Supporting Information: Exploitation of active site flexibility-low temperature activity relation for engineering broad range temperature active enzymes

Siva Dasetty,[†,¶] Jonathan Zajac,[‡] and Sapna Sarupria[*,‡,†]

†*Department of Chemical & Biomolecular Engineering, Clemson University, Clemson, SC 29634, USA*

‡*Department of Chemistry, University of Minnesota, Minneapolis, MN 55455, USA*

¶*Current address: Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL 60637, USA*

E-mail: sarupria@umn.edu

1 Molecular dynamics (MD) simulations details of replicate trajectories

Two additional MD simulations were performed at 280 K, 300 K, and 320 K to estimate sampling errors. GROMACS-2021.3[S1] was used for performing these additional simulations. The canonical velocity rescaling[S2] thermostat was used to control temperature for both equilibration and production runs. The Berendsen barostat[S3] was used to control pressure for equilibration runs, while the Parrinello-Rahman barostat[S4,S5] was used in the production runs. Temperature and pressure coupling constants were set to 0.2 ps and 4.0 ps, respectively.

2 Root mean square deviation of WT and mutants from WT crystal structure

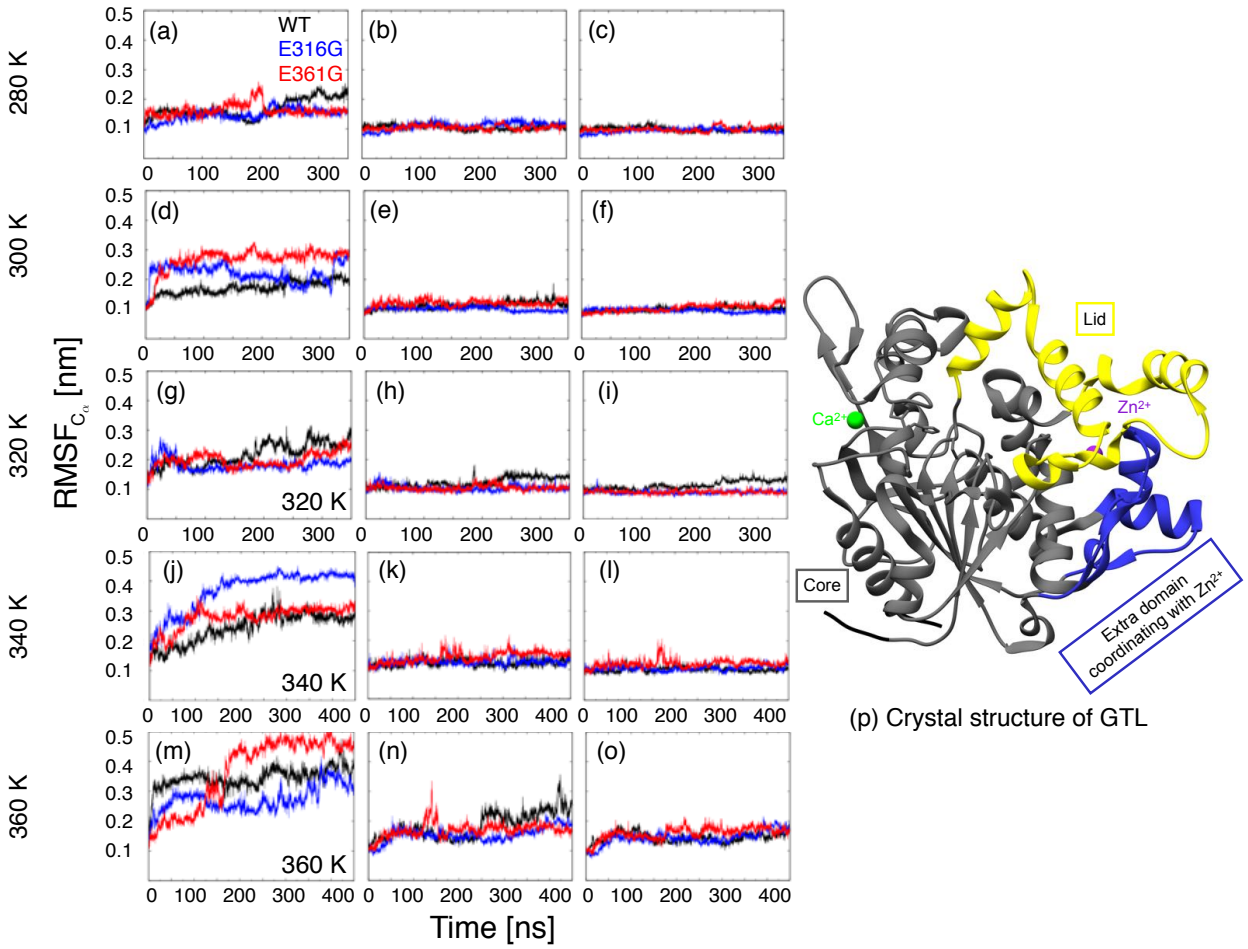

Figure S1: Root mean square deviation (RMSD) of one of the independent MD simulation trajectories from the energy minimized solvated WT crystal structure (reference) at different temperatures. Left column (a, d, g, j, m): RMSD of C_α of residues 6-384 i.e., excluding 5 residues at N- and C-terminus. Middle column (b, e, h, k, n): RMSD of $C_\alpha$residues 6-168 and 240-384 i.e, excluding lid domain and 5 residues at N- and C-terminus. Right column (c, f, i, l, o) shows RMSD of $C_\alpha$residues 6-73, 98-131, 152-168, and 240-384 that exclude both lid domain, Zn^{2+} domain and 5 residues at N- and C-terminus. (p) Illustration of crytal structure of WT GTL colored according to core (α-β hydrolaze fold - gray), lid domain (yellow) and the Zn^{2+} domain (blue).

3 $\Delta G^{\ddagger}$-T profiles

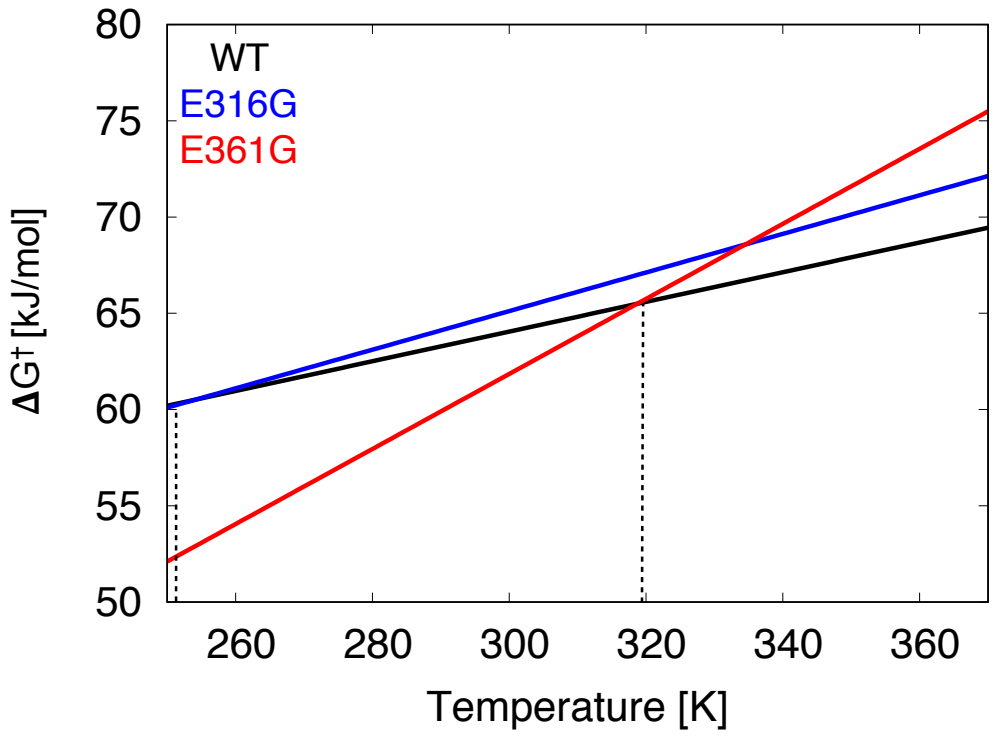

Figure S2: Variation of $\Delta G^{\ddagger}$ with temperature using $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ values estimated from Eyring plots with C4 as substrate obtained in wet-lab experiments.[S6] Dashed vertical lines in black are drawn at temperatures where $\Delta G^{\ddagger}$ of E316G ($\sim$255 K) and E361G ($\sim$319 K) are similar to that of WT.

