RESEARCH ARTICLE SUMMARY

ENZYME EVOLUTION

Parallel molecular mechanisms for enzyme temperature adaptation

Margaux M. Pinney*, Daniel A. Mokhtari, Eyal Akiva, Filip Yabukarski, David M. Sanchez, Ruibin Liang, Tzanko Doukov, Todd J. Martinez, Patricia C. Babbitt, Daniel Herschlag*

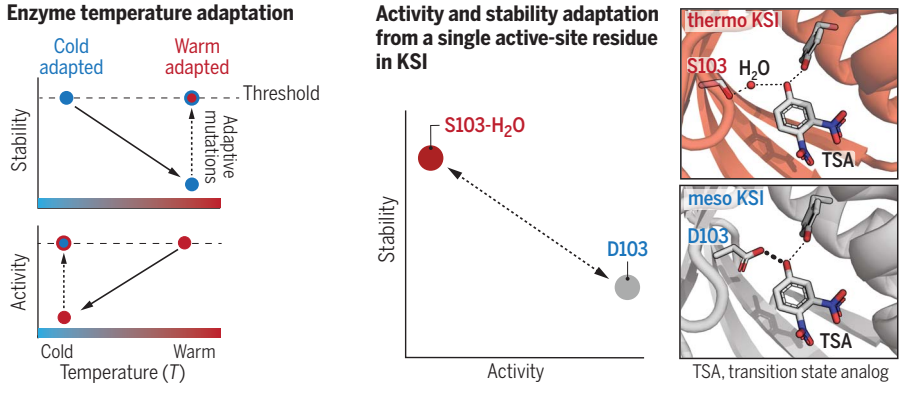
INTRODUCTION: Over billions of years, organisms and their enzymes have been evolving and adapting in response to selection pressures from their environments. In particular, livable temperature varies from about −15° to 121°C and exerts an evolutionary force that manifests in the adaptation of enzyme stabilities and activities: At increased temperatures, enzymes evolve greater stability to combat thermal denaturation and maintain a folded structure, whereas at decreased temperatures, nearly all chemical reactions necessary for life slow, placing evolutionary pressure on cold-adapted enzymes to be more active. An understanding of the molecular and evolutionary mechanisms that underlie enzyme temperature adaptation are integral to our understanding of how living systems have evolved and can reveal hidden aspects of how enzyme activity and stability is achieved, helping to define rules that can be used for enzyme design.

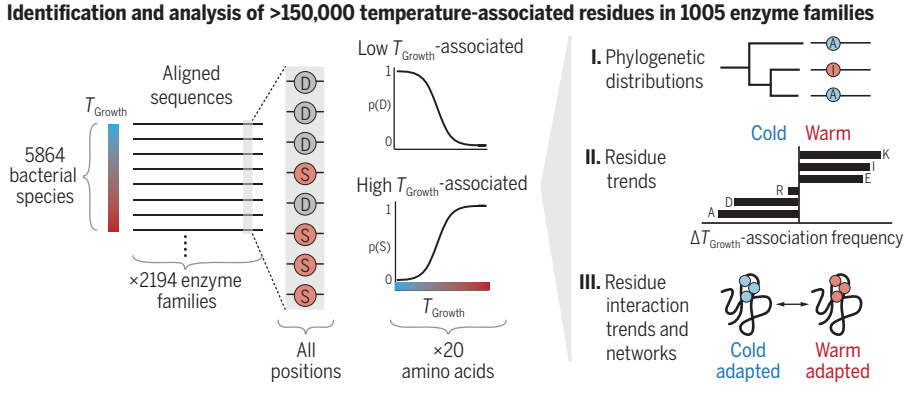
RATIONALE: We have dissected the molecular and evolutionary mechanisms underlying enzyme temperature adaptation both deeply and broadly. We first turned to the bacterial enzyme ketosteroid isomerase (KSI), combining mechanistic and structural studies with sequence and phylogenetic analyses to reveal the mechanisms underlying its activity and stability adaptation at the atomic and residue levels. Building on these results, we performed sequence and phylogenetic analyses, examining enzyme temperature adaptation in 2194 bacterial enzyme families to identify residue changes associated with growth temperature differences (referred to as "temperature-associated residues") and analyzing their physical properties and interactions.

RESULTS: We show that temperature adaptation in KSI arises primarily from a single active-site residue change with minimal epistasis. In cold-adapted KSI orthologs, a stronger active-site hydrogen-bond donor, protonated Asp^{103} (D103), improves activity. Conversely, warm-adapted KSI orthologs are stabilized by Ser^{103} (S103), which decreases activity but increases stability by removing the protonation-coupled folding of D103. Phylogenetic analyses showed that this active-site amino acid change (D103/S103) has occurred in diverse KSI sequence backgrounds from diverse bacteria, further supporting limited epistasis and suggesting parallel adaptation.

Our broad sequence and phylogenetic analyses revealed 158,184 statistically significant temperature-associated residues from 1005 enzyme families. Most of these residues are found in sequences from phylogenetically diverse bacteria, suggesting widespread temperature adaptation and parallel evolution. By mapping temperature-associated residues to structure, we found that these residues typically change with temperature on their own or with one other residue in physical contact, suggesting limited epistasis at these sites. Analyses of these temperature-associated residues reveal molecular and physical trends that test, hone, and revise nearly all prior mechanisms for enzyme temperature adaptation and identify networks of residues that appear to coadapt to temperature, perhaps cooperatively influencing catalysis stability, and/or allostery.

CONCLUSION: Our results broadly and deeply addressed enzyme temperature adaptation, revealing molecular mechanisms underlying the adaptation of KSI and identifying 158,184 temperature-associated residues; these data reveal physical trends and provide extensive data that can be further mined to understand molecular evolution and applied to enzyme design. These data further suggest that enzyme adaptation has repeatedly followed evolutionary paths of low epistasis, advancing our understanding of the evolutionary mechanisms that underly adaptation of nature's repertoire of enzymes. ■





Analyses of enzyme temperature adaptation. Enzymes adapt to low or high temperatures by modifying their activities or stabilities, respectively (top left). Temperature adaptation in KSI arises primarily from one active-site change (top right). Sequence analyses identify residues whose identity is associated with bacterial growth temperature (T_{Growth}) (bottom left). The phylogenetic distribution of these residues, their physical trends, and structural interactions were then analyzed (bottom right).

The list of author affiliations is available in the full article online.
*Corresponding author. Email: margauxp@stanford.edu (M.M.P.); herschla@stanford.edu (D.H.)
Cite this article as M. M. Pinney *et al.*, *Science* **371**, eaay2784 (2021). DOI: 10.1126/science.aay2784

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aay2784

RESEARCH ARTICLE

ENZYME EVOLUTION

Parallel molecular mechanisms for enzyme temperature adaptation

Margaux M. Pinney[1]*, Daniel A. Mokhtari[1], Eyal Akiva[2], Filip Yabukarski[1,3], David M. Sanchez[4,5], Ruibin Liang[4,5], Tzanko Doukov[6], Todd J. Martinez[4,5], Patricia C. Babbitt[2], Daniel Herschlag[1,7,8]*

The mechanisms that underly the adaptation of enzyme activities and stabilities to temperature are fundamental to our understanding of molecular evolution and how enzymes work. Here, we investigate the molecular and evolutionary mechanisms of enzyme temperature adaption, combining deep mechanistic studies with comprehensive sequence analyses of thousands of enzymes. We show that temperature adaptation in ketosteroid isomerase (KSI) arises primarily from one residue change with limited, local epistasis, and we establish the underlying physical mechanisms. This residue change occurs in diverse KSI backgrounds, suggesting parallel adaptation to temperature. We identify residues associated with organismal growth temperature across 1005 diverse bacterial enzyme families, suggesting widespread parallel adaptation to temperature. We assess the residue properties, molecular interactions, and interaction networks that appear to underly temperature adaptation.

Over billions of years, organisms and their enzymes have diversified and evolved in response to selection pressures and opportunities presented by their environments. The molecular mechanisms that underlie these adaptations and how these mechanisms shape the adaptive landscapes of enzymes are integral to how living systems have evolved. Conversely, evolutionary adaptations can reveal hidden aspects of enzyme function and contribute to our understanding of how enzymes work.

We address these questions through the lens of enzyme temperature adaptation. For most organisms, intracellular temperature matches their environment, requiring organisms and their enzymes to adapt over a range of about −15° to 121°C (*1*, *2*). Consequently, environmental temperature exerts a common and pervasive evolutionary force on enzymes, which manifests as temperatures of maximum activity that correlate strongly with the optimal growth temperatures (T_{Growth}) of their respective organisms (Fig. 1A and table S1) (*3*).

Changes in temperature directly influence the activities and stabilities of enzymes, presenting specific molecular challenges that must be overcome during adaptation. At increased temperatures, enzymes evolve greater stability to combat thermal denaturation and maintain a folded structure (*4*). At decreased temperatures, the rates of nearly all the chemical reactions necessary for life decrease, placing evolutionary pressure on cold-adapted enzymes to be more active. Consistent with this model, cold-adapted enzymes are often reported to be more active than their warm-adapted orthologs at a common temperature (*5*–*7*).

Enzyme temperature adaptation requires changes to the physical interactions responsible for activity and stability. For example, substitution of Thr^{26} (T26) in a low T_{Growth} ortholog of adenylate kinase (ADK) for the orthologous residue Ile^{26} (I26) found in many high T_{Growth} ADKs increased its melting temperature (T_m) by 8.1°C, apparently as a result of increased hydrophobic interactions with surrounding residues (*8*). Many molecular mechanisms for increasing enzyme stability have been described in the literature; however, the physical mechanisms underlying the reported increased activity of enzymes adapted to lower temperatures are less clear, with proposals to account for this behavior centered on enzymatic properties that correlate with organism T_{Growth}, such as enthalpy of activation ($\Delta H^{\ddagger}$), heat capacity of activation ($\Delta C_p^{\ddagger}$), or flexibility (*4*, *7*, *9*–*11*). Although these changes in enzyme behavior appear to be coincident with cold adaptation, their relationship to enhanced catalysis may be indirect or even coincidental, and the underlying physical processes remain elusive. Furthermore, different trends have been observed with different enzymes, complicating interpretation (supplementary text S1) (*9*, *11*, *12*).

The molecular mechanisms that are temperature adaptive in one enzyme sequence background may not be in another, depending on the extent that these mechanisms are influenced by specific physical interactions that are local or distributed. This context dependence is referred to as epistasis and is fundamental to our understanding of enzyme sequence-function landscapes (*13*). Although molecular epistasis has been widely discussed, the magnitude of its effects, the extent that epistasis is local versus distal, and the molecular mechanisms that underly epistasis remain to be broadly and deeply explored. Indeed, there is no consensus on the degree to which epistasis influences evolution, with some saying epistasis is the "primary factor" in protein evolution and others suggesting that its effects are limited (*13*–*15*). Furthermore, the extent of epistasis and the magnitude of epistatic effects may differ between sites depending on their interconnectivity and differ between enzymes depending on their functional and structural architecture, and the influence of epistasis on evolution will differ depending on the selective pressure(s) placed on the enzyme, organism, and population.

To dissect the molecular mechanisms of temperature adaptation and explore the extent to which epistasis has influenced the adaptive landscape of enzymes, we examined enzyme temperature adaptation both deeply and broadly. We performed deep mechanistic studies of the bacterial enzyme ketosteroid isomerase (KSI), revealing molecular mechanisms underlying its temperature adaptation. Notably, its stability and activity changes largely result from a single amino acid change, with epistasis limited to immediately surrounding residues. Phylogenetic analyses show that this amino acid change has occurred in diverse KSI backgrounds across bacteria, further supporting limited epistasis and suggesting parallel adaptation to temperature. To test the generality of this result and to identify molecular mechanisms of temperature adaptation in other enzymes, we carried out sequence and phylogenetic analyses for an additional 2194 bacterial enzyme families, identifying temperature-associated residues, trends in temperature-associated residue properties and interactions, and molecular mechanisms that appear to underlie their temperature adaptation for several case studies. We provide evidence that many temperature-associated residues are enriched in diverse bacteria with similar growth temperatures, suggesting parallel adaptation to temperature.

Results and discussion

Temperature adaptation of the activities and stabilities of KSI

We first turned to the enzyme KSI because the rich mechanistic and structural data from prior studies of this enzyme allowed for in-depth dissection of its temperature adaptation

[1]Department of Biochemistry, Stanford University, Stanford, CA 94305, USA. [2]Department of Bioengineering and Therapeutic Sciences and Quantitative Biosciences Institute, University of California, San Francisco, CA 94158, USA. [3]Chan Zuckerberg Biohub, San Francisco, CA 94110, USA. [4]Department of Chemistry, Stanford University, Stanford, CA 94305, USA. [5]Department of Photon Sciences, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [6]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [7]Department of Chemical Engineering, Stanford University, Stanford, CA 94305, USA. [8]Stanford ChEM-H, Stanford University, Stanford, CA 94305, USA.
*Corresponding author. Email: margauxp@stanford.edu (M.M.P.); herschla@stanford.edu (D.H.)

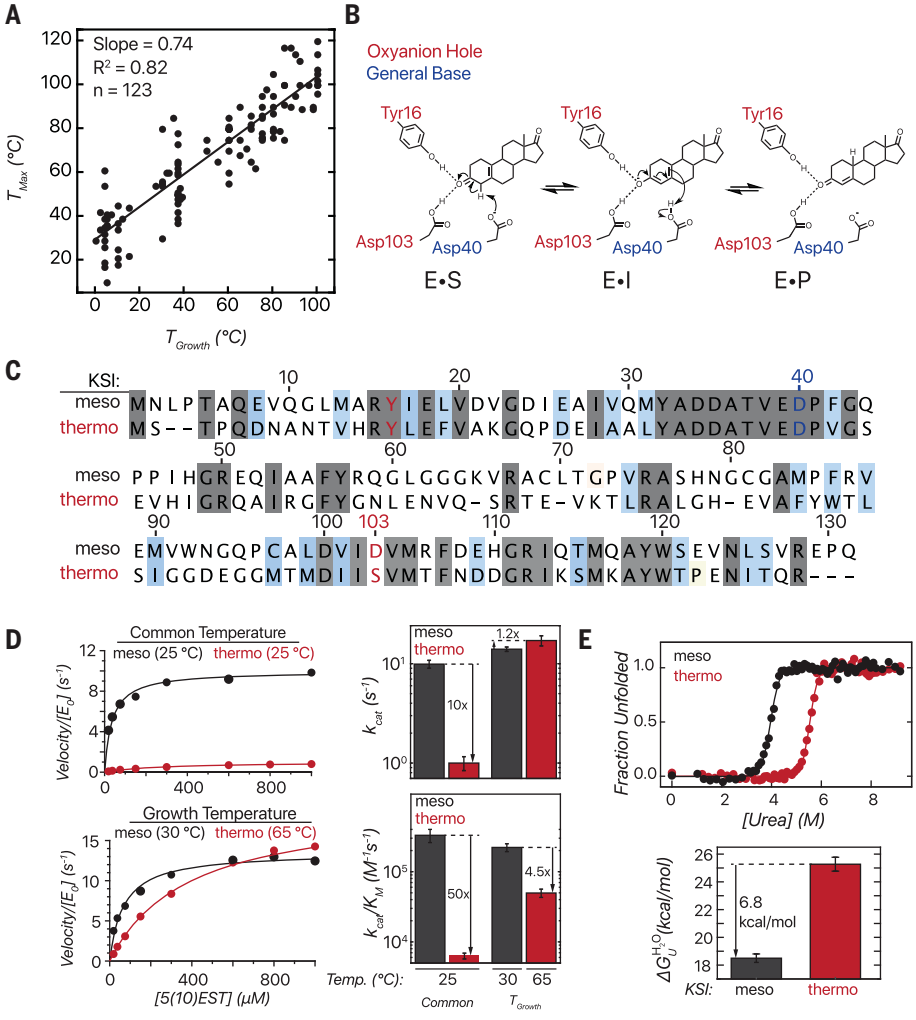


Fig. 1. Enzyme temperature adaptation through changes in activity and stability. (**A**) The temperature of the maximum enzymatic activity (T_{Max}) correlates with the median optimal growth temperature of its respective organism (T_{Growth}). Enzymes examined span all major enzyme classes and diverse cellular functions in both prokaryotes and eukaryotes (table S1). R^2, coefficient of determination. (**B**) The mechanism of isomerization of the steroid 5(10)-estren-3,17-dione [5(10)EST] by KSI, with residues and numbering from the *P. putida* ortholog (*16*). E, enzyme; S, substrate; I, intermediate; P, product. (**C**) Alignment of KSI sequences from *P. putida* (mesoKSI) and *M. hassiacum* (thermoKSI) (33% identical). Identical residues are highlighted in gray, and similar residues are highlighted in blue; the oxyanion hole hydrogen-bond donors Y16 and D103 (mesoKSI) or S103 (thermoKSI) are highlighted in red. MesoKSI numbering is used throughout for simplicity. (**D**) Activities (k_{cat} and k_{cat}/K_M) of mesoKSI (black) and thermoKSI (red) with 5(10)EST at a common temperature (25°C) and at their respective growth temperatures (30° and 65°C). See table S2 for compiled mesoKSI and thermoKSI activities. E_0, initial enzyme concentration. (**E**) Representative urea denaturation curves for mesoKSI (black) and thermoKSI (red) monitored by changes in internal tryptophan fluorescence (top) and stabilities extrapolated to 0 M urea ($\Delta G_U^{H_2O}$) (bottom). See table S3 for compiled mesoKSI and thermoKSI stabilities. Error bars in (D) and (E) represent ±1 SD from two to seven independent experiments. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

