EMI Series

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

Design, System, Application

Enzymes are eco-friendly and natural molecules with excellent properties. It is highly desirable to rationally engineer enzymes for a target application and its conditions. However, there are no clear design principles that enable this using rational site-directed mutagenesis approaches. In this work, we use molecular dynamics simulations to probe the activity-temperature relation used to explain the tradeoff between activity and stability in thermophilic and psychrophilic enzymes. Specifically, we investigate whether the conventional idea that higher active site flexibility leads to activity at low temperature in psychrophilic enzymes can be used as a design principle to incorporate low temperature activity in a thermophilic enzyme for engineering broad range temperature active enzymes. Our results indicate that simple design rules like reducing hydrogen bonding residues near the active site do not allow for changing active site flexibility in isolation and lead to changes in flexibility in the entire enzyme. Consequently, undesirable functional and specificity changes in the enzyme are observed. We demonstrate that studying residue-residue flexibility correlations can address this challenge and provide appropriate design guidelines to rationally engineer the active site flexibility. Hence, providing an ability to engineer broad range temperature active enzymes with potential applications in pharmaceutical, textile and food industries.

Journal Name

ARTICLE TYPE

Exploitation of active site flexibility-low temperature activity relation for engineering broad range temperature active enzymes[†]

Siva Dasetty,[a,c] Jonathan Zajac,[b] and Sapna Sarupria[*a,b]

Differences in the structural and thermodynamic properties of enzymes adapted to different temperatures indicate that broad range temperature active enzymes can be designed by incorporating cold activity in thermophilic enzymes. This is based on a concept that the cold activity and thermostability are not mutually exclusive and that cold activity in psychrophilic enzymes is associated with active site flexibility. In Wang et al. *Biochem. Eng. J.* **2021**, *174*, 10803[1], we identified two point mutants of *Geobacillus thermocatenulatus* lipase (GTL) which were screened to improve active site flexibility. Even though the identified thermophilic mutants had psychrophilic traits, we observed complex trends such as higher kinetic stability and substrate-dependent activity-temperature relation on further analysis. In this work, we apply molecular dynamics simulations and network theory to show that the changes in GTL properties with the selected mutations cannot be directly associated with active site flexibility. Our computational results indicate the mutations resulted in residues with both higher and lower flexibility, which are both proximal and away (> 1.5 nm) from the active site. We show that the intricate changes in the flexibility of residues distal from the mutation site can be rationalized by the altered dependency between residue-residue fluctuations with mutation. These alterations in residue-residue flexibility dependency are a consequence of the redistribution of the inter-residue interactions from the mutation site to other residues, which are driven by several tightly connected charged residues. This indicates design rules associated with residue-residue flexibility correlations are critical in applying site-directed mutagenesis to successfully exploit active site flexibility-activity relation for incorporating low temperature activity in thermophilic enzymes. Similarly, such correlations can be valuable in minimizing false positives in high-throughput screening methods based on directed evolution and/or machine learning-based engineering of enzyme activity-temperature relation.

1 Introduction

Enzymes are primarily proteins with critical biological functions as catalysts and numerous promising industrial applications[2–7] because of their excellent specificity, selectivity and environmentally friendly properties.[8] It is of both fundamental and com-

mercial interest to understand the design principles for engineering enzyme activity-temperature relation. This is because it helps in developing insights into the evolution of activity-stability of enzymes in extremophiles and also in engineering enzymes to function in a wide range of temperatures for applications such as biosensors in wearable bioelectronics,[9] detergents,[10] and in pharmaceutical industries.[11] To this end, differences in the structure of enzymes in extremophiles as well as their activity-temperature relation are widely studied.[8,10,12,13]

Enzymes in thermophiles that function at high ($\sim$ 318–394 K[1]) temperatures are typically characterized by tighter hydrophobic cores, shorter loops, a higher number of charged residues, and an overall rigid structure relative to their homologous enzymes in mesophiles and psychrophiles.[8,13–15] These traits are considered to help in maintaining high thermostability of thermophilic enzymes, which is required for functioning at high tem-

peratures.[1,13] In contrast, enzymes in psychrophiles have no selective pressure for thermostability and have to defy the exponential activity-temperature[16,17] relation to be active at low ($<$ $\sim$293 K[18]) temperatures.[13,19,20] The thermodynamic factor responsible for the weaker exponential dependence of activity on temperature for enzymes in psychrophiles can be rationalized by their weaker enthalpy of activation ($\Delta H^{\ddagger}$) at a given temperature than the homologous enzymes in thermophiles.[12,13,19,21] However, the structure-related traits responsible for low temperature activity in psychrophilic enzymes remain elusive. Current understanding points to a potential relationship between low temperature activity in psychrophilic enzymes and higher flexibility in their active site. This is supported by potential weaker interactions within the active site of enzymes in psychrophiles relative to mesophiles and thermophiles, as reflected by a relatively lower enthalpy ($\Delta H^{\ddagger}$) and more negative entropy ($\Delta S^{\ddagger}$) of activation[12] at a given temperature and a higher Michaelis-Menten (K_m) constant.[1,12,19]

In our previous work, we applied the knowledge of the structure and thermodynamic differences between enzymes in extremophiles to develop a better understanding of the design rules required for rationally controlling activity-temperature relation. Specifically, we relied on the concept that low temperature activity and thermostability are mutually exclusive[1,7] for designing robust industrial enzymes capable of functioning at a broad range of temperatures. The concept implies low temperature activity can be incorporated into thermostable thermophilic enzymes by simply modifying their active site flexibility. This is in contrast to the alternative popular strategy focused on engineering thermal stability of mesophilic enzymes by enhancing the rigidity of certain key residues to improve their high temperature function.[22]

Modulating active site flexibility without perturbing the arrangement of the catalytic residues is however challenging because proteins typically form small-world networks[23]. To test the concept and develop an effective approach to modulate active site flexibility, we considered a simple starting point by hypothesizing that reducing the number of hydrogen bonds in a given local region would increase its flexibility while preserving the secondary structure. For this, we screened the residues around the active site that would result in a lower number of hydrogen bonds when mutated to glycine using the crystal structure as a guide. We applied this approach to *Geobacillus thermocatenulatus* lipase (GTL), a model thermophilic enzyme that has numerous industrial applications[24–28].

From screening wet-lab soluble protein expression and activity experiments[1], we observed two point mutants E316G and E361G of GTL preserved the overall secondary structure and resulted in a better specific activity-temperature relation and thermostability than the wild-type (WT) GTL.[1] However, the change in specific activity was not restricted to low temperatures and the observed trend was specific to a short substrate p-nitrophenol butyrate. Interestingly, we observed traits of enzymes in psychrophiles such as lower $\Delta H^{\ddagger}$ and more negative $\Delta S^{\ddagger}$ with the shorter substrate in the two mutants relative to WT GTL. It is, however, unclear whether the changes in the activity-temperature relation observed in the two mutants E316G and E361G were a direct result of their

higher active site flexibility relative to WT.

In this work, our primary goal is to explore the correlation between active site flexibility and the specific activity-temperature relation observed in E316G and E361G mutants relative to WT GTL. For this, we applied all-atom molecular dynamics (MD) simulations to measure flexibility using root mean square fluctuation (RMSF) and constraint network analysis (CNA). We found no clear correlation between changes in specific activity and active site flexibility measured using RMSF at low temperatures where the mutants have better specific activity than the WT. While both the RMSF and CNA-based flexibility metrics revealed certain residues proximal to the active site have more flexibility on average, there are also residues that have less flexibility in the mutants relative to WT that are distributed throughout the enzyme.

To better understand the origin of the changes with the point mutations, we applied network theory approach and observed that the mutations resulted in both weaker and stronger inter-residue interactions that are distributed throughout the enzyme by a network of tightly connected charged residues. This revealed the likely reason for the unexpected changes in activity-temperature relation and kinetic stability of GTL with the mutations observed in Wang *et al.*[1]. In addition, this highlights that the context of the mutation site is critical for inducing predictable changes in active site flexibility. For this, information associated with correlations between residue-residue fluctuations is crucial along with models to predict changes in enzyme structure with point mutations. This can lead to the successful exploitation of active site flexibility-activity relation to rationally incorporate low temperature activity in thermophilic enzymes. In the following, we first present the details of our flexibility measurements and discuss the observed trends in flexibility with mutation. We then present our understanding of the changes in the structure of GTL with mutation and conclude by presenting our outlook for furthering hypothesis-driven approaches to rationally design broad-range temperature active enzymes.

2 Methods

2.1 Molecular dynamics simulations

2.1.1 Initial structure of WT GTL and the mutants

Fig. 1 shows the initial structure of the thermoalkalophilic WT GTL used in MD simulations that was obtained from the Protein Data Bank (PDB ID 2W22). WT GTL has a molecular weight of $\sim$43 kDa with optimal activity at 338 K and pH 8-10.[29] It contains 389 residues and has a α/β hydrolase fold with two cofactor metal ions (Zn^{2+} and Ca^{2+}), where the Zn^{2+} ion plays a role in the thermostability of the enzyme.[29] The core of WT GTL hydrolase fold contains seven sheets of β strands surrounded by α helices. This is covered by a lid domain protecting active site, which is comprised by an amphipathic helix-loop motif as highlighted in Fig. 1a. The three catalytic residues that are common to α/β hydrolases[8] and are often referred to as the catalytic triad are Ser114, His359, and Asp318 (Fig. 1b). The location of the residues used for generating the two point mutants E316G (Glu316 to Gly316) and E361G (Glu361 to Gly361) with better
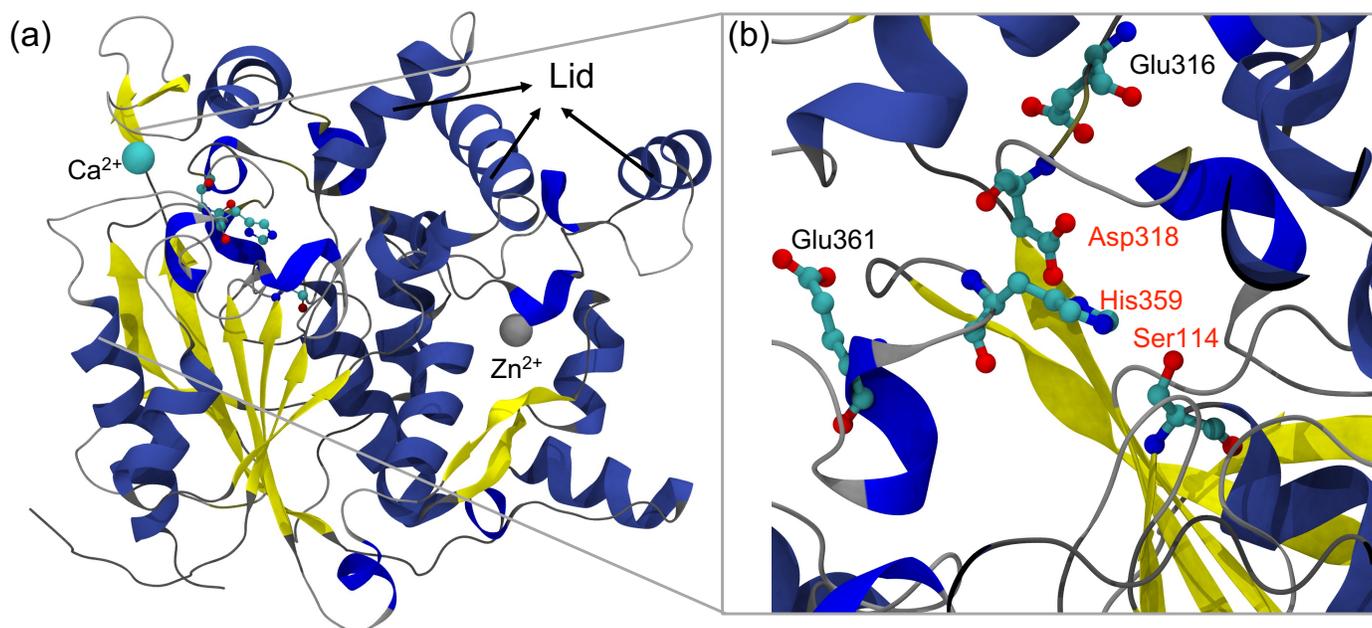
Fig. 1 Illustration of the crystal structure of WT GTL (PDB ID: 2W22) that was used for generating the structure of the two mutants E316G and E361G and subsequently as an initial structure for performing molecular dynamics simulations of WT and mutants at different temperatures. (a) Cartoon representation of the WT GTL with α-helices, β-sheets, and random coils colored in blue, yellow, and gray, respectively. The catalytic triad residues are represented by a ball and stick model with carbon, nitrogen, and oxygen atoms colored in cyan, blue, and red, respectively. The two co-factor metal ions Zn^{2+} and Ca^{2+} are represented by van der Waal spheres and colored in gray and cyan, respectively. The amphipathic helix-loop motif protecting the active site (i.e., the lid domain) is highlighted by arrows. (b) Magnification of the region surrounding the active site of GTL. The two point mutants E316G (Glu316 to Gly316) and E361G (Glu361 to Gly361) are separated by one residue from the catalytic residues Asp318 and His359, respectively and are shown using a ball and stick model.