4 RMSF difference between WT and mutants

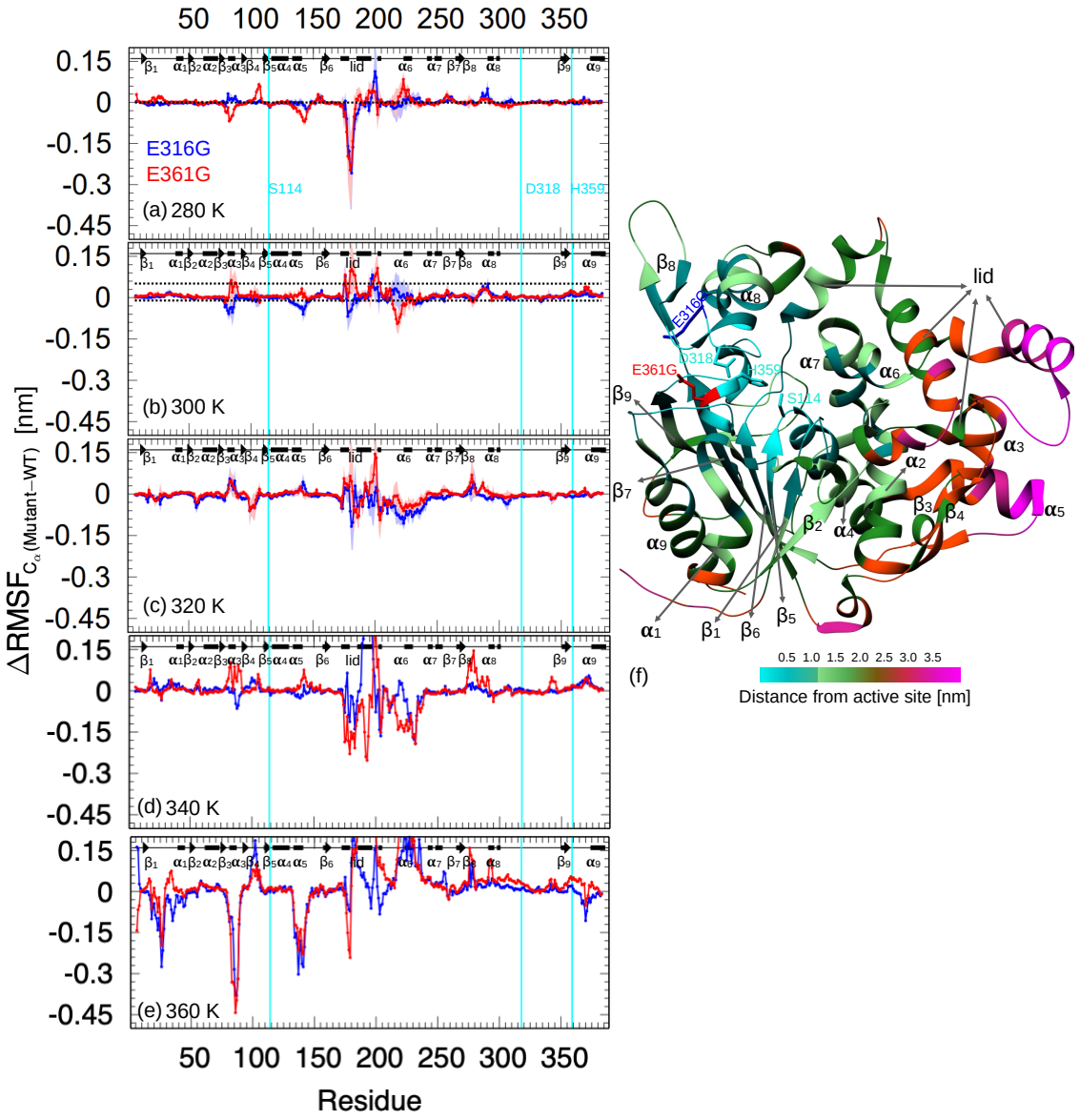

Figure S3: Difference in RMSF of C_α of each residue between mutants and WT at different temperatures. The first 50 ns of the production trajectory is considered part of the equilibrium cycle and excluded from the RMSF calculation. Difference in RMSF ($\Delta\text{RMSF}_{\text{Mutant}-\text{WT}}$) of E316G and WT (blue) and E361G and WT (red) at (a) 280 K, (b) 300 K, (c) 320 K, (d) 340 K, and (e) 360 K. For 280 K, 300 K, and 320 K, error bars are estimated from three independent trajectories. (f) Crystal structure of WT colored according to distance of C_α of residues from any of the active site residues C_α (S114, D318, and H359). The locations of catalytic residues are indicated by vertical cyan lines in (a-e).

5 Change in RMSF of residues with temperature

Change in RMSF of C_α of each residue in WT and mutants at a temperature T and 280 K calculated by ignoring the first 50 ns of the production trajectory.

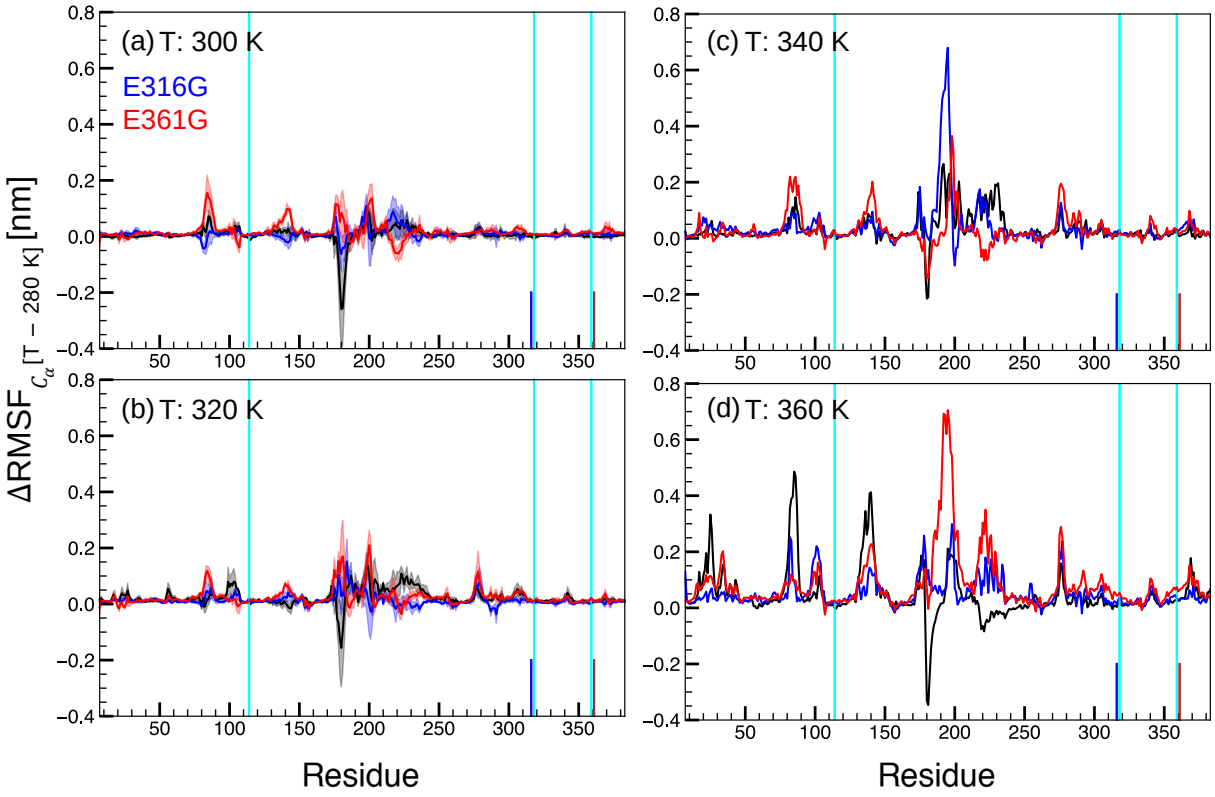

Figure S4: Difference in RMSF (ΔRMSF) of C_α of each residue in WT (black) and E316G (blue) or E361G (red) at a temperature T and 280 K. (a) 300 K and 280 K, (b) 320 K and 280 K, (c) 340 K and 280 K, and (d) 360 K and 280 K. For 300 K and 320 K, error bars are estimated from three independent trajectories. The locations of catalytic residues are indicated by vertical cyan lines.

