Title

Comparative analysis of thermal adaptations of extremophilic prolyl oligopeptidases.

Journal

Biophysical Journal, 123(18)

Authors

Diessner, Elizabeth

Takahashi, Gemma

Butts, Carter

et al.

Publication Date

2024-09-17

DOI

10.1016/j.bpj.2024.07.013

Peer reviewed

Biophysical Journal
Article

Biophysical Society

Comparative analysis of thermal adaptations of extremophilic prolyl oligopeptidases

Elizabeth M. Diessner,[1] Gemma R. Takahashi,[2] Carter T. Butts,[3,*] and Rachel W. Martin[1,2,*]
[1]Department of Chemistry, University of California, Irvine, Irvine, California; [2]Department of Molecular Biology & Biochemistry, University of California, Irvine, Irvine, California; and [3]Departments of Sociology, Statistics, Computer Science, and EECS, University of California, Irvine, Irvine, California

ABSTRACT Prolyl oligopeptidases from psychrophilic, mesophilic, and thermophilic organisms found in a range of natural environments are studied using a combination of protein structure prediction, atomistic molecular dynamics, and trajectory analysis to determine how the S9 protease family adapts to extreme thermal conditions. We compare our results with hypotheses from the literature regarding structural adaptations that allow proteins to maintain structure and function at extreme temperatures, and we find that, in the case of prolyl oligopeptidases, only a subset of proposed adaptations are employed for maintaining stability. The catalytic and propeller domains are highly structured, limiting the range of mutations that can be made to enhance hydrophobicity or form disulfide bonds without disrupting the formation of necessary secondary structure. Rather, we observe a pattern in which overall prevalence of bound interactions (salt bridges and hydrogen bonds) is conserved by using increasing numbers of increasingly short-lived interactions as temperature increases. This suggests a role for an entropic rather than energetic strategy for thermal adaptation in this protein family.

SIGNIFICANCE Prolyl oligopeptidases from extremophilic organisms have highly structured catalytic and propeller domains, which limit how the protease can adapt to function at extreme temperatures. In an in silico study of psychrophilic, mesophilic, and thermophilic POPs, we show that the number of potential hydrogen bonds and salt bridges increases with temperature, whereas the duration of bound interactions decreases, resulting in a conserved prevalence of bound interactions that provides the necessary structural stability for function of the enzyme. This strategy for thermal adaptation is compared with others found in the literature, such as additional disulfide bonds, an enlarged hydrophobic core, or increases in rigidity or packing at higher temperatures, for which we do not find support.

INTRODUCTION

Prolyl oligopeptidases (POPs), which belong to the MEROPS (http://www.ebi.ac.uk/merops) protease family S9 (1), are large, soluble proteins with an ancient fold found across all domains of life (2). They are typically characterized by a two-domain architecture consisting of an α/β-hydrolase domain and a seven-bladed propeller domain that occludes the active site (3,4). Unlike many other proteins with propeller domains, S9 proteases lack disulfide bonds or specific hydrogen bonds (H-bonds) holding together the first and last blades of the propeller; this domain is stabilized in large part by hydrophobic interactions (5). These enzymes cleave small-to medium-sized peptide substrates (up to 33 residues (6)) at a position C-terminal to proline, although detailed substrate preferences vary. Microbial POPs are very diverse and are thought to have spread and diversified via multiple instances of horizontal gene transfer (7). Sequence identity among members of this family is weak except around the active-site residues; however, the overall fold is conserved. Although the larger POP family also contains enzymes with other chemical activities, including dipeptidyl and tripeptidyl peptidases, acylaminoacyl peptidases, and carboxypeptidases (8), here we focus specifically on S9 proteases with prolyl endopeptidase activity.

As ancient and ubiquitous proteins with complex structure, S9 proteases offer a natural opportunity for comparative analysis; here, our focus is on variation associated with differences in thermal environment. We have collected a set of 51 POP sequences found in 13 extremophilic organisms from various thermal environments with observed environmental temperatures (OETs) ranging from 261 to 348 K. Of this set, 30 sequences are from psychrophiles (261–293 K), nine are from

Submitted January 25, 2024, and accepted for publication July 10, 2024.
*Correspondence: buttsc@uci.edu or rwmartin@uci.edu
Editor: Chris Neale.
https://doi.org/10.1016/j.bpj.2024.07.013

mesophiles (300–310 K), and 12 are from thermophiles (322–348 K). The majority of these proteins include a propeller domain; however, 12 proteins (one psychrophilic, five mesophilic, and six thermophilic) do not have propellers. Prior structural and computational studies have suggested that interactions between the propeller domain and the catalytic domain are critical for modulating substrate binding and specificity in POPs having both domains. Crystal structures show a range of conformations with varying gaps between the domains (9–11), and, in the well-studied case of POP from the thermophilic archaeon *Pyrococcus furiosus*, microsecond molecular dynamics (MD) trajectories show a relationship between the size of the opening and the temperature (12). Here, we employ structural modeling and atomistic MD to address hypotheses about the overall protein dynamics and cohesion, the role of the propeller, and the relationship between these properties and the OET by analyzing the correspondence between the observed thermal environment and structural adaptations in this protein set.

Our study proceeds as follows. First, we investigate the trends in amino acid composition with OET of the protein, and compare observed trends with similar analyses found in the literature for other sets of homologous enzymes. Similar to previous work on the S11 protease family (13), we find that the S9 family of proteins do not follow much of the conventional wisdom on amino acid composition but instead employ a unique set of molecular adaptations to accommodate differences in thermal environment. For example, it is often claimed that thermophilic proteins contain more disulfide bonds than their lower-temperature counterparts (14,15). Such an adaptation is not observed for our sample, as none of the thermophilic proteins contains disulfide bonds; indeed, we observe no disulfide bonds for any protein in our set, suggesting that this family does not employ disulfide-based stabilization. We therefore examine other specific intramolecular interactions. We find that salt bridges and H-bonds are particularly important for stabilizing the structures, consistent with prior work on other POPs (9–12). Our results show that, although enzymes from thermophiles have more residues that are capable of forming salt bridges and H-bonds than their lower-temperature counterparts, the actual number of these interactions mostly remains constant, suggesting that stabilization of the heat-tolerant enzymes is driven by a larger number of redundant interactions. Finally, examination of active-site dynamics suggests four characteristic conformations, with mesophilic (rather than thermophilic or psychrophilic) enzymes having the most distinct pattern of state occupancy.

MATERIALS AND METHODS

Sequence selection, alignment, and clustering

Thirteen microbial organisms that grow at different temperatures (representing psychrophiles, mesophiles, and thermophiles) were selected for this study. The preferred growth conditions of these organisms were determined based on the estimated average temperature of their documented habitats and descriptions in the NCBI database and publications associated with each entry. Serine proteases from these organisms were then identified in UniProt. For one organism (*Shewanella frigidimarina*), two proteomes were used because one (SHEFN) was derived from a less psychrophilic isolate in the North Sea near Aberdeen, Scotland, UK, whereas the other (SHEFR) was from an Antarctic isolate presumed to be adapted for growth at lower temperature. All organisms are bacteria except for *Halobacterium salinarum*, which is an archaeon.

The following proteomes were chosen based on the criteria described above. Metadata about the environment where each organism was found, including temperature and pH, were obtained from the respective NCBI entries. Measured intracellular pH values and salt concentrations were used in cases where they have been experimentally measured and reported. Otherwise, pH 7.2 and 200 mM salt (typical intracellular concentrations) were assumed.

Proteins with putative serine proteolytic activity (referred to here as serine proteases) were identified in these proteomes by searching UniProt (31) for all [[GO: serine-type] AND [NOT GO: inhibitor]] proteins. Proteins that did not satisfy these conditions were manually screened for putative proteolytic activity if they fitted any of the following criteria: [[GO: peptidase activity [0008233]] AND [Protein Name: serine]] or [[Protein Name: serine] AND [Protein Name: protease]] or [[Protein Name: serine] AND [Protein Name: peptidase]] or [[Protein Name: serine] AND [Protein Name: proteinase]]. At this point, proteins that contained transmembrane regions and/or intramembrane regions, predicted by either UniProt or Scampi2 (32), were flagged for further screening; remaining unflagged proteins were retained without further processing. Flagged proteins whose membrane regions were in the first 70 residues (inclusive) were unflagged if they also contained a signal sequence, predicted by SignalP-6.1 (33). Two Clp protease ATP-binding subunits (ClpA), which do not have proteolytic activity, were then removed from the dataset by hand. Remaining unflagged proteins were considered putative nonmembrane serine proteases and were compiled for further analysis. A BLAST search against SwissProt and InterProScan (34) was employed to identify sequences with annotations for MEROPS S9 serine proteases. This resulted in 51 S9 protease sequences: 10 from thermophiles, 11 from mesophiles, and 30 from psychrophiles. These were further clustered into five groups: S9A–S9E.

The above methods can be thought of as a series of datasets. In Table 1, each set builds into the final S9 protease set, set S9.

A generalization of Ward's method (35) was used to cluster the S9 serine proteases based on sequence dissimilarity into hierarchical clusters. Sequence alignments were produced by ClustalOmega (36) using the following settings: gap open penalty = 10.0, gap extension penalty = 0.05, hydrophilic residues = GPSNDQERK, and a BLOSUM weight matrix. Logo plots showing conservation of sequence regions were generated using WebLogo 3.0 (37).

Structure prediction

Initial structures were predicted by comparative modeling and all-atom refinement based on a simplified forcefield provided by the iTasser (38) software pipeline (a short discussion of structure prediction software choices can be found in the supplement). These preliminary models then underwent in silico maturation where signal sequences (identified using SignalP 6 (33)) and pro-sequences were removed. PROPKA3 (39) was used to adjust the protonation states at the relevant pH for each organism. Finally, modified structures were equilibrated in explicit solvent using atomistic MD, as described below. The PDB files of the in silico-matured structures are available in the supporting material. A search of the PDB for "POP" from organism classes "Bacteria" and "Archaea" yielded crystal structures of five unique mesophilic POPs, not including point mutations and inhibitor complexes (PDB: 3IUJ, 7VGB, 2BKL, 1YR2, and 4HVT (10,11,40,41) and one thermophilic POP (PDB: 5T88) (12) after manual

TABLE 1 Protease dataset search criteria

Dataset	Included	Input characteristics	Tools and resources
Set 1	selected proteomes	satisfied [[GO: serine-type] AND [NOT GO: inhibitor]]	UniProt search
Set 2	selected proteomes	satisfied [[GO: peptidase activity [0,008,233]] AND [Protein Name: serine]]	UniProt search
Set 3	selected proteomes	satisfied [[Protein Name: serine] AND [Protein Name: protease]]	UniProt search
Set 4	selected proteomes	satisfied [[Protein Name: serine] AND [Protein Name: peptidase]]	UniProt search
Set 5	selected proteomes	satisfied [[Protein Name: serine] AND [Protein Name: proteinase]]	UniProt search
Set 6	Unique sequences from set 1-set 5	did not contain transmembrane or intramembrane regions	UniProt, Scampi2
Set 6.5	Unique sequences from set 1 to set 5 that are not in set 6	did not make the cut for set 6, but did contain a signal sequence	SignalP-6.1
Set 7	set 6-set6.5	combined set 6 and set 6.5, but without two ClpA subunits	manual curation, UniProt
Set 9	set 7	identified as MEROPS S9 serine proteases	manual curation, BLAST, SwissProt, InterProScan

identification of enzymes with prolyl endopeptidase activity. There were no experimental structures of psychrophilic POPs; therefore, direct comparisons to experimental structures from all three thermal environments could not be made. UCSF Chimera (42) and VMD (43) were used to generate figures containing protein structures.

Molecular modeling and analysis

The proteases described above were modeled in explicit (TIP3P (44)) solvent using atomistic MD simulations, using PROPKA3 (39) to correct protonation states for the observed environmental pH. Systems were prepared in VMD (43) in cubic water boxes with a 15-Å minimum margin. Counterions were added to neutralize the structure, and additional ions (NaCl) were added to match the system ionic strength; where this was unknown, 200 mM salt was employed. Each system was hence modeled under conditions matching as closely as possible its known environment of origin. The prepared system was then equilibrated as follows: NAMD (45) was employed for all simulations using the CHARMM36m force field (46) under periodic boundary conditions in an *NpT* ensemble at 1 atm pressure with Nosé-Hoover Langevin piston pressure control (47,48) and Langevin temperature control (damping coefficient 1/ps). Representative structures for each protein were obtained by a 1-ns simulation at native temperature (following a short (10 ps) minimization and box size adjustment period), with the final frame being employed for visualization and analysis. Long trajectories were also obtained for each protein. These systems were prepared as above, but were simulated for 100 ns, with 5000 frames retained for analysis (1/20 ps). Convergence to equilibrium was verified using the Raftery-Lewis diagnostic (49) from the coda library for R (50), using the median RMSD from the initial frame as a target statistic (convergence criterion was 95% confidence of the sample median RMSD being within $\pm$ 1 Å of the true equilibrium median RMSD). As shown in Fig. S26, all 100-ns trajectories converged within the length of the simulation.

Salt-bridge data were obtained for each 100-ns simulation using the Salt Bridge Plugin in VMD with an oxygen-nitrogen cutoff of 3.2 Å. Specifically, charged residue pairs were regarded as having a salt bridge at some point within the trajectory if they met the 3.2-Å criterion within any frame. For purposes of bond occupancy analysis, we treated a salt bridge identified using this rule as being occupied in any given frame if the centers of mass for the respective residues were within 5.4 Å of each other; the center-of-mass cutoff was chosen to replicate the bond set identified by the atom-pair analysis. Hydrogen-bond data were obtained for each 100-ns simulation using the H-Bonds Plugin in VMD for unique H-bonds with a donor-acceptor distance of 3.0 Å and an angle cutoff of 20°. The surface area between the propeller domain and the catalytic domain was calculated using the Surf Plugin in VMD for each protein that had both domains by measuring the surface area of each domain independently, then subtracting the surface area of the whole protein from the sum of the two individual parts, and finally dividing the remaining value by two.