activity-temperature relation than WT observed in experiments[1] are highlighted in Fig. 1b. The residues used for creating the two point mutants E316G and E361G are situated one residue from the catalytic residues Asp318 and His359, respectively. Swiss-PdbViewer[30] was employed to generate the initial mutant structures of E316G and E361G.

2.1.2 System setup for molecular dynamics simulations

The protein atoms, and cofactor metal ions (Zn^{2+} and Ca^{2+}) were taken from the crystal structure of WT GTL. The missing N-terminus 1Met residue in the resulting structure was added using Swiss-PdbViewer[30]. The side chains with the lowest rotamer score in Swiss-PdbViewer were chosen for adding 1Met and for mutation of E316G and E361G. The protonation states of the charged residues Lys, Asp, and Glu were determined based on their pKa values in water and pH of ~ 8 used in experiments.[1,31,32] For certain His residues, we used the local protein environment based on its ability to form a hydrogen bond (defined using hydrogen-donor-acceptor angle and donor-acceptor distance) with surrounding residues to determine its protonation state.[31,32] Both WT GTL and the mutants were first placed in the center of a box with dimensions $9 \times 9 \times 9$ nm^3. The dimensions of the box were determined by assuming that the proteins will remain in the folded state during the course of the MD simulations based on experimental observations[1], while ensuring a minimum of 2.5 nm distance between a protein atom and the box edge to avoid system size effects. The resulting systems were

energy minimized by employing the steepest descent algorithm until convergence or to a maximum of 500000 steps using GROMACS 5.0.2[33] for stabilizing any high energy configurations. We employed the standard protein force field Amberff99SB-ILDN[34] for both WT and the mutants for simulating at all the target temperatures. The energy-minimized systems were then solvated by using the Amberff99SB-ILDN force field compatible water model TIP3P[35] with the `gmx solvate` tool in GROMACS 5.0.2[33]. Chloride ions were then added by replacing randomly selected water molecules using the `gmx genion` tool in GROMACS 5.0.2[33] to neutralize the net positive charge of each system. The resulting systems containing ~ 73070 atoms were then subjected to an additional energy minimization run to prevent any high-energy contacts between the water and enzyme atoms.

2.1.3 Molecular dynamics simulation details

A cutoff distance of 1.0 nm was used for calculating the short-range nonbonded Lennard-Jones and two-body Coulomb interactions. Particle mesh Ewald summation method[36] was employed to compute the long-range electrostatic interactions. We applied the LINCS algorithm[37] to constrain the covalent bonds involving hydrogen atoms in the enzyme, which allowed the usage of a timestep of 2 fs with the leap-frog integrator[38]. Neighbor lists were updated using Verlet list method[39] in GROMACS 5.0.2[33]. The SETTLE[40] algorithm was used to keep the water molecules rigid.

Following energy minimization during simulation setup, short

equilibration MD simulations of 1 ns in the NVT ensemble and 1 ns in the NPT ensemble were performed to prepare systems at a given target temperature and target pressure (1 bar). Five target temperatures (280 K, 300 K, 320 K, 340 K, 360 K) were used in this work to measure differences in flexibility at a given temperature and to analyze the change in flexibility with temperature. The canonical velocity rescaling[41] thermostat and Berendsen barostat[42] were used to control the temperature and pressure during the equilibration process. The final production simulations were performed in an NPT ensemble for 350 to 450 ns depending on the convergence of root mean square deviation (RMSD) with energy minimized initial structures (SI Fig. S1). The Nosé-Hoover thermostat[43,44] and Parinello-Rahman barostat[45,46] were used in production runs. Separate temperature coupling constants of 5 ps and 1 ps for protein and solvent, respectively, were applied to minimize temperature gradients for both equilibration and production runs as recommended by Lingenheil *et al.*[47]. Isotropic barostats with pressure coupling constant of 5 ps were utilized in both equilibration and production runs. The compressibility of the system was set to that of water (4.5×10^{-5} bar^{-1}) for both equilibration and production runs. Long-range dispersion corrections were applied for updating energy and pressure during the simulation.

The trajectories from the production simulations were saved every 5 ps for analysis. For temperatures 280 K, 300 K, and 320 K, three replicate simulations of length 500 ns were performed to estimate standard errors (S.E.). MD simulation details used for generating the replicate trajectories are described in SI Section 1. GROMACS-5.0.2[33] was utilized for preparing and performing the simulations. Each simulation of length 500 ns took $\sim$ 172.5 GPU-h on a node containing 2 GPUs and 24 CPUs. In total, we performed 33 simulations of an aggregated length of 16.5 μs with an estimated computational time of $\sim$ 5692.5 GPU-h. VMD-1.9.3[48] and UCSF Chimera[49] were used for visual analysis of the trajectories.

2.2 Analysis of enzyme flexibility

We applied two measures for quantifying the local flexibility of residues in the enzyme. Root mean square fluctuation (RMSF) of residues observed in MD trajectories is a common measure for quantifying flexibility of residues and to check its correlations with activity-stability of proteins.[50–54] We utilized the `gmx rmsf` tool in GROMACS 5.0.2[33] for calculating the RMSF of residues in both WT and mutants observed in MD simulations. We discarded the first 50 ns of the trajectory as a part of the equilibration cycle.

In addition to RMSF, flexibility of enzymes or proteins can be measured using linear response theory[55] and network theory-based metrics[56,57]. In this work, we apply a network theory approach, where atoms in proteins are represented by nodes in a graph. The edges connecting the nodes represent covalent and/or non-covalent bonds between the atoms.[56] According to network theory-based approaches, an atom in protein is considered flexible if it has free degrees of freedom. Alternatively, it can be interpreted as the ability of an atom in the network to move upon application of external force[58]. To identify and quantify the flexi-

bility of residues using network theory, we applied constraint network analysis (CNA)[57].

2.2.1 Quantification of flexibility using constraint network analysis

Body-and-bar is one of the common models used for constructing the protein constraint network, where the number of bars (constraints) between atoms are dependent on the nature of their interaction. For capturing the relative strength of interactions — six, five, and two bars are used for describing covalent, hydrogen-bond or salt-bridges, and hydrophobic interactions, respectively. Hydrogen-bonds and salt-bridges are included as constraints within the protein network when the strength of such interaction computed using energy expressions defined by Dahiyat *et al.*[59] is less than a cutoff value (E_{hb}). The hydrophobic interactions are identified by a distance criterion between carbon and/or sulfur atoms. While initial studies used thermal energy ($k_B T$) as E_{hb}, later studies[57] related E_{hb} to temperature. For this, the folding-unfolding transition is simulated and the phase transition temperature is calculated by constructing networks at different values of E_{hb}. Radestock and Gohlke[60] established the relation between E_{hb} and temperature by comparing the E_{hb} of transition and melting (or growth) temperatures of proteins in extremophiles calculated experimentally.

In this work, we applied the rigidity index (r_i) that is formulated by Radestock and Gohlke[60] to measure flexibility using network theory. r_i is calculated for each backbone bond of C_α atom$_i$ in the protein. The average r_i of all bonds connected to each atom$_i$ is then reported in the units of energy, which can be interpreted as the energy required to break a bond apart from any rigid cluster in the protein body-bar network. We utilized an ensemble of structures extracted every 0.5 ns starting at 50 ns from the MD trajectories at 300 K to obtain the statistics of r_i. The CNA parameters E_{hb} were set to a range of -25.104 kJ/mol to -0.418 kJ/mol with 0.418 kJ/mol increment. These parameters correspond to a target temperature range of interest $\sim$302 K to $\sim$420 K with 2 K increment and are selected using the E_{hb}-temperature relation reported in prior studies.[61–63] We used the web-server https://cpclab.uni-duesseldorf.de/cna/main.php developed by Gohlke's group to perform CNA.