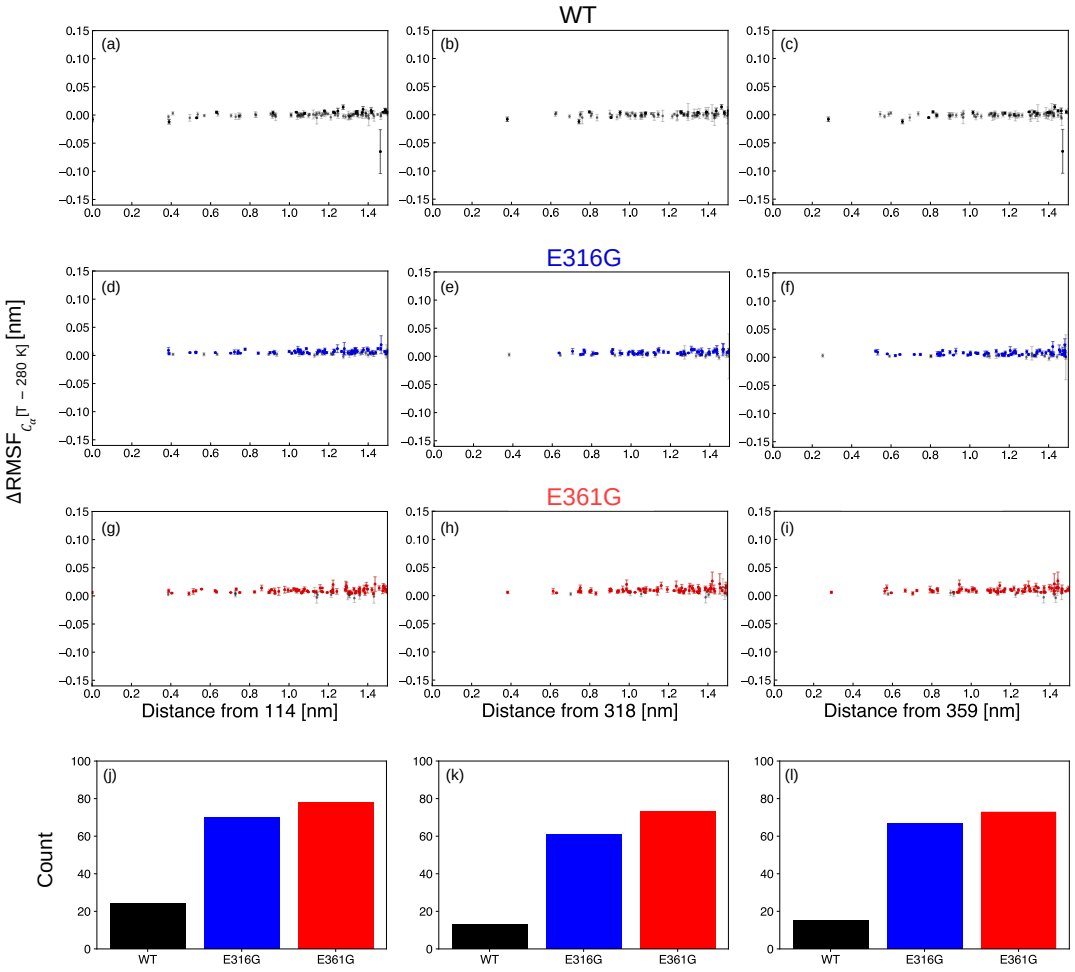

Figure S5: Scatter plot of difference in RMSF (ΔRMSF) at 300 K and at 280 K of WT (black; a, b, c), E316G (blue; d, e, f) and E361G (red; g, h, i) and residues within 1.5 nm of the active site residues — S114 (a, d, g), D318 (b, e, h), and H359 (c, f, i). The number of residues with ΔRMSF > 0 and within 1.5 nm from each of the active site residues 114 (j), 318 (k), and 359 (l). Error bars in the scatter plots are estimated from three independent trajectories.

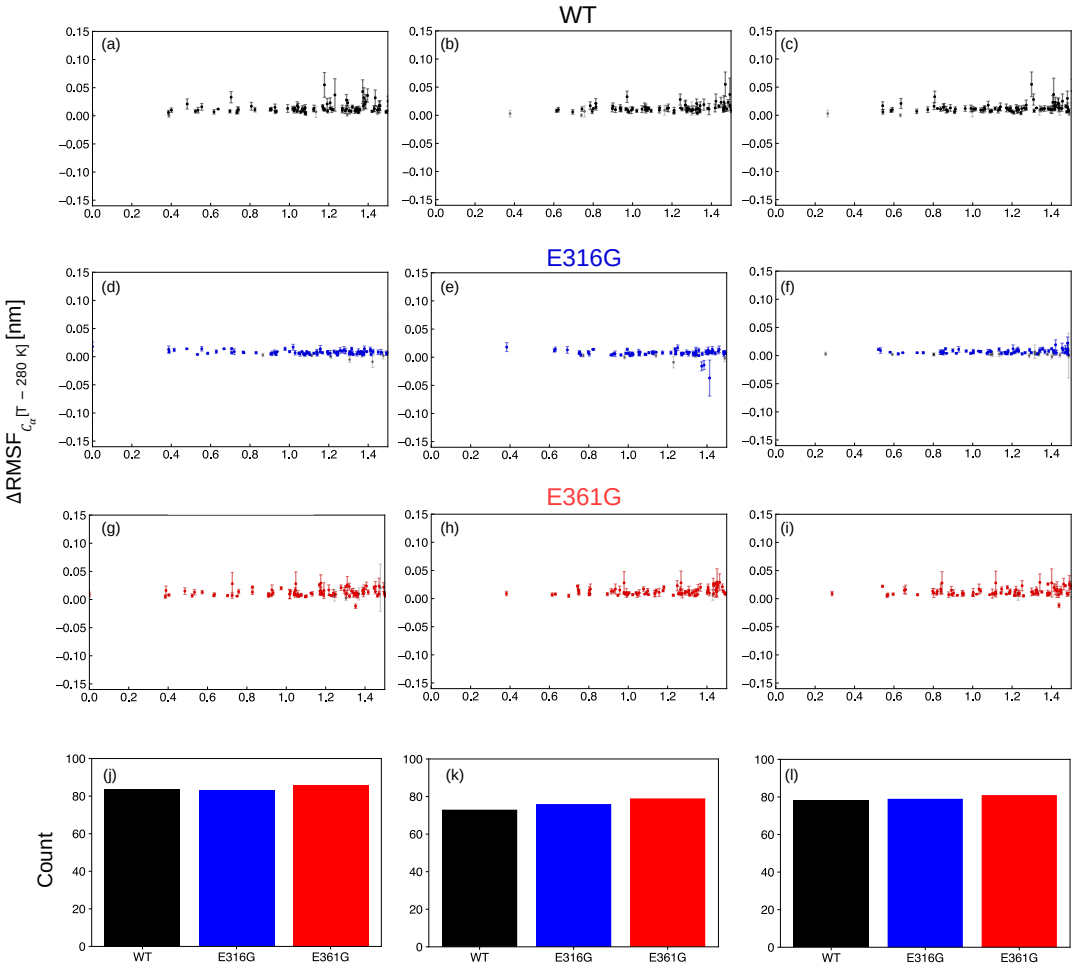

Figure S6: Scatter plot of difference in RMSF (ΔRMSF) at 320 K and at 280 K of WT (black; a, b, c), E316G (blue; d, e, f) and E361G (red; g, h, i) and residues within 1.5 nm of the active site residues — S114 (a, d, g), D318 (b, e, h), and H359 (c, f, i). The number of residues with ΔRMSF > 0 and within 1.5 nm from each of the active site residues 114 (j), 318 (k), and 359 (l). Error bars in the scatter plots are estimated from three independent trajectories.