Secondary structure information and the exposure of individual residues were analyzed using DSSP (51) for all frames of both the single structure and the 100-ns trajectories. Residue total surface areas were obtained from Pacios (52). A packing metric was calculated using the solvent accessible surface area (SASA) obtained from DSSP in combination with the total surface area. In the calculation of the packing metric, the total sum of SASA was subtracted from the total surface area of all residues in the sequence, leaving a value that is interpreted to be the total buried surface area. The ratio of the buried surface area with the total SASA for each sequence was then used as a rough packing metric.

To analyze the cohesion of the proteins, each frame of the trajectories was converted to a protein structure network (PSN) following the definitions for nodes set by Benson and Daggett where each node represents a chemical moiety within the proteins (53). Networks were constructed using the procedure of Butts et al. (54): nodes are defined to be in contact when at least one atom within each of the respective moieties are within 1.1 times the sum of their van der Waals radii. Statistics for degree and degree *k*-core were calculated for each PSN as measures of the number of contacts on each node, and the local structural cohesion (as suggested by Unhelkar et al.) (55), respectively. Construction and analysis of networks was completed using the sna (56) and network (57) libraries for R (58).

Random feature clustering and Markov analysis

Active-site dynamics of the S9 proteases were analyzed to compare thermal groups as well as the possible influence of the propeller domain where it is present. The active sites consist of the three catalytic residues (serine, lysine, and histidine) identified by MEROPS. Analysis was performed using random feature projections, in which the coordinates of the nitrogens and oxygens of each catalytic residue are extracted (using the bio3d library for R (59,58)) for each frame of each trajectory, creating an interatomic distance matrix for each frame in the set of trajectories. These matrices were converted to a single frame by distance matrix, which was then concatenated and projected onto a deep random feature space consisting of a stacked composite linear and arc-cosine kernel model (60). Input distance

vectors were given a constant intercept feature before being projected onto 500 random ReLU functions. A skip layer augmented the projections with a copy of the input distance matrix, all of which was then pruned to 25 features by principal-component analysis (PCA). Three total layers were used, with each subsequent layer taking the PCA scores from the previous layer as input. The PCA results of the final layer were used as the data embedding. Embedding was carried out in R, including some use of the Rcpp package (61). A scree plot (see Fig. S18 *A*) on the final PCA shows that 25 dimensions sufficiently describe the dynamics of the active-site residues.

Clustering of the conformations from the 25-dimensional embedding was done using k-means clustering (default R implementation, 75 restarts, 100 iterations). The method of total Markov error use by Grazioli et al. was used to find the number of clusters (62); the value of k was chosen as the largest number of clusters having a posterior predictive *Z* score for the reduction in mean root-mean-square error (see Fig. S18 *B*), with root-mean-square error calculated using the input interatomic distances, resulting in four clusters. A classification tree was fitted to predict cluster membership from raw distances using rpart (63) with a maximum depth of 3, cross-validation sample size of 150, and complexity parameter of 0.001. The distance thresholds that were indicated at branches of the dendrogram were then used to interpret the differences in conformational states of each cluster, as discussed below.

RESULTS

Clustering by active-site sequence motifs divides the S9 proteases into five subfamilies

The S9 serine proteases in this dataset were chosen from a selection of proteomes representing different thermal environments (Table 2). All the selected organisms are bacteria and were found in environments that are mild with respect to pH and salt concentration, except for the archaeon *H. salinarum*, which is also an extreme halophile that survives in a wide range of salt conditions (64). To assess the degree of thermal adaptations in overall amino acid sequence, the proteases were clustered by protein sequence similarity using Ward's algorithm implemented in R (58), resulting in the dendrogram shown in Fig. 1 (right). The first split separates a small group of seven psychrophilic sequences from a large, highly chained cluster containing the majority of the proteases from all three thermal groups, indicating substantial sequence diversity and a lack of strong sequence conservation among proteins from the same thermal environment, as might be expected for proteins under selection for particular sequence-based adaptations.

Clustering the S9 proteases found in this set by the subsequences surrounding the active S residue (spanning the range from two residues before and seven after the nucleophilic serine) divides the proteases into five categories (Fig. 1 *A–E*, left side of the figure). This sequence region was chosen because it contains both invariant residues (e.g., the GXSXG motif that includes the active serine) along with others whose variation characterizes different subtypes of this enzyme family (2). The corresponding sequence alignments are shown in the supporting material (see Figs. S1–S11). Color coding according to the sequence motif surrounding the active serine is maintained throughout the paper. Logo plots for the corresponding sequence motifs are shown as insets (except for cluster C, which contains only three sequences, making this type of visualization impractical). Each cluster has a characteristic variation of the sequence motif surrounding the active serine, which may be related to substrate preferences (Table 3). For each cluster, the centroid, or the sequence with the most similarity to all other sequences in that cluster, is indicated in bold, colored text. Only one set of sequences (cluster E on the left) is consistently observed using both local and global notions of sequence similarity: all but one of these proteins (Q3IKV5_PSET1) are also found in the first cluster split off from the rest of the full-sequence tree. This cluster is located at the bottom and colored dark blue in both panels.

Psychrophilic organisms have the most sequence diversity in this enzyme class, as measured both by active-site motifs and full sequences. In the active serine subsequence tree, clusters A and D contain sequences from all three thermal environments, whereas clusters B and E each contain one mesophilic sequence and are otherwise psychrophilic, and cluster C comprises only psychrophilic sequences. In the full-length sequence tree, all three clusters that are separate from the large chained cluster are composed almost entirely of psychrophilic sequences; the one exception is

TABLE 2 S9 sequences

Thermal group	Formal name	ID	UniProt ID	RefSeq ID	Reference
Psychrophiles	*Pseudoalteromonas translucida*	PSET1	UP000006843	GCF_000026085.1	Médigue et al. (16)
	Psychromonas ingrahamii	PSYIN	UP000000639	GCF_000015285.1	Copeland et al. (17)
	Shewanella frigidimarina	SHEFN	UP000000684	GCF_000014705.1	Copeland et al. (18)
	Shewanella frigidimarina	SHEFR	UP000055702	GCF_001529365.1	Parmeciano Di Noto et al. (19)
Mesophiles	*Bacillus subtilis*	BACSU	UP000001570	GCF_000009045.1	Kunst et al. (20)
	Escherichia coli	ECO57	UP000000558	GCF_000008865.2	Makino et al., Hayashi et al. (21,22)
	Pseudomonas fluorescens	PSEPF	UP000002704	GCF_000012445.1	Silby et al. (23)
Thermophiles	*Caldicellulosiruptor saccharolyticus*	CALS8	UP000000256	GCF_000016545.1	van de Werken et al., Copeland et al. (24,25)
	Dictyoglomus thermophilum	DICT6	UP000001733	GCF_000020965.1	Coil et al. (26)
	Geobacillus kaustophilus	GEOKA	UP000001172	GCF_000009785.1	Takami et al. (27)
	Halobacterium salinarum NRC1	HALSA	UP000000554	GCF_000006805.1	Ng et al. (28)
	Thermomicrobium roseum	THERP	UP000000447	GCF_000021685.1	Wu et al. (29)
	Thermus thermophilus HB8	THET8	UP000000532	GCF_000091545.1	Masui et al. (30)

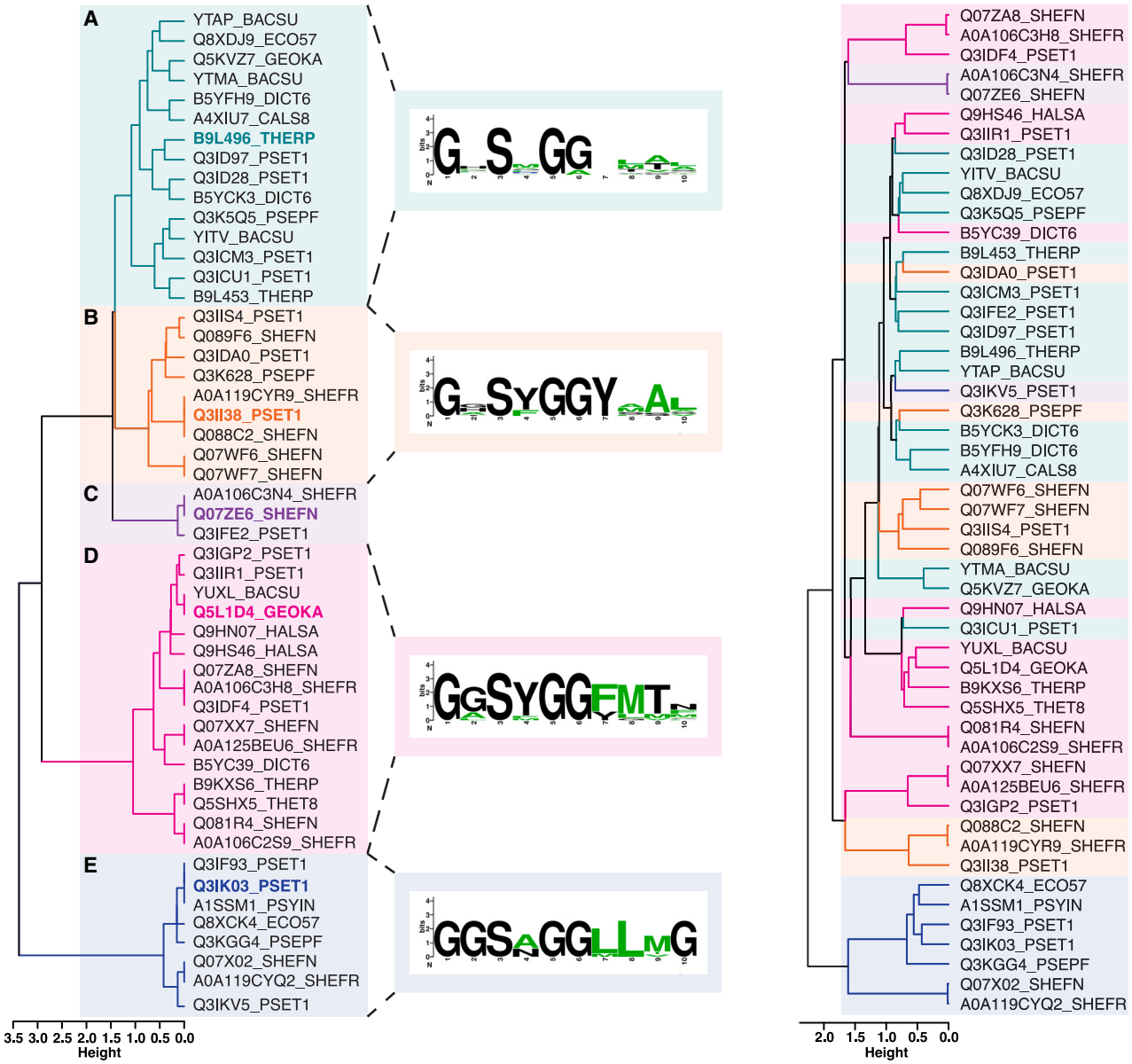

FIGURE 1 S9 protease sequence clusters. Left: S9 proteases from a variety of microbes from different thermal environments clustered by sequence similarity using a subsequence consisting of 10 residues surrounding the active serine. The clusters are labeled *A–E* for identification. Bold labels indicate the centroid of each cluster. The relevant sequence motifs are shown as logo plots in the insets. Right: the same sequences clustered by similarity of the full sequences. Color coding denotes the active serine subsequence clusters.

Q8XCK4_ECO57 from *Escherichia coli*, which clusters with a group of psychrophilic enzymes in both trees. Notably, each *Shewanella* isolate has multiple paralogs, six for SHEFR and nine for SHEFN, consistent with a previous observation that *Shewanella* species are particularly rich in these enzymes (7). PSET1 has 13. As in our previous investigation of S11 proteases, sequences from the mesophilic organism *Bacillus subtilis* (BACSU) are found in the clusters that include thermophilic enzymes, whereas the mesophilic proteases from *E. coli* (ECO57) and *Pseudomonas fluorescens* (PSEPF) are found in the groups that otherwise only contain psychrophilic proteases.

Molecular models show a $\alpha\beta$ hydrolase architecture

Molecular models representing the centroid proteins of each cluster (after in silico maturation) are shown in Fig. 2. Fig. 2 *A–E* are labeled according to the clusters in Fig. 1. The residues making up the catalytic triad (S, dark cyan; H, purple; and D, red) are drawn as space-filling models, and expanded views of them are shown in the insets. Examination of the positions of the active-site residues reveals three distinct types of conformations. The active conformation, where all three catalytic residues

TABLE 3 Properties of S9 proteases

Cluster	No.	Type(s)	Centroid protein	Active Ser Motif
A	15	thermo-, meso-, and psychrophiles	B9L496_THERP	GXSXXGGXXX
B	9	meso- and psychrophiles	Q3II38_PSET1	GSXYGGYXAX
C	3	psychrophiles	Q07ZE6_SHEFN	GHSW(G/A)GGYQS
D	16	thermo-, meso-, and psychrophiles	Q5L1D4_GEOKA	GGSYGG(F/Y)MTX
E	8	meso- and psychrophiles	Q3IK03_PSET1	GGS(A/N)GGLLMG

are aligned, can be observed in Fig. 2*B* and *C*. An open conformation, where all three catalytic residues are separated in space, is shown in Fig. 2*D*. Fig. 2 *A* and *F* show examples of partially open conformations, where the S and H are closely associated, but the D does not interact with the H on the other side.