3 Results and Discussion

3.1 Change in enzyme flexibility with mutation

3.1.1 Flexibility using RMSF

RMSF is calculated for each residue at different temperatures for WT and mutants by aligning the respective trajectories with the respective initial structures used for the simulations. RMSF and the related metric B-factors from crystal structures are commonly applied for understanding the correlation between flexibility and activity-stability of proteins.[50–54,65] Fig. 2 and Fig. 3 show the changes in RMSF of C_α atoms in each residue of the two mutants at 280 K and 320 K, respectively. 280 K is close to the lowest studied temperature in experiments and 320 K is around the optimal activity temperature of WT.[1] The influence of thermal energy at lower temperatures will be minimal and therefore can reveal in-
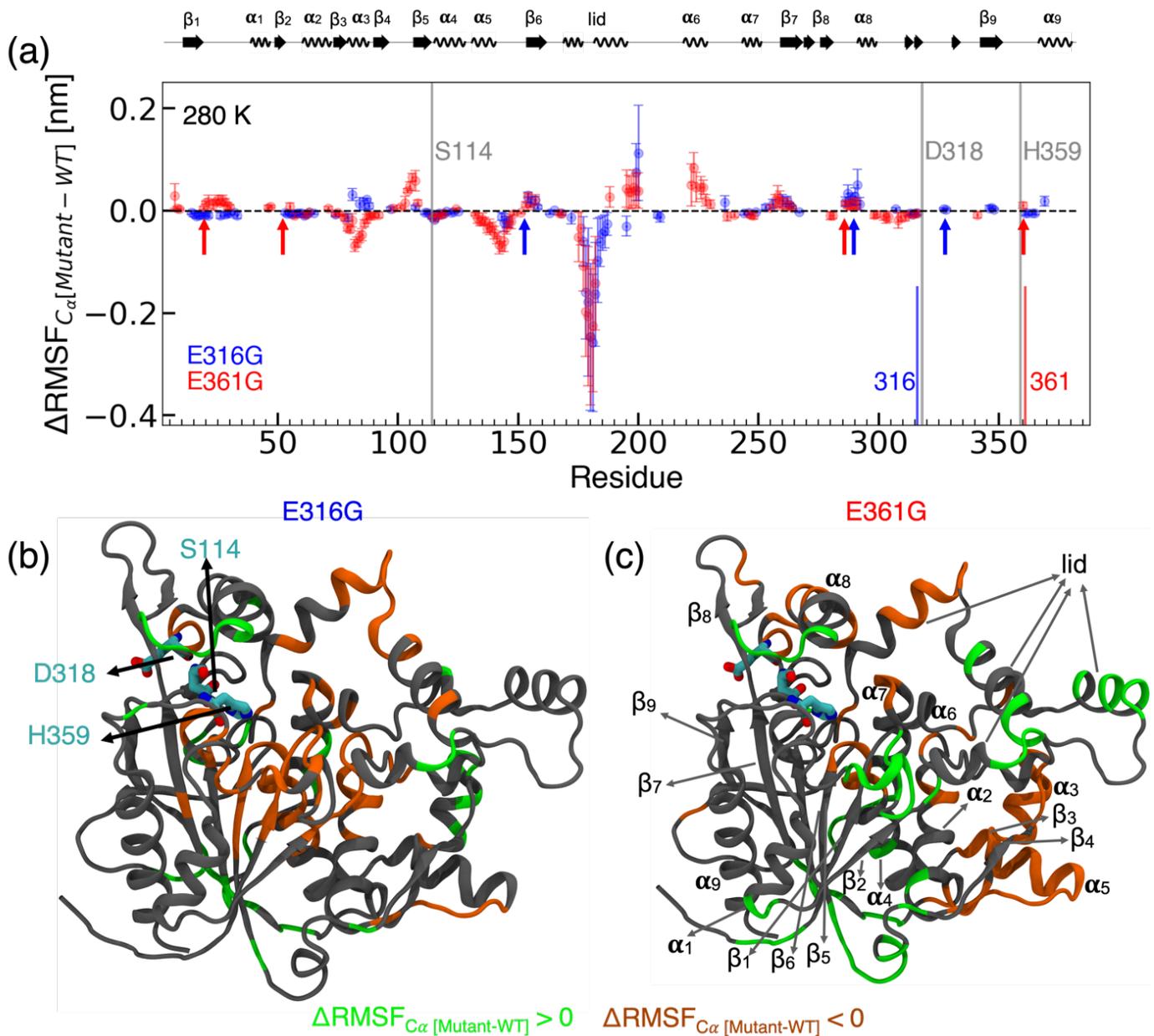
Fig. 2 Changes in the flexibility of GTL with E316G and E361G mutations measured using RMSF at 280 K. (a) Difference in RMSF of C_α ($\Delta RMSF_{C_\alpha(Mutant-WT)}$) of residues at 280 K in E316G and WT, and E361G and WT are shown in blue and red, respectively. Error bars represent standard errors obtained by measuring RMSF for WT and mutants using three independent MD simulation trajectories. Secondary structure of WT along the sequence was obtained using Biotite[64] and is shown at the top of (a). Arrows in blue and red refer to regions with higher RMSF in mutants than WT and that are within 1.5 nm from the active site residues in E316G and E61G, respectively. The locations of catalytic residues are shown using vertical lines in gray. For clarity, only residues with statistically significant changes in RMSF in WT and mutants are shown. (b) Residues in the crystal structure of WT GTL are colored according to their $\Delta RMSF_{C_\alpha(Mutant-WT)}$ of residues in E316G relative to WT, (c) Residues in the crystal structure of WT GTL are colored according to their $\Delta RMSF_{C_\alpha(Mutant-WT)}$ of residues in E361G relative to WT. In (b) and (c), residues with $\Delta RMSF_{C_\alpha(Mutant-WT)} > 0$ and $\Delta RMSF_{C_\alpha(Mutant-WT)} < 0$, i.e., more and less flexible in the mutant relative to WT are colored in green and orange, respectively.

herently flexible residues in the enzyme. On the other hand, the RMSF around optimal temperature of WT can be used for understanding the connection between optimal activity and flexibility of residues in the enzyme. In addition, 280 K and 320 K are close to the temperatures at which $\Delta G^{\ddagger}$ is similar for WT and E316G, and WT and E361G, respectively. We obtained these temperatures by extrapolating $\Delta G^{\ddagger}$-T curves linearly using the experimentally

measured $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ values (Fig. S2). A temperature at which $\Delta G^{\ddagger}$ is similar between WT and mutants is a good reference point to understand the molecular origin for the shift in the $\Delta H^{\ddagger}$-$\Delta S^{\ddagger}$ in the enzyme with mutation.[13]

In Fig. 2 and Fig. 3, negative and positive differences in RMSF of residues ($\Delta RMSF_{C_\alpha(Mutant-WT)}$) indicate decrease and increase in fluctuations of residues in the mutants rel-
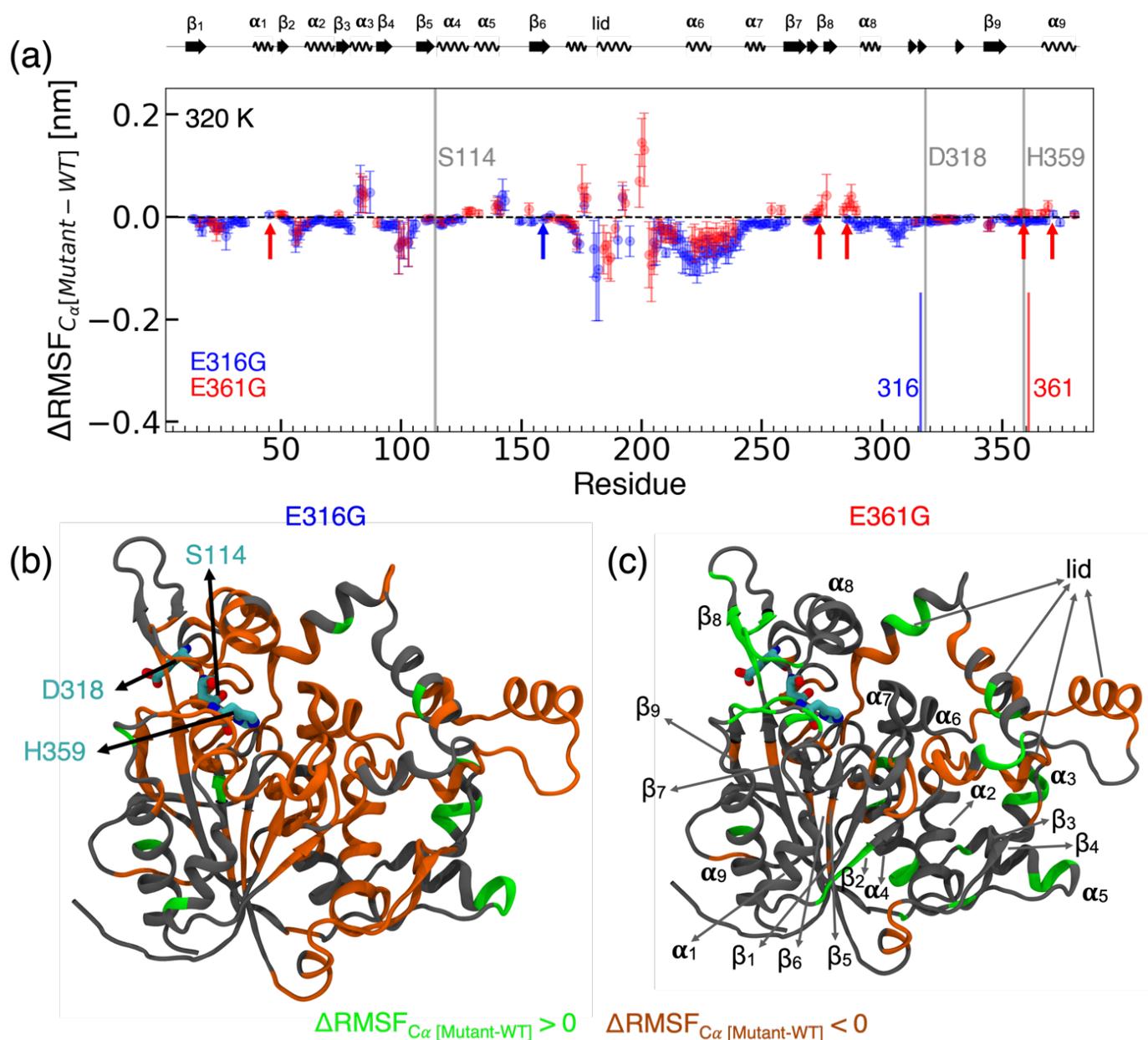
Fig. 3 Changes in the flexibility of GTL with E316G and E361G mutations measured using RMSF at 320 K. (a) Difference in RMSF of C_α ($\Delta\text{RMSF}_{C_\alpha(Mutant-WT)}$) of residues at 320 K in E316G and WT, and E361G and WT are shown in blue and red, respectively. Error bars represent standard errors obtained by measuring RMSF for WT and mutants using three independent MD simulation trajectories. Secondary structure of WT along the sequence was obtained using Biotite [64] and is shown at the top of (a). Arrows in blue and red refer to regions with higher RMSF in mutants than WT and that are within 1.5 nm from the active site residues in E316G and E361G, respectively. The locations of catalytic residues are shown using vertical lines in gray. For clarity, only residues with statistically significant changes in RMSF in WT and mutants are shown. (b) Residues in the crystal structure of WT GTL are colored according to their $\Delta\text{RMSF}_{C_\alpha(Mutant-WT)}$ of residues in E316G relative to WT, (c) Residues in the crystal structure of WT GTL are colored according to their $\Delta\text{RMSF}_{C_\alpha(Mutant-WT)}$ of residues in E361G relative to WT. In (b) and (c), residues with $\Delta\text{RMSF}_{C_\alpha(Mutant-WT)} > 0$ and $\Delta\text{RMSF}_{C_\alpha(Mutant-WT)} < 0$, i.e., more and less flexible in the mutant relative to WT are colored in green and orange, respectively.

ative to WT, respectively. We consider the residues in WT and mutant to have statistically different RMSF when |average of $\Delta\text{RMSF}_{C_\alpha(Mutant-WT)} \pm$ S.E.| > 0, where average and S.E are estimated using $\Delta\text{RMSF}_{C_\alpha(Mutant-WT)}$ from three independent simulations. Changes in RMSF profiles between RMSF of C_α atoms of residues in WT and mutants ($\Delta\text{RMSF}_{C_\alpha(Mutant-WT)}$) for all the studied temperatures are shown in Fig. S3.