(*16*–*18*). In a subset of bacteria, KSI is involved in a metabolic pathway that allows steroids to be used as carbon and energy sources (*19*). KSI catalyzes the isomerization of double bonds in steroid substrates that use a general base and an oxyanion hole, most often composed of a tyrosine and a protonated aspartic acid that donate hydrogen bonds to stabilize the increased negative charge on the transition states and intermediate (Fig. 1B). A mesophilic variant of KSI from *Pseudomonas putida* [mesoKSI, T_{Growth} = 30°C (*20*)] has been extensively studied (*16*–*18*). For comparison, we searched for thermophilic mesoKSI orthologs using BLAST and identified a single candidate (thermoKSI, 33% identical to mesoKSI) from *Mycobacterium hassiacum*, which can grow at 65°C (Fig. 1C) (*21*, *22*).

We began by determining the activities, represented by the Michaelis-Menten parameters k_{cat} and k_{cat}/K_M, of mesoKSI and thermoKSI. At a common temperature, 25°C, the mesoKSI-catalyzed reaction was faster than the thermoKSI-catalyzed reaction by 10-fold in k_{cat} (reaction rate with saturating substrate) and 50-fold in k_{cat}/K_M (reaction rate with subsaturating substrate), presumably reflecting the lower T_{Growth} of the mesophile and increased evolutionary pressure on mesoKSI activity (Fig. 1D and table S2) (*7*, *9*–*11*). This observation and the similar activities of mesoKSI and thermoKSI at their respective T_{Growth} values are consistent with reports for several other thermally adapted enzymes (Fig. 1D) (*7*, *9*–*11*). Moreover, thermoKSI was 6.8 kcal/mol more stable than mesoKSI, consistent with the higher T_{Growth} of thermoKSI and an increased evolutionary pressure for higher stability (Fig. 1E and table S3) (*7*, *9*–*11*). In contrast to prior proposals for enzyme temperature adaptation, mesoKSI and thermoKSI activities had indistinguishable $\Delta H^{\ddagger}$ values and near-zero values of $\Delta C_p^{\ddagger}$ (fig. S1 and supplementary text S1).

Differences in active-site residues between temperature-adapted KSIs

To identify the molecular mechanisms responsible for the activity and stability differences between mesoKSI and thermoKSI, we first solved the 1.25-Å resolution x-ray crystal structure of thermoKSI with a bound transition state analog, 3,4-dinitrophenolate (DNP) (Fig. 2A, fig. S2, and tables S4 and S5). The structure of thermoKSI•DNP was highly similar to mesoKSI•DNP [root mean square deviation (RMSD) = 0.76 Å for backbone atoms], despite sharing just 33% sequence identity (Fig. 2A). Intriguingly, we observed a change in a key active-site residue: Where mesoKSI has an aspartic acid hydrogen-bond donor at position 103, thermoKSI has a serine, S103, which binds a water molecule that acts as the oxyanion hole hydrogen-bond donor (Figs. 1C and 2B; mesoKSI numbering is used throughout for simplicity). Furthermore, the residues contacting position 103 in mesoKSI, Phe^{86} (F86) and Val^{101} (V101), were substituted by Trp^{86} (W86) and I101 in thermoKSI, and there were additional differences in the residues surrounding positions 86 and 101 (Fig. 2B and fig. S3).

Activity and stability adaptation in KSI arise predominantly from a single active-site change

Asp^{103} (D103) is expected to be a stronger hydrogen-bond donor than the S103-bound water owing to the increased polarity of a carboxylic acid relative to a hydroxyl donor (Fig. 2C and supplementary text S2) (*18*). This difference in hydrogen-bond donor strength could be responsible for the activity difference

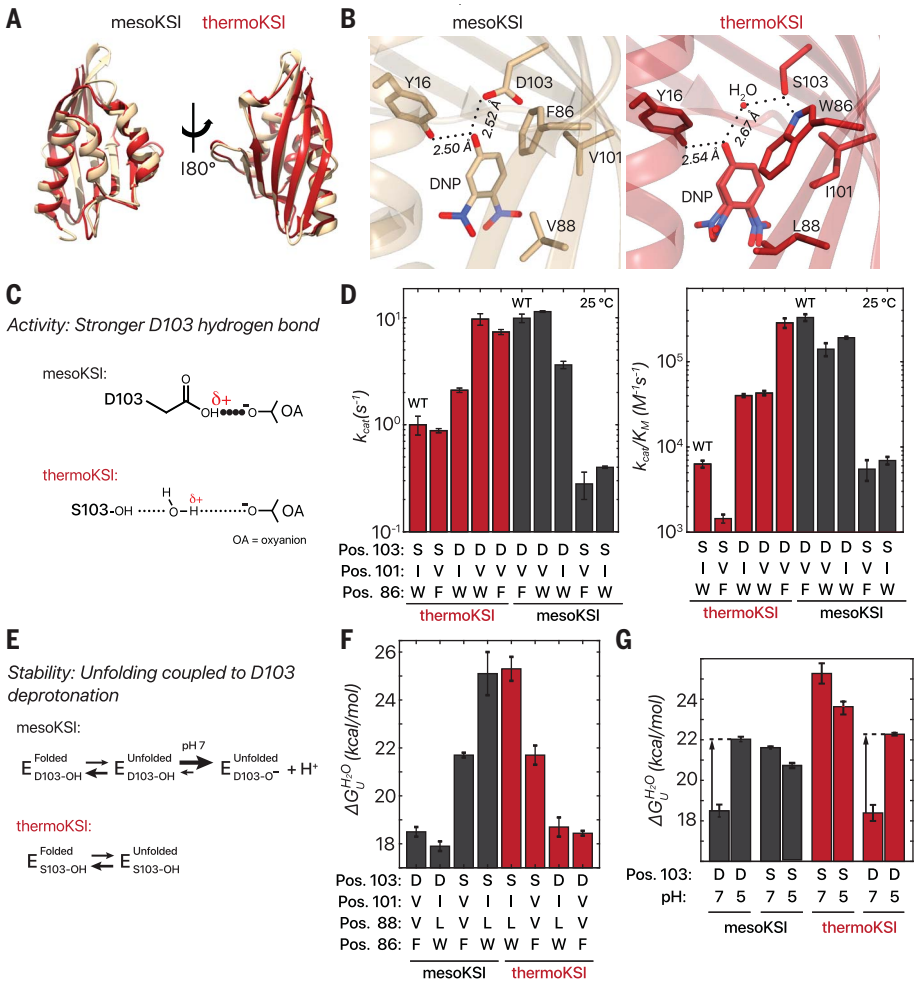


Fig. 2. Interconversion of D103 and S103, enhanced by one to three additional mutations to neighboring residues, accounts for differences in activity and stability between mesoKSI and thermoKSI. (**A**) The x-ray crystal structure (1.25 Å) of thermoKSI (red, PDB ID 6P44) bound to the transition state analog, DNP, aligned globally to the structure of mesoKSI bound to DNP (tan, PDB ID 6C17, RMSD = 0.76 Å for backbone atoms). (**B**) The active-site residues F86, V88, V101, and D103 in mesoKSI are replaced by W86, L88, I101, and S103 in thermoKSI, accompanied by the S103-bound water molecule. Dotted lines indicate hydrogen bonds. (**C**) D103 is expected to enhance activity relative to the S103-bound water owing to a higher partial positive charge density on the hydrogen-bond donor that interacts with the negatively charged transition states (supplementary text S2). (**D**) Reciprocal mutation of residues at positions 86, 101, and D103 interconvert the catalytic parameters, k_{cat} and k_{cat}/K_M, of mesoKSI and thermoKSI. (**E**) Coupled unfolding and D103 deprotonation in mesoKSI is predicted to destabilize its folded state, relative to thermoKSI, and result in a pH dependence in stability. (**F**) Reciprocal mutation at positions 86, 88, 101, and D103 interconvert the stabilities of mesoKSI and thermoKSI. (**G**) D103-containing variants of mesoKSI and thermoKSI are more stable at pH 5 than at pH 7, but S103-containing variants are not. Error bars in (D), (F), and (G) represent ±1 SD from two to seven independent experiments.

between mesoKSI and thermoKSI. Consistent with this model, the hydrogen bond between the S103-bound water and DNP is 0.15 Å longer than the D103-DNP hydrogen bond, a lengthening that predicts the weakening of this hydrogen bond and the observed 10-fold difference in activity between mesoKSI and thermoKSI based on systematic studies of KSI active-site hydrogen-bond energetics (Fig. 2B, fig. S4, and supplementary text S2) (*18*). Next, the hydrogen bonds donated by D103 versus S103-bound water were further investigated using molecular mechanics (MM) and quantum mechanics and molecular mechanics (QM/MM) calculations (see Materials and methods; figs. S5 and S6; and supplementary text S3). Consistent with D103 as a stronger hydrogen-bond donor, computed D103-DNP hydrogen-bond lengths were, on average, 0.11 Å shorter than the hydrogen bond between the S103-bound water and DNP in both mesoKSI and thermoKSI backgrounds. Finally, the Asp^{103}→Ser (D103S) mutation in mesoKSI decreased k_{cat} and k_{cat}/K_M, by 36- and 50-fold, respectively, and, more persuasively, the S103D mutation in thermoKSI increased activity by two- and sixfold in k_{cat} and k_{cat}/K_M, respectively (Fig. 2D).

To test if the remaining differences between mesoKSI and thermoKSI activities resulted from local epistasis, we made substitutions at two positions contacting position 103, positions 86 and 101. The I101V mutation in combination with S103D increased the activity of thermoKSI by 10-fold in k_{cat} (Fig. 2D). Adding in the W86F mutation increased k_{cat}/K_M by an additional fivefold, bringing the activity of thermoKSI to wild-type (WT) mesoKSI levels (Fig. 2D). Without the S103D mutation, the W86F/I101V mutations had no effect on thermoKSI k_{cat} and were deleterious to k_{cat}/K_M, indicating epistasis that presumably arises from the physical interactions between D103 and V101 and the interactions between W/F86 and the substrate (Fig. 2D and figs. S3 and S7). Conversely, mesoKSI D103S activity achieves WT thermoKSI levels by introduction of a tryptophan at position 86 along with S103 (mesoKSI F86W/D103S) (Fig. 2, B and D). These results are consistent with MM and QM/MM simulations that show that the addition of the F86W mutation stabilizes the native, active conformation of S103, limiting alternative rotamers of S103 that destabilize the native oxyanion hole configuration (supplementary text S3 and fig. S6). Thus, substituting the hydrogen-bond donor at position 103, enhanced by one or two neighboring mutations, quantitatively interconverts mesoKSI and thermoKSI activities, despite the overall sequence identity of only 33% for these enzymes.

These results suggest that the adaptation of KSI activity occurs through primarily one amino acid change, with modest local epistasis. To further test the model of limited epistatic effects on D/S103 mutations, we carried out analogous activity measurements for four additional KSI orthologs and their mutants (fig. S8). In all backgrounds, the D103 variants were faster than the corresponding S103 variants, regardless of whether the WT sequence originally contained D103 or S103 and despite sequence identities ranging from 28 to 58% (fig. S8).

We further predicted that the change from D103 to S103 would account for the observed stability difference between mesoKSI and thermoKSI. The pK_a of D103 (where K_a is the acid dissociation constant) is considerably perturbed (p$K_a^{D103} \geq 9$) from that of a solvent-exposed aspartate side chain [p$K_a^{Asp} \cong 3.7$ (*23*)], presumably the result of the surrounding hydrophobic environment and its proximity to the anionic D40 general base (Fig. 1B and fig. S3) (*16*). Thus, for mesoKSI to fold at pH 7, D103 must be protonated, an additional,

unfavorable step at neutral pH that would destabilize the enzyme (Fig. 2E, fig. S9, and supplementary text S4). This extra protonation step is absent with S103 so that, all else being equal, S103-containing KSIs would be more stable than D103-containing variants at pH 7, whereas S103- and D103-containing KSIs would have similar stabilities at low pH where the protonation of the D103 side chain is favored (Fig. 2E). Both of these thermodynamic predictions are met. First, S103-containing KSI variants (mesoKSI D103S and thermoKSI WT) were more stable than their corresponding D103-containing variants (mesoKSI WT and thermoKSI S103D; Fig. 2F). Second, D103-containing mesoKSI and thermoKSI variants were more stable at pH 5 than at pH 7, whereas the S103-containing variants were not (Fig. 2G). This result is the first evidence supporting the suggestion, made more than two decades ago by Shoichet and Matthews, that the removal of residues with perturbed pK_a values could be an evolutionary strategy to increase the stability of thermophilic enzymes in cases where the catalytic cost is not too high (*24*).

As we did for catalysis, we explored epistatic effects on stability by swapping residues at positions surrounding position 103. Specifically, combined substitution of thermoKSI residues at positions 86, 88, and 101 increased mesoKSI D103S stability by an additional 3.4 kcal/mol but had no stability effects in the mesoKSI D103 background (Fig. 2F). Taken together, these changes quantitatively accounted for the difference in mesoKSI and thermoKSI stability (Fig. 2F). Thus, despite 33% overall sequence identity, epistatic effects on stability were highly localized, because just three local substitutions, plus position 103, were sufficient to interconvert stabilities.