Fig. 2 *F* shows the same protein as in E (Q3IK03_PSET1), colored to show how its amino acid sequence contributes to its domain architecture, rendered in both ribbon (left) and surface (right) views. This protein is composed of two domains, a propeller domain and a catalytic domain. The N terminus of the protein begins with a flexible tail (gray), followed by a

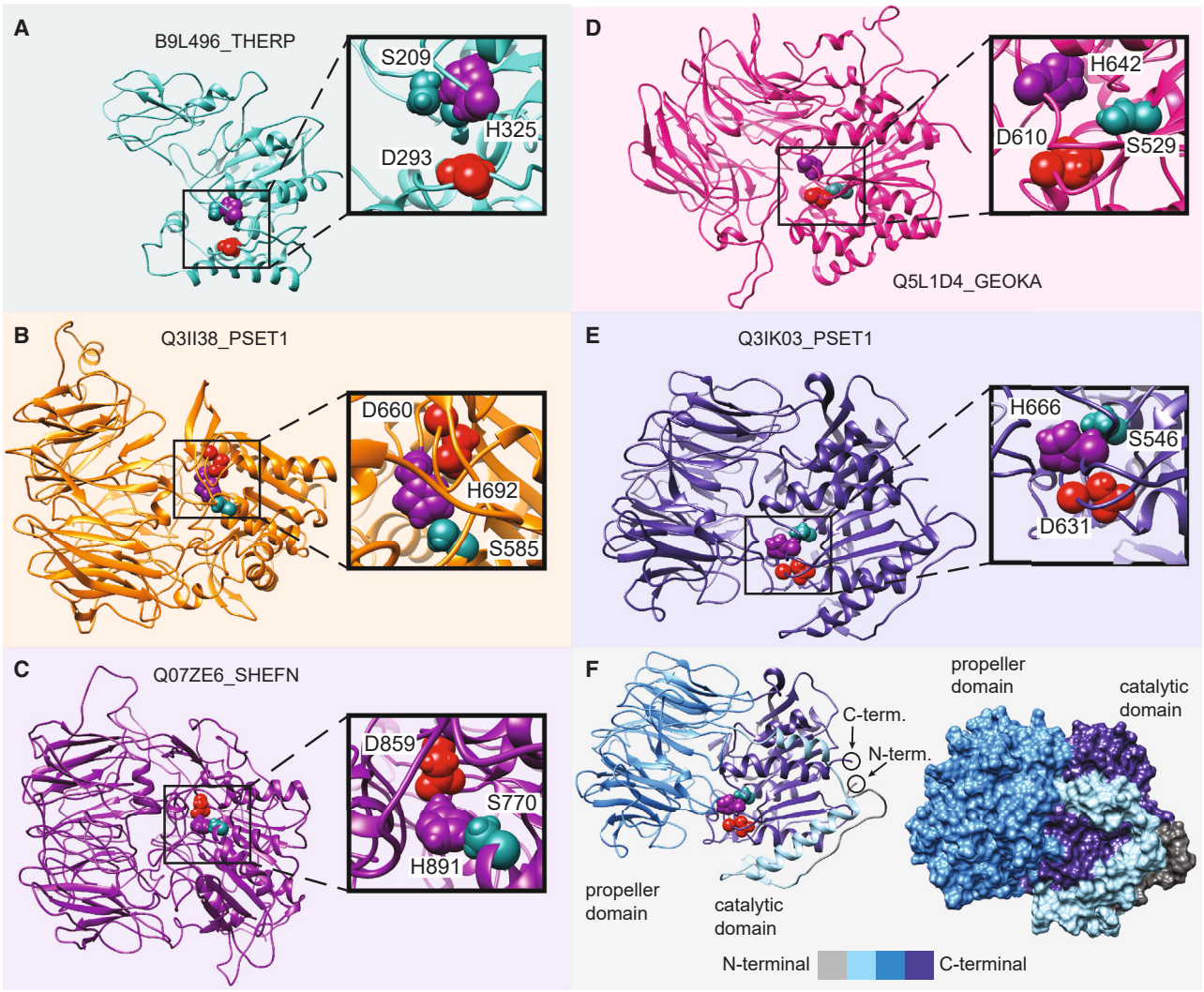

FIGURE 2 Predicted structures for representative S9 proteases. Molecular models of the proteins representing the centroids of each active-site subsequence cluster in Fig. 1, using the same labels and color coding for (*A*)–(*E*). In each case, an expanded view of the active-site residues (space-filling models: S, dark cyan; H, purple; and D, red) is shown in the inset. (*F*) Two additional views of Q3IK03_PSET1. Left: coloring indicates the relationship between the sequence position and the domain structure, which is complex. The N-terminal sequence region, comprising the propeller domain as well as two helices and two β strands that are part of the catalytic domain, is shown in light blue. The C-terminal sequence region, shown in dark blue, makes up the rest of the catalytic domain. Right: surface view of the same protein, illustrating how part of the N-terminal sequence region is integrated into the catalytic domain.

long helix, a linker region, and a shorter helix (light blue). Both N-terminal helices are packed against the catalytic domain, anchoring it to the propeller domain (medium blue). The rest of the catalytic domain (dark blue) comprises the C-terminal sequence region. The catalytic domain itself has an α/β-hydrolase fold, with the catalytic triad located near one edge of the inter-domain interface. With the exception of B9L496_THERP (Fig. 2 *A*), which lacks the N-terminal helices and most of the propeller domain (which has been previously observed in bacterial POPs) (7), all of these proteins have the same overall architecture. However, there is considerable variation in relative orientation between the two domains and in the lengths and degree of mobility of the loops in between the β strands comprising the propellers.

Detailed views of the propeller domain are shown in ribbon view and space-filling models in Fig. 3 *A* and *B*, respectively, and the catalytic domain is shown in Fig. 3 *C*. The points where the other domain has been removed for clarity are circled in each case. As is typical for this fold (65), the seven-bladed propeller is made up of twisted, four-strand β sheets arranged in a circle (numbered from N terminus to C terminus in Fig. 3 *A*). Unlike some other examples of similar proteins, where the gap between the first and the last blade is held together by backbone H-bonds as part of the β sheet structure, here the circle is closed with salt bridges, H-bonds, and hydrophobic interactions between side chains alone. Fig. 3 *B* shows the same view of the propeller with all side chains shown as space-filling models, illustrating that the central hole is not large enough to let substrates pass that way without significant rearrangement of side chains and potentially the loops at the point of each blade. Fig. 3 *C* shows the catalytic domain, with secondary structure elements labeled as in Fülöp et al.

Relative abundance does not vary monotonically with environmental temperature for most amino acids

An important consideration for protein adaptation to particular thermal environments is amino acid composition. A common assumption for enzymes is that orthologs from organisms adapted to different growth temperatures should maintain a similar structure and comparable dynamics to interact with substrates at similar rates despite the differences in temperature. To the extent that amino acid composition controls these factors, one would expect monotonic trends in amino acid composition that vary with the thermal environment. A more specific hypothesis is that proteases from organisms living in warmer thermal environments would be enriched in residues that enhance packing and rigidity to maintain a given structure at high temperature. This could manifest simply as more disulfide bonds (and thus more cysteines), or, more subtly, as more salt bridges (and thus more charged residues). Another possibility is increased stability due to the hydrophobic effect (an increased proportion of hydrophobic residues). Berezovsky et al. have found in a study of triosephosphate isomerase homologs that both hydrophobic and charged residues are

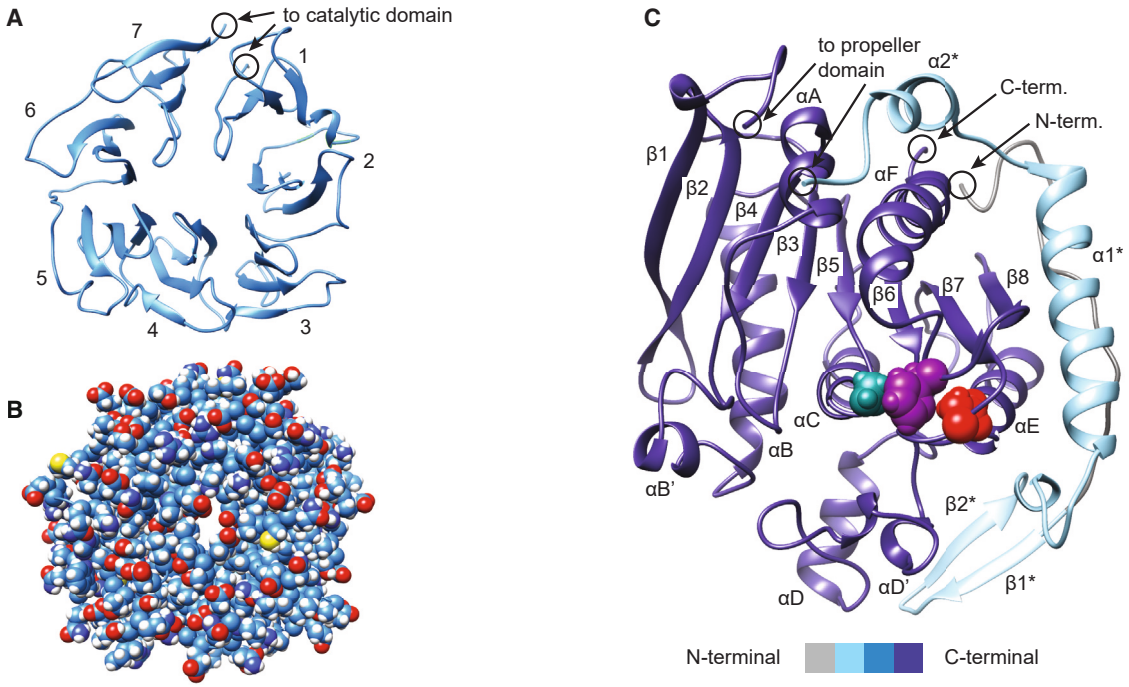

FIGURE 3 Structural details of the two domains for a representative example, Q3IK03_PSET1. (*A*) Ribbon representation of the propeller domain. Blades are numbered 1–7 from N terminus to C terminus. (*B*) Propeller domain with all side chains shown using space-filling models. (*C*) Ribbon representation of the catalytic domain, using the numbering scheme of Fülöp et al. for secondary structure elements.

enriched in thermophilic proteins, which they describe as thermal stability "from both ends of the hydrophobicity scale" (66). They find further that enrichment in a particular set of amino acids (IVYWREL) correlates strongly with optimal growth temperature for the prokaryotic proteomes in their sample (67). Conversely, one might expect fewer stabilizing interactions in psychrophiles, where flexibility is paramount; at low temperature, stability is less of a concern than having the ability to bind and release substrate at all. Another complicating factor is that the hydrophobic effect contributes less to stabilization of the folded state at low temperature.

Contra these hypotheses, this set of S9 proteases does not show monotonic trends for the relative abundance of most amino acids, as shown in Fig. 4. For example, the fraction of cysteine, which would need to be elevated in thermophiles to produce increased thermal stability via disulfide bonds, is similarly low in all three thermal groups. Many amino acids (notably, H, F, L, V, and I) have nonmonotonic temperature dependences. In others, (e.g., R, K, and N), mesophiles and thermophiles have similar proportions. Compared to the other two thermal groups, psychrophiles are enriched in K and N but exiguous in R. The enrichment of K may be explained by the increased entropy afforded in the folded state due to the rotameric degrees of freedom provided by K relative to R (68). The latter observation can also be explained in terms of arginine's propensity to make strong salt bridges with carboxylic acids such as the side chains of D and E, potentially leading to rigidifying salt bridges. At lower temperatures, this effect combined with arginine's hydrophobicity can lead to undesirable intermolecular interactions, such as liquid-liquid phase separation (69,70). Both sequences from *H. salinarum* are particularly enriched in negatively charged residues as well as having much more R than K, consistent with the idea of simultaneous adaptation for high-temperature and high-salt