3.1.1.1 RMSF of E316G at 280 K Statistically significant increase in RMSF of residues in mutant relative to WT following E316G mutation at 280 K (Fig. 2 a and b) are observed for α_3-helix residues (81, 84, 86-88), β_4-sheet residue 97, residue 107 between β_4 and β_5, β_6-sheet residues (154-155, 157-158), lid domain residues (199-200, 236), α_7-helix (253-256), residues between α_7 and β_7 (260, 262-264), residues between β_8 and
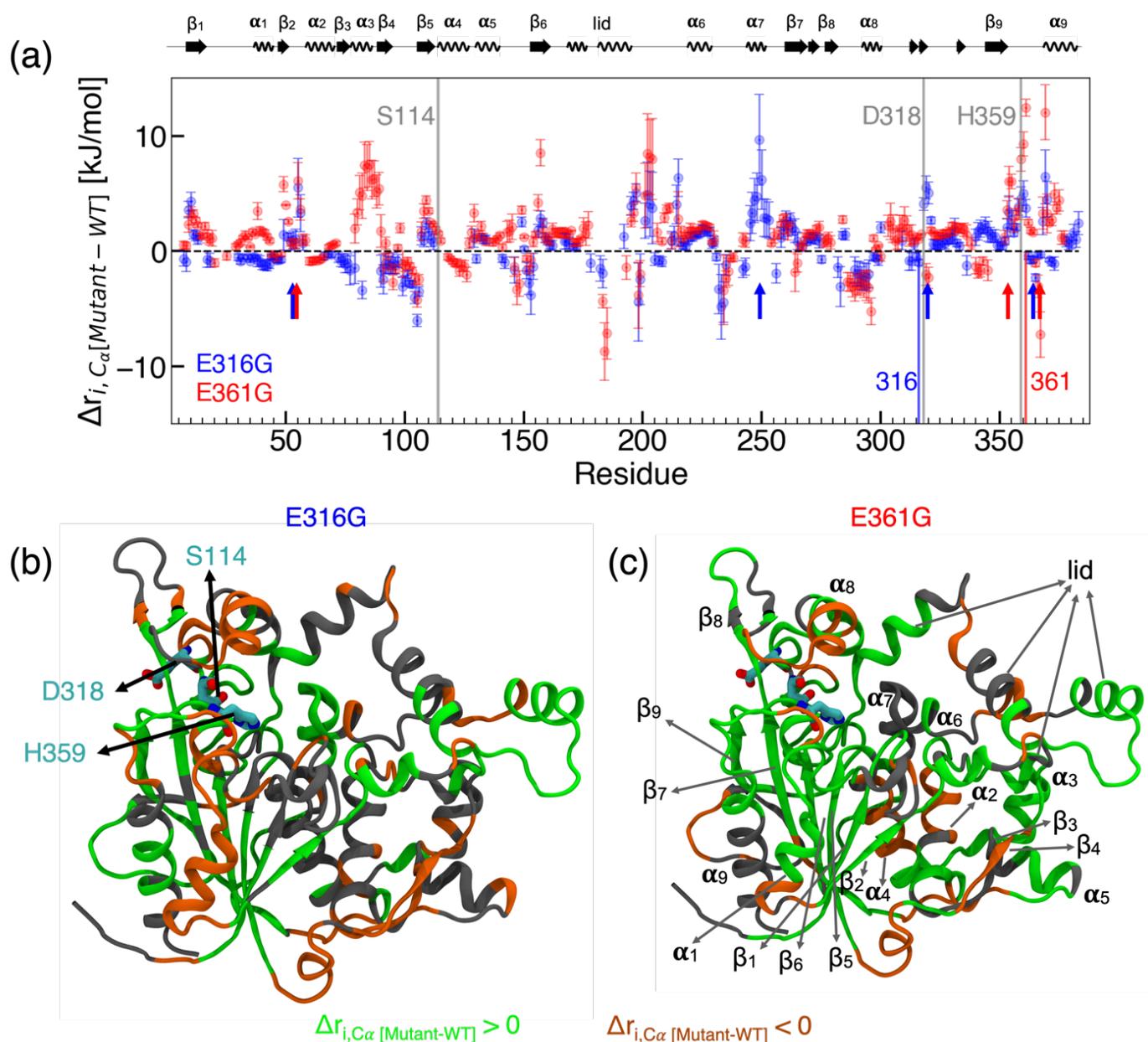
Fig. 4 Changes in the flexibility of GTL with E316G and E361G mutations measured using CNA. (a) Difference in r_i ($\Delta r_{i,(Mutant-WT)}$) of residues in E316G and WT, and E361G and WT are shown in blue and red, respectively. Error bars represent standard errors obtained by measuring r_i for WT and mutants using three independent simulation trajectories. Secondary structure of WT along the sequence was obtained using Biotite [64] and is shown at the top of (a). Arrows in blue and red refer to regions with higher r_i in mutants than WT and that are within 1.5 nm from the active site residues in E316G and E361G, respectively. The locations of catalytic residues are shown using vertical lines in gray. For clarity, only residues with statistically significant changes in r_i in WT and mutants are shown. (b) Residues in the crystal structure of WT GTL are colored according to their $\Delta r_{i,(Mutant-WT)}$ of residues in E316G relative to WT, (c) Residues in the crystal structure of WT GTL are colored according to their $\Delta r_{i,(Mutant-WT)}$ of residues in E361G relative to WT. In (b) and (c), residues with $\Delta r_{i,(Mutant-WT)} > 0$ and $\Delta r_{i,(Mutant-WT)} < 0$, i.e., more and less flexible in the mutant relative to WT are colored in green and orange, respectively.

α_8 (286-292), between α_8 and β_9 (207, 327-328, 346-348) and residue 369 between β_9 and α_9. Of these residues, 165, 288, 292, and 327-328 are within 1.5 nm of at least one of the catalytic residues (S114, D318, and H359) as shown in Fig. 2 a and b.

Statistically significant decrease in RMSF of residues in mutant relative to WT following E316G mutation at 280 K (Fig. 2 a and b) are observed for residues between β_1 and α_1 (14-21, 26, 28-30,

32-33), around β_2/α_2 (53-63, 65), β_3 residues (74, 79), residues around β_5/α_5 (111-120, 123, 125), residues between α_5 and β_6 (143-146), residue 163 of β_6, lid domain residues (168, 172, 177-187, 195, 208-209), α_7-helix (244, 246, 249), residue 267 of β_7, residues between α_8 and β_9 (313-316), and residues between β_9 and α_9 (361-362, 364-365). Residues near the active site with decreased RMSF for E316G are found in regions between β_1 and α_1,

between β_2 and α_2, around β_5/α_5, β_6, lid domain, α_7, between α_8 and α_9, and between α_8 and β_9 as highlighted in Fig. 2b.

3.1.1.2   RMSF of E361G at 280 K   Statistically significant increase in RMSF of residues in mutant relative to WT following E361G mutation at 280 K (Fig. 2 a and c) are observed in regions near β_1 (7-9), between β_1 and α_1 (19-21, 23-30), between α_1 and β_2 (46-47), residue 55 between β_2 and α_2, α_2 residues (70-71), between β_4 and β_5 (98, 101-108), α_4 residues (124-125), β_5 residues (153-158), lid domain (188, 195-200, 222-228, 230), between α_7 and β_7 (255, 257-260, 262-264), between β_8 and α_8 (286-291), residue 347 between α_8 and β_9, and between β_9 and α_9 (360, 366) (Fig. 2c). Of these residues, 19-20, 30, 55, 288-289, 291, 360, and 366 are within 1.5 nm of at least one of the catalytic residues (S114, D318, and H359) as highlighted in Fig. 2 a and c.

Residues with significant RMSF decreases following E361G mutation are located near β_1 (11), between β_2 and α_2 (57, 61-63), part of β_3, α_3, and β_4 (74-90, 92), around α_4 and α_5 (114-119, 132-148), residue 151 between α_5 and β_6, β_6 (160, 164), lid domain (167-168, 175-182, 237), α_7 (240, 246-248), residue 267 of β_7, between β_8 and α_8 (279-281), α_8 (298-299), and between α_8 and β_9 (302-304, 308-309, 311-312, 314-316, 322, 324, 341, 349) (Fig. 2c). Residues near the active site with decreased RMSF for E361G are found in regions near β_1, between β_2 and α_2, around β_5 and α_5, β_6, lid domain, between α_7 and β_7, and between α_8 and β_9.

Compared to the E316G mutation, E361G results in residues with decreased RMSF in the α_3-, α_5- and α_6-helices, along with an increase in RMSF near the β_5-sheet. In comparison to E316G, there are more residues in E361G (109 total in E316G and 149 total in E361G) with statistically significant changes in RMSF relative to WT at 280 K.

3.1.1.3   RMSF of E316G and E361G at 320 K   Statistically significant changes in RMSF of C_α of residues in mutant and WT at 320 K are shown in Fig. 3. Residue 162 is the only residue within 1.5 nm of the active site with a statistically significant increase in RMSF for the E316G mutation at 320 K (Fig. 3b). For the E361G mutation, residues with statistically significant increase in RMSF include residue 52 of the β_2-sheet, residues 272, 279, 288, and 289 between β_8 and α_8, and residues 357-359, 361 and 366 near β_9.

Both mutations result in residues with statistically significant decrease in RMSF located between β_2 and α_2, around β_5 and α_5, within β_6, in the lid domain, between α_7 and β_7, and between α_8 and β_9. E316G additionally has residues with significant RMSF reductions compared to WT in the α_7-helix (240, 246-247), while E361G includes residues near β_1 (16, 20) and β_9 (354).

Overall, the E316G mutant has more residues with significant RMSF changes (237 for E316G compared to 132 for E361G) at 320 K. The differences are largely from the higher number of residues in the lid domain and between α_8 and β_9 with RMSF changes for E316G compared to E361G mutant.

3.1.1.4   Change in RMSF with temperature   Analyzing RMSF of residues with change in temperature reveals the degree of flexibility of each residue and therefore provides another perspective

on RMSF changes with mutation. In other words, the degree of flexibility can be interpreted as the energy required to change the flexibility of a residue. Fig. S4 shows the changes in RMSF of WT and the two mutants with change in temperature starting from 300 K to 360 K using RMSF of the respective trajectories at 280 K as reference.

In WT and mutants, we observed both an increase and decrease in RMSF of residues that are distributed throughout the enzyme at any temperature T $>=$ 300 K relative to 280 K. We observed that the number of residues within $\sim$1.5 nm of active site that have higher RMSF at temperatures T $>=$ 300 K relative to 280 K are greater (except at 320 K) in the mutants than WT (Fig. S5-Fig. S8). This implies the degree of flexibility of residues closer to active site in mutants is higher than WT.

Together with the changes in RMSF of residues around the active site observed in Fig. 2 and Fig. 3, this indicates the mutations resulted in higher active site flexibility relative to WT. However, the simultaneous reductions in RMSF of residues away from the active site makes the establishment of a concrete relation with the complex changes in enzyme activity-temperature with mutation observed in experiments [1] difficult.