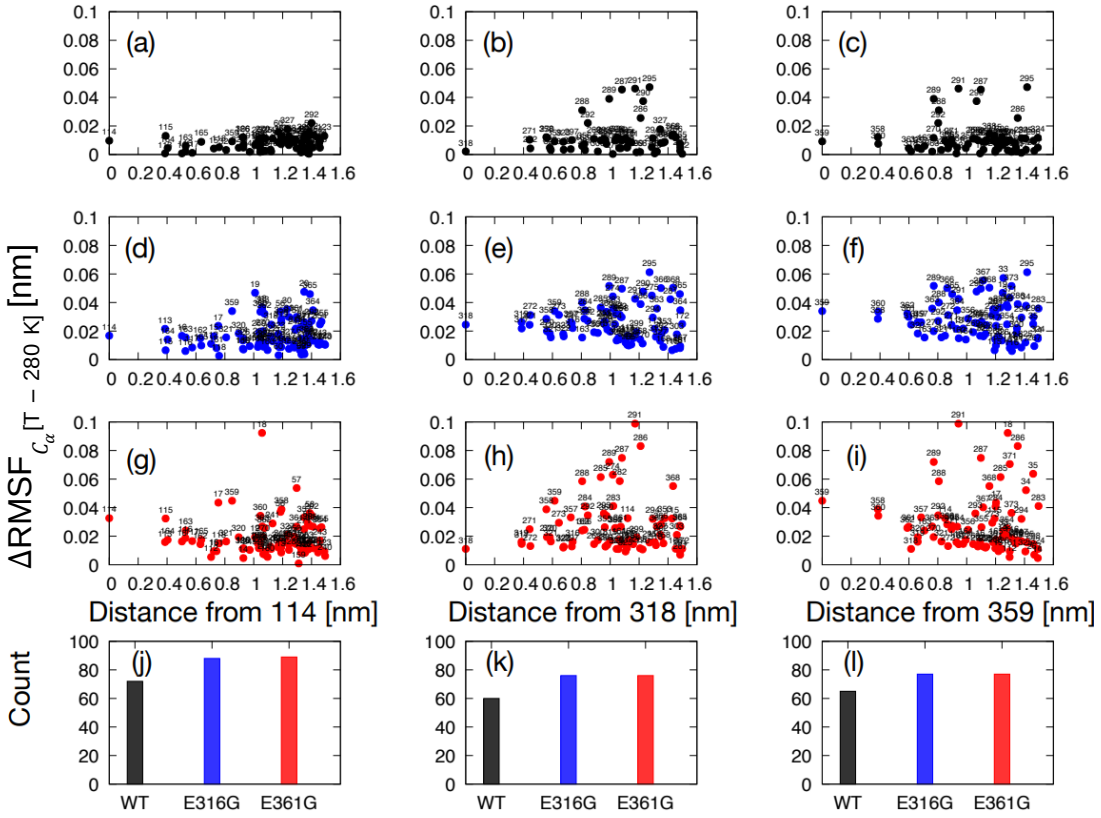

Figure S7: Scatter plot of difference in RMSF (ΔRMSF) a 340 K and 280 K of WT (black; a, b, c), E316G (blue; d, e, f) and E361G (red; g, h, i) and residues within 1.5 nm of the active site residues — S114 (a, d, g), D318 (b, e, h), and H359 (c, f, i). The number of residues with ΔRMSF > 0 and within 1.5 nm from each of the active site residues 114 (j), 318 (k), and 359 (l).

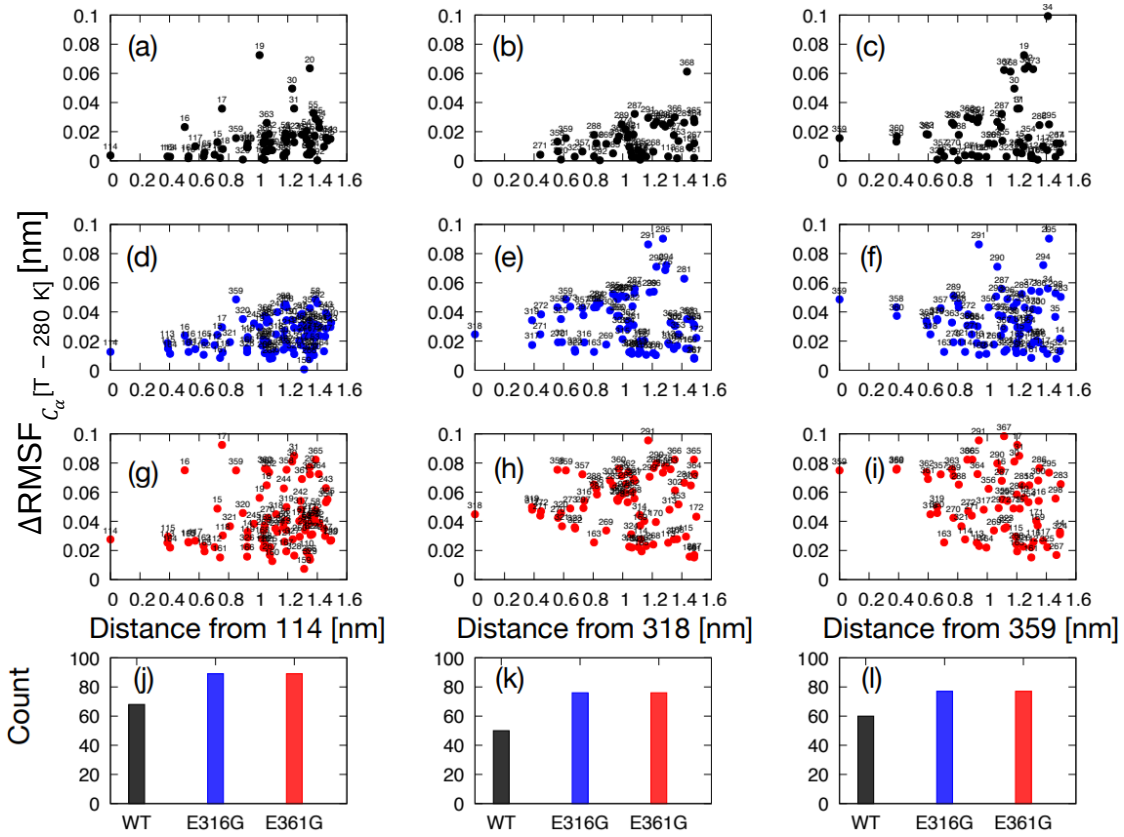

Figure S8: Scatter plot of difference in RMSF (ΔRMSF) a 360 K and 280 K of WT (black; a, b, c), E316G (blue; d, e, f) and E361G (red; g, h, i) and residues within 1.5 nm of the active site residues — S114 (a, d, g), D318 (b, e, h), and H359 (c, f, i). The number of residues with ΔRMSF > 0 and within 1.5 nm from each of the active site residues 114 (j), 318 (k), and 359 (l).

6 Estimated folding-unfolding transition temperatures from CNA

Estimates of folding-unfolding transition temperature obtained by tracking the variation of different global indices with E_{hb} in constraint network analysis (CNA).[S7] These estimates are obtained by applying CNA over the ensemble of structures in one of the independent MD simulation trajectories. The relation between E_{hb} and T is only approximate, and is obtained by comparing E_{hb} transition of different extremophilic proteins and their melting (or growth) temperatures.[S8]

Table S1: Folding-unfolding transition

Enzyme	Global Index	Average $E_{hb,transition}$ (kJ/mol) ± [1]SD	Average [2]T (K) ± SD
WT	$P_{\infty,type1}$	-7.740 ± 0.084	337.03 ± 0.49
	$P_{\infty,type2}$	-7.657 ± 0.084	336.60 ± 0.49
	H_{type1}	-5.52 ± 0.084	326.85 ± 0.40
	H_{type2}	-11.46 ± 0.126	354.83 ± 0.57
E316G	$P_{\infty,type1}$	-8.66 ± 0.084	341.33 ± 0.45
	$P_{\infty,type2}$	-8.577 ± 0.084	341.10 ± 0.45
	H_{type1}	-6.64 ± 0.084	330.74 ± 0.45
	H_{type2}	-11.967 ± 0.126	357.18 ± 0.51
E361G	$P_{\infty,type1}$	-6.90 ± 0.084	331.01 ± 0.39
	$P_{\infty,type2}$	-6.82 ± 0.084	332.70 ± 0.39
	H_{type1}	-5.272 ± 0.084	325.25 ± 0.40
	H_{type2}	-10.418 ± 0.126	349.85 ± 0.58

[1] SD refers to standard deviation obtained from the ensemble based constraint network analysis.
[2] T[S8] $= -20$ (K/(kcal mol^{-1}))$\times E_{hb,transition} + 300$ K
[3] $P_{\infty,type1}$[S9] = Order parameter that tracks the fraction of network belonging to giant percolating cluster.
[4] $P_{\infty,type2}$[S9] = Similar to $P_{\infty,type1}$ but uses the actual largest rigid cluster.
[5] H_{type1}[S9] = Monitors the degree of disorder in the network and is sensitive to initial transitions in unfolding.
[6] H_{type2}[S9] = Similar to H_{type1} but is sensitive to later transitions of folding-unfolding.

