alpha$-glucosidase MalL (sucrase-isomaltase-maltase) of Bacillus subtilis, J. Bacteriol. **180**, 2574 (1998).

[6] C. Lankatillake, S. Luo, M. Flavel, G. B. Lenon, H. Gill, T. Huynh, and D. A. Dias, Screening natural product extracts for potential enzyme inhibitors: Protocols, and the standardisation of the usage of blanks in $\alpha$-amylase, $\alpha$-glucosidase and lipase assays, Plant Methods **17**, 3 (2021).

[7] S. Arrhenius, On the reaction rate of the inversion of non-refined sugar upon souring, Z. Phys. Chem. **4**, 226 (1889).

[8] H. Eyring, The activated complex in chemical reactions, J. Chem. Phys. **3**, 107 (1935).

[9] J. K. Hobbs, W. Jiao, A. D. Easter, E. J. Parker, L. A. Schipper, and V. L. Arcus, Change in heat capacity for enzyme catalysis determines temperature dependence of enzyme catalyzed rates, ACS Chem. Biol. **8**, 2388 (2013).

[10] J. Åqvist, J. Sočan, and M. Purg, Hidden conformational states and strange temperature optima in enzyme catalysis, Biochemistry **59**, 3844 (2020).

[11] M. W. van der Kamp, E. J. Prentice, K. L. Kraakman, M. Connolly, A. J. Mulholland, and V. L. Arcus, Dynamical origins of heat capacity changes in enzyme-catalysed reactions, Nat. Commun. **9**, 1177 (2018).

[12] V. L. Arcus, E. J. Prentice, J. K. Hobbs, A. J. Mulholland, M. W. Van der Kamp, C. R. Pudney, E. J. Parker, and L. A. Schipper, On the temperature dependence of enzyme-catalyzed rates, Biochemistry **55**, 1681 (2016).

[13] H. A. Bunzel, H. Kries, L. Marchetti, C. Zeymer, P. R. E. Mittl, A. J. Mulholland, and D. Hilvert, Emergence of a negative activation heat capacity during evolution of a designed enzyme, J. Am. Chem. Soc. **141**, 11745 (2019).

[14] A. Golubev, Exponentially modified peak functions in biomedical sciences and related disciplines, Comput. Math. Meth. Med. **2017**, 1 (2017).

[15] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Series in Systems Science (McGraw-Hill, New York, 1965).

[16] J. Pitman, *Probability*, Springer Texts in Statistics (Springer-Verlag, New York, 1993).

[17] N. A. Nikonenko, D. K. Buslov, N. I. Sushko, and R. G. Zhbankov, Investigation of stretching vibrations of glycosidic linkages in disaccharides and polysaccharides with use of IR spectra deconvolution, Biopolymers **57**, 257 (2000).

[18] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd ed. (Pergamon Press, New York, 1980), Vol. 5.

[19] R. Mazo, Temperature fluctuations in thermodynamic equilibrium, Physica **25**, 57 (1959).

[20] S. J. Blundell and K. M. Blundell, *Concepts in Thermal Physics* (Oxford University Press, New York, 2009).

[21] H. Hernandez, Analysis of temperature fluctuations in ideal gases: From the macroscopic to the molecular scale, Tech. Rep., ForsChem Research, 2017.

[22] D. A. McQuarrie, *Statistical Mechanics* (Harper and Row, New York, 1976).

[23] X. Yu and D. M. Leitner, Vibrational energy transfer and heat conduction in a protein, J. Phys. Chem. B **107**, 1698 (2003).

[24] I. V. Rubtsov and A. L. Burin, Ballistic and diffusive vibrational energy transport in molecules, J. Chem. Phys. **150**, 020901 (2019).

[25] J. E. Elenewski, K. A. Velizhanin, and M. Zwolak, Topology, landscapes, and biomolecular energy transport, Nat. Commun. **10**, 4662 (2019).

[26] M. Kanou, K. Nakanishi, A. Hashimoto, and T. Kameoka, Influences of monosaccharides and its glycosidic linkage on infrared spectral characteristics of disaccharides in aqueous solutions. Appl. Spectrosc. **59**, 885 (2005).

[27] E. Hutter and D. Maysinger, Gold-nanoparticle-based biosensors for detection of enzyme activity, Trends Pharmacol. Sci. **34**, 497 (2013).