3.1.2   Flexibility using rigidity index r_i

As another measure, we apply CNA metrics to capture the local flexibility of the enzyme. CNA relies on the 3D structure utilized to construct the body-and-bar model. For this reason, we used an ensemble of structures in the folded state for both WT and mutants by extracting frames every 0.5 ns starting at 50 ns, which corresponds to the equilibrated region of the simulation trajectory (Fig. S1) at 300 K.

CNA-based flexibility measures are more sensitive to local connectivity than RMSF, which does not take local neighborhood information directly. [57] We set the E_{hb} range from -25.104 kJ/mol to -0.418 kJ/mol with 0.418 kJ/mol increment following prior work on related lipases and other proteins [61–63] for obtaining the phase transition temperature by mimicking folding-unfolding transition with CNA. The applied E_{hb} range is equivalent to a temperature range of approximately 302–420 K with 2 K increment [60], where the maximum temperature is well beyond the optimal activity temperatures of WT GTL ($\sim$ 320 K) and mutants ($\sim$ 333 K) observed in experiments [1].

3.1.2.1   Trends in folding-unfolding transition   According to CNA, the global indices capturing the folding-unfolding of the enzyme (SI Tab. S1), the transition of average E_{hb} follows the order E361G > WT > E316G. The lower transition E_{hb} or equivalently a higher folding-unfolding transition temperature implies a higher stability of the E316G mutant, but not E361G mutant, compared to WT. The higher stability of E316G mutant, but not E361G mutant, from CNA correlates with the higher kinetic stability measured with C4 substrate in wet-lab experiments [1]. We note that a caveat in this interpretation is that higher kinetic stability can also result from lower aggregation propensity [66] that is not captured in CNA, and therefore need not have any direct relation with the stability of the structure.
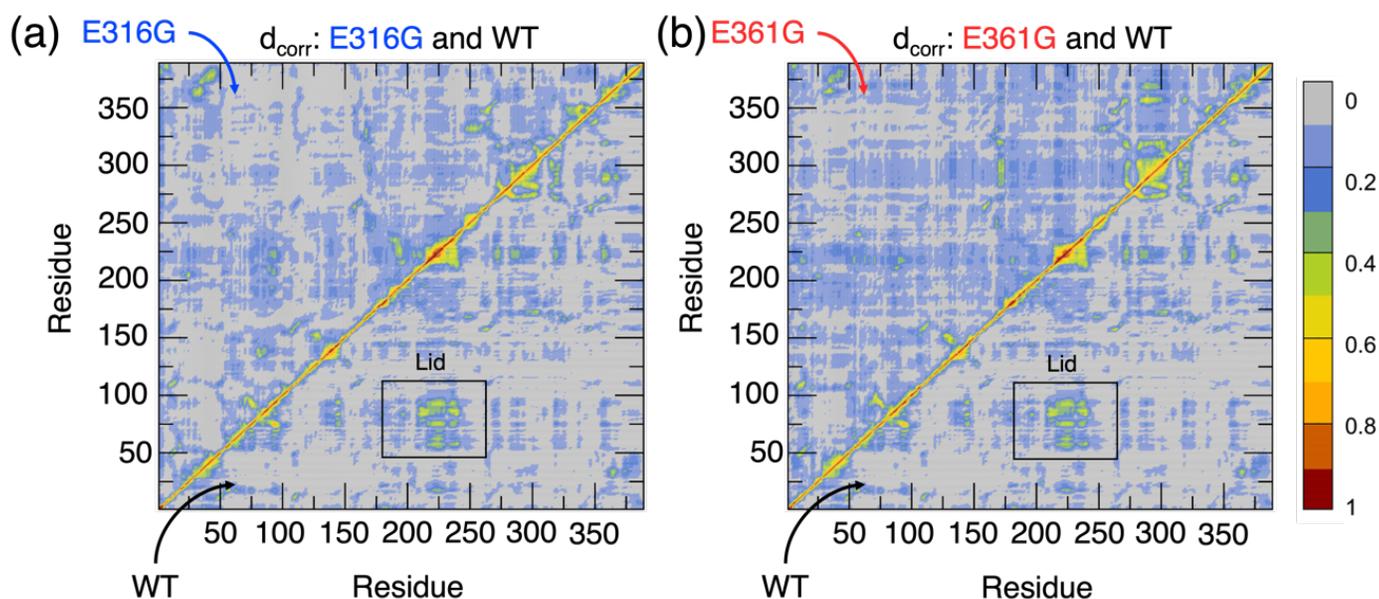
Fig. 5 Distance correlation, d_{corr}, between RMSF of residue pairs in WT, E316G, or E361G. (a) d_{corr} of residues in E316G (upper diagonal) and WT (lower diagonal) at 280 K, (b) d_{corr} of residues E361G (upper diagonal) and WT (lower diagonal) at 280 K. Higher d_{corr} values between residue pairs implies stronger dependency between residue-residue fluctuations. The residues in the lid domain are highlighted by a box in both (a) and (b).

3.1.2.2 Rigidity index r_i

Fig. 4 shows the differences in Δr_i of mutants and WT. A non-zero Δr_i on average indicates mutations altered flexibility of the enzyme. The positive and negative values of the average Δr_i imply the residues in mutants are relatively more and less flexible than WT, respectively. The standard error of Δr_i in Fig. 4a is indicative of the differences in the distribution of r_i in the WT and the mutants. Several residues with altered r_i are in similar locations along the enzyme sequence to that identified by RMSF analysis (Fig.2 and Fig.3).

3.1.2.3 Trends in r_i of E316G

Residues proximal to the active site (Fig. S9) with a significant difference in r_i with E316G mutation include 55, 249-250, 319-320, and 360 (Fig. 4b). Most of the residues of loop regions in E316G have negative Δr_i on an average indicating enhanced rigidity or a higher E_{hb} is required to segregate from a rigid cluster. These changes in loop rigidity of the mutants correlate with their higher kinetic stability relative to WT observed in experiments [1].

3.1.2.4 Trends in r_i of E361G

In the E361G mutant, residue 55 in β_2 and residues in the loop between β_9-α_9 (354-355, 359-361) are relatively more flexible, i.e., break apart from a rigid cluster at less negative E_{hb} on an average than WT (Fig. 4c). These residues are <1.5 nm from the active site and are highlighted in Fig. 4a. This suggests the ability of E361G to break hydrogen-bonds or salt-bridges is relatively easier on average at a lower temperature than WT. These identified flexible regions around the active site (Fig. S9) could be associated with the large changes in $\Delta H^{\ddagger}$-$\Delta S^{\ddagger}$ and specific activity at lower T (T < ~325 K) of E361G relative to WT [1].

However, similar to the RMSF changes at a given temperature (Fig. 2, 3) and with temperature, the CNA-based metric r_i also revealed certain residues around the active site are

more flexible than in WT. However, there are also residues with lower and higher r_i that are distributed throughout the enzyme, which implies a non-trivial relation between flexibility changes and activity-temperature profiles or $\Delta H^{\ddagger}$-$\Delta S^{\ddagger}$ changes observed in experiments [1] with mutations.

3.2 Understanding differences in enzyme flexibility and structure with mutation

We have established that both mutations resulted in certain flexible residues around the active site, but there are also residues with both lower and higher flexibilities that are distributed throughout the enzyme. These concomitant changes in flexibility of residues with mutation that are distributed throughout the enzyme make it difficult to study the relation between active site flexibility and activity-temperature relation observed in experiments. In addition, both the mutations also resulted in modified organization of the catalytic triad, which are the key residues for catalyzing the hydrolysis reaction (Fig. S10). Another complication is that in experiments, the two point mutants have resulted in a substrate dependent activity-temperature relation and altered (higher) kinetic stability than WT. [1] With the shorter substrate p-nitrophenol butyrate, we observed improved activity in the mutants relative to WT mainly at high temperatures. [1] In contrast, the mutants have lower activity than WT at both lower and higher temperatures than WT with the longer substrates p-nitrophenol octanoate and p-nitrophenol laurate. [1]

To better understand the changes in the enzyme structure and the leading modes of variation with mutation, we employed principal component analysis (PCA). SI Fig. S14 shows the projection of each configuration of thermophilic WT (black) and its improved mutants (E316G–blue, E361G–red) with psychrophilic
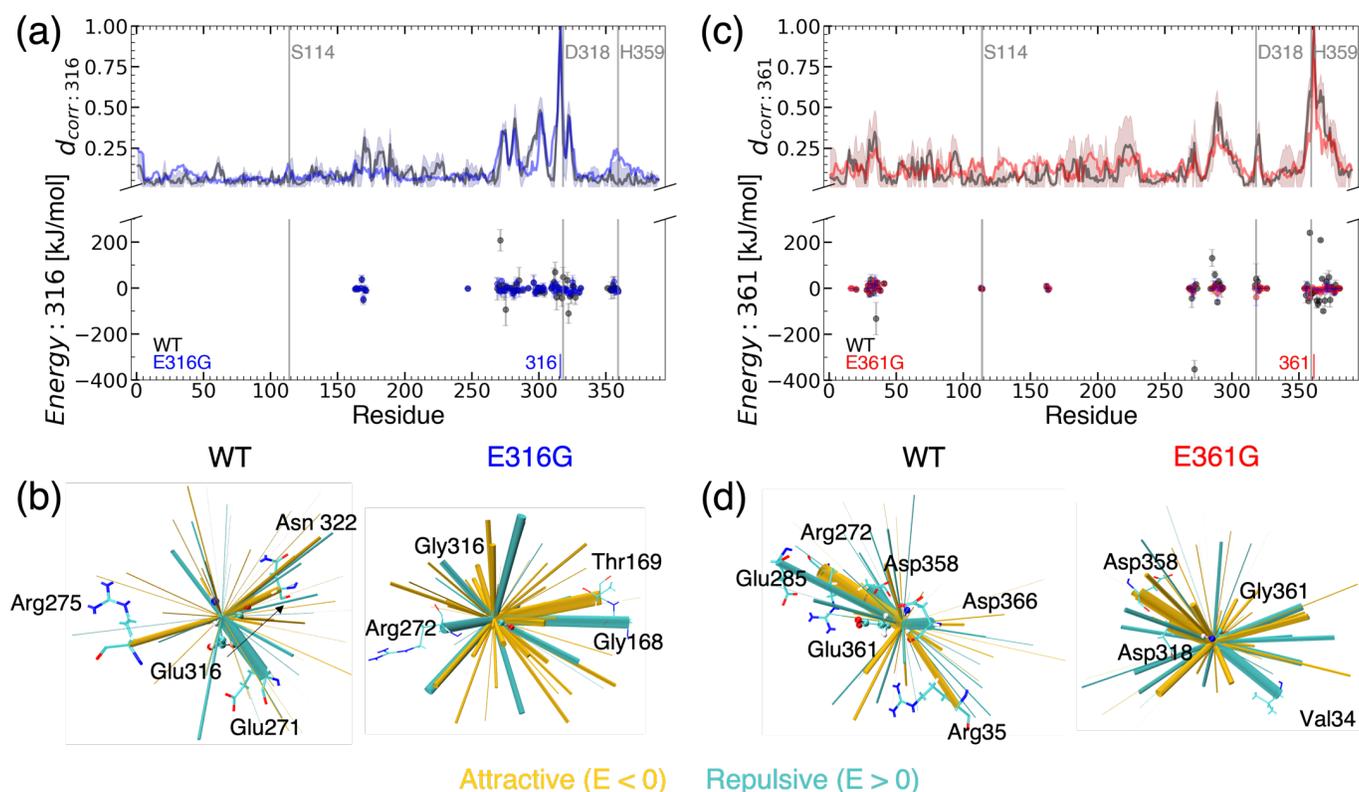
Fig. 6 Connection between distance correlation d_{corr} between the RMSF of residue pairs and the residue-residue interaction energies. (a) average d_{corr} of residue 316 in WT (black) and E316G (blue) and the average total non-bonded interaction energy (E, Energy) with all the residues in WT and E316G at 280 K, (b) the average total E between residue 316 and other residues shown in (a) mapped onto the energy minimized structure of WT and E316G. The cylinders thickness is proportional to E. (c) average d_{corr} of residue 361 in WT (black) and E361G (red) and the average total non-bonded interaction energy (E, Energy) with all the residues in WT and E361G at 280 K, (d) the average total E between residue 361 and other residues shown in (c) mapped onto the energy minimized structure of WT and E361G. The cylinder thickness is proportional to E. Residue-residue interaction energies in WT and mutants that are attractive (E < 0) or repulsive (E > 0) are colored in orange or cyan, respectively. In (a) and (c), the vertical gray lines correspond to the three active site residues (Ser114, Asp318, and His359). Error bars represent standard errors and are obtained from three independent simulation trajectories.