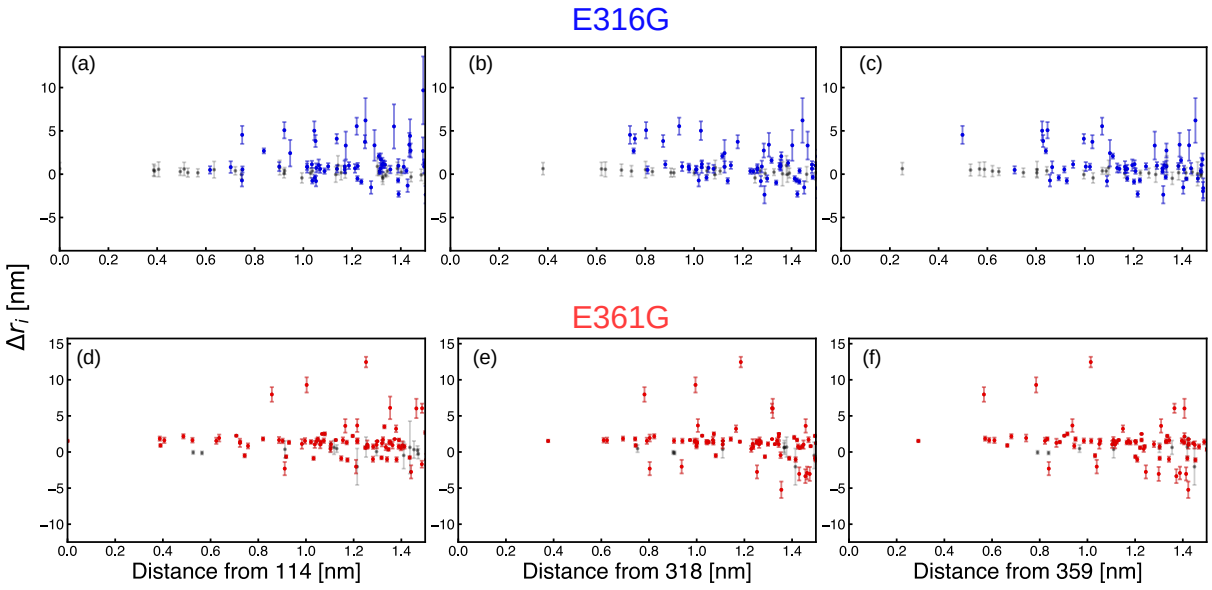

Figure S9: Scatter plot of the difference in rigidity index (r_i) of mutant (b, c, d - E316G; e, f, g - E361G) and WT and distance of residue from the one of the active site residues are shown. Residues with statistically significant differences in r_i are shaded in blue and red for E316G and E361G, respectively.

7 Organization of active site

MD trajectories indicate modified organization of the active site (Fig. S10) in both E316G and E361G when compared to that of WT.

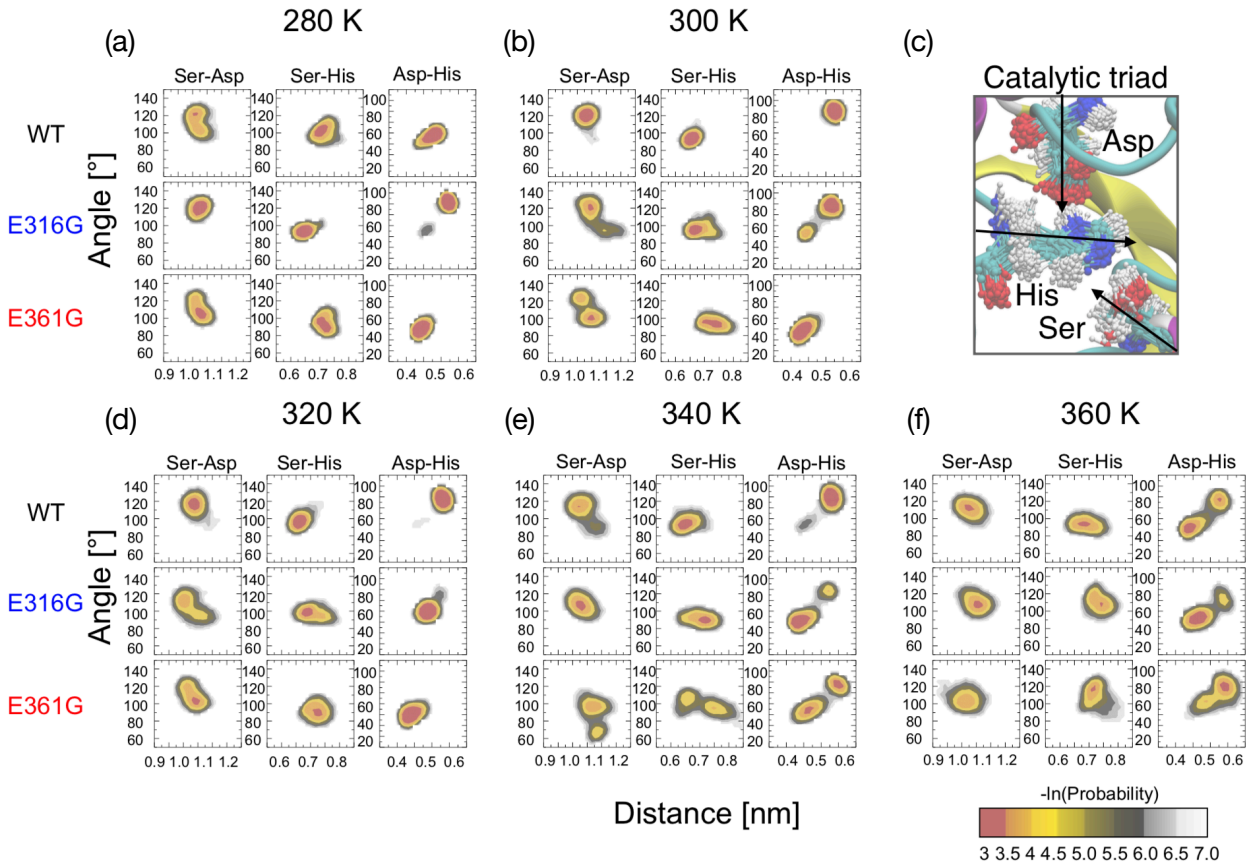

Figure S10: Organization of active site measured by the distance between C_α atoms of active site residues and the angle formed between the side chain vectors as shown in (c) calculated in one of the independent MD simulation trajectory. The line joining the C_α atom of a residue and the center of mass of side chain of the residue forms the side chain vector. The 2D distribution of the distance between C_α atoms of active site residues and the angle formed between the side chain vectors are show for each pair of active site residues at (a) 280 K, (b) 300 K, (d) 320 K, (e) 340 K, and (f) 360 K. Brown and white regions in the heat maps indicate high and low probable regions, respectively.

8 Distance correlation between residue-residue fluctuations at different temperatures

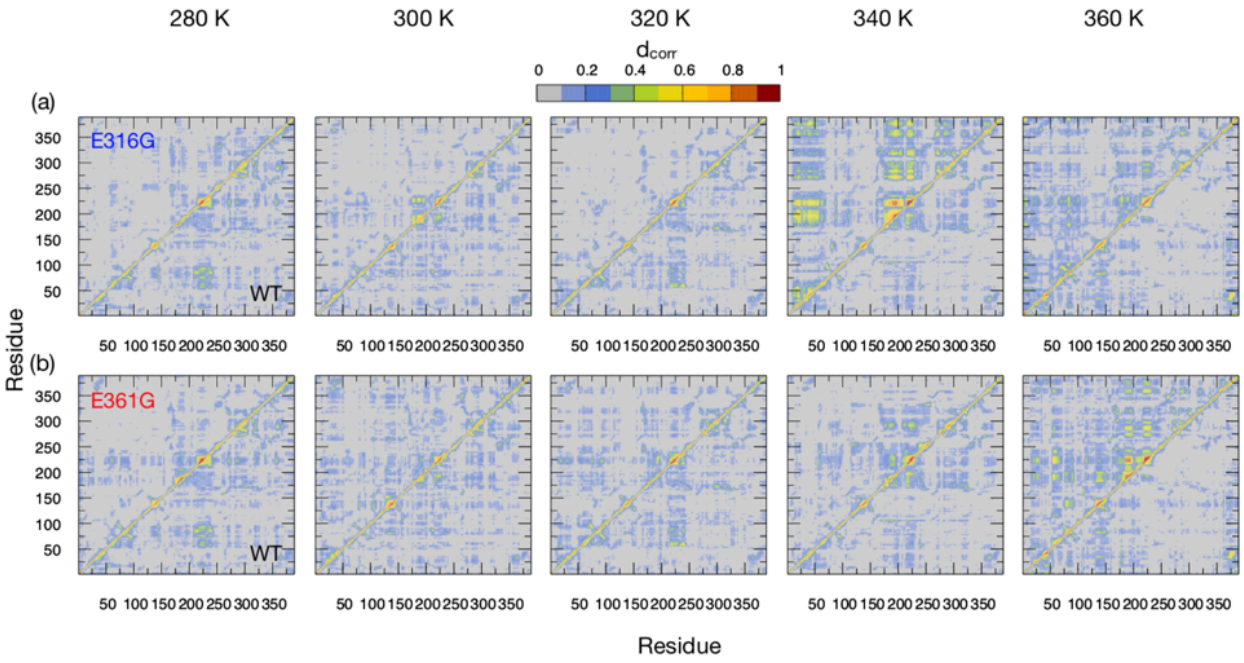

Figure S11: (a) d_{corr} between residues in E316G (upper diagonal) and WT (lower diagonal) at different temperatures, (b) d_{corr} between residues in E361G (upper diagonal) and WT (lower diagonal) at different temperatures. These profiles are generated using one of the independent MD simulation trajectories.