Favorable mutational pathways between temperature-adapted KSIs

Epistasis between residues can restrict the molecular outcomes and mutational pathways of proteins (*25*). We predicted that the probability of KSI temperature adaptation through substitutions at position 103 and surrounding residues would be greater than other potential solutions that require more amino acid changes, provided that there is a pathway between temperature-adapted states without highly unfit intermediates. To investigate the mutational pathways for KSI temperature adaptation, we performed all possible combinations of residue swaps at positions 86, 101, and 103 for activity and at positions 86, 88, 101, and 103 for stability in both mesoKSI and thermoKSI backgrounds. For both activity or stability, we observed mutational paths in mesoKSI and thermoKSI backgrounds without unfit intermediates, suggesting available pathways to temperature adaptation (figs. S10 and S11).

We additionally considered changes at the nucleotide level, because two nucleotide changes are needed to interconvert F/W and D/S codons in bacteria: F86W proceeds through Leu^{86} (L86) or Cys^{86} (C86) intermediates and D103S proceeds through Ala^{103} (A103), Gly^{103} (G103), Asn^{103} (N103), or Tyr^{103} (Y103) intermediates (figs. S12 and S13). Of these additional amino acid intermediates, Y103 and C86 were highly deleterious, decreasing stability by 3.9 to 11.5 kcal/mol (fig. S13). As predicted based on these large deleterious effects, sequences with Y103 or C86 were not observed in the 1140 KSI genomic and metagenomic sequences, whereas the remaining potential codon intermediates were observed, consistent with natural variants avoiding large decreases in stability (table S6). Intriguingly, N103, a hydrogen-bond donor similar to S103-H_2O, is disfavored in natural KSI variants, perhaps because N103 is less active than D103 and less stable than S103 (figs. S12 to S16); other factors such as environmental nitrogen limitation and/or Asn deamidation could also play a role (*26*, *27*).

An interplay between selective pressures in temperature adaptation of protonated Asp and Glu residues

In substituting protonated D103 for S103, KSIs sacrifice a 10-fold advantage in k_{cat} for a 6.8 kcal/mol increase in stability. Basic thermodynamic principles predict that protonated Asp and Glu residues will be similarly destabilizing across different enzymes, whereas their effect on activity will depend on their catalytic role. Thus, we expect that protonated Asp and Glu (Asp/Glu-OH) residues would be substituted by other, more stabilizing residues in high-T_{Growth} orthologs, contingent on their importance for activity. This trend emerges for the three enzyme families with Asp/Glu-OH residues identified in the literature with both stability and activity data. The Glu-OH in glycosidases, a general acid directly involved in catalysis, gives large rate decreases when mutated and is conserved at all values of T_{Growth} (fig. S14). An Asp-OH in the active site of thioredoxin has an activity effect of 10- to 100-fold and is also conserved across T_{Growth} (fig. S14). By contrast, the Glu-OH in triosephosphate isomerase is located far from the active site at the dimer interface, has no catalytic effect when mutated to Gln, and is not found in high-T_{Growth} organisms (above 40°C; fig. S14).

Although we observed an activity-stability trade-off in KSI and such trade-offs are often proposed to be widespread in enzyme temperature adaptation, there are many compelling examples of changes that enhance stability without resulting in deleterious functional effects (*9*, *10*, *28*). Our observations for temperature adaptation involving Asp/Glu-OH residues demonstrate the basic interplay between competing selection pressures placed on enzymes: Substitution of Asp/Glu-OH in temperature adaptation occurs when the catalytic sacrifice is not too large.

Given the limited direct evidence for Asp/Glu-OH residues in the literature, we searched the Protein Data Bank (PDB) for Asp or Glu residues predicted to be protonated based on their local environment (see Materials and methods). Of the 67 putative Asp/Glu-OH residues identified, 14 decrease in abundance in high-T_{Growth} orthologs (table S7). We predict that these 14 Asp/Glu-OH residues will have smaller functional roles than the remaining 53 Asp/Glu-OH residues. Other factors may also be at play, including whether there are traversable mutational pathways and the extent of epistasis in these systems, and exceptions to this trend may help us understand the molecular underpinnings of these additional factors.

Sequence and phylogenetic analysis suggest parallel adaptation to temperature in KSI

KSI temperature adaptation, in both activity and stability, can occur through primarily one amino acid change, D103 versus S103, with minimal, local epistasis and along mutational paths without highly unfit intermediates between temperature-adapted states. Most simply, our results predict that the presence of D or S at position 103 will depend on the temperature that the organism is adapted to, with a preponderance of D103 in low-T_{Growth} organisms imparting higher activity, and a preponderance of S103 in high-T_{Growth} organisms imparting greater stability for KSI. Considering 1140 KSIs from genomic and metagenomic sources, these data meet this prediction: The fraction of KSI sequences with D103 decreased with T_{Growth}, whereas the fraction of sequences with S103 increased with T_{Growth}, such that D103 is favored below 40°C and S103 is favored above 50°C (Fig. 3). Analysis of the phylogenetic tree of KSI sequences shows that D103 and S103 occur in diverse KSI sequences and diverse bacteria, suggesting that changes between these residues have occurred multiple times during KSI evolution, as expected for parallel temperature adaptation (i.e., common residue acquisition or maintenance in diverse bacteria with similar values of T_{Growth}) and in line with the minimal epistasis observed experimentally (figs. S15 and S16).

Identifying tens of thousands of temperature-associated residues in thousands of enzyme families

To broadly explore enzyme temperature adaptation, we searched each position in thousands of sequences from thousands of enzyme families to determine if the same residues at these positions are enriched in orthologs from diverse bacteria with similar values of T_{Growth}.

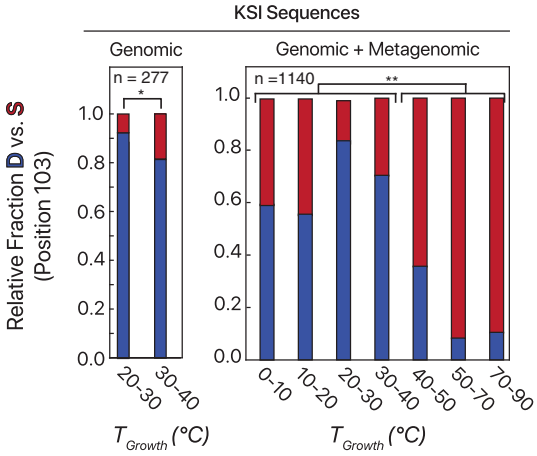


Fig. 3. D103 and S103 are favored in diverse low- and high-T_{Growth} sequences, respectively. Relative fraction of D103 versus S103 in KSI sequences from bacterial genomes (left) and combined genomic and metagenomic sources (right) at different values of T_{Growth} (*p < 0.01 and **p < 0.001 by chi-square test). For metagenomic sequences, where the species is unknown, we used environmental temperatures as an estimate for T_{Growth} (see Materials and methods). Intriguingly, D103 appears less favored when T_{Growth} < 20°C. Perhaps the relative stability of KSI also decreases at low temperatures because of cold denaturation (*99*). Alternatively, additional selection pressures may be present in the environments and organisms sampled during these low-temperature metagenomic sequencing experiments, such as increased pH, decreasing the preference for D103.

These systematic analyses have suggested widespread parallel adaptation to temperature, allowed us to test and refine prior models for temperature adaptation, and allowed us to construct a rich dataset for future mechanistic dissection.

To identify residues that are enriched in lineages with low and high T_{Growth}, we computed associations between the presence of each of the 20 amino acids and T_{Growth} at each position in each of 2194 enzyme families across 5852 different bacterial species, for a total of 1.7×10^7 positional comparisons (Fig. 4, A to C). Notably, nearly half of these enzyme families (46%; 1005 of the 2194) had one or more positions where the presence of a specific amino acid correlates with either increasing or decreasing T_{Growth}, for a total of 158,184 statistically significant residue positions after correction for multiple hypothesis testing (Fig. 4C and figs. S17 and S18). We refer to these as temperature-associated residues. Temperature adaptation of the remaining 1189 enzymes may be more diverse and perhaps lineage-specific, which future analyses at different evolutionary depths will help elucidate.

Temperature-associated residues suggest widespread parallel adaptation

We focused on bacterial enzymes, because there are a myriad of sequences available from diverse bacteria that have undergone temperature adaptation independently to achieve growth temperatures ranging from 4° to 80°C (fig. S19). This large dataset enabled us to distinguish residues that are enriched in diverse bacteria with similar values of T_{Growth}, as expected for parallel adaptation, from clade-specific changes that may require a specific local sequence background to be adaptive, or to low-probability solutions to temperature adaptation.

To distinguish between these possibilities, we mapped each temperature-associated residue onto the phylogenetic tree of bacterial species and computed its phylogenetic signal ($D_{Residue}$), which measures the degree to which a residue is clustered versus dispersed on a phylogenetic tree; low values of $D_{Residue}$ correspond to residues clustered within a narrow clade, and high values of $D_{Residue}$ correspond to residues distributed across many clades (Fig. 4D and figs. S20 and S21) (*29*). We observed that most temperature-associated residues (51%) had $D_{Residue}$ scores indicative of dispersal throughout the phylogenetic tree, rather than extensive clustering within a particular clade (Fig. 4, D and E). These data suggest parallel adaptation to temperature through the acquisition and/or maintenance of temperature-associated residues along diverse bacterial lineages with similar values of T_{Growth} (fig. S22). As an example, Ile residues at position 452 of phosphate acetyltransferase are distributed across the bacterial tree and strongly enriched in diverse high-T_{Growth} organisms, with I452 present in all but one bacterial species with T_{Growth} values above 56°C (Fig. 4, C and F). Because this residue is located at the enzyme's dimer interface, I452 may stabilize intersubunit association (Fig. 4C). The enrichment and dispersal of most of our 158,184 temperature-associated residues further suggests that these residues represent highly probable adaptive mechanisms. This work extends prior focused studies that observed parallel mutations during the adaption of several proteins (*30*–*34*) and a broad study that observed parallel mutations in eukaryotic proteins where the underlying selection pressures were unknown (*35*).

Testing and refining models for temperature adaptation through examination of 158,184 temperature-associated residues

Hundreds of published studies have examined differences in bulk amino acid compositions between mesophilic and thermophilic organisms, differences in sequences between mesophilic and thermophilic proteins, and differences in structural properties between mesophilic and thermophilic proteins and have suggested principles for temperature adaptation (*4*). However, because these trends were obtained from a limited number of organismal, sequence, and structural comparisons (typically 2 to 10) of differing phylogenetic relatedness, they could represent evolutionary noise, or clade or background-specific adaptations that are not general. Indeed, many specific stabilizing residue changes identified from comparative enzyme studies are not significantly enriched in diverse thermophilic bacteria (table S8).

Our dataset allowed us to systematically interrogate temperature adaptation at four levels of sequence and structural complexity and address nearly all previous models for temperature adaptation. Specifically, we assessed (i) preferences in the types of amino acids associated with high or low T_{Growth} (Fig. 5A and table S9), (ii) the identities and frequencies of site-specific residue changes across T_{Growth} (Fig. 5B and table S10), (iii) the identities and frequencies of physically interacting temperature-associated residue pairs (Fig. 5C and table S11), and (iv) networks of temperature-associated residues that change concomitantly with T_{Growth} (Fig. 6). Through these analyses, we have reconsidered, refined, and expanded previous models and, in the process, revealed new relationships, as outlined below and in table S12. In this section we discuss two insights that arise from analyses (i) to (iii) above; network analyses (iv) are described in a later section.

Several groups have postulated that thermophilic proteins frequently gain stability by increasing the amount of branched-chain amino acid residues (Ile, Leu, and Val) in their hydrophobic cores, thereby improving packing and increasing the hydrophobic effect (*4*). Our data support yet refine this model. We observed that Ile is more often associated with high T_{Growth} than low T_{Growth}, but Leu and Val are not, suggesting that Ile is stabilizing in diverse contexts, whereas the other branched hydrophobic amino acids are not, or may be stabilizing but in more limited instances or specific contexts (Fig. 5A). Further underscoring the distinction between the branched-chain hydrophobic residues, the two most frequent site-specific residue changes observed throughout enzyme families are between branched-chain residues: from Leu and Val in low-T_{Growth} orthologs to Ile in

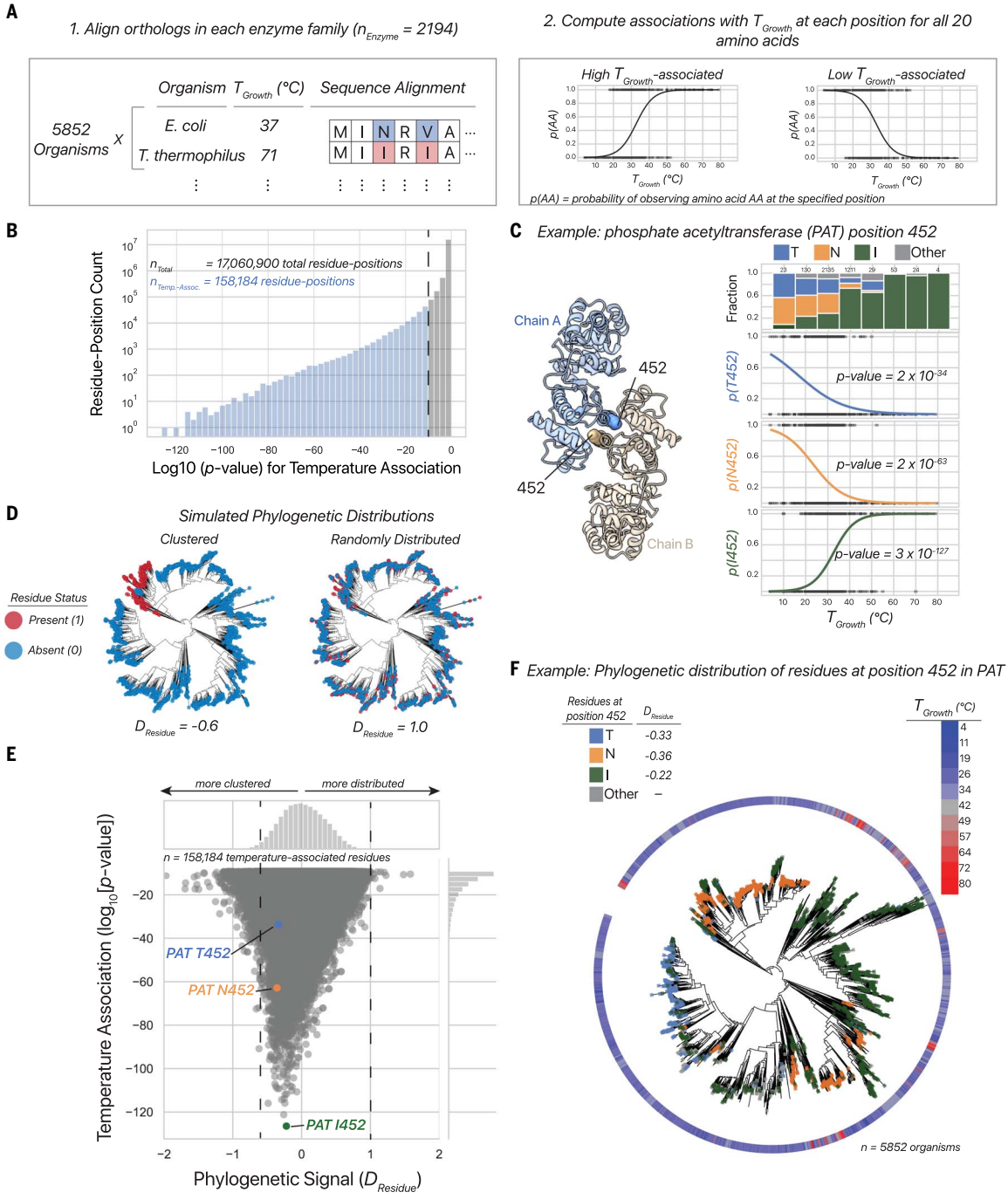


Fig. 4. Identifying tens of thousands of temperature-associated residues in thousands of enzyme families and assessing their distributions across the bacterial tree of life. (**A**) Simplified workflow for identifying temperature-associated residues (see Materials and methods). For each enzyme family, orthologous sequences from up to 5852 bacteria, with T_{Growth} values ranging from 4° to 80°C, were aligned (left). Then, for each position in each enzyme, logistic regressions were computed, modeling the probability of observing a specific amino acid at that position [p(AA)] across T_{Growth} (right). *T. thermophilus*, *Thermus thermophilus*. (**B**) Histogram of *p* values for temperature association from logistic regression models. The vertical dashed line corresponds to the *p* value cutoff for temperature association, corrected for multiple hypothesis testing ($p < 0.01/17{,}000{,}000 < 5.9 \times 10^{-10}$). (**C**) Observed temperature associations for residues at position 452 in phosphate acetyltransferase (PAT), which is located at the dimer interface. (**D**) Simulated residue distributions on the phylogenetic tree of bacterial species [red indicates residue present (1); blue indicates residue absent (0)] and their corresponding phylogenetic signals ($D_{Residue}$) (*29*). Greater clustering of the residue of interest (red) is associated with more negative values of $D_{Residue}$ (left), dispersal is associated with values of $D_{Residue} > 0$, and a random distribution corresponds to $D_{Residue} = 1$ (right) (*29*). See fig. S20 for additional simulations and their corresponding values of $D_{Residue}$. (**E**) $D_{Residue}$ versus *p* value for T_{Growth} association for each temperature-associated residue identified herein. Vertical dashed lines correspond to the $D_{Residue}$ values simulated in (D). We anticipated that temperature-associated residues would be similarly dispersed across the phylogenetic tree as T_{Growth}. Indeed, values of $D_{Residue}$ are centered around 0.02 ± 0.31, similar to the underlying phylogenetic signal for high T_{Growth} ($D_{TGrowth>40°C}$ = −0.05; fig. S21). (**F**) Distributions for residues at position 452 in PAT across the bacterial species tree. Each leaf (bacterial species) is colored by the residue present at position 452 in the PAT ortholog from that organism.