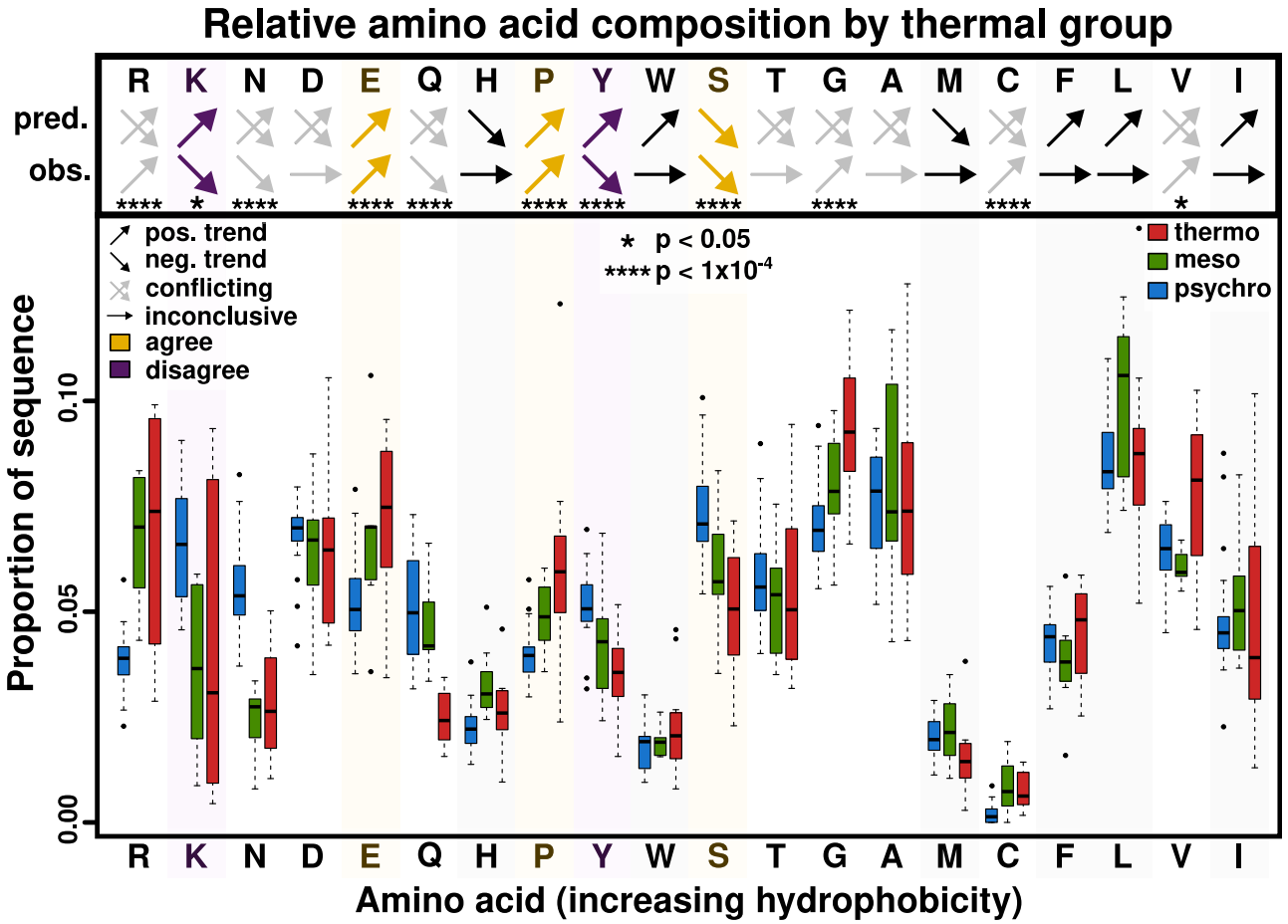

FIGURE 4 Amino acid compositions normalized by sequence length and grouped by thermal group (shown as boxplots). Few amino acids follow monotonic trends with respect to temperature, although relative proportions of the amino acids R, E, P, G, and C are observed to significantly (denoted by a star) increase with temperature, irrespective of monotonicity. The relative proportions of amino acids N, Q, Y, and S are observed to decrease significantly with increased temperature. Results from the literature are conflicting for nearly half of all amino acids (nine out of 20) and are found to contradict our observations in the cases of K and Y. However, the trends we observe for E, P, and S agree with the literature consensus. Our measurements for H, W, M, F, L, and I showed no significant trend (See Figs. S12 and S13 for detailed plots).

environments. In the latter, enrichment of negative charges can be driven by destabilization of the unfolded state (71). For Q and M, the proportions in mesophiles are more similar to those of psychrophiles, with reduced abundance in thermophiles.

The amino acids E, P, Y, S, and G do follow monotonic trends with respect to thermal environment; the fractions of E, P, and G increase with growth temperature, whereas the fractions of Y and S decrease. Thus, for this set of enzymes, the prediction of increased hydrophobic residues as a function of temperature is not fully supported (see Figs. S12 and S13 for more information). Thermophiles do tend to have more V than either psychrophiles or mesophiles, but, in this case, mesophiles have a lower proportion than psychrophiles, disrupting the trend with temperature. In fact, none of the hydrophobic residues show a significant trend with temperature, and, in general, these proteins appear not to follow the trends reported by Berezovsky et al. (see Fig. S14). The abundance of the structure breakers P and G in thermophiles and their increase with increasing temperature, likewise, is inconsistent with the argument that enzymes in warmer thermal environments require adaptations that increase rigidity.

Combining amino acids into groups based on the interactions they can form gives a clearer picture of overall trends, as shown in Fig. 5. Hydrophobic residues are again present in higher proportions overall in mesophiles, resulting in a significant trend with temperature despite the lower abundance in thermophiles. Other residue groups, such as aromatic and positively charged residues, do not show any significant trends with temperature. Negatively charged residues show a significant increase with temperature, and polar residues are present in higher proportions as temperature decreases. The abundance of polar residues in psychrophiles may be an adaptation that helps to control the local hydration environment surrounding the active site that may then guide and orient substrate toward the catalytic residues, taking advantage of inherent structural stability that results from the lower kinetic energy of the system. Such a structural adaptation would

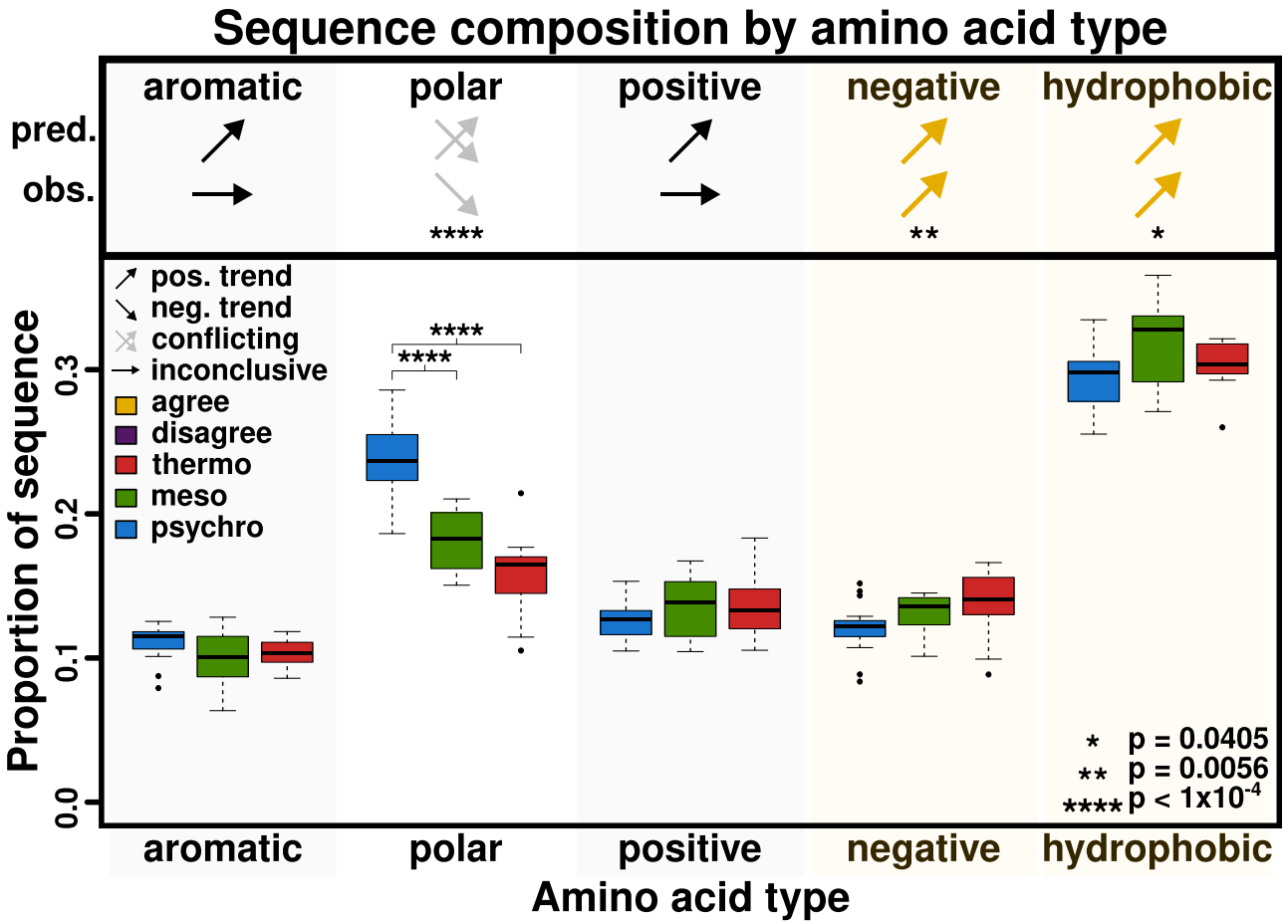

FIGURE 5 Differences in overall composition by amino acid type (proportions shown as boxplots). Mesophiles have a higher proportion of hydrophobic residues than either thermophiles or psychrophiles. Thermophiles are relatively exiguous in uncharged polar amino acids and enriched in charged residues. On the other hand, psychrophiles are relatively enriched in uncharged polar residues; however, their composition differs considerably by sequence cluster (See Figs. S15 and S16 for detailed plots).

not necessarily be applicable for thermophiles and may end up increasing the disorder of the system. Likewise, thermophiles may require higher proportions of charged residues to contribute to salt bridges or to minimize the formation of local regions of nonnative secondary structure that lead to aggregation (66).

Thermophilic S9 proteases are stabilized by H-bonds and salt bridges

Beyond the hydrophobic effect, another hypothesis for increased stability at high temperatures is an increase in disulfide bonds and/or salt bridges. As shown in Fig. 4, the S9 have few cysteine residues and, relatedly, are not observed to contain any disulfide bonds. Instead, salt bridges and H-bonds appear to do the work of "stapling" the structure together to prevent unraveling and maintain function. By tabulating the counts of salt bridges by the pairs of amino acids that can form them (see Table S1), we observe differences in the frequency of specific pairs of salt bridges with changes in the OET. The increase in relative amounts of E and R with temperature is consistent with an increase in salt bridges. Further, when controlling for sequence length, salt bridges containing arginine and/or glutamic acid appear more frequently in the structures of thermophilic proteins. In addition, when focusing on only those cross-domain salt bridges, we find a trend toward larger amino acids at higher OET. Fig. 6 *A* shows the relationship between temperature and the frequency of each of five sets of salt-bridge pairs: E-R, D-R, E-K, D-K, and D/E-H. Pairs with H are grouped together due to the small number of protonated H residues in a position to form salt bridges. In general, the frequency of salt bridges increases with temperature, with the only exception being the pair with the smallest side chains, D-K. Additionally, salt bridges between residues with larger side chains, such as R and/or E, are significantly more frequent in thermophiles than in psychrophiles. The longer and more tightly bound (due to the guanidino group of R) salt bridges may be favored at higher temperatures due to their ability to accommodate more local fluctuations in structure. The longer side chain may also help re-form the salt bridge sooner after it is broken by a transient local unfolding event. This strongly suggests that salt bridges play a major role in preventing the structure from either unraveling or unfolding by destruction of the inter-domain interface.

The occupancy of salt bridges, shown in Fig. 6 *B*, was determined by the sum of frames in which the centers of mass of the residues within a given salt-bridge residue pair were within a threshold distance of 5.4 Å, divided by the total number of frames in the trajectory. The average occupancy of a salt bridge decreases with increasing temperature, indicating a more transient relationship between salt-bridge pairs. The combination of the decrease in salt-bridge occupancy and the increase in salt-bridge frequency with temperature indicates the role of salt bridges in maintaining structure—at higher temperatures, individual salt bridges are less stable, but more potential bridges are available. The net effect is to maintain a roughly constant prevalence of bound interactions, with compensation occurring by increasing the number of potential salt bridges to offset the reduced probability of any given bridge being present at any given time.

To assess the role of salt bridges in the interface between the propeller domain and the catalytic domain, a similar analysis was done for salt-bridge pairs that included one residue from each domain, thus providing an inter-domain connection. Results are shown in Fig. 6 *C* and *D*. The E-R and D-R pair are again observed to be significantly more frequent in thermophiles. This is again likely due to the strength of the bond between the carboxylic acid moiety of D or E with the guanidino group of R. Occupancy follows the same trend as above, with more transient salt bridges at higher temperatures. The structures in Fig. 6 *E* and *F* show the salt bridges that span the two domains, with side chains drawn as lines colored to match the plots in Fig. 6 *C* and the charged N and O atoms shown with blue and red spheres, respectively. The salt bridges holding the two domains together appear to have a pattern of increased density of salt bridges containing arginine around the active-site residues. The interactions between residues D/E, and R, are stronger than those with K, again due to the resonance afforded by the guanidino group of R. The additional strength of these interactions appears to lend increased stability around the active site. This also hints at the functional dynamics of the protein; the two domains are heavily interconnected such that a gap can be formed to allow substrate to reach the active site without large-scale and potentially irreversible separation of the domains.