9 Distance correlation between residue-residue fluctuations of mutant residues 316 and 361 with all other residues

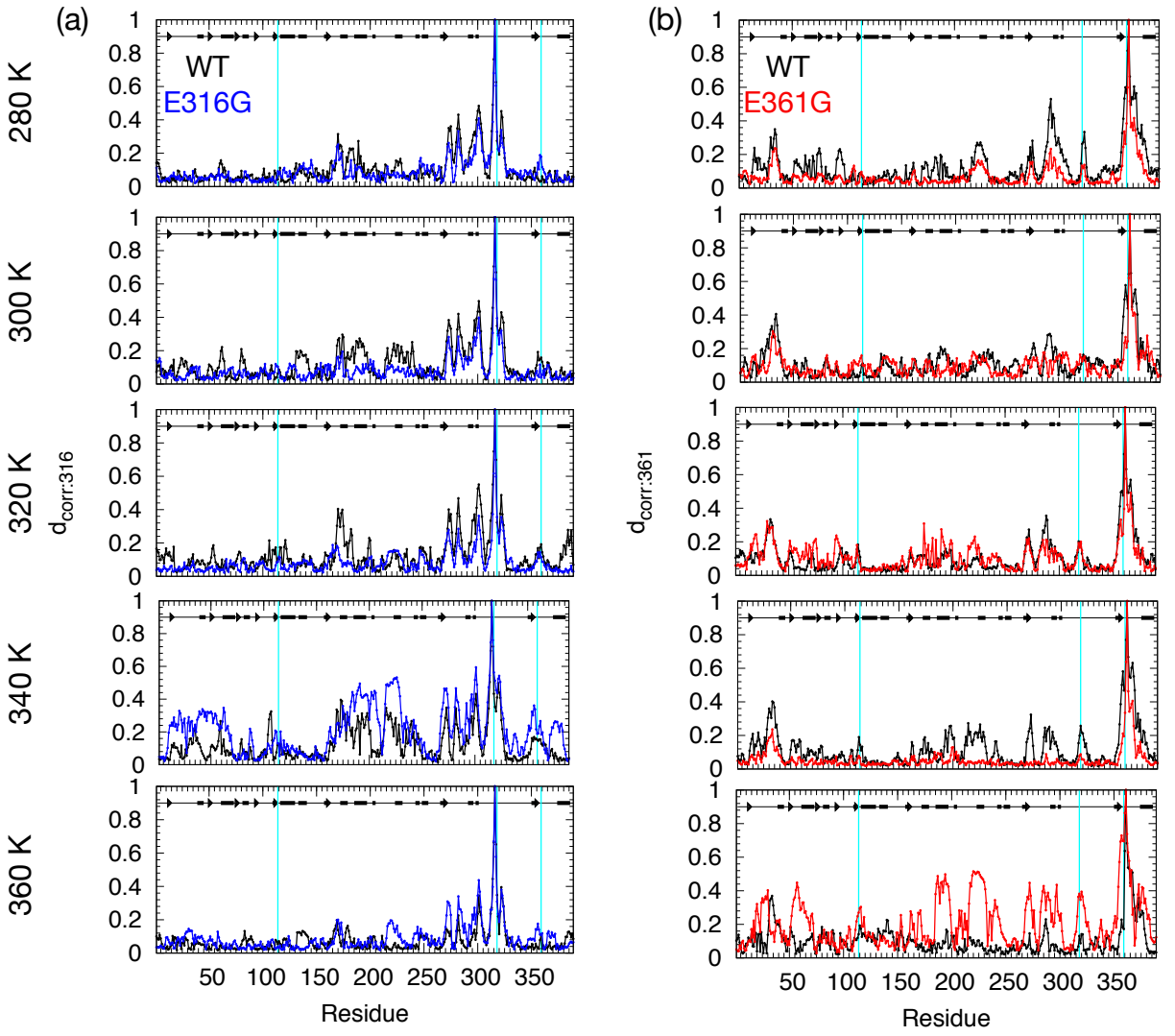

Figure S12: (a) d_{corr} between E316 and other residues in E316G (blue) and WT (black) at different temperatures , (b) d_{corr} between E361 and other residues in E361G (red) and WT (black) at different temperatures. These profiles are generated using one of the independent MD simulation trajectories.

10 Correlation between residue-residue fluctuations and change in interaction energies at 280 K

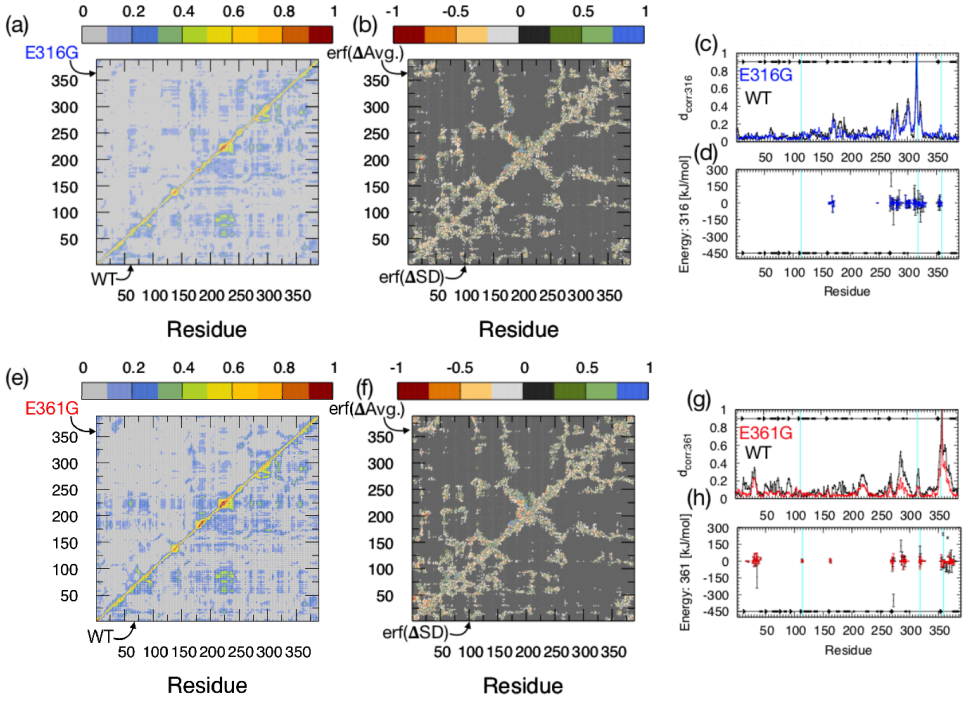

Figure S13: d_{corr} between fluctuation of residue pairs in (a) E316G (upper diagonal) and WT (lower diagonal), (b) Difference in average interaction energy between E316G and WT (upper diagonal) and standard deviation of interaction energy between E316G and WT (lower diagonal). The error function is applied on the average interaction energies to normalize the scale of the energies between -1 and 1 kJ/mol. (c) d_{corr} between fluctuation of residue 316 and other residues in E316G (blue) and WT (black), (d) average interaction energy of residue 316 and other residues in E316G (blue) and WT (black), d_{corr} between fluctuation of residue pairs in (e) E361G (upper diagonal) and WT (lower diagonal), (f) Difference in average interaction energy between E361G and WT (upper diagonal) and standard deviation of interaction energy between E316G and WT (lower diagonal) scaled by the error function, (g) d_{corr} between fluctuation of residue 316 and other residues in E361G (red) and WT (black), (d) average interaction energy of residue 361 and other residues in E361G (red) and WT (black). These profiles are generated using one of the independent MD simulation trajectories.