traits sampled in MD simulations at 280 K along the top five leading shared principal components. We obtain the shared principal components for the three variants by performing PCA over the combined trajectories of WT and the mutants. Each configuration in the trajectory was featurized by the backbone C_α positions for performing the PCA. 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 (SI Fig. S16).

Upon further analysis, we observed variations in the configurations of mutants compared to WT along PC1–PC2 could be attributed to the differences in the collective motions in the lid domain relative to the core $\alpha-\beta$ hydrolase fold of GTL (SI Fig. S15). This is consistent with the RMSF and r_i analysis (Fig. 2, 3, 4) which showed that the mutations largely influence the lid domain fluctuations indicating a potential link to the specific activity differences between WT and the mutants. Further details on PCA are provided in SI.

To minimize such drastic changes in the enzyme structure and alter the flexibility of the enzyme in a predictable manner using

rational point mutations, it is crucial to understand the origin of the differences in enzyme structure and flexibility with mutation. For this reason, we first focus on understanding the dependency between the RMSF of a residue with that of another residue by measuring pairwise distance correlation coefficient (d_{corr})[67] between the RMSF of two residues using the trajectories at 280 K, where both WT and mutants are active.[1] We expect the influence of thermal energy to be relatively low at 280 K. Therefore, analysis at this temperature could reveal differences in the inherent flexibility correlations between residues in the enzyme. We then focus on understanding the origin of the differences in flexibility of residues that are located far from the mutation site in the two mutants.

3.2.1 Distance correlation, d_{corr}, between residue-residue RMSF

Distance correlation, d_{corr},[67] captures non-linear correlations between two variables. A higher d_{corr} implies a stronger dependency between the two variables. Fig. 5a shows the d_{corr} between RMSF of two residues of E316G and WT at 280 K in the upper and lower diagonal of the heat map, respectively. d_{corr} trends at other
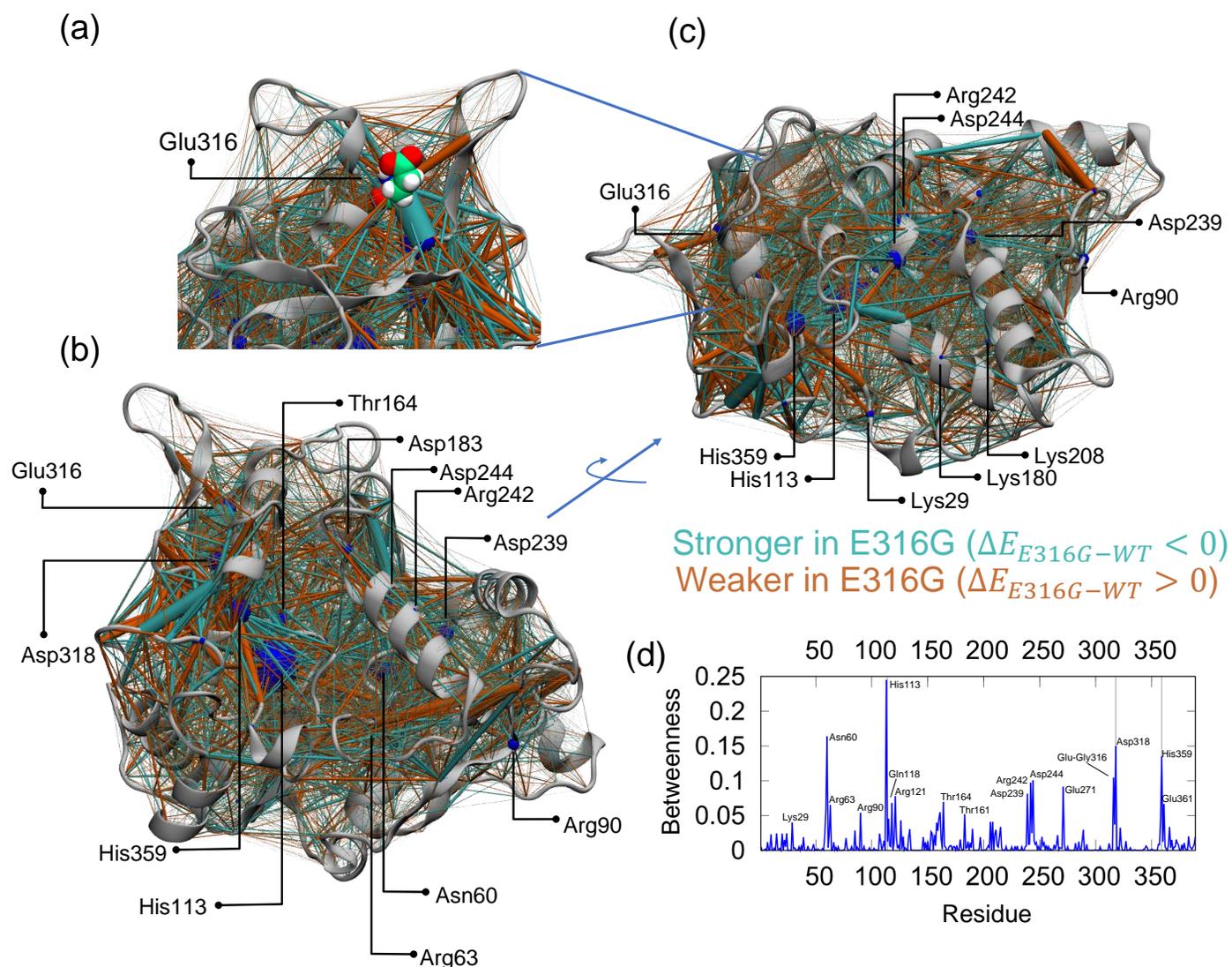
Fig. 7 Cylinders with radius proportional to the difference in average total E between residues of mutant and WT ($\Delta E_{E316G-WT}$) are used for showing the differences in the strength of interactions with mutation E316G at 280 K. Cylinder radii are normalized by the minimum value of $\Delta E_{E361G-WT}$. Thus, edges connecting C_α nodes with strong attractive energy in the mutant relative to WT are represented by the cylinders with the highest radii. (a) Closer view of residues surrounding the site of mutation. In (b) and (c), the changes in the residue-residue interactions are shown in two different perspectives focusing on the binding site. Cyan and orange cylinders represent interactions that are stronger ($\Delta E_{E316G-WT} < 0$) and weaker ($\Delta E_{E316G-WT} > 0$) on average in E316G than WT, respectively. (d) Betweenness value of the residues in the residue-residue network, where the value is dependent on the number of shortest paths passing through the residue. The shortest path is defined as the path with minimum sum of weights, and therefore we have used $\frac{1}{|E_{E316G-WT}|}$ for edge weights. This provides more weight to edges with higher $|E_{E316G-WT}|$ in determining shortest paths. In (a-c), the betweenness values of each residue are shown with blue spheres using the radius of the sphere. The betweenness values are computed from average $|E_{E316G-WT}|$ at 280 K, which is estimated from three independent trajectories.

temperatures are shown in Fig. S11 and Fig. S12.

The d_{corr} results indicate strong dependency between RMSF of residues ~50-90 and part of the lid domain (i.e. residues ~210-240) in the WT but not in E316G. In addition, the E316G mutation resulted in altered dependency between various other residue pairs with mostly lower (gray regions in Fig. 5a) d_{corr} when compared to WT at 280 K.

Similar changes in d_{corr} can be observed in E361G mutant relative to WT at 280 K (Fig. 5b). However, the changes are greater in E361G compared to E316G at 280 K. The changes in dependency resulting from mutation indicate modified connectivity be-

tween residues in the mutants compared to WT at 280 K. These changes in principle are not desired according to the hypothesis and are likely associated with the unexpected changes to enzyme properties such as kinetic stability and substrate-dependent activity-temperature relation observed in the experiments.[1]