We observe a similar result for trends in the frequency of H-bonds with OET (i.e., counts of donor-acceptor pairs observed to form H-bonds at some point in the MD trajectory). Increased numbers of potential H-bonding pairs in thermophiles points to their utility for accommodating the increased range of conformational states that the protein samples at higher temperatures. The additional "stickiness" provided by the H-bonds may counter the effect of excess thermal energy and prevent stepwise unfolding. As shown in Figs. 7A and 7C, the frequency of H-bonds also increases significantly with temperature for both the overall protein structure and for H-bonds between the two domains. The occupancy of H-bonds, measured as the fraction of frames in which the interacting atoms are within 3.0 Å, again shows a significantly decreasing trend with temperature (shown in Fig. 7 *B* and *D*), supporting the hypothesis that the instability of individual bonds at higher temperatures is compensated for by increasing the number of potential bonds. These results also support the hypothesis that proteins at higher

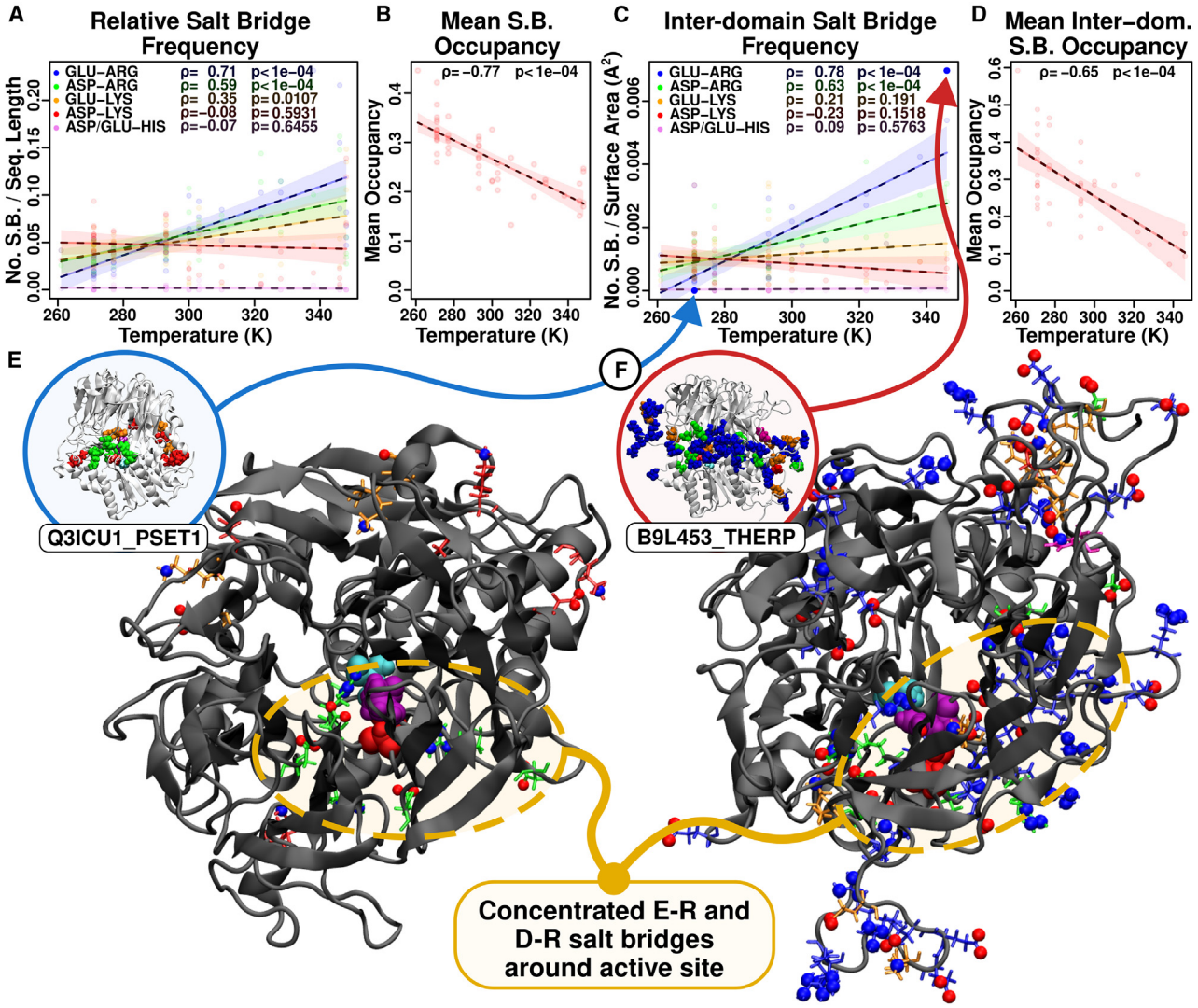

FIGURE 6 The frequency and occupancy of salt bridges as calculated in VMD using the distance between the center of mass of the oxygens and nitrogens. (*A*) Frequency of salt bridges (SBs) by OET given by the total count of SB pairs, normalized by sequence length. SB pair types (E-R, D-R, E-K, D-K, and D/E-H) were counted separately. (*B*) Mean occupancy of all SBs in each protein by OET given as the fraction of frames in which the respective residue centers of mass within 5.4 Å. (*C*) Frequency of SB pairs formed by residues from each domain, normalized by the inter-domain surface area calculated by VMD. (*D*) Mean occupancy of SBs spanning the two domains, using a 5.4-Å cutoff. OLS fits shown by central lines; shaded area indicates 95% confidence bands. (*E*) The psychrophilic enzyme Q3ICU1_PSET1 has the lowest frequency of inter-domain E-R pairs. An overlay shows the backbone as ribbons with SBs as spheres colored to match (*C*). The larger structure looks through the propeller domain toward the catalytic domain, with active-site residues drawn as spheres colored by amino acid (teal, purple, and red representing S, H, and D, respectively) and SB pairs drawn as lines colored to match (*C*) with blue and red spheres representing the interacting nitrogen and oxygen atoms, respectively. (*F*) The thermophilic enzyme B9L453_THERP has the highest frequency of inter-domain E-R pairs. Style matches (*E*). Yellow ovals highlight an increase in SB pairs containing R in the active-site region.

temperatures may experience more transient intraprotein interactions as they explore a larger conformational space, necessitating a larger number of available H-bonds (and salt bridges) to accommodate the thermally induced fluctuations, preventing the structure from unraveling. Psychrophiles, by contrast, do not require additional bonds to maintain structure and could even be disadvantaged by the reduced flexibility that comes with an overabundance of salt bridges and H-bonds.

As shown in Fig. 7 *E* and *F*, the pattern of H-bonds that span the two domains reveals a region between the active site and the N terminus that has fewer interactions. The lack of H-bond interactions in this region, combined with the increased density of salt bridges in an adjacent region surrounding the active site, may point to the functional dynamics of the protein. This interaction pattern suggests that the N terminus should remain less tightly bound to the rest of the protein; this may act as an entry

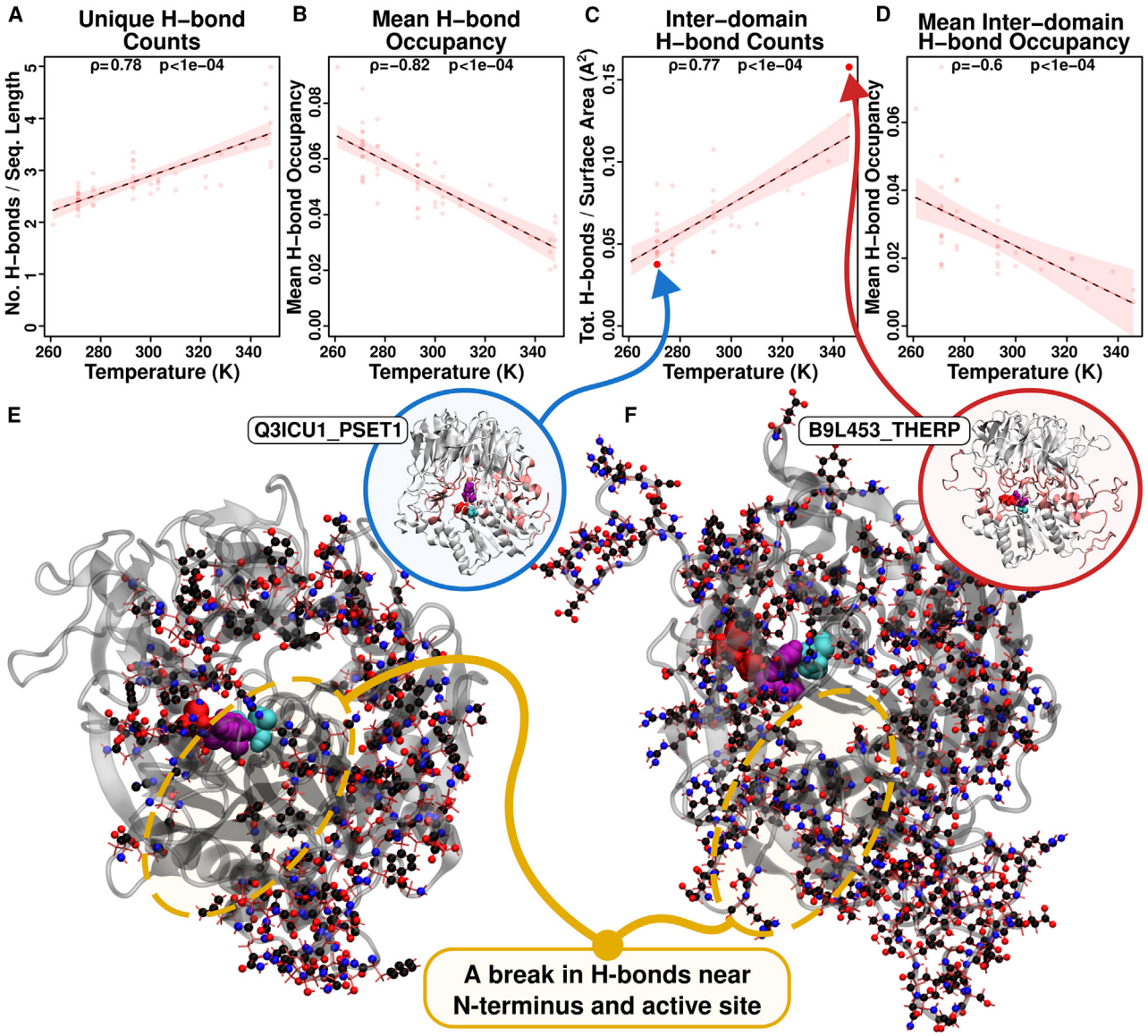

FIGURE 7 The frequency and occupancy of H-bonds as calculated in VMD. (*A*) Frequency of unique H-bonds by OET. Frequency is the total count of unique H-bonds normalized by sequence length. (*B*) Mean occupancy of all unique H-bonds in each protein as a function of OET, calculated as fraction of frames in which the H-bonds is occupied. (*C*) Frequency of H-bonds containing a residue in each domain, normalized by the inter-domain surface area calculated in VMD. (*D*) Mean occupancy of H-bonds spanning the two domains. OLS fits shown by central lines; shaded area indicates 95% confidence bands. (*E*) Psychrophile Q3ICU1_PSET1 is used to allow for comparison with SB patterns, although it does not have the minimum H-bond frequency. An overlay shows the protein backbone as ribbons with light red coloring indicating a residue found in an inter-domain H-bond. A larger structure is drawn as in Fig. 6 *E*, with residues that are found in H-bonds drawn as lines with spheres representing heavy atoms. (*F*) Thermophile B9L453_THERP has the highest frequency of inter-domain H-bonds. Illustration style matches that of (*E*). Yellow ovals highlight a decreased number of inter-domain H-bonds near the active-site region.

point for substrate to access the active site, and it may also possibly promote the unfolding of substrates by decoupling their salt bridges and H-bonds, as hypothesized by Kiss-Szemán et al. (72). The increased density of salt bridges containing arginine around the active site could also provide an anchor that prevents the N terminus from unfolding away from the protein any more than necessary, maintaining a shield over the aggregation-prone beta-edge of the hydrolase core β sheets (4) but still allowing slight local unfolding to admit the substrate.

Salt bridges compensate for increased fluctuations and maintain inter-domain interactions

The function and specificity of POP enzymes require that the two domains separate enough to allow substrate to reach the active site but not so much as to cause unfolding. Given the location of the active site at the inter-domain interface, it is worthwhile to understand how stable the environment is surrounding the catalytic residues and whether

such an environment is affected by OET. Fig. 8 shows a comparison between the mean density of salt bridges across the entire protein and the mean density of salt bridges that span the two domains. Here, density is defined as the fraction of potential salt bridges that are occupied (the two residues of each pair are within 5.4 Å) in each frame of the 100-ns simulation. The density of salt bridges throughout the protein decreases significantly with temperature. By contrast, when focusing only on the inter-domain salt bridges, we find no change with temperature. This consistency in interactions between the domains, despite the structural fluctuations in the rest of the protein, strongly suggests the importance of maintaining the local environment of the inter-domain region.

Similarly, when we examine the mean duration of each salt bridge (i.e., the number of consecutive frames in which a salt bridge is occupied), we observe a significant decrease in salt-bridge duration with temperature. Combining this result with the results for mean occupancy from Fig. 6 tells us that the salt bridges in thermophiles are not simply being occupied fewer times throughout the simulation; the occupancy also lasts a significantly shorter amount of time. The combination of a decrease in occupancy, duration, and density of salt bridges with an increase in the proportion of salt bridges suggests a hypothesis describing the role of salt bridges in holding the protein together. The protein appears to be adapted to make up for the rapid decrease in intramolecular interactions by increasing the proportion of charged residues. This compensation (also noted above) is globally imperfect but is well maintained for the inter-domain interface. This may stem from the lack of a decline in salt-bridge duration for salt bridges spanning the inter-domain region. This again points to the importance of maintaining the local environment between the two domains. Due to compensation, the density of salt bridges between the two domains does not vary significantly by temperature, meaning there are the same number of occupied salt bridges per surface area of the domain interface regardless of thermal group, whereas the overall occupancy of inter-domain salt bridges significantly decreases (Fig. 6 *D*). Thus, the larger number of salt bridges must rapidly "take turns" to maintain the local structure as the protein responds to thermal fluctuations.