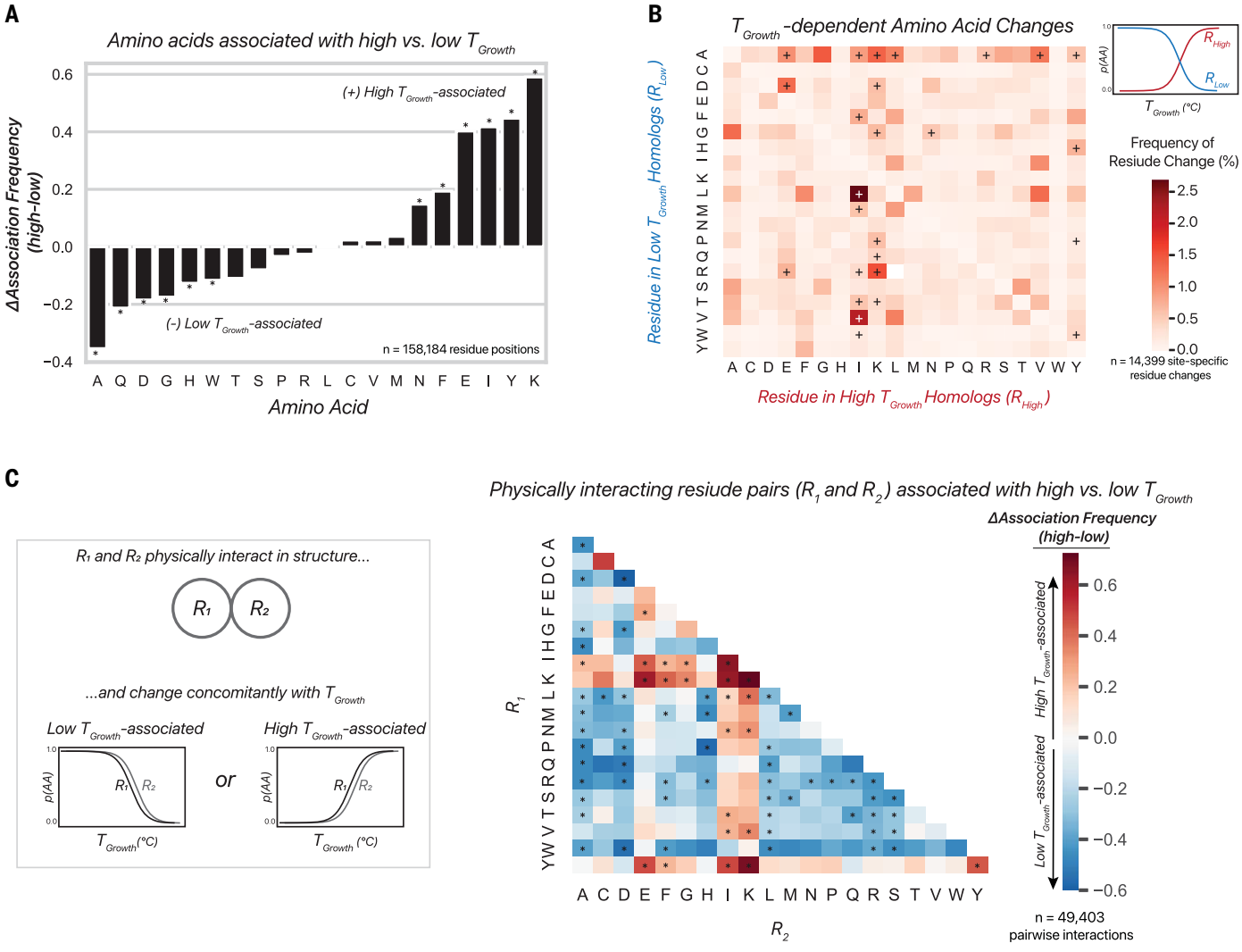


Fig. 5. Examination of temperature-associated residues and their interactions. (**A**) For each amino acid, the difference in frequency of association with high versus low T_{Growth} is shown. An asterisk indicates that the difference in association frequency with low T_{Growth} versus high T_{Growth} is statistically significant by chi-square test (table S9). (**B**) The frequency of observed site-specific changes in temperature-associated residues. Counts and *p* values for the significance of the directionality of these changes are in table S10. A "+" indicates that the frequency of a residue change (e.g., L→I) is significant relative to the reciprocal residue change (e.g., I→L; table S10). (**C**) The difference in frequency of association with high versus low T_{Growth} for all possible physically interacting pairs of residues (made up of residues R_1 and R_2) that change concomitantly with T_{Growth}. Residue pairs that are darker blue are more frequently associated with low T_{Growth}, whereas residue pairs that are darker red are more frequently associated with high T_{Growth}. An asterisk indicates that the difference in association frequency for a residue pair with low T_{Growth} versus high T_{Growth} is statistically significant (table S11).

high-T_{Growth} orthologs (Fig. 5B). These changes are two- and threefold more common than reciprocal Val-to-Ile changes ($p = 4.6 \times 10^{-14}$) and Leu-to-Ile changes ($p = 7.4 \times 10^{-32}$), respectively (Fig. 5B and table S10). Notably, we observed that nearly all pairwise physical interactions that involve Ile show preferences for association with high T_{Growth} and, particularly, interactions involving two Ile residues (Fig. 5C). A strong preference for Ile in high-T_{Growth} orthologs was previously observed for two enzymes: Comparisons between mesophilic and thermophilic orthologs of glyceraldehyde-3-phosphate dehydrogenase and glutamate dehydrogenase showed 6 and 12 Val/Leu-to-Ile changes, respectively (*36*, *37*). Perhaps Ile is favored over Val because its larger size provides greater hydrophobicity and surface area for packing and favored over Leu because its less restricted side-chain rotamer preferences allow it to explore more conformations to fill packing voids in protein interiors (*36*, *38*, *39*). Consistent with this model, Ile residues associated with high T_{Growth} populate dihedral angles rarely observed for Leu residues (fig. S23).

A second common model postulates that thermophilic proteins gain stability by increasing the number of salt bridges and hydrogen-bonding interactions with charged residues. (*4*) These and other models are based on apparent increases in the charged residues Lys, Arg, Asp, and Glu and observations of salt bridges in thermophilic proteins (table S12) (*4*). As above, our data support yet refine this model. We observed frequent association of Lys and Glu with high T_{Growth} but not Arg and Asp (Fig. 5A). Highlighting the preference for Lys over Arg, the third most common site-specific change is from Arg to Lys in high-T_{Growth} enzymes, a change that is threefold more common than the reciprocal Lys-to-Arg change ($p = 3.5 \times 10^{-19}$; Fig. 5B and table S10). Similarly, site-specific changes from Asp to Glu with increased T_{Growth} are

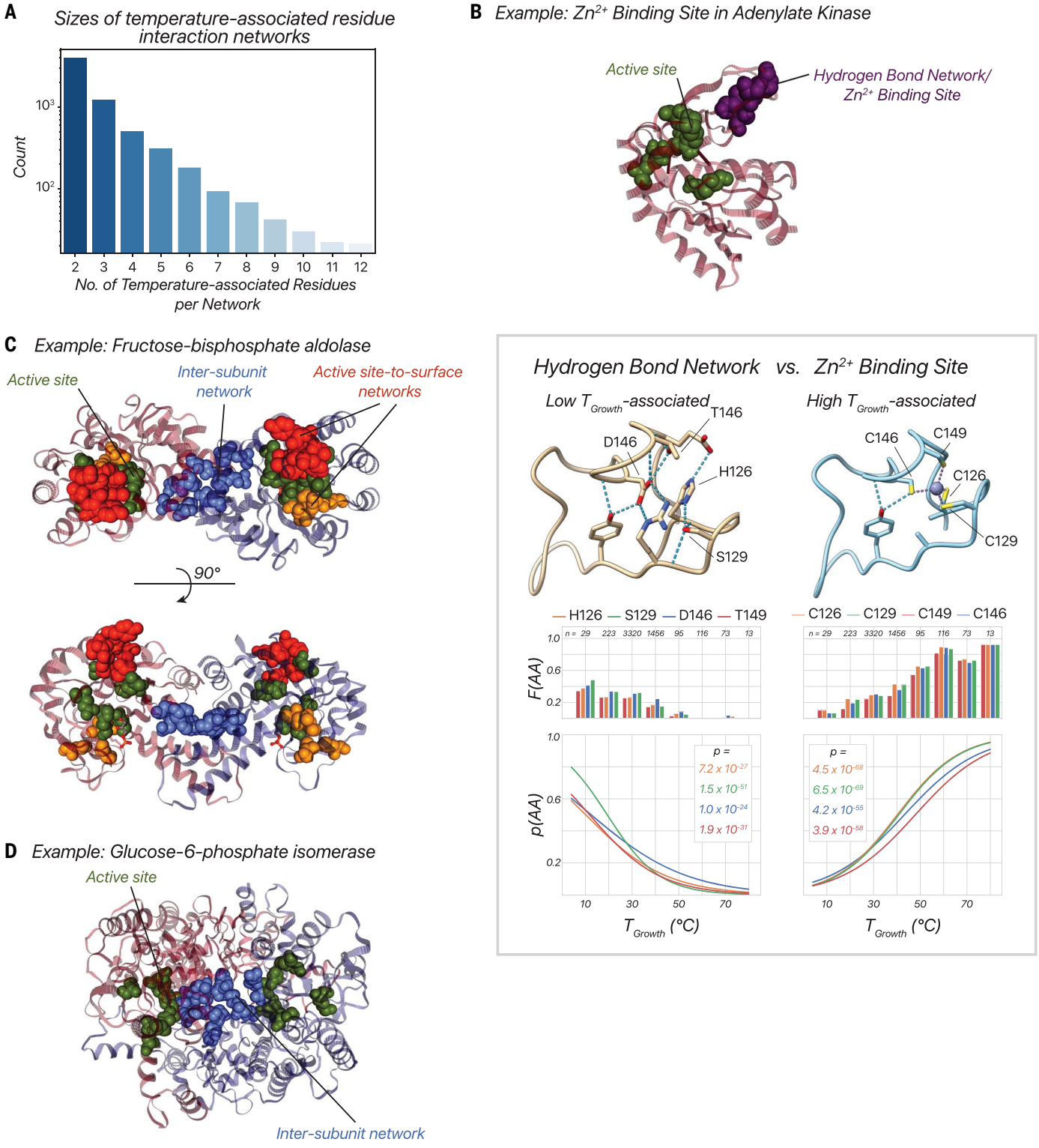


Fig. 6. Networks of temperature-associated residues and their involvement in temperature adaptation. (**A**) The distribution of sizes of networks of temperature-associated residues (n = 6612 networks). Counts for larger networks (>12 residues) are shown in fig. S31. (**B**) Representative structural snapshots for the hydrogen-bond network (PDB ID 1AKE) and zinc binding site (PDB ID 1ZIP) in the lid domain (purple) of ADK (top). Bar plots of the fraction of sequences with the noted residues [F(AA)] at positions 126 (H or C; orange), 129 (S or C; green), 146 (D or C; blue), and 149 (T or C; red) at increasing values of T_{Growth}; and logistic regression models of the probability [p(AA)] of observing a specific amino acid at positions 126, 129, 146, and 149 across T_{Growth} are shown at the bottom. (**C** and **D**) Networks of temperature-associated residues ($p < 10^{-50}$) in fructose-bisphosphate aldolase (C) and glucose-6-phosphate isomerase (D) (see also figs. S28 and S29). Active-site residues (green) assigned from the Mechanism and Catalytic Site Atlas (*43*) are shown. Temperature-associated residues involved in intersubunit networks (blue) and active site–to-surface networks (orange) are shown in space-filling representation.

threefold more common than reciprocal Glu-to-Asp changes ($p = 1.8 \times 10^{-13}$; Fig. 5B and table S10). Furthermore, physical interactions between Lys-Glu show the strongest preference for association with high T_{Growth} of any pairwise interaction, whereas Arg-Asp interactions are more often associated with low T_{Growth} (Fig. 5C). Thus, adaptation to high T_{Growth} apparently does not arise simply from an increased number of salt bridges; rather, particular features of the residues that constitute the salt bridges are important. For example, salt bridges containing Lys are less conformationally restricted than those with Arg, and the more flexible side chain of Glu, relative to Asp, may allow more effective conformational exploration to find interaction partners (*38*, *40*). Relative to Arg, Lys exhibits a smaller reduction in side-chain entropy upon protein folding, making Lys more entropically favored (*41*). Lys, which has the largest median hydrophobic accessible surface area of any residue, may allow for additional hydrophobic packing (*42*). In addition, the multiple positioned hydrogen-bond acceptors needed to fulfill interactions with the Arg $NH_{(1/2)}$ groups may render its favorable incorporation more dependent on surrounding residues (*40*). This greater dependency, in turn, would predict greater epistasis and less general utilization of Arg in adaptation to high T_{Growth}, as is observed (Fig. 5C).

Frequent temperature association of positions contacting catalytic residues

Catalytic residues directly participating in enzymatic reactions are known to be highly conserved, and their mutagenesis is typically highly deleterious. As expected, in the 360 enzymes in our dataset with annotated catalytic residues, we did not observe temperature-dependent changes at any of the 1347 catalytic positions (*43*).

Changes in second-shell residues have been suggested to be important in maximizing the efficiency of primary catalytic residues, consistent with changes observed in the maturation of a catalytic antibody and the directed evolution of a designed enzyme (*44*, *45*). Indeed, we observed that 69% of catalytic residues (926/1347) have at least one temperature-associated residue at a second-shell position.