3.2.2 Changes in total nonbonded interaction energy, E with mutation

To further investigate the origin of the changes in d_{corr}, we calculated the total nonbonded interaction energies (E) between every residue-residue pair in WT and mutants at 280 K. The to-
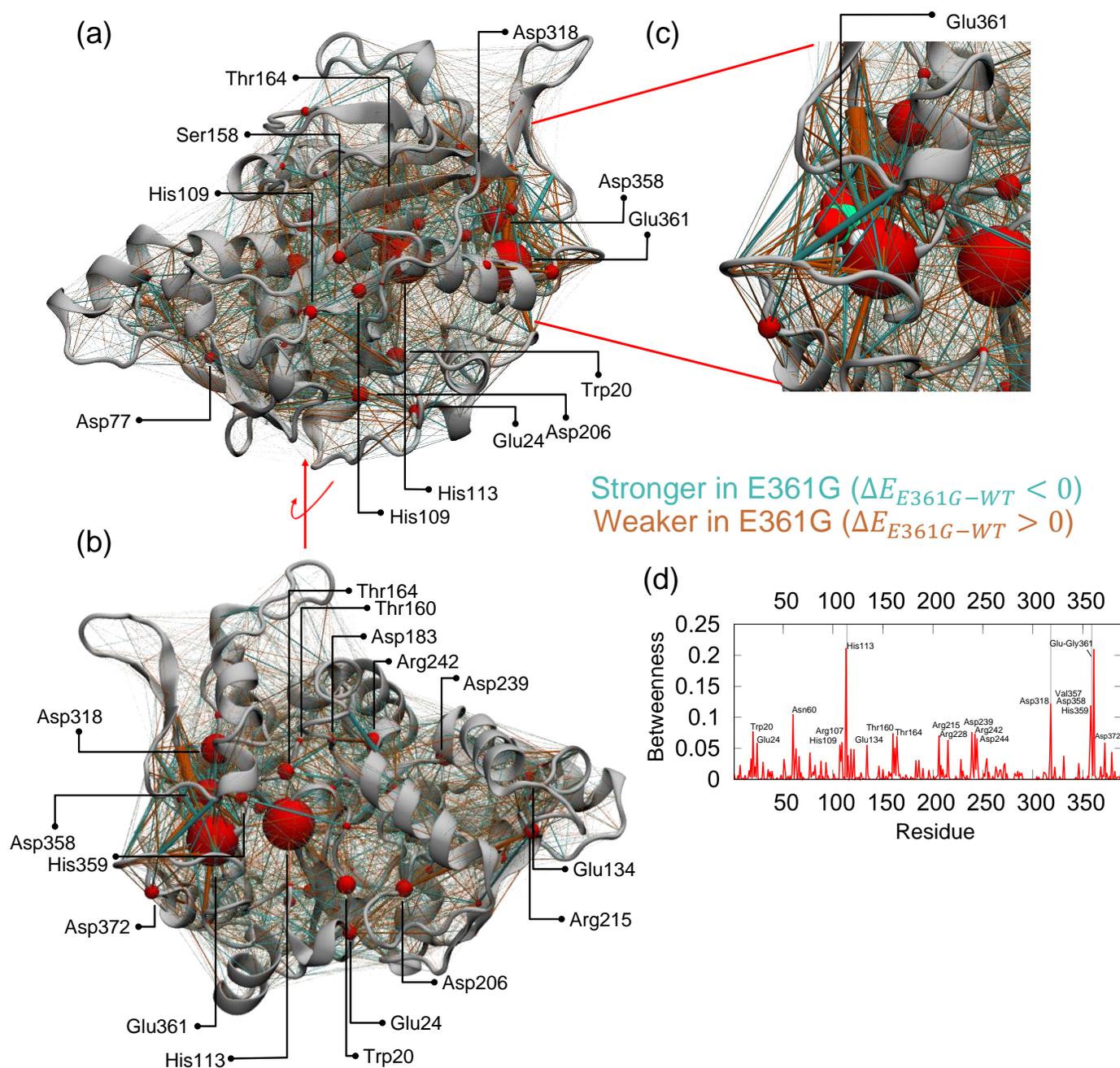
Fig. 8 Cylinders with radius proportional to the difference in average total E between residues of mutant and WT ($\Delta E_{E361G-WT}$) are used for showing the differences in the strength of interactions with mutation E361G at 280 K. Cylinder radii are normalized by the minimum value of $\Delta E_{E361G-WT}$. Thus, edges connecting C_α nodes with strong attractive energy in the mutant relative to WT are represented by the cylinders with the highest radii. (a) One perspective of residues surrounding the site of mutation. In (b) and (c), the changes in the residue-residue interactions are shown in two other perspectives focusing on the mutation site. Cyan and orange cylinders represent interactions that are stronger ($\Delta E_{E361G-WT} < 0$) and weaker ($\Delta E_{E361G-WT} > 0$) on an average in E361G than WT, respectively. (c) Closer view surrounding the site of mutation. (d) Betweenness value of the residues in the residue-residue network, where the value is dependent on the number of shortest path passing through the residue. The shortest path is defined as the path with minimum sum of weights, and therefore we have used $\frac{1}{|E_{\textbf{E361G-WT}}|}$ for edge weights. This provides more weight to edges with higher $|E_{\textbf{E316G-WT}}|$ in determining shortest paths. In (a-c), the betweenness values of each residue are shown with red spheres using the radius of the sphere. The betweenness values are computed from average $|E_{\textbf{E316G-WT}}|$ at 280 K, which is estimated from three independent trajectories.

tal nonbonded interaction energy E was calculated as the sum of Lennard-Jones and Coulomb interactions. Normalized differences in interaction energy E between WT and mutants scaled to -1 and 1 are shown in SI Fig. S13. The results indicate that the mutations caused both large and subtle differences in the interaction energies between various residue pairs that are spread throughout the enzyme.

While most of the residue pairs with strong interactions

(Fig. S13) have high d_{corr} (Fig. 5), there are exceptions where the residues that are not directly interacting have high d_{corr}. These residue pairs show dependency in d_{corr} via indirect interactions mediated by either other residues and/or water. To illustrate this, we explored the connection between d_{corr} and total E of all the residues in the enzyme with the mutant residue 316 (Fig. 6a-b) and 361 (Fig. 6c-d) in WT and the mutants.

3.2.2.1 Connection between d_{corr} and interaction energy E of the mutant residues Few residues, especially in the lid domain, have higher d_{corr} or dependency with residue 316 in WT than in the mutant E316G (Fig. 6a). Some of these residues are directly interacting with residue 316, as observed from their non-zero interaction energies (Fig. 6a), indicating spatial proximity. Other residues are indirectly interacting, as there is no interaction energy observed yet dependencies in d_{corr} are observed. The residues proximal in sequence space to 316 have higher d_{corr} in both WT and mutant E316G with minor differences between them, which is also reflected in the change in energy (Fig. 6a). To emphasize the differences in the interaction energy E in WT and E316G on an average, we projected the strengths proportional to the cylinder radius connecting residues on the structure (Fig. 6b). The projected strengths demonstrate the redistribution of the interaction energies after mutation. For example, E316 has strong repulsive interactions (cyan) with E271 in WT but not in the E316G mutant. The E316G mutation largely influenced the interaction energies with the polar and charged residues in the immediate vicinity, including the loss of strong attractive interactions (orange) with R275 following mutation.

In WT, there are several residues along the sequence that are correlated (although weakly) with the residue 361 (Fig. 6c) than the residue 316 (Fig. 6a). The mutation E361G resulted in stronger as well as weaker correlations with the mutated residue 361 relative to the corresponding residue in the WT. A major loss in the attractive interaction energy between residue E361 and R272 is observed in WT after E361G mutation (Fig. 6c). These changes in E surrounding the mutant residue is the likely origin in the redistribution of E across every residue pair, and thus the dependency in the fluctuations d_{corr} (Fig. 5) of the residue. These are also seen in the visual plots with energy E projected by cylinder radius connecting mutant residue 361 with other residues in WT and E361G (Fig. 6d).

3.3 Understanding changes in enzyme flexibility with mutation

To further track the propagation of the changes in enzyme flexibility from the point of origin (mutation), we applied network analysis to the ensemble of structures explored in MD simulations. Network analysis is a powerful approach that can enable access to various aspects of a protein such as the global topological connectivity between residues[68–71], local variation in the connectivity of each residue with its neighboring residues[72], allosteric interactions[73], and identification of critical residues[23,74]. In protein structure network analysis, residue interaction networks are constructed by representing each residue as a node. The edges between the residues or nodes are typically constructed based on

the spatial proximity of the residues using a user-defined distance cutoff.[68,72–74] In order to capture the dynamical information associated with an ensemble of structures accessible by MD simulations, correlation between the residue fluctuations can be utilized to define the weight of each edge connecting different residue pairs.[73] In our work, we use a simple alternative approach, where we define edges between all pairs of residues in the enzyme and utilize the change in average non-bonded interaction energies E between residues in mutant and WT ($\Delta E_{Mutant-WT}$) at 280 K as weights to set their relative importance. In this way, connections between all residue pairs are retained, and because the weights are based on average $\Delta E_{Mutant-WT}$, it enables direct access to the changes in topological connectivity of the enzyme with mutation observed in the ensemble of structures sampled in MD simulations. For estimating the edge weights, average $\Delta E_{Mutant-WT}$ at 280 K, we used the average non-bonded interaction energies obtained from three independent simulations for WT and mutants.

Figs. 7(a-c) and 8(a-c) show the resulting weighted network mapped on the energy-minimized WT crystal structure using cylinders proportional to the weight in the mutant E316G and E361G, respectively. These weighted networks reveal the changes in the total interaction energy E between every residue-residue pair with mutation. It shows the interaction strength between certain residue pairs increased (cyan) as well as decreased (orange) with mutation. These networks can be used to rationalize the simultaneous increase and decrease in RMSF of the enzyme with mutation at 280 K. For example, for the E316G mutant, there are several residue pairs with large changes in their interaction energy relative to WT far from the mutation site as reflected in the radius of the edges (cyan and orange cylinders) connecting nodes in Figs. 7(a-c).

To identify residues that are strongly influenced by mutation, we calculated betweenness centrality using NetworkX[75–77] on the network. For this, we utilize the inverse of the absolute difference in interaction energy E between mutant and WT ($\frac{1}{|\Delta E_{Mutant-WT}|}$) as edge weights for computing shortest paths and betweenness centrality in the resulting network. This allows the usage of standard shortest path algorithms[75–77], where the shortest paths are those that have minimum sum total edge weights and therefore, applying an inverse to ($|\Delta E_{Mutant-WT}|$) provides higher weightage to edges with larger differences in the interaction energy E. Betweenness centrality for each node k is defined as $\sum_{i,j\in N} \frac{\sigma(i,j|k)}{\sigma(i,j)}$, where N is the set of nodes. $\sigma(i,j|k)$ represents the number of shortest paths between s and t passing through a node k and $\sigma(s,t)$ refers to the total number of shortest paths between two nodes s and t.[75,76]

Betweenness centrality provides a value proportional to the number of shortest paths passing through it. Residues with higher betweenness can be considered as hubs for the change in the total E with mutation as the number of shortest paths passing through them are higher. According to the betweenness of residues, a majority of the hubs are charged residues in both E316G and E361G (Figs. 7d and 8d). These residues are connected mainly to other charged residues in the enzyme, which suggests a large role of electrostatic interactions for the observed changes in the structure-dynamics with mutation.

The weighted networks also demonstrate differences in the mechanism of changes in the enzyme's internal network with each mutation. In E316G (Fig. 7), the residue H113 is the major hub that resulted in large changes near the lid domain and other regions beyond 1.0 nm from the mutation site E316. The charged residues such as D318, H359 and the polar residue N60 acted as intermediate hubs. For E361G (Fig. 8), the changes with the mutation are propagated via several charged residues such as D318, D358, H359, and certain polar/apolar (N60 and V357) residues adjoining E361 and H113.

4   Conclusions

Given low temperature activity in enzymes from psychrophiles can be uniquely related to their active site flexibility[1,8,12,19], it implies broad range temperature active enzymes can be designed by incorporating active site flexibility in a thermophilic enzyme. This is in contrast to the alternative strategy of thermostabilization of psychrophilic and mesophilic enzymes by engineering rigidity of certain key residues to enable their high temperature function.[8,22] In our prior study[1], we exploited the potential relation between active site flexibility and low temperature activity for designing broad-range temperature active enzymes. With a simple human design approach, we attempted to identify point mutants using crystal structures that can improve active site flexibility in a model thermophilic enzyme GTL. Initial wet-lab tests revealed two point mutants with better low temperature activity than WT GTL.[1] However, we observed complex trends such as an increase in high temperature activity, substrate-dependent activity-temperature relation, and kinetic stability in the two mutants upon further analysis.[1]

In this study, we applied MD simulations and network theory to understand the origin of the complex trends observed in the experiments. We showed that the mutations resulted in both an increase and decrease in the flexibility of residues in the enzyme. We used two measures to quantify flexibility – a conventional RMSF approach and an alternative network theory-based technique. We found changes in the flexibility of residues both proximal and distal to the active site. This shows the limitations of the applied human design approach to increase active site flexibility using only the crystal structure of GTL. To discover the additional design rules and understand the origin of the differences in enzyme flexibility with mutation, we analyzed the correlation between the flexibility of enzyme residues and the changes in enzyme structure with mutation.