11 Principal Component Analysis

Principal component analysis (PCA) is a linear dimensionality reduction technique that involves the diagonalization of the covariance matrix of an input feature matrix. It is commonly employed to describe the global motions of proteins in an essential dynamics space.[S10] We employed PCA to understand the general structural differences between WT GTL, the two mutants E316G and E361G at 280 K. We featurized the ensemble of structures of the WT and the mutants obtained from MD simulations using Cartesian coordinates of the backbone C_α atoms. The resulting covariance matrix is given by $C_{ij} = \langle x_i - \langle x_i \rangle (x_j - \langle x_j \rangle) \rangle$, where x is the set of cartesian coordinates that describes atom i or j, and $\langle ... \rangle$ represents an ensemble average. We used the positions of C_α atoms in both WT and the mutants from two of the replicate simulations. Extracted Cartesian coordinates were concatenated into a single feature covariance matrix, such that the principal component space was representative of all three variants. In total, this resulted in a covariance matrix of size 1167 by 1167 with the number of rows and columns equal to the number of C_α atoms (= 389) times the number of dimensions (=3). To estimate the ensemble average, we used 100000 frames from two of the replicate trajectories of both WT and the mutants. The essential subspace of GTL was then obtained by projecting C_α atoms onto eigenvectors 1-5, which captured ∼57% of the total variance. We performed the covariance analysis and the projection of input features onto principal components using `gmx covar` and `gmx anaeig` in GROMACS-2021.3,[S1] respectively.

Fig. S14 shows the projection of each configuration of thermophilic WT (black) and its improved mutants (E316G – blue, E361G – red) with psychrophilic traits sampled in MD simulations at 280 K along the top five leading principal components. The WT and the mutants share certain regions along the leading five PCs. The diverging areas indicate that the leading PCs capturing the collective linear modes of the enzymes can be used for differentiating the WT and the mutants (Fig. S14)

To understand the differences in the configurations of WT and mutants as well as mea-

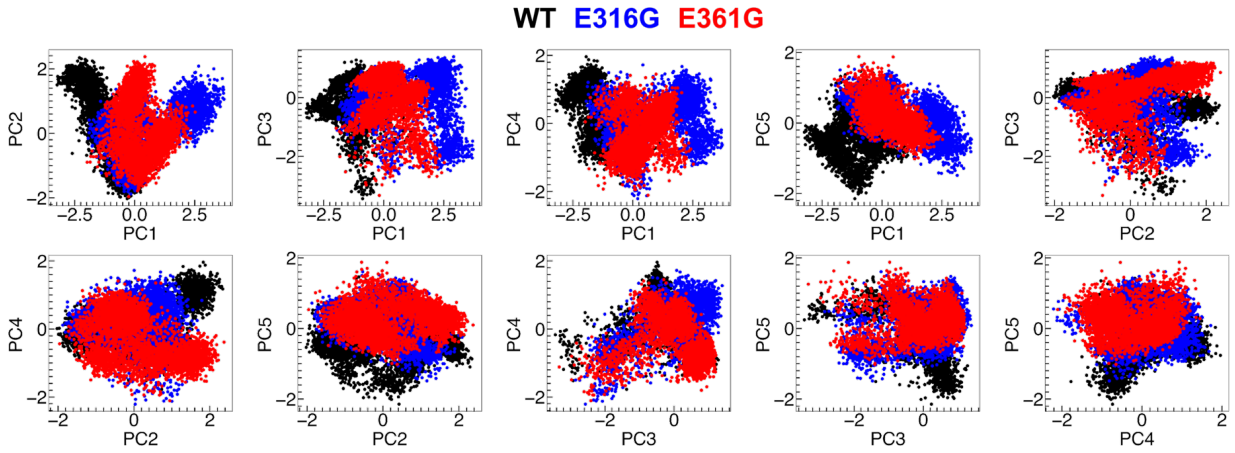

Figure S14: Scatter plots showing the 2D projections of WT and the mutant ensembles along every unique pair of the five leading principal components, PC1-PC5. Scatter markers corresponding to WT, E316G, E361G are in black, blue, and red, respectively.

sure their relative favorability along the principal components, we mapped the free energy landscapes (FEL) along the leading two principal components PC1 and PC2. FEL of both WT GTL and the two mutants E316G and E361G along the leading two principal components PC1 and PC2 are shown in Fig. S15. We obtained these FELs by computing the joint probability distribution along PC1 and PC2 and setting the minimum free energy, G to 0 kT. Free energies of both within combined WT/mutant ensembles as well as across the WT and mutant ensembles are shown in Fig. S15a. Both WT and the mutants share some common favorable regions with similar free energies along PC1-PC2 (Fig. S15a), which correspond to the overlapped regions in the 2D scatter plots shown in Fig. S14.

To better understand the differences in the free energies within the individual ensembles of WT and mutants, we calculated FEL of WT and variants along PC1 and PC2 separately (Fig. S15b-S15d). Both WT and the mutants have favorable regions in different locations along PC1 and PC2 (Fig. S15b-S15d). While these regions have similar free energies, the configurations are different. Representative all-atom configurations in cartoon form corresponding to such favorable regions along PC1 and PC2 in WT, E316G and E361G are marked as S1-S2, S3-S4, and S5-S6 in Fig. 2b, Fig. S15c, and Fig. S15d, respectively. These representative configurations are illustrated in Fig. S15 S1-S6.

The differences in the configurations of WT corresponding to S1 and S2 marked in Fig. S15b are mainly associated with the collective motions along the lid region of the enzyme. This can be seen in the superimposed structures of S2, where the marked points S1 and S2 in Fig. S15b are shown in transparent and opaque, respectively. A similar difference is observed in the favorable structures of the mutants E316G and E361G relative to the WT. The primary difference in the mutants is that the lid has a different arrangement relative to the core α/β hydrolase fold of the enzyme GTL compared to the WT. This is consistent with our analysis of RMSF (main text Fig. 2-4), which showed that the mutations mainly influenced the lid domain fluctuations and its origin can be rationalized by the changes in the lid arrangement and the resulting residue-residue correlations.

To further gain insights into the differences of the configurations of WT and mutants along the leading principal components PC1-PC5, we checked correlations with typical structural metrics such as root mean square deviation and radius of gyration. We observed that the differences in the ensembles of WT and the mutants captured by the leading principal component PC1 and PC2 correlate positively with the RMSD from the initial WT structure and negatively with the radius of gyration of the enzyme. In addition, we observed PC1 and PC2 negatively correlate with the distance between residues F181 and G219 around the lid domain, which is known to undergo significant conformational changes upon activation[S11] as shown in Fig. S16. Together with the changes in lid domain arrangement, the correlation with distance of F181-G219 indicates potential changes to the shape of the binding site and consequently potential changes to the specific activity of the WT enzyme with mutation.