Most simply, second-shell residues could alter the electronic properties or help position catalytic groups. Catalytic Zn^{2+} ions provided an opportunity to look for possible cases of electronic tuning, because ligands can alter the electronic properties of Zn^{2+} ions (*46*). Seven enzyme families that use an active-site Zn^{2+} to activate catalytic water molecules and/or stabilize charged transition states exhibit T_{Growth}-dependent changes in their Zn^{2+} ligands (fig. S24 and table S13). In each case, Cys ligands replace other residues in high-T_{Growth} orthologs (table S13). The strong interaction between Cys and Zn^{2+} may help maintain Zn^{2+} binding at high temperature while reducing the catalytic ability of Zn^{2+} by partially neutralizing its charge (*46*). Consistent with this model, Cys ligands are more common at structural relative to catalytic Zn^{2+} binding sites (*46*, *47*).

Although we cannot yet broadly assess the effects of these second-shell changes on active-site positioning, in cases where motions are involved in the enzyme's rate-limiting step, we might observe changes in the residues involved in these motions. For *Escherichia coli* dihydrofolate reductase (DHFR), where changes in conformation of the Met20 loop (residues 9 to 24) across the DHFR reaction cycle limit the rate of catalysis (*48*), we observe that residues at 4 of the 16 positions in the Met20 loop are temperature-associated (fig. S25). More generally, temperature-associated residue changes may help identify residues that influence functional conformational dynamics for in-depth study.

Networks of interactions associated with temperature adaptation

In many proteins, groups of amino acids work together to influence enzyme stability and function (*4*, *47*, *49*, *50*). We therefore sought to identify networks of physically contiguous, temperature-associated residues because these may function together to stabilize the enzyme or modulate its function. Specifically, we identified residues that co-occur in sequence, change concomitantly with T_{Growth}, and contact one another in x-ray crystal structures. Our criteria were met by networks in 525 of the 1005 enzyme families with temperature-associated residues. Most networks involved only two residues, consistent with the increased evolutionary probability for strategies with a smaller number of changes (Fig. 6A). Nevertheless, 422 enzyme families had networks of three or more residues (Fig. 6A and fig. S26). We anticipated that residues in these larger networks could work cooperatively to enhance stability or alter functionally important conformational changes, because the rates and equilibria of conformational changes are likely sensitive to temperature. Below we present examples from our dataset that are consistent with these roles.

The introduction of metal-ion binding sites in thermophilic enzymes has been widely discussed as a stabilization mechanism but is based on a small number of anecdotal examples and might be improbable because it would typically require the introduction of four or more residues to serve as ligands (*4*, *47*, *51*). To test whether metal sites are a common mechanism for temperature adaptation, we searched for bound divalent metal ions in x-ray crystal structures of the 422 enzymes with networks of at least three temperature-associated residues (Fig. 6A). We identified 10 cases where multiple metal ligands were associated with high T_{Growth} (table S14). In all 10 cases, the sites contained Zn^{2+} and, in nearly all cases, four Cys ligands, perhaps because of the high stability of the sulfur-Zn^{2+} interaction (*46*). As an example, ~80% of the most thermophilic orthologs of ADK contain a Zn^{2+} binding site in the enzyme's lid domain, but only ~10% of the psychrophiles have a Zn^{2+} binding site (Fig. 6B). Prior experiments confirm that replacing the hydrogen-bond network found in many low-T_{Growth} ADKs with the Zn^{2+} site stabilizes a mesophilic ADK without altering its activity (*52*, *53*). Additional examples are shown in the supplement (fig. S27).

Prior studies have proposed the strengthening of existing subunit interactions as a stabilization mechanism for thermophilic proteins (*4*). Of the 1005 enzymes with temperature-dependent residues, 548 were multimeric, and 154 of those (28%) had networks of temperature-associated residues that extend across subunit interfaces. For example, low-T_{Growth} variants of fructose-bisphosphate aldolase and glucose-6-phosphate isomerase and have five primarily hydrophobic intersubunit residues that are replaced by several salt bridges and hydrogen bonds in high-T_{Growth} variants (Fig. 6, C and D, and figs. S28 and S29).

Long-range allosteric communication in proteins is common, but the residues constituting the underlying networks are difficult to identify and can be difficult to link to biological function and evolutionary adaptation (*49*). Allosteric communication occurs from an enzyme's surface, where modifications are made or ligands bind, to the active site; additionally, the function of many enzymes involves intersubunit coordination (*49*). We therefore searched for networks that extended from active sites to surfaces or between active sites across subunit interfaces (i.e., active site–to–active site networks). We found that 142 of the 360 enzymes (39%) with annotated active-site residues had active site–to-surface networks, and 43 of the 192 multimeric enzymes (22%) with annotated active-site residues had active site–to–active site networks. Fructose-bisphosphate aldolase, noted above, provides an example of active site–to-surface networks (Fig. 6C and fig. S28), and glucose-6-phosphate isomerase, also discussed above, provides an example of active site–to–active site networks (Fig. 6D and fig. S29).

Conclusions and implications

We have broadly and deeply addressed enzyme temperature adaptation, revealing two molecular mechanisms underlying the temperature adaptation of KSI and identifying 158,184 temperature-associated residues and their physical trends and interaction networks that occur among 1005 enzyme families. These

data can be further mined to deepen our understanding of molecular evolution and to aid in uncovering additional mechanisms and principles that can be applied to enzyme engineering. In addition, this dataset may help identify and dissect allosteric networks, sites of oligomerization, and sites of protein-protein association. As an example of the latter, singular temperature-associated residue changes observed at many surface residues with no obvious functional role may represent positions of protein or other ligand association or may have a role in preventing nonspecific interactions or aggregation (fig. S30). Finally, and of broadest biological importance, the observation of widespread parallel adaptation to temperature and minimal epistasis at temperature-associated sites suggests that enzyme adaptation repeatedly follows evolutionary paths of low epistasis, advancing our understanding of the evolutionary mechanisms that drive the adaptation of nature's repertoire of enzymes.

Materials and methods

Materials

All reagents were of the highest purity commercially available (≥97%). Type I deionized water (≥18.2 megohm·cm) was generated using an Aqua Solutions 2121BL system and used to prepare all aqueous solutions. Crystallization materials were purchased from Hampton Research. Any materials not specified below were purchased from Sigma-Aldrich (now MilliporeSigma) or Fisher Scientific.

Collection of temperatures of optimal activity for enzymes from the literature

The temperature of optimal activity (T_{opt}) values in Fig. 1 and table S1 were collected from articles reviewing enzymatic temperature adaptation (*4*, *54*) and from articles identified from a PubMed search for the terms "thermophile," "psychrophile," "mesophile," "activity vs. temperature," "activity versus temperature," and "temperature adaption." The T_{opt} values listed in table S1 are the temperatures of maximal activity from activity versus temperature experiments for the stated enzyme.

KSI sequence identification, cloning, mutagenesis, expression, and purification

The sequence of thermoKSI was obtained from a protein BLAST search (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the sequence of *P. putida* KSI sequence (referred to as "mesoKSI" herein). Default BLAST parameters were used (Database = nr, Expect Threshold = 10, word size = 6, matrix = BLOSSUM62, Gap Costs = Existence: 11 Extension: 1, Compositional Adjustments = Conditional compositional score matrix adjustment) except for "Max Target Sequences," which was changed from 500 to 1000 to increase the number of KSI orthologs in the output. Hits were cross-referenced with the literature to identify KSI orthologs from organisms that grow at high temperatures. The most thermophilic organism identified in this set was *M. hassiacum* (referred to as "thermoKSI" herein), which grows at 65°C (Fig. 1C) (*21*, *22*). The gene encoding the thermoKSI protein sequence (MSTPQDNANTVHRYLEFVAKGQPDEIAALYADDATVEDPVGSEVHIGRQAIRGFYGNLENVQSRTEVKTLRALGHEVAFYWTLSIGGDEGGMTMDIISVMTFNDDGRIKSMKAYWTPENITQR) was ordered from GenScript and cloned into a pET-21c vector using the Gibson Assembly Protocol (New England BioLabs). The gene encoding mesoKSI was previously cloned into a pET21c vector and used here (*55*). QuikChangeII (Agilent) site-directed mutagenesis was used to incorporate mutations into these thermoKSI and mesoKSI pET21c plasmids. Mutations were confirmed by sequencing miniprep DNA from DH5α cells on an ABI3730xl capillary sequencer (Elim Biopharmaceuticals). The constructs were expressed in *E. coli* BL21(DE3) cells and purified as previously described (*55*).

KSI kinetic measurements

The KSI substrate 5(10)-estren-3,17-dione [5(10)EST] was purchased from Steraloids (Newport, RI). Reactions of purified mesoKSI and thermoKSI with 5(10)EST were monitored continuously at 248 nm using a Perkin Elmer Lambda 25 UV/Vis spectrometer with an attached VWR digital temperature-controlled circulating water bath. Temperatures within the cuvettes were checked before and after reaction using a platinum electrode thermistor (Omega Engineering), and the temperature of the circulating water bath was modified to maintain a constant internal cuvette temperature between reactions. Reactions were conducted in 10 mM potassium phosphate (KP_i) buffer, pH 7.2, 0.1 mM disodium EDTA, with 2% dimethyl sulfoxide (DMSO) as a cosolvent to maintain substrate solubility. The kinetic parameters k_{cat} and K_M were determined by fitting the observed initial velocity of each reaction as a function of 5(10)EST concentration (9 to 1000 μM; six to eight different substrate concentrations per experiment) to the Michaelis-Menten equation. Rates of the background reaction [reaction of 5(10) EST in 10 mM KP_i buffer, pH 7.2, 0.1 mM disodium EDTA, with 2% DMSO without enzyme] were determined at 25°, 30°, and 65°C and were orders of magnitude slower than our slowest mutant at each temperature, indicating minimal involvement of the background reaction relative to the enzyme-catalyzed reaction in our experiments. Reported values of k_{cat} and K_M in table S2 are the average plus or minus one standard deviation of two to seven independent experiments with at least two different enzyme concentrations varied by at least 2.5-fold from different enzyme preps and on different days.

KSI stability measurements

The stabilities of WT and mutant mesoKSI and thermoKSI variants were determined by monitoring the change in intrinsic tryptophan fluorescence upon equilibrium unfolding in the denaturant urea. A stock solution of KSI was diluted into different concentrations of urea (final [urea] = 0 to 9 M in 40 mM potassium phosphate buffer, pH 7.0 or 40 mM acetic acid-sodium acetate buffer, pH 5.0; 24 to 64 urea concentrations per variant) in a 96-well plate. Plates were sealed and incubated at room temperature for 18 to 366 hours. For each variant, multiple, consecutive measurements were made on each sample, hours apart, to ensure that the system was at equilibrium. Tryptophan residues were excited at 290 nm, and their fluorescence emission was monitored at 370 to 420 nm (10-nm steps) on a Tecan Infinite M200 Pro Plate Reader. Fluorescence intensity at all emission wavelengths examined showed a two-state transition with increasing [urea]. The data were fit to a model of an equilibrium between folded dimer (F_2) and unfolded monomer (U; Eq. 1):

$$F_2 \rightleftharpoons 2U \tag{1}$$

where K_U, the equilibrium constant for KSI unfolding, equals

$$K_U = 2P_T\left[\frac{f_U{}^2}{1-f_U}\right] \tag{2}$$

and

$$\begin{aligned}\Delta G_U &= -RT\ \ln(K_U)\\ &= \Delta G_U^{H_2O} - m[\text{urea}]\end{aligned} \tag{3}$$

$$\Delta G_U^{H_2O} = ([\text{urea}]_{50} * m) - RT\ \ln(P_T) \tag{4}$$

with P_T being the total enzyme concentration, f_U being the fraction of unfolded enzyme, R being the ideal gas constant (1.99×10^{-3} kcal mol^{-1} K^{-1}), T being the temperature (298 K), $\Delta G_U^{H_2O}$ being the Gibbs free-energy change in the absence of urea, m being the dependence of the free energy of unfolding on denaturant concentration, and $[\text{urea}]_{50}$ being the urea concentration at the midpoint of the unfolding curve. See supplementary text S4 for further discussion of KSI stability measurements.

In support of reversible KSI unfolding in urea, the same $\Delta G_U^{H_2O}$ was obtained from folding curves starting with unfolded mesoKSI or thermoKSI at 10 M urea followed by dilution into different final concentrations of urea. $\Delta G_U^{H_2O}$ values reported in table S3 are the average from at least two independent experiments using different final enzyme concentrations varied by at least 2.5-fold in the range of 1 to 25 μM enzyme. Reported errors are the standard deviations of these values.

X-ray crystallography

ThermoKSI WT crystals were grown in presence of the transition state or intermediate analog DNP. In a second enzyme construct, the D38N mutation (D40N with mesoKSI numbering) was introduced to mimic the protonated general base in the enzyme-intermediate complex and to increase the affinity of the enzyme for DNP relative to thermoKSI WT, as previously described with mesoKSI (*17*, *56*). ThermoKSI (1 μl at 1 to 2 mM in 40 mM potassium phosphate, pH 7.2), preincubated with 4 mM DNP, was mixed with 1 μl of precipitant solution (1.4 to 2 M ammonium sulfate, 10 mM potassium phosphate, pH 7.2). Crystals formed after incubation at room temperature in 1 to 2 weeks. All single-crystal x-ray diffraction data was collected at Stanford Synchrotron Radiation Lightsource's (SSRL's) BL9-2 at 100 K and using a Pilatus 6M detector in shutterless mode. Before data collection, all crystals' mother liquor was exchanged with paratone oil and crystals were mounted on loops and flash-frozen in liquid nitrogen. All diffraction datasets were processed using the XDS package (*57*) and the programs pointless (*58*) and aimless (*59*), as implemented in the autoxds in-house processing script at SSRL (*57*).

The thermoKSI WT structure was solved experimentally. A native dataset was collected at 14,000 eV (λ = 0.88557 Å) to 1.57-Å resolution. A second dataset of exceptional quality and strong anomalous signal was collected at 6000 eV (λ = 2.06633 Å; see table S4). Anomalous signal was obtained from native methionine residues and sulfate ions bound to the protein (table S5). The graphical software HKL2MAP (*60*) using SHELXC/D/E programs (*60*, *61*) was used to process the 6000-eV anomalous dataset and the 14,000-eV native dataset together. SHELXD was looking for 10 sites (five Met for each monomer in the crystallographic asymmetric unit) at 4.00-Å resolution and found 12 sites, including ordered sulfate ions from the crystallization mix (CFOM = 67.2, CCall = 47.2, CCwk = 20.1). Refining the sites and phases with an initial estimated solvent content of 40%, combined with automated model building as implemented in HKL2MAP, resulted in a clear map with a model built in the $P4_132$ space group (fig. S2). This initial model was used for further automatic model building with Buccaneer (*62*) and was subsequently used for molecular replacement of the native dataset using Phaser (*63*). The native dataset was used to further build and refine the structure (see below). The asymmetric unit contained two monomers. The N-terminus methionine, the loop 85-91, and the last two residues in the C terminus were disordered and not included in the WT model. No clear electron density for bound DNP was observed, and this structure was primarily used as a molecular replacement model (see below).

ThermoKSI D38N-DNP diffraction data was collected using a wavelength of 0.88557 Å and to a resolution of 1.25 Å. Initial phases were obtained by molecular replacement using the experimentally solved thermoKSI WT structure as a search model and using Phaser (*63*).