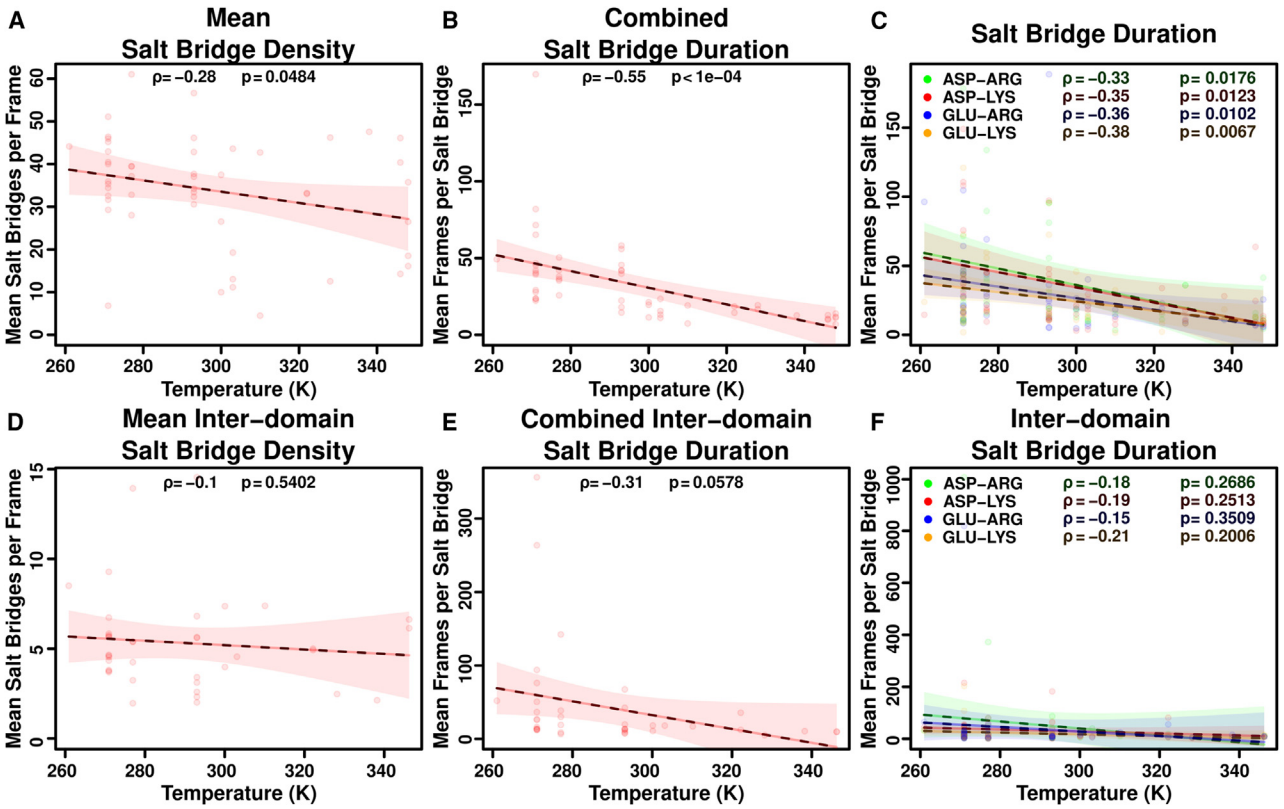

FIGURE 8 The density and duration of SBs throughout each protein, compared with those interactions between the two domains. (*A*) SB density is the number of SB pairs that are occupied (residue centers of mass are within 5.4 Å) during a single frame, averaged over all 5000 frames, and shows a significant decrease with OET. (*B*) Duration is the number of consecutive frames in which an SB is occupied, averaged across all SBs, and shows a significant decrease with OET. (*C*) SB duration is measured for each SB pair type, showing significant decreases for each with OET. (*D*) When only considering SBs that span the domain interface, SB density is not observed to be dependent on OET. (*E*) Inter-domain SBs are observed to have similar duration across OET. (*F*) When broken down by SB pair type, there is no difference observed in the duration of inter-domain SBs with OET. The interactions between the propeller domain and the catalytic domain are maintained across temperatures, despite significant changes in interaction profiles throughout the rest of the protein structure. OLS fits shown by central lines; shaded area indicates 95% confidence bands.

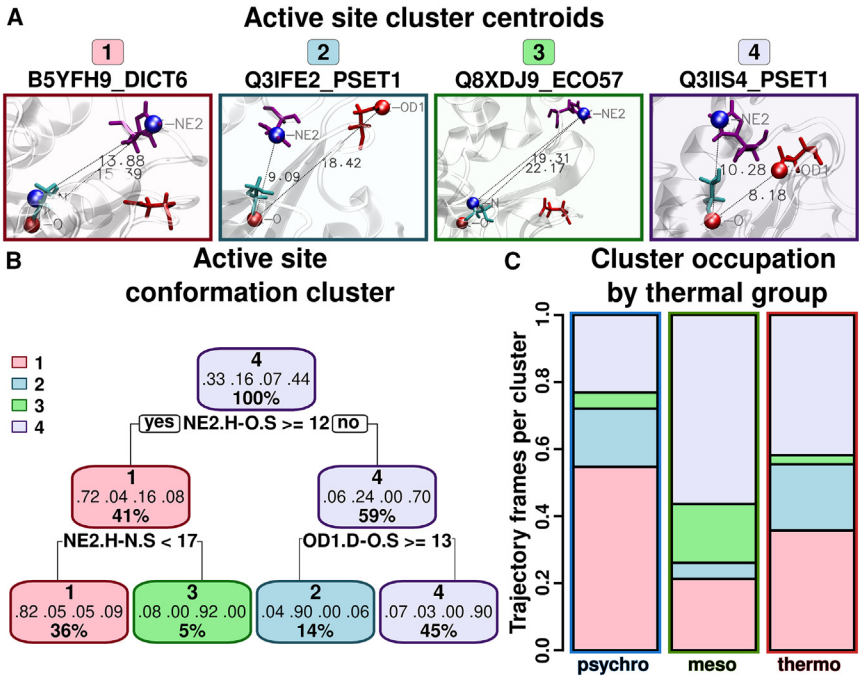

FIGURE 9 Active site conformations for representative S9 proteases. (*A*) Active-site conformations for the centroid of each conformation cluster; the catalytic residues (S, D, and H) shown in licorice representation, with larger spheres highlighting atoms selected by classification tree to distinguish conformations. Distances are in angstroms and were calculated by VMD. (*B*) Classification tree to determine cluster membership using differences in interatomic distances. Nodes are colored and labeled by the dominant cluster for points within the node, followed by the fractions of points from each cluster within the node, as well as the fraction of the total dataset found within the node. A small number of key distances approximately characterize the cluster states. (*C*) Stacked bars show the proportion of frames from each thermal group that sample each of the four active-site conformation clusters, as depicted by their respective centroids in (*A*).

POP active sites have four characteristic active-site conformations: Open, closed, and two intermediate states

The stability provided by salt bridges and H-bonds in the region between the two domains is likely to have an effect on the range of conformations available to the catalytic residues. We analyzed the range of conformational states of the active site by clustering of deep random feature embedding based on distances between noncarbon heavy atoms of the catalytic residue side chains and backbones that were observed over 5000 frames of each 100-ns simulation (see Figs. S17 and S18 for further information). This resulted in four main conformation clusters, the centroids of which are depicted in Fig. 9 *A* with distances labeled in angstroms using the measuring tools in VMD. The protein to which the centroid conformation belongs is indicated above each frame. Each thermal group is represented by at least one cluster centroid, with psychrophiles having representative conformations in both clusters 2 and 4.

The dendrogram in Fig. 9 *B* describes the distinguishing features that differentiate the four clusters. It should be noted that the first branch is determined by the distance between the reactive N of the catalytic His and the backbone O of the catalytic Ser. This distance is also the distinguishing feature when clustering is performed with only those proteins with propellers, as well as with only those proteins without propellers (see Fig. S19), indicating that all proteins in this set have two major active-site conformations regardless of the presence of the propeller: an open conformation and a closed conformation. The precise distances used as cutoffs for branches of the dendrogram are specific to the given data and clustering method, and may not be generalizable, but indicate the relative size of boundaries between clusters describing the dominant conformations that are observed.

Comparison of the centroid conformations in Fig. 9 *A* with the dendrogram of Fig. 9 *B* shows that the two clusters represented by psychrophiles, 2 and 4, tend to have closed active-site conformations. These two clusters are then further distinguished by the distance between the Asp side chain and the catalytic Ser, where cluster 2 contains the conformation with the Asp at a further distance from Ser than cluster 4 conformations, in which the Asp is nearer the Ser. Worth noting is that the two psychrophile proteins depicted as the centroids are from the same organism but are very different in size (see Fig. S20). Q3IFE2_PSET1 is one of the largest proteins in the set, with 903 residues, whereas Q3IIS4_PSET1 has a length closer to the median (632 residues). The additional 271 resides of Q3IFE2_PSET1 are found in unstructured loops that extend from the β sheets of the propeller, as well as from the α helices of the catalytic domain. These longer loops mingle in the inter-domain region and potentially result in more interactions with the catalytic residues that pull them out of alignment for proper functioning. By contrast, Q3IIS4_PSET1 has few additional loops hanging from the propeller domain, limiting the interactions that might perturb the catalytic residues. The "open/closed" state of the monomer has been shown to be linked to the geometry of the catalytic triad

(73); however, the influence of the inter-domain loops is not clear.

Clusters 1 and 3 are represented by the thermophilic B5YFH9_DICT6 and the mesophilic Q8XDJ9_ECO57, respectively, neither of which carries a propeller domain. Q8XDJ9_ECO57 is the smallest protein of the set, with only 249 residues, whereas B5YFH9_DICT6 is only slightly longer with 256 residues. Both have a larger distance between the catalytic His and Ser, distinguishing them from clusters 2 and 4. Clusters 1 and 3 are further distinguished from each other by an additional measurement of the distance between the catalytic His and the backbone of the catalytic Ser, where cluster 3, represented by a mesophilic enzyme, has the most open active-site conformation. This open conformation accounts for only 5% of the 5000 frames from each of 51 proteins, indicating that this state is rarely visited.

Fig. 9 *C* shows the relative proportion of simulation frames from each thermal group that are found in each active-site conformation cluster. Proteins from psychrophiles are most commonly found in cluster 1, followed by clusters 4, 2, and 3, respectively. This is counterintuitive, but the similarity between thermophilic and psychrophilic structures may be indicative of the stabilizing effect of the large numbers of salt bridges and H-bonds in thermophiles. Although cluster 1 is represented by a thermophile, it describes an open active-site conformation in which the distance between the catalytic His and Ser is constrained to be less than 17 Å, as shown in Fig. 9 *B*. Mesophilic enzymes, on the other hand, sample cluster 4 most frequently, followed by clusters 1, 3, and 2, respectively. Cluster 4 is described by the most closed active site, whereas clusters 1 and 3 are more open. This describes a set of proteins that sample a large range of active-site conformations and have few constraints on their structure. Thermophiles, like psychrophiles, favor clusters 4 and 1, followed by cluster 2, and very few samples of cluster 3.

It is interesting to note that the more open conformation of cluster 3 is sampled more frequently by mesophiles rather than thermophiles, as one might expect a thermophilic protein to visit a larger range of conformations than a mesophilic counterpart. This discrepancy points to the effect of additional salt bridges and H-bonds in the thermophilic proteins, which may provide stability by increasing the overall cohesion of the protein, damping the structural fluctuations caused by the increased thermal energy. Mesophiles, on the other hand, are not constrained by either extreme low temperature or excess salt bridges and H-bonds. Alternatively, mesophiles may benefit from their relative abundance of hydrophobic residues, which may provide the necessary structural buffer against thermal fluctuation by utilizing the hydrophobic effect but still allowing for flexibility.

DISCUSSION

Hypotheses about extremophilic enzymes

One of the objectives of this study is to investigate which, if any, previously reported hypotheses about adaptations in proteins adapted to different temperatures are applicable to S9 proteases. Table 4 summarizes literature findings for thermal adaptation, which represent a set of hypotheses that may be tested on a set of proteins, along with a comparison with our findings for the S9 proteases. A common hypothesis suggests that enzymes from psychrophiles are more rigid, to prevent cold denaturation due to the hydrophobic effect being less effective at low temperatures (74). The hydrophobic effect is also hypothesized to increase packing in thermophiles and hyperthermophiles by creating a more compact hydrophobic core (75–79). Berezovsky and Shakhnovich describe this mode of thermostabilization as structure based, whereas sequence-based thermostability, which is more commonly found in bacterial species, may be observed as substitutions that affect polar interactions between residues. We did not observe any trend in rigidity or packing of POPs with changes in temperature, even when testing the propeller and catalytic domains separately (see Figs. S21–S23). The reason for this is intuitive in the case of S9 proteases: the structure of either domain is largely dependent on the secondary structures of β sheets and α helices, leaving little room for modifications that would disrupt those secondary structures from forming. The hypothesis of increased packing due to the hydrophobic effect also does not hold for the S9 proteases for similar reasons: the propeller domain lacks a globular core, consistent with a lack of structure-based thermostability. Additionally, although we do see an increasing trend of hydrophobic residues with temperature (Fig. 5), the trend is driven by the increase in hydrophobic residues in *mesophiles* compared to psychrophiles, with thermophiles having similar proportions of hydrophobic residues to psychrophiles.

Other researchers have suggested that thermophilic enzymes are more rigid at room temperature than their mesophilic homologs (and are equally flexible at their target temperatures) (84). This can be accomplished with fewer, shorter, and more rigid loops(74, 76–78). This set of S9 proteases does show a significant decrease in loop length with temperature (see Fig. S24). The majority of loops appear between the two domains as extensions of the β turns in the propeller domain. However, longer loops on psychrophilic enzymes may function to create a local environment that encourages substrate to interact with the active site, increasing the rate of catalytic activity in a thermal environment that inherently slows the rate of interactions between molecules. This function would not be necessary in thermophiles, and it could possibly be disadvantageous by increasing the conformational space, thereby encouraging local unfolding events or unfavorable interactions.