Our analysis revealed that the changes in the residue fluctuations with mutation are a result of the changes in the residue-residue interactions. These changes comprise both increase and decrease in the strength of residue-residue interactions, and are distributed throughout the enzyme. This explains the changes in the flexibility of residues distal from the site of mutation. Our network analysis revealed that several charged residues, distributed throughout the enzyme, act as mediators or hubs in propagating the changes in residue-residue interactions from the site of the mutation. Collectively, our results imply information associated with correlations between residue-residue fluctuations could be crucial in designing rational mutations for altering flexibility

in a targeted region of an enzyme. This is mainly necessary for the successful implementation of hypothesis driven engineering techniques requiring control over active site flexibility using site directed mutagenesis. To this end, our plan for future work is to utilize the discussed insights and related dynamic flexibility metrics[55,78] to construct a computational screening approach for predicting active site flexibility inducing mutations and thereby, successfully exploit active site flexibility-activity relation for rationally engineering low temperature activity into thermophilic enzymes.

Author Contributions

S.D.: Conceptualization, Methodology, Software, Data Curation, Investigation, Validation, Formal Analysis, Visualization, Writing - Original Draft, Writing - Review and Editing; J.Z.: Investigation, Data Curation, Validation, Formal Analysis, Visualization, Writing - Review and Editing; S.S.: Conceptualization, Methodology, Investigation, Validation, Formal Analysis, Visualization, Supervision, Project Administration, Funding Acquisition, Resources, Writing - Original Draft, Writing - Review and Editing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 W. Wang, S. Dasetty, S. Sarupria and M. Blenner, *Biochem. Eng. J.*, 2021, **174**, 108093.

2 P. A. Fields, *Comp. Biochem. Physiol., Part A: Mol. Integr. Physiol.*, 2001, **129**, 417–431.

3 D. Georlette, V. Blaise, T. Collins, S. D'Amico, E. Gratia, A. Hoyoux, J.-C. Marx, G. Sonan, G. Feller and C. Gerday, *FEMS Microbiol. Rev.*, 2004, **28**, 25–42.

4 E. Papaleo, M. Olufsen, L. De Gioia and B. O. Brandsdal, *J. Mol. Graphics Modell.*, 2007, **26**, 93–103.

5 G. V. Isaksen, J. Åqvist and B. O. Brandsdal, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, 201605237.

6 F. Sarmiento, R. Peralta and J. M. Blamey, *Front. Bioeng. Biotechnol.*, 2015, **3**, 148.

7 K. S. Siddiqui, *Crit. Rev. Biotechnol.*, 2017, **37**, 309–322.

8 S. Dasetty, M. A. Blenner and S. Sarupria, *Mater. Res. Express*, 2017, **4**, 114008.

9 J. Kim, I. Jeerapan, J. R. Sempionatto, A. Barfidokht, R. K. Mishra, A. S. Campbell, L. J. Hubble and J. Wang, *Acc. Chem. Res.*, 2018, **51**, 2820–2828.

10 C. Struvay and G. Feller, *Int. J. Mol. Sci.*, 2012, **13**, 11643–11665.

11 B. Van Den Burg, *Curr. Opin. Microbiol.*, 2003, **6**, 213–218.

12 T. Lonhienne, C. Gerday and G. Feller, *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, 2000, **1543**, 1–10.

13 J. Åqvist, G. V. Isaksen and B. O. Brandsdal, *Nat. Rev. Chem.*, 2017, **1**, 0051.

14 F. Sterpone and S. Melchionna, *Chem. Soc. Rev.*, 2012, **41**, 1665–1676.

15 S. Kumar and R. Nussinov, *ChemBioChem*, 2002, **3**, 604–617.

16 H. Eyring, *Chem. Rev.*, 1935, **17**, 65–77.

17 M. G. Evans and M. Polanyi, *Trans. Faraday Soc.*, 1935, **31**, 875–894.

18 A. Hamdan, *S. Afr. J. Sci.*, 2018, **114**, 1–6.

19 G. Feller and C. Gerday, *Nat. Rev. Microbiol.*, 2003, **1**, 200.

20 D. Perl and F. X. Schmid, *ChemBioChem*, 2002, **3**, 39–44.

21 P. A. Fields and G. N. Somero, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 11476–11481.

22 M. Rahban, S. Zolghadri, N. Salehi, F. Ahmad, T. Haertlé, N. Rezaei-Ghaleh, L. Sawyer and A. A. Saboury, *Int. J. Biol. Macromol.*, 2022, 642–654.

23 M. Vendruscolo, N. V. Dokholyan, E. Paci and M. Karplus, *Phys. Rev. E*, 2002, **65**, 061910.

24 R. D. Schmid and R. Verger, *Angew. Chem., Int. Ed.*, 1998, **37**, 1608–1633.

25 A. Svendsen, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.*, 2000, **1543**, 238.

26 K.-E. Jaeger and T. Eggert, *Curr. Opin. Biotechnol.*, 2002, **13**, 390–397.

27 A. Houde, A. Kademi and D. Leblanc, *Appl. Biochem. Biotechnol.*, 2004, **118**, 155.

28 R. Fernandez-Lafuente, *J. Mol. Catal. B: Enzym.*, 2010, **62**, 197–212.

29 C. Carrasco-López, C. Godoy, B. de las Rivas, G. Fernández-Lorente, J. M. Palomo, J. M. Guisán, R. Fernández-Lafuente, M. Martínez-Ripoll and J. A. Hermoso, *J. Biol. Chem.*, 2009, **284**, 4365–4372.

30 N. Guex and M. C. Peitsch, *Electrophoresis*, 1997, **18**, 2714–2723.

31 *gmx pdb2gmx*, https://manual.gromacs.org/documentation/current/onlinehelp/gmx-pdb2gmx.html (accessed April 2023).

32 M. Abraham, D. van der Spoel, E. Lindahl and B. Hess, *GROMACS User Manual version 5.0.2*, 2014, www.gromacs.org.

33 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith,

B. Hess and E. Lindahl, *SoftwareX*, 2015, **1**, 19–25.

34 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 1950–1958.

35 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

36 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

37 B. Hess, *J. Chem. Theory Comput.*, 2008, **4**, 116–122.

38 R. W. Hockney, S. Goel and J. Eastwood, *J. Comput. Phys.*, 1974, **14**, 148–158.

39 S. Páll and B. Hess, *Comput. Phys. Commun.*, 2013, **184**, 2641–2650.

40 S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952–962.

41 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

42 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

43 S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.

44 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695.

45 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

46 S. Nosé and M. L. Klein, *Mol. Phys.*, 1983, **50**, 1055–1076.

47 M. Lingenheil, R. Denschlag, R. Reichold and P. Tavan, *J. Chem. Theory Comput.*, 2008, **4**, 1293–1306.

48 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

49 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.

50 P. Gatti-Lafranconi, A. Natalello, S. Rehm, S. M. Doglia, J. Pleiss and M. Lotti, *J. Mol. Biol.*, 2010, **395**, 155–166.

51 M. Z. Kamal, T. A. S. Mohammad, G. Krishnamoorthy and N. M. Rao, *PLoS One*, 2012, **7**, e35188.

52 G. V. Isaksen, J. Åqvist and B. O. Brandsdal, *PLoS Comput. Biol.*, 2014, **10**, e1003813.

53 J. M. Saavedra, M. A. Azócar, V. Rodríguez, C. A. Ramírez-Sarmiento, B. A. Andrews, J. A. Asenjo and L. P. Parra, *Biotechnol. J.*, 2018, **13**, 1700669.

54 Y.-w. Dong, M.-l. Liao, X.-l. Meng and G. N. Somero, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 1274–1279.

55 P. Campitelli, T. Modi, S. Kumar and S. B. Ozkan, *Annu. Rev. Biophys.*, 2020, **49**, 267–288.

56 D. J. Jacobs, A. J. Rader, L. A. Kuhn and M. F. Thorpe, *Proteins: Struct., Funct., Bioinf.*, 2001, **44**, 150–165.

57 S. M. A. Hermans, C. Pfleger, C. Nutschel, C. A. Hanke and H. Gohlke, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2017, **7**, e1311.

58 A. Karshikoff, L. Nilsson and R. Ladenstein, *FEBS J.*, 2015, **282**, 3899–3917.

59 B. I. Dahiyat, D. Benjamin Gordon and S. L. Mayo, *Protein*

Sci., 1997, **6**, 1333–1337.

60 S. Radestock and H. Gohlke, *Proteins: Struct., Funct., Bioinf.*, 2011, **79**, 1089–1108.

61 S. Radestock and H. Gohlke, *Eng. Life Sci.*, 2008, **8**, 507–522.

62 P. C. Rathi, K.-E. Jaeger and H. Gohlke, *PloS One*, 2015, **10**, e0130289.

63 P. C. Rathi, A. Fulton, K.-E. Jaeger and H. Gohlke, *PLoS Comput. Biol.*, 2016, **12**, e1004754.

64 P. Kunzmann and K. Hamacher, *BMC Bioinf.*, 2018, **19**, 1–8.

65 Z. Sun, Q. Liu, G. Qu, Y. Feng and M. T. Reetz, *Chem. Rev.*, 2019, **119**, 1626–1665.

66 E. Timucin and O. U. Sezerman, *Biochemistry*, 2015, **54**, 3901–3910.

67 G. J. Székely, M. L. Rizzo and N. K. Bakirov, *Ann. Stat.*, 2007, **35**, 2769–2794.

68 A. Aszói and W. R. Taylor, *Bioinformatics*, 1993, **9**, 523–529.

69 S. Bhakat, A. J. Martin and M. E. Soliman, *Mol. BioSyst.*, 2014,

70 V. T. Duong, M. H. Unhelkar, J. E. Kelly, S. H. Kim, C. T. Butts and R. W. Martin, *Integr. Biol.*, 2018, **10**, 768–779.

71 T. J. Cross, G. R. Takahashi, E. M. Diessner, M. G. Crosby, V. Farahmand, S. Zhuang, C. T. Butts and R. W. Martin, *Biochemistry*, 2020, **59**, 3741–3756.

72 G. Grazioli, R. W. Martin and C. T. Butts, *Front. Mol. Biosci.*, 2019, **6**, 42.

73 K. A. James and G. M. Verkhivker, *PLoS One*, 2014, **9**, e113488.

74 D. K. Brown and Ö. T. Bishop, *Glob. heart*, 2017, **12**, 151–161.

75 A. A. Hagberg, D. A. Schult and P. J. Swart, Proc. Python Sci. Conf., Pasadena, CA USA, 2008, pp. 11 – 15.

76 U. Brandes, *J. Math. Sociol.*, 2001, **25**, 163–177.

77 L. C. Freeman, *Sociometry*, 1977, 35–41.

78 I. C. Kazan, P. Sharma, M. I. Rahman, A. Bobkov, R. Fromme, G. Ghirlanda and S. B. Ozkan, *eLife*, 2022, **11**, e67474.

10, 2215–2228.