For both structures, after molecular replacement, model building was carried out with the program Buccaneer (*62*) and manually in Coot (*64*). Models were refined manually after visual inspection with Coot and using phenix.refine (*65*). Torsion-angle simulated annealing (as implemented in phenix.refine) was used during the initial stages of refinement. Ligand restraints were generated using the Elbow (AM1) (*66*) and the Grade server (http://grade.globalphasing.org/cgi-bin/grade/server.cgi). Model quality was checked by MolProbity (*67*) and gave an overall score of 1.53 and 1.36, for thermoKSI WT and thermoKSI D38N•DNP, respectively. Data and model statistics are presented in table S4.

Two very similar electron densities of unknown molecular identity were observed on the protein surface, next to Glu^{43} and Asp^{95} in both structures. These electron densities were modeled as guanidine, which was used to wash purification columns and is expected to be present in the purification buffer in trace amounts.

Simulations of WT and mutant mesoKSI and thermoKSI variants

The initial mesoKSI WT structure was based on the monomeric unit in the 1.1-Å resolution crystal structure of the mesoKSI D40N mutant complexed with DNP. As described above, the D40N mutation mimics the protonated general base in the enzyme-intermediate complex and increases the affinity of the enzyme for the transition state analog DNP relative to mesoKSI D40 (*17*, *56*). As was done previously, the dimeric KSI structure in fig. S6A was created using the "symmetry mates" utility in PyMol to generate an inverted image, nearest neighbor, monomer 4 Å from the mesoKSI WT monomer (*18*). The mesoKSI F86W, mesoKSI D103S, and mesoKSI F86W/D103S mutants were created using the Maestro software package with Schrödinger by mutating the 86 and 103 positions of the mesoKSI WT structure to tryptophan and serine, respectively (*68*). The initial structure of thermoKSI WT was based on the dimeric unit in the 1.25-Å resolution crystal structure of thermoKSI D40N complexed with DNP solved herein (mesoKSI numbering used throughout). Missing loops in the thermoKSI WT structure were completed using the Modeller software package (*69*). As was done with mesoKSI, the thermoKSI W86F, thermoKSI S103D, and thermoKSI W86F/S103D mutants were created by mutating the 86 and 103 positions of the thermoKSI WT to phenylalanine and aspartic acid, respectively.

MM simulations

To prepare the mesoKSI and thermoKSI WT dimers for MM, their initial structures were stripped of all surface waters and hydrogens from their crystal protein databases (PDB). There were no structural waters present in the mesoKSI WT, whereas the active-site water molecule observed between S103 and DNP in the thermoKSI crystal structure was preserved for the thermoKSI WT simulation. The protonation states for both proteins were determined at a pH of 7.0 by using the web-based protonation prediction tool, H++ (v. 3.0) (*70*). Side-chain titration curves predicted the expected protonation states for each residue in the mesoKSI and thermoKSI. To prepare the eight solvated mesoKSI and thermoKSI dimer systems, the tleap utility in the AmberTools 16 suite (*71*) was used with the ff14SB (*72*) and the generalized Amber Force Field (GAFF). (*73*, *74*) The DNP ligand charges were computed with the restricted electrostatic potential (RESP) scheme from R.E.D Server (*75*). The aforementioned mesoKSI and thermoKSI mutants were generated from these WT structures described above. The proteins were solvated by ~31,000 explicit SPC/Fw water molecules with 12 Na^+ ions to maintain charge neutrality, resulting in a near cubic simulation box with dimensions of ~100 Å by 100 Å by 100 Å. Using the GPU-accelerated classical molecular dynamics code OpenMM (*76*), minimizations were performed where the mesoKSI and thermoKSI backbones were held rigid and the explicit solvent and nonbackbone atoms were allowed to relax. These optimized structures were then heated over 50 ps to 300 K using a Langevin thermostat with a friction coefficient of 1 ps^{-1} in the constant isovolume-isothermal (NVT) ensemble. Constant isobaric-isothermal (NPT) ensemble simulations were performed for 10 ns with the mesoKSI and thermoKSI backbones harmonically constrained (force constant, k = 1.0 kcal/mol) using a Monte Carlo barostat at a pressure and temperature of 1 bar and 300 K, respectively. The constraints on the mesoKSI and thermoKSI backbones were lifted, and the entire system was allowed to propagate freely according to the classical equations of motion for 10 ns for all mutants. Lastly, production level NPT dynamics were performed for 100 ns without any constraints for all eight protein systems.

MM and QM/MM optimizations

To compute the hydrogen-bond distances in the active sites of the mesoKSI and thermoKSI structures, representative snapshots were chosen from the 100-ns NPT classical molecular dynamics trajectories (fig. S5A). As was performed in our previous studies on mesoKSI

WT, QM/MM optimizations were performed on 20 snapshots for each mutant chosen equidistant in time from the NPT trajectories (*18*). Using the Visual Molecular Dynamics (VMD) (*77*) software package, spherical cutouts for QM/MM optimizations were generated from these frames by selecting all molecules (protein and water) within 40 Å from the ligand. These truncated structures were first minimized using the Amber force field with no constraints and then with QM/MM, where the QM atoms were described with the B3LYP exchange-correlation functional and Grimme's D3 dispersion correction with the 6-31G** basis set (*73*, *74*, *78*). For the QM/MM optimizations, the MM atoms in the protein and water were described with the Amber force field and the SPC/Fw water model, respectively. (*73*, *74*) Previously, convergence of the D103-DNP and Y16-DNP hydrogen-bond distances in mesoKSI WT was shown with a QM region containing the sidechains of Y16, Y32, Y57, and D103 and the ligand DNP; further increases in the size of the QM region had no meaningful effect on the lengths of the D103-DNP and Y16-DNP hydrogen bonds (*18*). Thus, in mesoKSI mutants, a QM region including the side chain of residues D103/S103, F86/W86, Y16, Y57, Y32, and DNP was chosen (fig. S5A). The equivalent residues were chosen in the thermoKSI mutants (fig. S5A). Water molecules that were involved in the hydrogen-bond network around the ligand were also included in the QM region for both protein systems.

Literature search for protonated Asp and Glu residues

Bacterial enzymes with protonated Asp/Glu residues were identified from a review article on perturbed amino acid pK_a values (*79*) and from a PubMed search for the terms: "perturbed pK_a," "protonated Asp," "protonated Glu," "glutamic acid," and "aspartic acid" (fig. S13). Enzymes not found in bacteria with known culture temperatures were excluded from further analysis.

Obtaining growth temperatures (T_{Growth})

Median organism optimal growth temperatures for bacteria in culture were obtained from https://doi.org/10.5281/zenodo.1175608 (*3*). For the analyses below, we searched specifically for orthologous enzyme sequences from bacteria with growth temperatures in this dataset.

Identification of putative protonated Asp and Glu residues in the PDB and computation of their association with T_{Growth}

We used a previously published dataset of Asp-Asp, Asp-Glu, and Glu-Glu interactions identified from the PDB, with carboxyl O–O distances <3.0 Å. The close proximity of these two Asp/Glu residues, enforced by the surroundings, increases the likelihood that one of these carboxylate residues is protonated (*80*). We restricted our analysis to hydrogen bonds within a protein monomer, because intermolecular hydrogen bonds may be a consequence of nonphysiological crystal contacts, and to those with crystallization pH values ≥6 to avoid artifactual Asp/Glu interactions that result from the low-pH conditions. The remaining positions were manually inspected, and metal ligands were excluded from further analysis because these Asp/Glu residues will be deprotonated despite close proximity to another Asp/Glu residue. We further restricted our dataset to Asp/Glu residues found in at least 80% of the orthologous sequences in the lowest 10°C temperature bin. This additional filtering step increases the likelihood that we are examining sites where protonated Asp/Glu residues provide some fitness advantage (presumably activity) at low T_{Growth}, allowing us to address the interplay between section pressures on activity and stability.

For each PDB ID in the dataset above, we gathered the corresponding UniProtKB accession and standard protein names and obtained the corresponding full-length (WT) protein sequence. We then mapped the PDB-defined index for each potential carboxylic acid-carboxylate hydrogen bonding residue to the corresponding position in the UniProt sequence and removed duplicate entries. See section "Identifying temperature-associated residues using logistic regression models" below for a description of how we obtained sequences for each protein from organisms with known T_{Growth}. We performed correlations between the presence of these protonated Asp and Glu residues and T_{Growth} as in the section "Identifying temperature-associated residues using logistic regression models" below, with two differences: (i) For each protein with a protonated Asp and Glu residue, the reference sequence used for alignment was the WT sequence obtained from UniProtKB, and (ii) enzymes with filtered and dereplicated sequence sets smaller than 200 sequences were excluded from further analysis to ensure enough sequences for robust statistical analyses. Significance threshold, corrected for multiple hypothesis testing, was $p < 1.5 \times 10^{-4}$ (Bonferroni correction; $p < 0.01/67$ positions).

Acquisition of genomic KSI sequences

KSI sequences were obtained from the National Center for Biotechnology Information's (NCBI's) protein database using the recommended enzyme name from the BRENDA database: "steroid delta-isomerase" (*81*, *82*). The resulting sequences were cross-referenced against those of organisms with known values of T_{Growth} [see section "Obtaining growth temperatures (T_{Growth})] (*3*). The resulting temperature-annotated sequences were aligned to thermoKSI using BioPython *pairwise2* module (matrix = blosum62, gap open penalty = −10, gap extension penalty = −0.5; similar results are observed if alignments are performed to mesoKSI). Duplicate sequences were removed. To remove sequences with poor alignments, we generated 1000 random permutations of thermoKSI and performed pairwise alignment against the original thermoKSI sequence (again using the BioPython *pairwise2*). The median and standard deviation of the distribution of simulated randomly permuted alignment scores were calculated and temperature-annotated KSI sequences with alignment scores less than 40 (median + 4*standard deviations) were excluded from further analysis. Only sequences that contained the KSI general base (D40 in mesoKSI) were used in further analyses (Fig. 1B).

Acquisition and analysis of metagenomic KSI sequences

All translated open reading frames from the IMG/M (Mar 2018) database of metagenomes (*83*) were searched for sequences that fit a KSI sequence pattern using the NERSC computing services (*84*), enabled by the FICUS program (https://jgi.doe.gov/user-programs/program-info/ficus-overview/). The sequence pattern was based on a manually curated, structure-based (*85*) multiple sequence alignment of canonical KSIs. This search yielded 659,576 sequences from 5694 metagenomic samples. We proceeded to analyze either metagenome samples with annotated collection temperatures or samples taken from environments with approximately constant temperatures available in the literature (e.g., hot pools). The list of sequences was thus reduced to 1114 sequences associated with environmental temperatures that range from 0° to 90°C. The multiple sequence alignment was then used to predict the amino acid present in positions corresponding to amino acid 103 of mesoKSI. This computation was based on adding each metagenomic sequence to the MSA by using a specific module of MAFFT (*86*). Only sequences that contained the KSI general base (D40 in mesoKSI) were used in further analyses (Fig. 1B).

Generation of phylogenetic trees

Three types of phylogenetic trees are used herein:

1. Enzyme sequence-based trees. Multiple sequence alignments of orthologous sequences of a given enzyme type were generated using MAFFT (*86*). Alignment positions that contained gaps in >75% of sequences were trimmed using the program TrimAl (*87*). The program IQ-Tree was used to generate maximum likelihood phylogenetic trees from the resulting multiple sequence alignment using default parameters (*88*). Selected examples of the resulting phylogenetic trees were visualized

and plotted with the Interactive Tree of Life (iTOL) (*89*). An example for KSI is shown in Fig. 3B, where each leaf corresponds to a different KSI sequence.

2. Bacterial species tree. We constructed the bacterial species phylogenetic tree consisting solely of organisms with known values of T_{Growth} by subsetting the genome-based phylogenetic tree from the Genome Taxonomy Database (GTDB) (*3*, *90*). Leaf names (accession indexes) were mapped to NCBI organism names, and these organism names were parsed to genus and species (using BioPython's *Phylo* module and the R package *ape)* (*91*, *92*). Redundant genus- and species-mapped tree leaves (introduced upon dropping strain and subspecies information) were removed, and leaves associated with organisms absent in our growth temperature–associated list of organisms were removed. The trimmed phylogenetic tree was visualized and plotted with iTOL (fig. S19) (*89*).

3. Bacterial species tree (BacDive subsetted). We constructed an additional bacterial species phylogenetic tree consisting of organisms with experimentally derived values of T_{Growth} exclusively from the BacDive database (a subset of the larger dataset of bacterial T_{Growth} values from https://doi.org/10.5281/zenodo.1175608) (*3*, *20*). This tree was constructed as described for the full bacterial species tree, except subsetted to the smaller, but similarly diverse, set of bacterial genus and species identifiers. This tree was used in the calculation of phylogenetic signals ($D_{Residue}$) because of time limitations from using the larger tree.

Identifying temperature-associated residues using logistic regression models

We obtained a list of every enzyme within the BRENDA database from www.brenda-enzymes.org/ and extracted the recommended names(s) for each enzyme (entry) in the database (n = 6531 nonredundant enzymes). To obtain sequences for each protein from organisms with values of T_{Growth}, we used NCBI Entrez queries [via the Entrezpy package (*93*)] to obtain Entrez unique identifiers (UIDs) for each protein from NCBI's protein database (*63*) for each organism. Entrez queries were of the following structure: ["protein name"][PROTEIN] AND ["organism"][ORGANISM]. Collected UIDs [GenInfo identifiers (GIs)] were mapped to UniProtKB accessions via a mapping database and script provided by the NCBI (ftp://ftp.ncbi.nlm.nih.gov/genbank/livelists/gi2acc_mapping/), and sequences were extracted from the NCBI nonredundant protein database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz) with a custom python script.

Sequences were then dereplicated and subsetted to those containing only canonical amino acid one-letter codes (i.e., the standard 20 amino acids) and were additionally subsetted to those corresponding to organisms represented in the large bacterial species tree (tree type 2; 5852 organisms). Finally, enzyme records were subsetted to those containing at least 20 distinct organisms to provide enough organisms for statistically robust analyses (see below). A total of 2194 enzymes contained representatives for further analysis following these procedures.

We performed pairwise alignments to a sequence representative to generate a multiple sequence alignment for each enzyme. First, to select a reference sequence for these pairwise alignments, pairwise all-by-all blast (via the NCBI BLAST+ 2.7.1 command line application blastp, using the following nondefault parameters: "-outfmt 6 -max_hsps 1 -evalue 10e10 -max_target_seqs 100000 -num_threads 20") was performed on a locally generated BLAST database of a subset of the filtered and dereplicated sequences. This subset was obtained by sampling up to 250 sequences with a length within 5% of the median sequence length for each enzyme. If more than 250 sequences met this sequence-length constraint, a random sample of 250 of them were used for all-by-all blast. Then, cumulatively summed bit-scores were calculated for each candidate representative sequence, and the sequence with the largest summed bit-score was selected as the pairwise alignment representative. Second, to remove sequences with poor alignments, we generated 500 random permutations of each reference sequence and performed pairwise alignment against the assigned reference sequence (via the BioPython Bio.Align.PairwiseAligner class with the following parameters: *open_gap_score*=-10, *extend_gap_score*=-0.5, *substitution_matrix*=Bio.SubsMat.MatrixInfo.matlist.blosum62). The mean and standard deviation of the permutation alignment scores were calculated, and the mean+3*standard deviation was subsequently used as a per-enzyme alignment score lower threshold to remove sequences with low similarity to the reference (below).