TABLE 4 Selection of literature claims regarding protein structure vs. temperature

References	Examined proteins	Results	Comparison with current study
Number of Specific Amino Acid Types			
Kannan and Vishveshwara (80)	24 meso. and thermo. homologs	increase of aromatic networks/ clusters w/increased temp.	disagrees
Vieille and Zeikus (81)	eight meso. and seven hyper-thermo. organisms	increase EGIKPRVWY w/increased temp.	disagrees
Kumar et al. (82)	six each psychro., meso., and thermophilic β-D-galactosidases	more AGSR in psychro., more VQEFTY in thermo.	disagrees
Berezovsky et al. (66)	triosephosphate isomerase	more IVYWREL in thermophiles	disagrees
Packing and Flexibility			
Karshikoff and Ladenstein (83)	80 meso. proteins and 24 thermo. proteins	packing density similar between meso. and thermo	agrees
Radestock and Gohlke (84)	19 meso./thermo. protein homologs	increased rigidity in thermophiles	disagrees
Wells et al. (85)	citrate synthase	increased rigidity in thermophiles	disagrees
Amadei et al. (86)	57 thermo./meso. pairs	decreased density w/ increased temp.	disagrees
Sen and Sarkar (87)	17 thermo./meso. and 18 psychro./meso. pairs	no difference in average packing	agrees
Disulfide Bonds			
Appleby et al. (88)	5′-deoxy-5'-methylthioadenosine phosphorylase *Solfolobus solfataricus*	disulfide bonds increase thermal stability	disagrees
Weak Interactions and Salt Bridges			
Szilàgyi and Zàvodszky (89)	64 meso., 29 thermo. homologs	more ion pairs w/ higher growth temp.	agrees
D'Amico et al. (90)	psychro. α-amylase	decreased weak interactions in psychrophiles	agrees
Diessner et al. (13)	S11 proteases	more salt bridges in thermophiles	agrees
Chan et al. (91)	thermo. ribosomal protein L30e	increased salt bridges stabilize thermophiles	agrees

Related to the idea that thermophilic enzymes are more rigid, some researchers have observed that thermophilic enzymes have more disulfide bonds (75,76). Although that is true for some proteins, it is not relevant for these enzymes, all of which lack disulfide bonds. In contrast, the observation that many thermophilic enzymes have more salt bridges (13,74–76,78), particularly at interfaces (92), is supported. Here, the increase in the number of salt bridges appears to balance reduced occupancy in enzymes from organisms with higher growth temperatures. Similarly, we observe an increase in the number of H-bonds with temperature, a trend that is also supported in the literature (76,74) and that may also contribute to stability by providing additional points of interaction to prevent local unfolding in turbulent thermal environments. When it comes to aromatic residues and π-π interactions, however, the S9 family exhibits no trend in the proportion of aromatic residues with temperature (Fig. 5), and it even shows a significant decrease in the proportion of tyrosine with increased OET. This differs from findings on the contribution of aromatic residues and π-π interactions toward stability in thermophiles (74–76,80).

Conventional wisdom on protein adaptation to extreme thermal environments may not generalize to all enzyme classes

The most striking result from these analyses is the observation that, although the total number of H-bonds and salt bridges increases with OET, the occupancy for any given bond falls with temperature in a way that tends to preserve the expected number of bonds active at a given time. This can be understood in terms of redundancy of interactions preventing unraveling, or stepwise unfolding starting from local structural fluctuations. Protein unfolding is often considered to follow a two-state model characterized by an abrupt transition from the folded to the unfolded state, as is often observed for small proteins. However, experimental data show that the unfolding behavior of many larger or multi-domain proteins is inconsistent with this model. Protein folding is not an all-or-nothing proposition, and, in some cases, unfolding proceeds via a series of pre-molten globule and molten globule states with varying degrees of hydration (93,94). Because proteins are held together by

different interactions, there are also different unfolding modes specific to particular structures. Parts of the protein can exist in native-like states, whereas other parts are unfolded (95). For example, this has been observed in the structural crystallins of vertebrate eye lenses, which are highly stable proteins characterized by a two-domain double Greek key structure. In these proteins, the mechanism of unfolding is complex and critically depends on the solution conditions. Thermodynamically, the N-terminal domain is less stable than the C-terminal domain (96,97), and it unfolds first under conditions where the inter-domain interface is disrupted. However, the C-terminal domain unfolds first under milder denaturing conditions (98). MD simulations suggest that domain swapping of the last three C-terminal β strands may be an important early intermediate on the pathway to cataract formation (99). We hypothesize that, for the S9 family, thermal adaptations have focused on an *entropic* rather than *energetic* stability strategy: rather than increase the number and/or strength of bound interactions in cross-section (and hence the energy of unfolding), thermophilic S9s are adapted to have larger numbers of dynamically available bound interactions (reducing the relative number of conformational states in which all bonds are broken at once). One advantage of this entropic adaptation may be the ability to maintain similar levels of local flexibility across a wide temperature range while still maintaining resistance to unfolding at high temperatures. This may be particularly important for organisms that experience large variations in temperature over their lifespans or whose lineages may frequently be swept into micro-environments with very different temperatures (thus exposing them to large variations in temperature over evolutionary time): such variable conditions are common among extremophiles (100), and may select for "soft" thermal adaptations that are easily reversible (e.g., larger numbers of dynamically exchanging H-bonds) rather than "hard" adaptations that impose all-or-nothing constraints on structure and dynamics (e.g., disulfide bonds). If so, differences in selection for variable versus consistent thermal conditions may potentially explain some of the divergent observations in the literature on thermal adaptation.

CONCLUSIONS

Comparative modeling of S9 proteases from selected organisms that thrive in different thermal environments suggests that enhanced stability in thermophilic S9 proteases is driven by an increase in the number of salt bridges and H-bonds compensating for the lower occupancy of each one. This suggests a model of protein stability where the dominant failure mode is not catastrophic unfolding but rather local unraveling that can be protected against by having a multitude of individually weak interactions, similar to the strength of Kevlar and other nanocomposite materials (101,102). This strategy of "entropic stabilization" complements more widely appreciated, energetic stabilization strategies such as the use of disulfide bonds and may represent a broader category of thermal adaptations of particular relevance for populations exposed to widely varying conditions over evolutionary time.

This study also illustrates the value of comparative MD simulations from large ensembles of proteins to examine general patterns of structure and dynamics in unusual thermal environments. Further comparisons incorporating systematic variation of salt concentration, pH, and other variables are also potentially insightful but, at present, are limited by the high cost of long atomistic trajectories in explicit solvent. As ongoing improvements in both GPU (graphics processing unit) technology and methods for accelerated forcefield calculations accumulate, it is hoped that the range of feasible comparative studies will correspondingly expand. Likewise, although seeding MD simulations with predicted structures allows for systematic sampling of protein populations, it also depends on the accuracy of structural predictions. Further systematic experimental studies of extremophilic enzymes are thus important both to validate model-based studies and to provide improved accuracy for future studies of this kind.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2024.07.013.

AUTHOR CONTRIBUTIONS

C.T.B. and R.W.M. designed the study. G.R.T. retrieved data. C.T.B. and E.M.D. wrote code for simulation and analysis. E.M.D. carried out all simulations. E.M.D., G.R.T., R.W.M., and C.T.B. analyzed the data. E.M.D. drafted the initial manuscript. All authors contributed to editing the article.

ACKNOWLEDGMENTS

This work was supported by NASA award 80NSSC20K0620 to R.W.M. and C.T.B. G.R.T. was supported by GAAN P200A210024.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Rawlings, N. D., M. Waller, …, A. Bateman. 2014. MEROPS: The Database of Proteolytic Enzymes, Their Substrates and Inhibitors. *Nucleic Acids Res.* 42:D503–D509.

2. Venäläinen, J. I., R. O. Juvonen, and P. T. Männistö. 2004. Evolutionary Relationships of the Prolyl Oligopeptidase Family Enzymes: Evolutionary Relationships of the POP Family Enzymes. *Eur. J. Biochem.* 271:2705–2715.

3. Rawlings, N. D., L. Polgár, and A. J. Barrett. 1991. A New Family of Serine-Type Peptidases Related to Prolyl Oligopeptidase. *Biochem. J.* 279:907–908.
4. Kiss-Szemán, A. J., V. Harmat, and D. K. Menyhárd. 2019. Achieving Functionality Through Modular Build-up: Structure and Size Selection of Serine Oligopeptidases. *Curr. Protein Pept. Sci.* 20:1089–1101.
5. Polgár, L. 2002. The Prolyl Oligopeptidase Family. *Cell. Mol. Life Sci.* 59:349–362.
6. Shan, L., T. Marti, …, C. Khosla. 2004. Comparative Biochemical Analysis of Three Bacterial Prolyl Endopeptidases: Implications for Coeliac Sprue. *Biochem. J.* 383:311–318.
7. Kaushik, S., and R. Sowdhamini. 2014. Distribution, Classification, Domain Architectures and Evolution of Prolyl Oligopeptidases in Prokaryotic Lineages. *BMC Genom.* 15:985.
8. Yadav, P., V. D. Goyal, …, R. D. Makde. 2019. Carboxypeptidase in prolyl oligopeptidase family: Unique enzyme activation and substrate-screening mechanisms. *J. Biol. Chem.* 294:89–100.
9. Fülöp, V., Z. Böcskei, and L. Polgár. 1998. Prolyl Oligopeptidase: An Unusual β-Propeller Domain Regulates Proteolysis. *Cell.* 94:161–170.
10. Shan, L., I. I. Mathews, and C. Khosla. 2005. Structural and Mechanistic Analysis of Two Prolyl Endopeptidases: Role of Interdomain Dynamics in Catalysis and Specificity. *Proc. Natl. Acad. Sci. USA.* 102:3599–3604.
11. Li, M., C. Chen, …, T. K. Chiu. 2010. Induced-Fit Mechanism for Prolyl Endopeptidase. *J. Biol. Chem.* 285:21487–21495.
12. Ellis-Guardiola, K., H. Rui, …, J. C. Lewis. 2019. Crystal Structure and Conformational Dynamics of Pyrococcus Furiosus Prolyl Oligopeptidase. *Biochemistry.* 58:1616–1626.
13. Diessner, E. M., G. Takahashi, …, C. T. Butts. 2023. Comparative Modeling and Analysis of Extremophilic D-Ala-D-Ala Carboxypeptidases. *Biomolecules.* 13:328.
14. Beeby, M., B. D. O'Connor, …, T. O. Yeates. 2005. The Genomics of Disulfide Bonding and Protein Stabilization in Thermophiles. *PLOS Bio.* 3:e309.
15. Ladenstein, R., and B. Ren. 2006. Protein Disulfides and Protein Disulfide Oxidoreductases in Hyperthermophiles. *FEBS J.* 273:4170–4185.
16. Médigue, C., E. Krin, …, A. Danchin. 2005. Coping with Cold: The Genome of the Versatile Marine Antarctica Bacterium Pseudoalteromonas Haloplanktis TAC125. *Genome Res.* 15:1325–1335.
17. Copeland, A., S. Lucas, …, P. Richardson. 2007. Complete Sequence of Psychromonas Ingrahamii 37, Technical report.
18. Copeland, A., S. Lucas, …, P. Richardson. 2006. Complete Sequence of Shewanella Frigidimarina NCIMB 400, Submitted (Aug. Sept. 2006). Technical report, Released 09/14/2006 by the DOE Joint Genome Institute.
19. Parmeciano Di Noto, G., S. C. Vázquez, …, C. Quiroga. 2016. Draft Genome of Shewanella Frigidimarina Ag06-30, a Marine Bacterium Isolated from Potter Peninsula, King George Island, Antarctica. *Genome Announc.* 4:e00289-16.
20. Kunst, F., N. Ogasawara, …, A. Danchin. 1997. The Complete Genome Sequence of the Gram-Positive Bacterium Bacillus Subtilis. *Nature.* 390:249–256.
21. Makino, K., K. Ishii, …, H. Shinagawa. 1998. Complete Nucleotide Sequences of 93-Kb and 3.3-Kb Plasmids of an Enterohemorrhagic Escherichia Coli O157:H7 Derived from Sakai Outbreak. *DNA Res.* 5:1–9.
22. Hayashi, T., K. Makino, …, H. Shinagawa. 2001. Complete Genome Sequence of Enterohemorrhagic Escherichia Coli O157:H7 and Genomic Comparison with a Laboratory Strain K-12. *DNA Res.* 8:11–22.
23. Silby, M. W., A. M. Cerdeño-Tárraga, …, N. R. Thomson. 2009. Genomic and Genetic Analyses of Diversity and Plant Interactions of Pseudomonas Fluorescens. *Genome Biol.* 10:R51.
24. van de Werken, H. J. G., M. R. A. Verhaart, …, S. W. M. Kengen. 2008. Hydrogenomics of the Extremely Thermophilic Bacterium Caldicellulosiruptor Saccharolyticus. *Appl. Environ. Microbiol.* 74:6720–6729.
25. Copeland, A., S. Lucas, …, P. Richardson. 2007. Genome Sequence of the Thermophilic Hydrogen-Producing Bacterium Caldicellulosiruptor Saccharolyticus DSM 8903, Technical report, Released 09/14/2006 by the DOE Joint Genome Institute.
26. Coil, D. A., J. H. Badger, …, J. A. Eisen. 2014. Complete Genome Sequence of the Extreme Thermophile Dictyoglomus Thermophilum H-6-12. *Genome Announc.* 2:e00109-14. https://doi.org/10.1128/genomea.00109–14.
27. Takami, H., Y. Takaki, …, I. Uchiyama. 2004. Thermoadaptation Trait Revealed by the Genome Sequence of Thermophilic Geobacillus Kaustophilus. *Nucleic Acids Res.* 32:6292–6303.
28. Ng, W. V., S. P. Kennedy, …, S. DasSarma. 2000. Genome Sequence of Halobacterium Species NRC-1. *Proc. Natl. Acad. Sci. USA.* 97:12176–12181.
29. Wu, D., J. Raymond, …, J. A. Eisen. 2009. Complete Genome Sequence of the Aerobic CO-Oxidizing Thermophile Thermomicrobium Roseum. *PLoS One.* 4:e4207.
30. Masui, R., K. Kurokawa, …, Kuramitsu. 2022. Complete Genome Sequence of Thermus Thermophilus HB8.
31. The UniProt Consortium. 2023. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* 51:D523–D531.
32. Peters, C., K. D. Tsirigos, …, A. Elofsson. 2016. Improved Topology Prediction Using the Terminal Hydrophobic Helices Rule. *Bioinformatics.* 32:1158–1162.
33. Teufel, F., J. J. Almagro Armenteros, …, H. Nielsen. 2022. SignalP 6.0 Predicts All Five Types of Signal Peptides Using Protein Language Models. *Nat. Biotechnol.* 40:1023–1025.
34. Jones, P., D. Binns, …, S. Hunter. 2014. InterProScan 5: Genome-Scale Protein Function Classification. *Bioinformatics.* 30:1236–1240.
35. Szekely, G. J., and M. L. Rizzo. 2005. Hierarchical Clustering via Joint Between-Within Distances: Extending Ward's Minimum Variance Method. *J. Classif.* 22:151–183.
36. Sievers, F., A. Wilm, …, D. G. Higgins. 2011. Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* 7:539.
37. Crooks, G. E., G. Hon, …, S. E. Brenner. 2004. WebLogo: A Sequence Logo Generator. *Genome Res.* 14:1188–1190.
38. Yang, J., and Y. Zhang. 2015. I-TASSER Server: New Development for Protein Structure and Function Predictions. *Nucleic Acids Res.* 43:W174–W181.
39. Olsson, M. H. M., C. R. Søndergaard, …, J. H. Jensen. 2011. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theor. Comput.* 7:525–537.
40. Huang, P., A. Lv, …, S. Yang. 2022. The structure and molecular dynamics of prolyl oligopeptidase from *Microbulbifer arenaceous* provide insights into catalytic and regulatory mechanisms. *Acta Crystallogr.* 78:735–751.
41. Abendroth, J., B. Sankaran, …, B. Staker. 2012. Structure of a Post-proline Cleaving Enzyme from *Rickettsia Typhi*, *To be published*.
42. Meng, E. C., T. D. Goddard, …, T. E. Ferrin. 2023. UCSF ChimeraX: Tools for Structure Building and Analysis. *Protein Sci.* 32:e4792.
43. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* 14:33–38.
44. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 79:926–935.
45. Phillips, J. C., D. J. Hardy, …, E. Tajkhorshid. 2020. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem. Phys.* 153:044130.