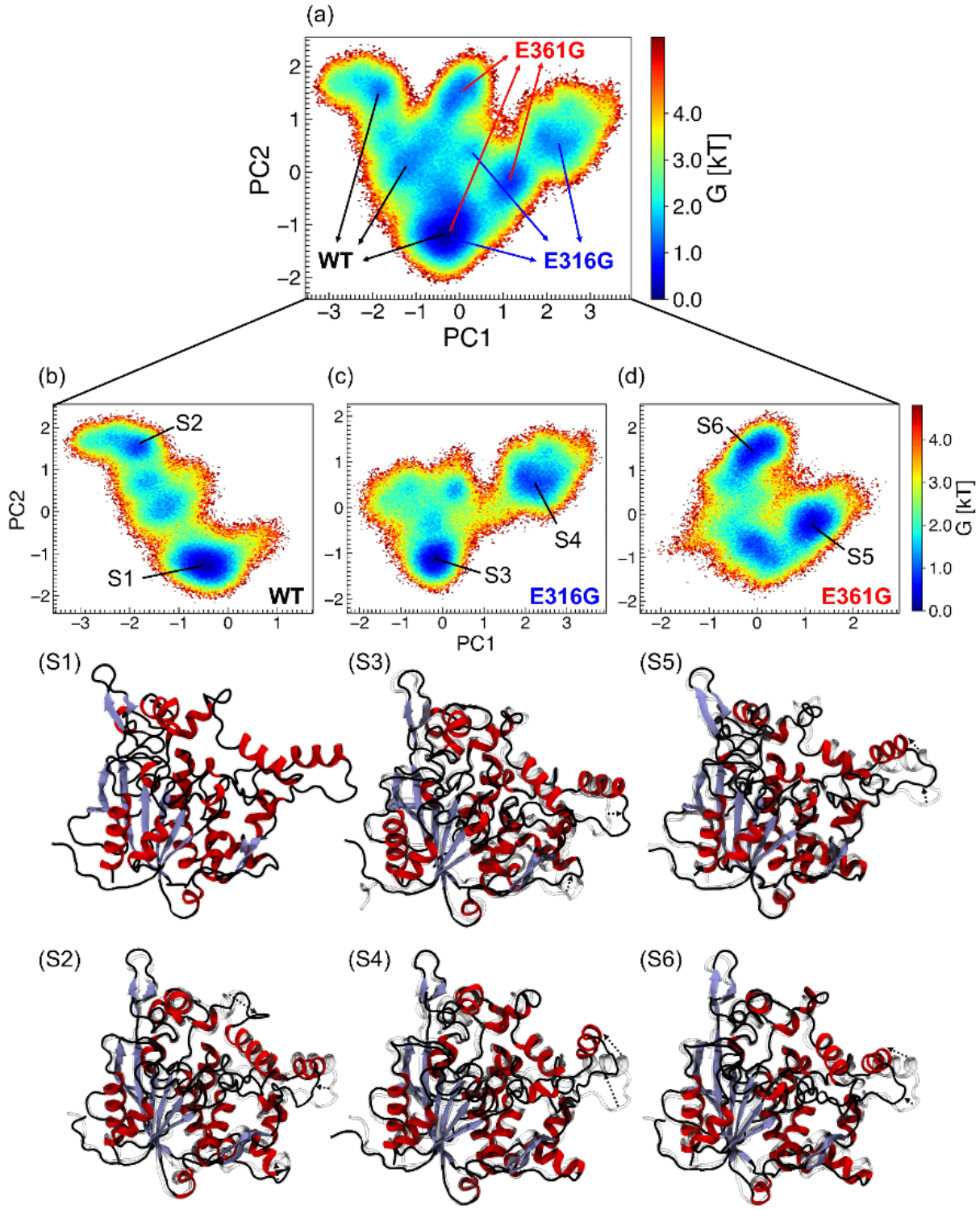

Figure S15: Free energy landscapes of the WT GTL and the two mutants E316G and E361G along the leading two principal components PC1 and PC2 and the representative minimum structures corresponding to the favorable regions. (a) Free energies (G) of both WT and the mutants along PC1 and PC2, showing the free energies within the WT ensemble and the mutant ensembles as well as the relative free energies between the ensembles of WT and the

Figure S15: (continued) two mutants. The favorable regions corresponding to WT (black), E316G (blue), and E361G (red) are denoted by arrows. (b) Free energies of WT ensemble along PC1 and PC2, (c) Free energies of E316G ensemble along PC1 and PC2, (d) Free energies of E361G ensemble along PC1 and PC2. In (a)-(d), the free energy landscapes are shifted to align the minimum free energy G at 0 kT. Representative minimum structures corresponding to favorable regions in (b), (c) and (d) are marked using labels S1-S6 and illustrated in S1-S6. (S1) Representative minimum structures corresponding to the favorable region marked as S1 in panel (b) for WT are illustrated in opaque with helices in red, beta strands in blue, and loops/coils in black. (S2) Representative minimum structures corresponding to the favorable region marked as S2 in panel (b) for WT are illustrated in opaque with identical color code as S1. The transparent snapshots illustrate the favorable WT minimum structure marked as S1 in panel (b). The arrows from the transparent to opaque structure highlight the transition from structure in S1 to the opaque structure in S2. (S3) Representative minimum structures corresponding to the favorable region marked as S3 in panel (c) for E316G are illustrated in opaque with identical color code as S1. The transparent snapshots illustrate the favorable WT minimum structure marked as S1 in panel (b). The arrows from the transparent to opaque structure highlight the changes in structure in S1 corresponding to WT to the opaque structure in S3 corresponding to the mutant E316G. (S4) Similar to S3 but the opaque structure illustrates the configuration corresponding to S4 marked in panel (c). (S5)-(S6) Similar to representative structures shown in (S3)-(S4) but those marked as S5 and S6 in panel (d) corresponding to the mutant E361G. In (S1)-(S6), the differences in the structures (highlighted by dashed arrows) show that configurations obtained in PC1 or PC2 regions are largely associated with changes in the lid domain arrangement relative to the core α/β hydrolase of the enzyme.

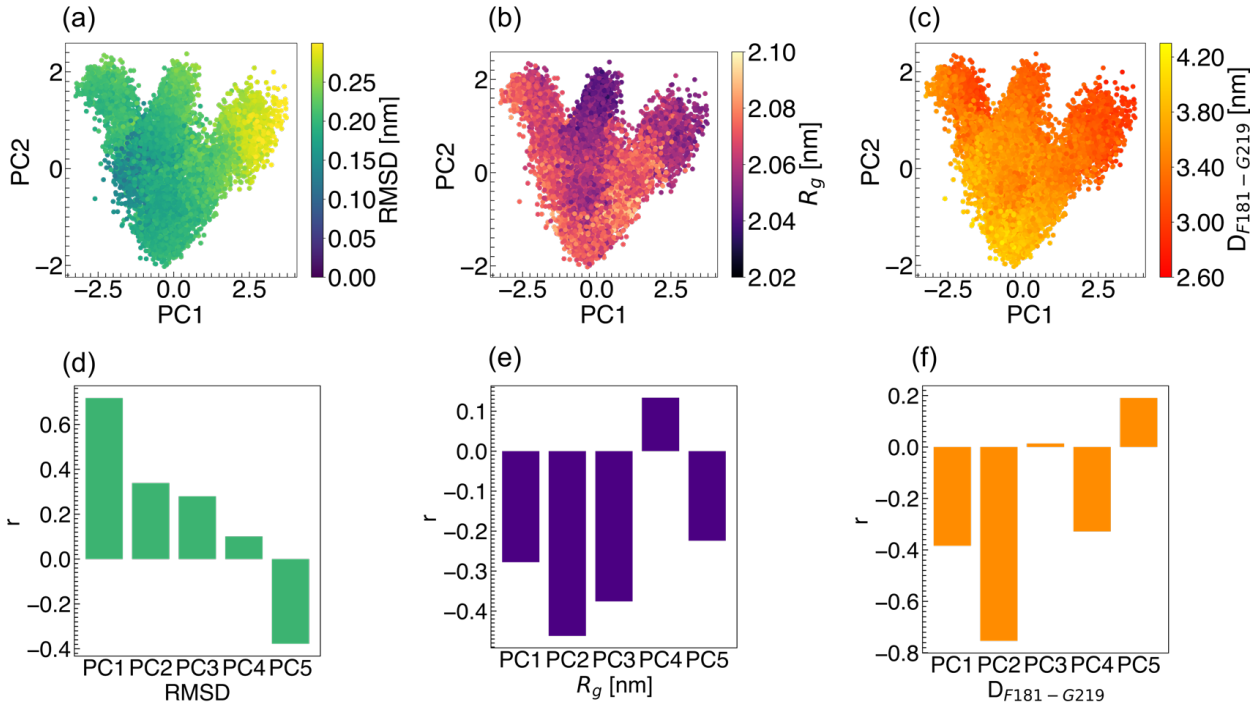

Figure S16: Scatter plots of the projection of the WT and mutant configurations sampled in molecular dynamics simulations along the leading two principal components (PCs) colored according to various structural properties and their Pearson correlation coefficient with the leading five principal components. (a) Scatterplots of PC1-2, colored according to RMSD of both WT and mutants relative to the initial WT structure. PC1 is positively linearly correlated with RMSD. (b) Scatterplots of PC1-2, colored according to radius of gyration Rg of WT and mutants. PC2 is negatively linearly correlated with Rg, indicating it inversely varies with the size of the enzyme. (c) Scatterplots of PC1-5 colored according to the distance separating residues F181 and G219 around the lid domain. PC2 is also linearly negatively correlated with this distance, indicating the potential origin of the differences in the structures of the enzyme during the simulations. (d-f) The Pearson correlations between the principal components and the measured structural properties are shown below their respective plots.

References

(S1) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(S2) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(S3) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(S4) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(S5) Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055–1076.

(S6) Wang, W.; Dasetty, S.; Sarupria, S.; Blenner, M. Rational engineering of low temperature activity in thermoalkalophilic Geobacillus thermocatenulatus lipase. *Biochem. Eng. J.* **2021**, *174*, 108093.

(S7) Krüger, D. M.; Rathi, P. C.; Pfleger, C.; Gohlke, H. CNA web server: rigidity theory-based thermal unfolding simulations of proteins for linking structure,(thermo-) stability, and function. *Nucleic Acids Res.* **2013**, *41*, W340–W348.

(S8) Radestock, S.; Gohlke, H. Protein rigidity and thermophilic adaptation. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 1089–1108.

(S9) Pfleger, C.; Radestock, S.; Schmidt, E.; Gohlke, H. Global and local indices for characterizing biomolecular flexibility and rigidity. *J. Comput. Chem.* **2013**, *34*, 220–233.

(S10) Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K. Principal Component Analysis and Long Time Protein Dynamics. *The Journal of Physical Chemistry* **1996**, *100*, 2567–2572.

(S11) Carrasco-López, C.; Godoy, C.; de las Rivas, B.; Fernández-Lorente, G.; Palomo, J. M.; Guisán, J. M.; Fernández-Lafuente, R.; Martínez-Ripoll, M.; Hermoso, J. A. Activation of bacterial thermoalkalophilic lipases is spurred by dramatic structural rearrangements. *J. Biol. Chem.* **2009**, *284*, 4365–4372.