Global pairwise alignments between each sequence from NCBI (query) to our reference sequence (subject) were computed and the full alignment was trimmed to nongapped positions in the alignment reference sequence to generate per-enzyme trimmed multiple sequence alignments (again via the BioPython Bio.Align.PairwiseAligner class with the following parameters: *open_gap_score*=-10, *extend_gap_score*=-0.5, *substitution_matrix* = Bio.SubsMat.MatrixInfo.matlist.blosum62).
Aligned sequences with alignment scores below the per-enzyme alignment score threshold were removed from the raw multiple sequence alignment. To avoid statistical bias from over representation of particular bacteria in the alignment, sequences corresponding to the same organism were collapsed into a consensus sequence derived from the most frequent nongapped residue identity at each aligned position (or a gap if no other residue identities were observed).

These per-enzyme consensus alignments were then parsed and joined with the growth temperature dataset to yield delimited files in which a given organism name, reference-indexed alignment position, consensus residue identity, and associated growth temperature (the T_{Growth} of the organism from which it came) constituted a single observation (row). Each amino acid at each position in each set of temperature-associated alignments for each enzyme were subjected to a logistic regression model, corresponding to ~17,000,000 total regressions performed.

For all logistic regressions computed, we treated T_{Growth} as the continuous predictor and the presence or absence of one of the 20 amino acids (encoded as 1 or 0, respectively) as the response. To prepare the temperature-associated alignments for regression, gapped positions and those corresponding to organisms not present in the bacterial species phylogenetic tree (prepared as per the section "Generation of phylogenetic trees") were removed. Logistic regressions were preformed using the glm function of base R (with all default parameters except *family*=binomial). Correlations between the presence of a specific amino acid at a specific position in a specific enzyme and T_{Growth} were deemed significant if the p value for the logistic regression was $<5.9 \times 10^{-10}$ (Bonferroni correction; $p < 0.01/17{,}000{,}000$ residue positions).

Calculations of phylogenetic signal for temperature-associated residues

We used the method of Fritz and Purvis to measure the phylogenetic signal ($D_{Residue}$) for binary presence or absence (coded as 1 or 0, respectively) for each temperature-associated residue on the bacterial species tree (*29*). We performed this analysis with a larger set of temperature-associated residues, defined at $p < 0.05/17{,}000{,}000 = 2.9 \times 10^{-9}$. This more permissive threshold was used to include positions at the edge of our original p value threshold ($p < 5.9 \times 10^{-10}$) and yield a more conservative inference when assessing the contribution of horizontal gene transfer to apparent temperature-associated residue changes (see fig. S22). We calculated phylogenetic signal using both the enzyme sequence-based trees (tree type 1) and the bacterial species tree (BacDive subsetted)(tree type 3); the former subsetted to the leaves found in the latter for each enzyme. Values of $D_{Residue}$ for a given residue generally agreed for the two types of trees used, suggesting that the contribution of horizontal gene transfer to apparent temperature-associated residue changes is not dominating our observed trends (see fig. S22).

For a given enzyme, alignment position, and residue identity meeting the p value threshold defined, we calculated the phylogenetic signal using a binary encoding (1 or 0) for presence of the residue of interest as the organismal trait. This calculation was first performed using the bacterial species tree (BacDive subsetted) trimmed to contain only species represented in the alignment of that site. The calculation was executed using the phylo.d function in the Comparative Analyses of Phylogenetics and Evolution in R (caper) R package (*94*) with all default parameters except *permut*=1000. Subsequently, each significant enzyme, alignment position, and residue identity set was subjected to an analogous calculation (as before, using the phylo.d function with *permut*=1000) using the enzyme sequenced-based tree subsetted to leaves also present in the trimmed bacterial species tree (BacDive subsetted) (subsetting the two trees so they contain the same leaves allows for head-to-head comparison of phylogenetic signal values, as this parameter is sensitive to tree size) (*29*). Finally, calculated phylogenetic signal values using the two trees were collated and joined on the enzyme, alignment position, and residue identity as a unique index.

Identifying site-specific changes in temperature-associated residues

We identified enzyme positions where one amino acid was associated with low T_{Growth} and another with high T_{Growth}. In cases where multiple amino acids are associated with low and/or high T_{Growth}, we used the amino acid with the lowest p value for temperature association. The amino acid identities of these residue changes are shown in Fig. 5B and in table S10.

Identifying representative enzyme structures

To identify x-ray crystal structures in the PDB that correspond to the enzymes in our dataset, we used NCBI's command line application to BLAST each representative enzyme sequence against the PDB. For enzymes with annotated catalytic residues in the Mechanism and Catalytic Site Atlas (M-CSA), downstream analyses were performed on the structure with the highest bit score with annotated catalytic sites in M-CSA (*43*). For enzymes not represented in the M-CSA, the structure with the highest bit score was used, with a minimum bit score cutoff of 200.

Identifying temperature-associated metal ligands

To identify metal sites whose ligands are temperature-associated, we searched the representative structures of each enzyme in our dataset (identified above) for bound metal ions using the Bio.PDB module of Biopython (*95*). All representative structures with bit scores >200 were analyzed to increase the coverage of metal-containing orthologs. By explicitly requiring structural confirmation of metal sites, we have increased confidence in the sites discussed herein but may be missing metal sites that are not represented in the PDB. For our analysis of active-site Zn^{2+} ions, we restricted our dataset to only enzymes whose structures contained Zn^{2+} and whose Zn^{2+} ligand(s) are temperature-associated. The structures of each Zn^{2+}-containing enzyme with temperature-associated ligands was then manually examined and cross-referenced with the literature to distinguish active-site Zn^{2+} ions from structural sites (table S13).

Structural metal binding sites are typically formed by at least four metal ligands (*47*, *51*). Thus, to identify metal binding sites acquired at high growth temperature, we looked specifically for cases where ≥3 metal ligands are associated with high T_{Growth}. (We chose a cutoff of three metal ligands to include sites where one metal ligand is conserved and acts as a founder for the rest of the site or is weakly temperature-associated.) Structures of the resulting metal sites were then manually examined and cross-referenced with the literature to distinguish structural from active-site metal binding sites (table S14).

Identifying pairwise residue interactions and networks

The Bio.PDB module of Biopython was used to analyze the representative structures of each enzyme identified above (*95*). We mapped each temperature-associated residue to structure by pairwise alignment between the reference sequence identified in the section "Identifying temperature-associated residues using logistic regression models" and the sequence of the PDB structure using the pairwise2 module of Biopython (any N-terminal tags included in the PDB sequence were excluded from the alignment). Temperature-associated residues that could not be mapped to structure (e.g., not modeled, or part of a protein domain not crystalized) were excluded from future analysis. Contacts with each enzyme position with a temperature-associated residue or contacts with M-CSA-annotated catalytic residues were identified using the NeighborSearch class of Bio.PDB, considering only side-chain atoms with a contact distance <5 Å. Temperature-associated residues at the enzyme surface were identified using the SurfaceDepth class of Bio.PDB, which uses Michel Sanner's MSMS program for molecular surface calculations (*96*).

Co-occurrence between temperature-associated residues was assessed by chi-square test. For each sequence in our dataset, the residue identities at the two positions being compared were recorded. Two amino acids were considered to co-occur if the frequency of observing these amino acids in the same sequence at the positions being compared was significantly different than the expected frequencies of observing these two amino acids by chance [$p < 0.01/(20*20$ possible amino acid pairs)]. Expected frequencies of amino acid pairs were calculated as the product of the individual observed frequencies of each amino acid in bacteria (*97*). Contacting temperature-associated residues were considered to "change together" with temperature if the direction of their temperature-associated trend was the same [i.e., they are both associated with either low or high T_{Growth} (Fig. 4A)] and the transition points (i.e., the value of T_{Growth} where the probability of observing the temperature-associated residue is half its maximum value) of their logistic regression fits were within ±10°C of one another. Pairwise interactions in Fig. 5C and networks can be downloaded from Dryad (*98*) and include temperature-associated residues with p values of $<1 \times 10^{-15}$. This more stringent p value cutoff increases our confidence in defining these networks.

REFERENCES AND NOTES

1. K. Kashefi, D. R. Lovley, Extending the upper temperature limit for life. *Science* **301**, 934 (2003). doi: 10.1126/science.1086823; pmid: 12920290
2. N. C. S. Mykytczuk *et al*., Bacterial growth at −15°C; molecular insights from the permafrost bacterium *Planococcus halocryophilus* Or1. *ISME J.* **7**, 1211–1226 (2013). doi: 10.1038/ismej.2013.8; pmid: 23389107
3. M. K. M. Engqvist, Correlating enzyme annotations with a large set of microbial growth temperatures reveals metabolic adaptations to growth at diverse temperatures. *BMC Microbiol.* **18**, 177 (2018). doi: 10.1186/s12866-018-1320-7; pmid: 30400856
4. C. Vieille, G. J. Zeikus, Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **65**, 1–43 (2001). doi: 10.1128/MMBR.65.1.1-43.2001; pmid: 11238984
5. G. N. Somero, A. C. Giese, D. E. Wohlschlag, Cold adaptation of the Antarctic fish *Trematomus bernacchii*. *Comp. Biochem. Physiol. B* **26**, 223–233 (1968). doi: 10.1016/0010-406X(68)90327-7
6. T. Bullock, Compensation for temperature in the metabolism and activity of poikilotherms. *Biol. Rev. Camb. Philos. Soc.* **30**, 311–342 (1955). doi: 10.1111/j.1469-185X.1955.tb01211.x
7. G. Feller, C. Gerday, Psychrophilic enzymes: Hot topics in cold adaptation. *Nat. Rev. Microbiol.* **1**, 200–208 (2003). doi: 10.1038/nrmicro773; pmid: 15035024
8. S. Moon, J. Kim, J. Koo, E. Bae, Structural and mutational analyses of psychrophilic and mesophilic adenylate kinases highlight the role of hydrophobic interactions in protein thermal stability. *Struct. Dyn.* **6**, 024702 (2019). doi: 10.1063/1.5089707; pmid: 31111079
9. V. Nguyen *et al*., Evolutionary drivers of thermoadaptation in enzyme catalysis. *Science* **355**, 289–294 (2017). doi: 10.1126/science.aah3717; pmid: 28008087
10. G. N. Somero, Adaptation of enzymes to temperature: Searching for basic "strategies". *Comp. Biochem. Physiol. B* **139**, 321–333 (2004). doi: 10.1016/j.cbpc.2004.05.003; pmid: 15544958
11. V. L. Arcus *et al*., On the temperature dependence of enzyme-catalyzed rates. *Biochemistry* **55**, 1681–1688 (2016). doi: 10.1021/acs.biochem.5b01094; pmid: 26881922
12. R. Jaenicke, Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity? *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2962–2964 (2000). doi: 10.1073/pnas.97.7.2962; pmid: 10737776
13. T. N. Starr, J. W. Thornton, Epistasis in protein evolution. *Protein Sci.* **25**, 1204–1218 (2016). doi: 10.1002/pro.2897; pmid: 26833806
14. M. S. Breen, C. Kemena, P. K. Vlasov, C. Notredame, F. A. Kondrashov, Epistasis as the primary factor in molecular