46. Huang, J., S. Rauscher, …, A. D. MacKerell. 2017. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods.* 14:71–73.

47. Martyna, G. J., D. J. Tobias, and M. L. Klein. 1994. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* 101:4177–4189.

48. Feller, S. E., Y. Zhang, …, B. R. Brooks. 1995. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* 103:4613–4621.

49. Raftery, A. E., and S. M. Lewis. 1992. How Many Iterations in the Gibbs Sampler? *In* Bayesian Statistics, *4*. J. M. Bernardo, J. O. Berger, and …, A. F. M. Smitheds. Oxford University Press, Oxford, pp. 763–773.

50. Plummer, M., N. Best, …, K. Vines. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R. News.* 6:7–11. https://journal.r-project.org/archive/.

51. Touw, W. G., C. Baakman, …, G. Vriend. 2015. A Series of PDB-Related Databanks for Everyday Needs. *Nucleic Acids Res.* 43:D364–D368.

52. Pacios, L. F. 2001. Distinct Molecular Surfaces and Hydrophobicity of Amino Acid Residues in Proteins. *J. Chem. Inf. Comput. Sci.* 41:1427–1435.

53. Benson, N. C., and V. Daggett. 2012. A Chemical Group Graph Representation for Efficient High-Throughput Analysis of Atomistic Protein Simulations. *J. Bioinf. Comput. Biol.* 10:1250008.

54. Butts, C. T., X. Zhang, …, R. W. Martin. 2016. Sequence Comparison, Molecular Modeling, and Network Analysis Predict Structural Diversity in Cysteine Proteases from the Cape Sundew, *Drosera Capensis*. *Comput. Struct. Biotechnol. J.* 14:271–282.

55. Unhelkar, M. H., V. T. Duong, …, R. W. Martin. 2017. Structure Prediction and Network Analysis of Chitinases from the Cape Sundew, Drosera Capensis. *BBA - Gen. Subjects.* 1861:636–643.

56. Butts, C. T. 2008. Social Network Analysis with Sna. *J. Stat. Software.* 24:1–51.

57. Butts, C. T. 2008. **Network** : A Package for Managing Relational Data in *R*. *J. Stat. Software.* 24.

58. R Core Team. 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

59. Grant, B. J., L. Skjærven, and X.-Q. Yao. 2021. The Bio3D Packages for Structural Bioinformatics. *Protein Sci.* 30:20–30.

60. Cho, Y., and L. K. Saul. 2011. Analysis and Extension of Arc-Cosine Kernels for Large Margin Classification. Preprint at arXiv. https://doi.org/10.48550/arXiv.1112.3712.

61. Eddelbuettel, D., and R. Francois. 2011. Rcpp: Seamless R and C++ Integration. *J. Stat. Software.* 40:1–18.

62. Grazioli, G., R. W. Martin, and C. T. Butts. 2019. Comparative Exploratory Analysis of Intrinsically Disordered Protein Dynamics Using Machine Learning and Network Analytic Methods. *Front. Mol. Biosci.* 6:42.

63. Therneau, T., B. Atkinson; B. R. p. o. t. i. R. port, and maintainer 1999-2017. 2022. Rpart: Recursive Partitioning and Regression Trees.

64. Vauclare, P., F. Natali, …, B. Franzetti. 2020. Surviving salt fluctuations: stress and recovery in *Halobacterium salinarum*, an extreme halophilic Archaeon. *Sci. Rep.* 10:3298:Article Number 3298.

65. Pereira, J., and A. N. Lupas. 2022. New β-Propellers Are Continuously Amplified From Single Blades in All Major Lineages of the β-Propeller Superfamily. *Front. Mol. Biosci.* 9:895496.

66. Berezovsky, I. N., K. B. Zeldovich, and E. I. Shakhnovich. 2007. Positive and Negative Design in Stability and Thermal Adaptation of Natural Proteins. *PLoS Comput. Biol.* 3:e52.

67. Zeldovich, K. B., I. N. Berezovsky, and E. I. Shakhnovich. 2007. Protein and DNA Sequence Determinants of Thermophilic Adaptation. *PLoS Comput. Biol.* 3:e5.

68. Berezovsky, I. N., W. W. Chen, …, E. I. Shakhnovich. 2005. Entropic Stabilization of Proteins and Its Proteomic Consequences. *PLoS Comput. Biol.* 1:e47–e0332.

69. Bierma, J. C., K. W. Roskamp, …, R. W. Martin. 2018. Controlling Liquid–Liquid Phase Separation of Cold-Adapted Crystallin Proteins from the Antarctic Toothfish. *J. Mol. Biol.* 430:5151–5168.

70. Hong, Y., S. Najafi, …, D. S. Hwang. 2022. Hydrophobicity of Arginine Leads to Reentrant Liquid-Liquid Phase Separation Behaviors of Arginine-Rich Proteins. *Nat. Commun.* 13:7326.

71. Amangeldina, A., Z. W. Tan, and I. N. Berezovsky. 2024. Living in Trinity of Exteremes: Genomic and Proteomic Signatures of Halophilic, Thermophilic, and pH Adaptation. *Curr. Res. Str. Biol.* 7

72. Kiss-Szemán, A. J., P. Stráner, …, A. Perczel. 2022. Cryo-EM Structure of Acylpeptide Hydrolase Reveals Substrate Selection by Multimerization and a Multi-state Serine-protease Triad. *Chem. Sci.* 13:7132–7142.

73. Harmat, V., K. Domokos, …, L. Polgár. 2011. Structure and Catalysis of Acylaminoacyl Peptidase: Closed and Open Subunits of a Dimer Oligopeptidase. *J. Biol. Chem.* 286:1987–1998.

74. Goldstein, R. A. 2007. Amino-Acid Interactions in Psychrophiles, Mesophiles, Thermophiles, and Hyperthermophiles: Insights from the Quasi-Chemical Approximation. *Protein Sci.* 16:1887–1895.

75. Pucci, F., and M. Rooman. 2017. Physical and Molecular Bases of Protein Thermal Stability and Cold Adaptation. *Curr. Opin. Struct. Biol.* 42:117–128.

76. Kumar, S., A. K. Dangi, …, S. K. Khare. 2019. Thermozymes: Adaptive Strategies and Tools for Their Biotechnological Applications. *Bioresour. Technol.* 278:372–382.

77. Tych, K. M., M. Batchelor, …, L. Dougan. 2016. Differential Effects of Hydrophobic Core Packing Residues for Thermodynamic and Mechanical Stability of a Hyperthermophilic Protein. *Langmuir.* 32:7392–7402.

78. Robinson-Rechavi, M., A. Alibés, and A. Godzik. 2006. Contribution of Electrostatic Interactions, Compactness and Quaternary Structure to Protein Thermostability: Lessons from Structural Genomics of Thermotoga Maritima. *J. Mol. Biol.* 356:547–557.

79. Berezovsky, I. N., and E. I. Shakhnovich. 2005. Physics and Evolution of Thermophilic Adaptation. *Proc. Natl. Acad. Sci. USA.* 102:12742–12747.

80. Kannan, N., and S. Vishveshwara. 2000. Aromatic Clusters: A Determinant of Thermal Stability of Thermophilic Proteins. *Protein Eng.* 13:753–761.

81. Vieille, C., and G. J. Zeikus. 2001. Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Microbiol. Mol. Biol. Rev.* 65:1–43.

82. Kumar, V., N. Sharma, and T. C. Bhalla. 2014. In Silico Analysis of β-Galactosidases Primary and Secondary Structure in Relation to Temperature Adaptation. *J. Amino Acids.* 2014:e475839.

83. Karshikoff, A., and R. Ladenstein. 1998. Proteins from Thermophilic and Mesophilic Organisms Essentially Do Not Differ in Packing. *Protein Eng.* 11:867–872.

84. Radestock, S., and H. Gohlke. 2011. Protein Rigidity and Thermophilic Adaptation. *Proteins.* 79:1089–1108.

85. Wells, S. A., S. J. Crennell, and M. J. Danson. 2014. Structures of Mesophilic and Extremophilic Citrate Synthases Reveal Rigidity and Flexibility for Function. *Proteins.* 82:2657–2670.

86. Amadei, A., S. Del Galdo, and M. D'Abramo. 2018. Density Discriminates between Thermophilic and Mesophilic Proteins. *J. Biomol. Struct. Dyn.* 36:3265–3273.

87. Sen, S., and M. Sarkar. 2022. Insights on Rigidity and Flexibility at the Global and Local Levels of Protein Structures and Their Roles in Homologous Psychrophilic, Mesophilic, and Thermophilic Proteins: A Computational Study. *J. Chem. Inf. Model.* 62:1916–1932.

88. Appleby, T. C., I. I. Mathews, …, S. E. Ealick. 2001. Three-Dimensional Structure of a Hyperthermophilic 5'-Deoxy-5'-Methylthioadenosine Phosphorylase from *Sulfolobus solfataricus*. *J. Biol. Chem.* 276:39232–39242.

89. Szilágyi, A., and P. Závodszky. 2000. Structural Differences between Mesophilic, Moderately Thermophilic and Extremely Thermophilic

Protein Subunits: Results of a Comprehensive Survey. *Structure.* 8:493–504.

90. D'Amico, S., C. Gerday, and G. Feller. 2001. Structural Determinants of Cold Adaptation and Stability in a Large Protein. *J. Biol. Chem.* 276:25791–25796.
91. Chan, C.-H., T.-H. Yu, and K.-B. Wong. 2011. Stabilizing Salt-Bridge Enhances Protein Thermostability by Reducing the Heat Capacity Change of Unfolding. *PLoS One.* 6:e21624.
92. Ma, B.-G., A. Goncearence, and I. N. Berezovsky. 2010. Thermophilic Adaptation of Protein Complexes Inferred from Proteomic Homology Modeling. *Structure.* 18:819–828.
93. Pande, V., and D. Rokhsar. 1998. Is the Molten Globule a Third Phase of Proteins? *Proc. Natl. Acad. Sci. USA.* 95:1490–1494.
94. Gupta, M. N., and V. N. Uversky. 2023. Pre-Molten, Wet, and Dry Molten Globules En Route to the Functional State of Proteins. *Int. J. Mol. Sci.* 24:2424.
95. Regan, L. 2003. Molten Globules Move into Action. *Proc. Natl. Acad. Sci. USA.* 100:3553–3554.
96. Kosinski-Collins, M. S., a. K. S. L. Flaugh, and J. King. 2004. Probing Folding and Fluorescence Quenching in Human gammaD Crystallin Greek Key Domains Using Triple Tryptophan Mutant Proteins. *Protein Sci.* 13:2223–2235.
97. Mills, I. A., S. L. Flaugh, …, J. A. King. 2007. Folding and stability of the isolated Greek key domains of the long-lived human lens proteins gammaD-crystallin and gammaS-crystallin. *Protein Sci.* 16:2427–2444.
98. Mills-Henry, I. A., S. L. Thol, …, J. A. King. 2019. Kinetic Stability of Long-Lived Human Lens γ-Crystallins and Their Isolated Double Greek Key Domains. *Biophys. J.* 117:269–280.
99. Das, P., J. A. King, and R. Zhou. 2011. Aggregation of γ-Crystallins Associated with Human Cataracts via Domain Swapping at the C-Terminal β-Strands. *Proc. Natl. Acad. Sci. USA.* 108:10514–10519.
100. Brininger, C., S. Spradlin, …, C. Evilia. 2018. The more adaptive to change, the more likely you are to survive: Protein adaptation in extremophiles. *Semin. Cell Dev. Biol.* 84:158–169.
101. Tanner, D., J. A. Fitzgerald, and B. R. Phillips. 1989. The Keviar Story—an Advanced Materials Case Study. *Adv. Mater.* 1:151–156.
102. Guan, Y., W. Li, …, Y. Wang. 2017. Aramid Nanofibers and Poly (Vinyl Alcohol) Nanocomposites for Ideal Combination of Strength and Toughness via Hydrogen Bonding Interactions. *Compos. Sci. Technol.* 144:193–201.