evolution. *Nature* **490**, 535–538 (2012). doi: 10.1038/nature11510; pmid: 23064225
15. M. B. Doud, O. Ashenberg, J. D. Bloom, Site-specific amino acid preferences are mostly conserved in two closely related protein homologs. *Mol. Biol. Evol.* **32**, 2944–2960 (2015). doi: 10.1093/molbev/msv167; pmid: 26226986
16. R. M. Pollack, Enzymatic mechanisms for catalysis of enolization: Ketosteroid isomerase. *Bioorg. Chem.* **32**, 341–353 (2004). doi: 10.1016/j.bioorg.2004.06.005; pmid: 15381400
17. D. A. Kraut *et al*., Testing electrostatic complementarity in enzyme catalysis: Hydrogen bonding in the ketosteroid isomerase oxyanion hole. *PLOS Biol.* **4**, e99 (2006). doi: 10.1371/journal.pbio.0040099; pmid: 16602823
18. M. M. Pinney *et al*., Structural coupling throughout the active site hydrogen bond networks of ketosteroid isomerase and photoactive yellow protein. *J. Am. Chem. Soc.* **140**, 9827–9843 (2018). doi: 10.1021/jacs.8b01596; pmid: 29990421
19. M. Horinouchi, T. Kurita, T. Hayashi, T. Kudo, Steroid degradation genes in *Comamonas testosteroni* TA441: Isolation of genes encoding a Δ4(5)-isomerase and 3α- and 3β-dehydrogenases and evidence for a 100 kb steroid degradation gene hot spot. *J. Steroid Biochem. Mol. Biol.* **122**, 253–263 (2010). doi: 10.1016/j.jsbmb.2010.06.002; pmid: 20554032
20. L. C. Reimer *et al*., BacDive in 2019: Bacterial phenotypic data for high-throughput biodiversity analysis. *Nucleic Acids Res.* **47**, D631–D636 (2019). doi: 10.1093/nar/gky879; pmid: 30256983
21. K. H. Schröder, L. Naumann, R. M. Kroppenstedt, U. Reischl, *Mycobacterium hassiacum* sp. nov., a new rapidly growing thermophilic mycobacterium. *Int. J. Syst. Bacteriol.* **47**, 86–91 (1997). doi: 10.1099/00207713-47-1-86; pmid: 8995808
22. I. Tiago *et al*., Genome sequence of *Mycobacterium hassiacum* DSM 44199, a rare source of heat-stable mycobacterial proteins. *J. Bacteriol.* **194**, 7010–7011 (2012). doi: 10.1128/JB.01880-12; pmid: 23209251
23. R. L. Thurlkill, G. R. Grimsley, J. M. Scholtz, C. N. Pace, pK values of the ionizable groups of proteins. *Protein Sci.* **15**, 1214–1218 (2006). doi: 10.1110/ps.051840806; pmid: 16597822
24. B. K. Shoichet, W. A. Baase, R. Kuroki, B. W. Matthews, A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 452–456 (1995). doi: 10.1073/pnas.92.2.452; pmid: 7831309
25. D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006). doi: 10.1126/science.1123539; pmid: 16601193
26. L. Shenhav, D. Zeevi, Resource conservation manifests in the genetic code. *Science* **370**, 683–687 (2020). doi: 10.1126/science.aaz9642; pmid: 33154134
27. H. Yang, R. A. Zubarev, Mass spectrometric analysis of asparagine deamidation and aspartate isomerization in polypeptides. *Electrophoresis* **31**, 1764–1772 (2010). doi: 10.1002/elps.201000027; pmid: 20446295
28. P. L. Wintrode, F. H. Arnold, Temperature adaptation of enzymes: Lessons from laboratory evolution. *Adv. Protein Chem.* **55**, 161–225 (2000). doi: 10.1016/S0065-3233(01)55004-4; pmid: 11050934
29. S. A. Fritz, A. Purvis, Selectivity in mammalian extinction risk and threat types: A new measure of phylogenetic signal strength in binary traits. *Conserv. Biol.* **24**, 1042–1051 (2010). doi: 10.1111/j.1523-1739.2010.01455.x; pmid: 20184650
30. H. J. Pelz *et al*., The genetic basis of resistance to anticoagulants in rodents. *Genetics* **170**, 1839–1847 (2005). doi: 10.1534/genetics.104.040360; pmid: 15879509
31. L. Menéndez-Arias, Molecular basis of human immunodeficiency virus drug resistance: An update. *Antiviral Res.* **85**, 210–231 (2010). doi: 10.1016/j.antiviral.2009.07.006; pmid: 19616029
32. S. B. Powles, Q. Yu, Evolution in action: Plants resistant to herbicides. *Annu. Rev. Plant Biol.* **61**, 317–347 (2010). doi: 10.1146/annurev-arplant-042809-112119; pmid: 20192743
33. S. Yokoyama, T. Tada, H. Zhang, L. Britt, Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13480–13485 (2008). doi: 10.1073/pnas.0802426105; pmid: 18768804
34. K. G. McCracken *et al*., Parallel evolution in the major haemoglobin genes of eight species of Andean waterfowl. *Mol. Ecol.* **18**, 3992–4005 (2009). doi: 10.1111/j.1365-294X.2009.04352.x; pmid: 19754505
35. A. Rokas, S. B. Carroll, Frequent and widespread parallel evolution of protein sequences. *Mol. Biol. Evol.* **25**, 1943–1953 (2008). doi: 10.1093/molbev/msn143; pmid: 18583353
36. K. L. Britton *et al*., Insights into thermal stability from a comparison of the glutamate dehydrogenases from *Pyrococcus furiosus* and *Thermococcus litoralis*. *Eur. J. Biochem.* **229**, 688–695 (1995). doi: 10.1111/j.1432-1033.1995.tb20515.x; pmid: 7758464
37. P. Zwickl, S. Fabry, C. Bogedain, A. Haas, R. Hensel, Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaebacterium *Pyrococcus woesei*: Characterization of the enzyme, cloning and sequencing of the gene, and expression in *Escherichia coli*. *J. Bacteriol.* **172**, 4329–4338 (1990). doi: 10.1128/JB.172.8.4329-4338.1990; pmid: 2165475
38. M. V. Shapovalov, R. L. Dunbrack Jr.., A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure* **19**, 844–858 (2011). doi: 10.1016/j.str.2011.03.019; pmid: 21645855
39. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982). doi: 10.1016/0022-2836(82)90515-0; pmid: 7108955
40. J. E. Donald, D. W. Kulp, W. F. DeGrado, Salt bridges: Geometrically specific, designable interactions. *Proteins* **79**, 898–915 (2011). doi: 10.1002/prot.22927; pmid: 21287621
41. A. J. Doig, M. J. Sternberg, Side-chain conformational entropy in protein folding. *Protein Sci.* **4**, 2247–2251 (1995). doi: 10.1002/pro.5560041101; pmid: 8563620
42. L. Lins, A. Thomas, R. Brasseur, Analysis of accessible surface of residues in proteins. *Protein Sci.* **12**, 1406–1417 (2003). doi: 10.1110/ps.0304803; pmid: 12824487
43. A. J. M. Ribeiro *et al*., Mechanism and Catalytic Site Atlas (M-CSA): A database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res.* **46**, D618–D623 (2018). doi: 10.1093/nar/gkx1012; pmid: 29106569
44. P. A. Patten *et al*., The immunological evolution of catalysis. *Science* **271**, 1086–1091 (1996). doi: 10.1126/science.271.5252.1086; pmid: 8599084
45. R. Obexer *et al*., Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* **9**, 50–56 (2017). doi: 10.1038/nchem.2596; pmid: 27995916
46. Y. M. Lee, C. Lim, Physical basis of structural and catalytic Zn-binding sites in proteins. *J. Mol. Biol.* **379**, 545–553 (2008). doi: 10.1016/j.jmb.2008.04.004; pmid: 18462757
47. I. L. Alberts, K. Nadassy, S. J. Wodak, Analysis of zinc binding sites in protein crystal structures. *Protein Sci.* **7**, 1700–1716 (1998). doi: 10.1002/pro.5560070805; pmid: 10082367
48. D. D. Boehr, D. McElheny, H. J. Dyson, P. E. Wright, The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638–1642 (2006). doi: 10.1126/science.1130258; pmid: 16973882
49. N. M. Goodey, S. J. Benkovic, Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* **4**, 474–482 (2008). doi: 10.1038/nchembio.98; pmid: 18641628
50. C. Vetriani *et al*., Protein thermostability above 100°C: A key role for ionic interactions. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12300–12305 (1998). doi: 10.1073/pnas.95.21.12300; pmid: 9770481
51. M. M. Harding, The architecture of metal coordination groups in proteins. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 849–859 (2004). doi: 10.1107/S0907444904004081; pmid: 15103130
52. V. Perrier, S. Burlacu-Miron, S. Bourgeois, W. K. Surewicz, A. M. Gilles, Genetically engineered zinc-chelating adenylate kinase from Escherichia coli with enhanced thermal stability. *J. Biol. Chem.* **273**, 19097–19101 (1998). doi: 10.1074/jbc.273.30.19097; pmid: 9668094
53. S. Burlacu-Miron, V. Perrier, A.-M. Gilles, E. Pistotnik, C. T. Craescu, Structural and energetic factors of the increased thermal stability in a genetically engineered *Escherichia coli* adenylate kinase. *J. Biol. Chem.* **273**, 19102–19107 (1998). doi: 10.1074/jbc.273.30.19102; pmid: 9668095
54. M. Elias, G. Wieczorek, S. Rosenne, D. S. Tawfik, The universality of enzymatic rate-temperature dependency. *Trends Biochem. Sci.* **39**, 1–7 (2014). doi: 10.1016/j.tibs.2013.11.001; pmid: 24315123
55. D. A. Kraut, P. A. Sigala, T. D. Fenn, D. Herschlag, Dissecting the paradoxical effects of hydrogen bond mutations in the ketosteroid isomerase oxyanion hole. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1960–1965 (2010). doi: 10.1073/pnas.0911168107; pmid: 20080683
56. I. P. Petrounia, R. M. Pollack, Substituent effects on the binding of phenols to the D38N mutant of 3-oxo-Δ^5-steroid isomerase. A probe for the nature of hydrogen bonding to the intermediate. *Biochemistry* **37**, 700–705 (1998). doi: 10.1021/bi972262s; pmid: 9425094
57. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010). pmid: 20124692
58. P. R. Evans, An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282–292 (2011). pmid: 21460446
59. P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013). doi: 10.1107/S0907444913000061; pmid: 23793146
60. T. Pape, T. R. Schneider, HKL2MAP: A graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Cryst.* **37**, 843–844 (2004). doi: 10.1107/S0021889804018047
61. G. M. Sheldrick, Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485 (2010). pmid: 20383001
62. K. Cowtan, The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006). pmid: 16929101
63. A. J. McCoy *et al*., Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007). doi: 10.1107/S0021889807021206; pmid: 19461840
64. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics.*Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004). pmid: 15572765
65. P. V. Afonine *et al*., Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012). pmid: 22505256
66. N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, electronic Ligand Builder and Optimization Workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080 (2009). pmid: 19770504
67. V. B. Chen *et al*., MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010). pmid: 20057044
68. Schrödinger LLC, Maestro (Schrödinger, New York, 2020).
69. B. Webb, A. Sali, Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **54**, 5.6.1–5.6.37 (2016). doi: 10.1002/cpbi.3; pmid: 27322406
70. R. Anandakrishnan, B. Aguilar, A. V. Onufriev, *H*++ 3.0: automating p*K* prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **40**, W537–W541 (2012). doi: 10.1093/nar/gks375; pmid: 22570416
71. R. Salomon-Ferrer, D. A. Case, R. Walker, An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* **3**, 198–210 (2013). doi: 10.1002/wcms.1121
72. J. A. Maier *et al*., ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015). doi: 10.1021/acs.jctc.5b00255; pmid: 26574453
73. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004). doi: 10.1002/jcc.20035; pmid: 15116359
74. J. Wang, W. Wang, P. A. Kollman, D. A. Case, Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260 (2006). doi: 10.1016/j.jmgm.2005.12.005; pmid: 16458552
75. E. Vanquelef *et al*., R.E.D. Server: A web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **39**, W511–W517 (2011). doi: 10.1093/nar/gkr288; pmid: 21609950
76. P. Eastman *et al*., OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Comput. Biol.* **13**, e1005659 (2017). doi: 10.1371/journal.pcbi.1005659; pmid: 28746339
77. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28 (1996). doi: 10.1016/0263-7855(96)00018-5; pmid: 8744570
78. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010). doi: 10.1063/1.3382344; pmid: 20423165
79. T. K. Harris, G. J. Turner, Structural basis of perturbed pK_a values of catalytic groups in enzyme active sites. *IUBMB Life* **53**, 85–98 (2002). doi: 10.1080/15216540211468; pmid: 12049200

80. J. Lin, E. Pozharski, M. A. Wilson, Short carboxylic acid-carboxylate hydrogen bonds can have fully localized protons. *Biochemistry* **56**, 391–402 (2017). doi: 10.1021/acs.biochem.6b00906; pmid: 27989121
81. L. Jeske, S. Placzek, I. Schomburg, A. Chang, D. Schomburg, BRENDA in 2019: A European ELIXIR core data resource. *Nucleic Acids Res.* **47**, D542–D549 (2019). doi: 10.1093/nar/gky1048; pmid: 30395242
82. E. W. Sayers *et al*., Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **47**, D23–D28 (2019). doi: 10.1093/nar/gky1069; pmid: 30395293
83. I. A. Chen *et al*., IMG/M v.5.0: An integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* **47**, D666–D677 (2019). doi: 10.1093/nar/gky901; pmid: 30289528
84. W. Arndt, paper presented at the IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Vancouver, BC, Canada, 21 to 25 May 2018.
85. O. O'Sullivan, K. Suhre, C. Abergel, D. G. Higgins, C. Notredame, 3DCoffee: Combining protein sequences and structures within multiple sequence alignments. *J. Mol. Biol.* **340**, 385–395 (2004). doi: 10.1016/j.jmb.2004.04.058; pmid: 15201059
86. K. Katoh, M. C. Frith, Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* **28**, 3144–3146 (2012). doi: 10.1093/bioinformatics/bts578; pmid: 23023983
87. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009). doi: 10.1093/bioinformatics/btp348; pmid: 19505945
88. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015). doi: 10.1093/molbev/msu300; pmid: 25371430
89. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019). doi: 10.1093/nar/gkz239; pmid: 30931475
90. D. H. Parks *et al*., A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018). doi: 10.1038/nbt.4229; pmid: 30148503
91. E. Talevich, B. M. Invergo, P. J. A. Cock, B. A. Chapman, Bio. Phylo: A unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. *BMC Bioinformatics* **13**, 209 (2012). doi: 10.1186/1471-2105-13-209; pmid: 22909249
92. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004). doi: 10.1093/bioinformatics/btg412; pmid: 14734327
93. J. P. Buchmann, E. C. Holmes, Entrezpy: A Python library to dynamically interact with the NCBI Entrez databases. *Bioinformatics* **35**, 4511–4514 (2019). doi: 10.1093/bioinformatics/btz385; pmid: 31077305
94. C. D. L. Orme, R. P. Freckleton, G. H. Thomas, T. Petzoldt, S. A. Fritz, The caper package: comparative analysis of phylogenetics and evolution in R (2012); http://caper.r-forge.r-project.org.
95. T. Hamelryck, B. Manderick, PDB file parser and structure class implemented in Python. *Bioinformatics* **19**, 2308–2310 (2003). doi: 10.1093/bioinformatics/btg299; pmid: 14630660
96. M. F. Sanner, A. J. Olson, J. C. Spehner, Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **38**, 305–320 (1996). doi: 10.1002/(SICI)1097-0282(199603)38:3<305::AID-BIP4>3.0.CO;2-Y; pmid: 8906967
97. A. Moura, M. A. Savageau, R. Alves, Relative amino acid composition signatures of organisms and environments. *PLOS ONE* **8**, e77319 (2013). doi: 10.1371/journal.pone.0077319; pmid: 24204807
98. M. M. Pinney *et al*., Parallel molecular mechanisms for enzyme temperature adaptation, Version 1. Dryad (2020); https://doi.org/10.5061/dryad.3ffbg79h2.
99. P. L. Privalov, Cold denaturation of proteins. *Crit. Rev. Biochem. Mol. Biol.* **25**, 281–306 (1990). doi: 10.3109/10409239009090612; pmid: 2225910

ACKNOWLEDGMENTS

We thank C. Sabatti for statistics advice; J. Eisen for advice on bacterial phylogenetic trees; M. Wilson and E. Pozharskiy for access to the dataset of Asp and Glu interactions from the PDB; C. Sitron for experimental advice; R. Alegado, P. Fordyce, H. McShea, K. Peichel, Y. Simons, C. Stark, and J. Winnicoff for thoughtful discussions; C. Stark and A. Powers for sharing code for analyzing activity and stability data; and M. Laub, A. Nabhan, J. Thornton, N. Tokuriki, C. Walsh, and Herschlag lab members for manuscript comments. **Funding:** This work was funded by a National Science Foundation (NSF) grant (MCB-1714723) and a National Institutes of Health (NIH) grant (R01 GM60595). This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy (DOE) Office of Science User Facility operated under contract no. DE-AC02-05CH11231, as well as resources obtained from the FICUS program (https://jgi.doe.gov/user-programs/program-info/ficus-overview/), proposal 503369. Use of the Stanford Synchrotron Radiation Lightsource (SSRL), SLAC National Accelerator Laboratory, is supported by the U.S. DOE, Office of Science, Office of Basic Energy Sciences under contract no. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the NIH, National Institute of General Medical Sciences (NIGMS) (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. M.M.P. was supported by NSF Graduate Research and Gerald J. Lieberman Fellowships. D.A.M. was supported by the Stanford Medical Scientist Training Program and a Stanford Interdisciplinary Graduate Fellowship (anonymous donor) affiliated with Stanford ChEM-H. D.M.S. was supported by an NSF Graduate Research Fellowship. F.Y. was supported by a Human Frontiers Postdoctoral Fellowship. **Author contributions:** M.M.P. and D.H. designed the study. M.M.P. performed experiments and crystallized thermoKSI. F.Y. and T.D. performed x-ray crystallography data collection and structure refinement. E.A. and P.C.B. acquired and analyzed metagenomic KSI sequence data with help from M.M.P. D.A.M. and M.M.P. analyzed enzyme sequence data from bacterial genomes. D.M.S. performed and analyzed molecular mechanics simulations with help from R.L. and T.J.M. M.M.P. and D.H. wrote the manuscript with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Atomic coordinates and structure factors for thermoKSI are deposited in the PDB (IDs 6P3L and 6P44). The starting datasets, analysis code, and results (*p* values and parameters) for logistic regression models applied to 17,060,900 residue positions in 2194 enzymes and phylogenetic signals for temperature-associated residues can be downloaded from Dryad (*98*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6533/eaay2784/suppl/DC1
Supplementary Text
Figs. S1 to S31
Tables S1 to S14
References (*100*–*258*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

5 June 2019; resubmitted 23 August 2020
Accepted 4 January 2021
10.1126/science.aay2784

Parallel molecular mechanisms for enzyme temperature adaptation

Margaux M. Pinney, Daniel A. Mokhtari, Eyal Akiva, Filip Yabukarski, David M. Sanchez, Ruibin Liang, Tzanko Doukov, Todd J. Martinez, Patricia C. Babbitt and Daniel Herschlag

Science **371** (6533), eaay2784.
DOI: 10.1126/science.aay2784

Some like it hot, others not

Enzymes strike a delicate balance between features that enhance chemical reactivity and those that contribute to stable structure. Both features are important and can be unrelated or antagonistic. Pinney *et al.* combined rich experimental work on thermophilic and mesophilic variants of the enzyme ketosteroid isomerase (KSI) with bioinformatic data from a diverse set of bacterial enzymes to reveal the molecular determinants of thermal adaptation in enzymes. For KSI, they observed a trade-off between activity and thermal stability that comes down to a single active-site residue. With their larger dataset, they identified patterns of individual amino acid substitutions that are favored at higher temperatures, and also consider how networks of stabilizing interactions develop.

Science, this issue p. eaay2784