PLOS | ONE

RESEARCH ARTICLE

Comparing Residue Clusters from Thermophilic and Mesophilic Enzymes Reveals Adaptive Mechanisms

Deanne W. Sammond[1], Noah Kastelowitz[2], Michael E. Himmel[1], Hang Yin[2], Michael F. Crowley[1], Yannick J. Bomble[1]*

1 Biosciences Center, National Renewable Energy Laboratory, Golden, Colorado, 80401, United States of America, **2** Department of Chemistry & Biochemistry and the BioFrontiers Institute, University of Colorado, Boulder, Colorado, 80309, United States of America

* Yannick.Bomble@nrel.gov

CrossMark
click for updates

OPEN ACCESS

Citation: Sammond DW, Kastelowitz N, Himmel ME, Yin H, Crowley MF, Bomble YJ (2016) Comparing Residue Clusters from Thermophilic and Mesophilic Enzymes Reveals Adaptive Mechanisms. PLoS ONE 11(1): e0145848. doi:10.1371/journal.pone.0145848

Editor: Gideon Schreiber, Weizmann Institute of Science, ISRAEL

Received: September 2, 2015

Accepted: December 9, 2015

Published: January 7, 2016

Copyright: © 2016 Sammond et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was funded by the US Department of Energy's Bioenergy Technologies Office (DOE-BETO), contract no. DE-AC36-08GO28308 with the National Renewable Energy Laboratory and the National Institutes of Health, NIH R01GM103843 (HY), F30CA180249 (NK).

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Understanding how proteins adapt to function at high temperatures is important for deciphering the energetics that dictate protein stability and folding. While multiple principles important for thermostability have been identified, we lack a unified understanding of how internal protein structural and chemical environment determine qualitative or quantitative impact of evolutionary mutations. In this work we compare equivalent clusters of spatially neighboring residues between paired thermophilic and mesophilic homologues to evaluate adaptations under the selective pressure of high temperature. We find the residue clusters in thermophilic enzymes generally display improved atomic packing compared to mesophilic enzymes, in agreement with previous research. Unlike residue clusters from mesophilic enzymes, however, thermophilic residue clusters do not have significant cavities. In addition, anchor residues found in many clusters are highly conserved with respect to atomic packing between both thermophilic and mesophilic enzymes. Thus the improvements in atomic packing observed in thermophilic homologues are not derived from these anchor residues but from neighboring positions, which may serve to expand optimized protein core regions.

Introduction

Enzymes have evolved to function in a wide range of conditions, including temperatures up to 130°C [1]. Highly thermostable enzymes are of industrial interest since performing processes at higher temperatures offers benefits such as decreased risk of contamination, increased substrate solubility and higher reaction rates [2]. Understanding adaptive evolutionary response under the selective pressure of high temperature promises to provide a set of rules that can impart desired thermostability to any target protein. In a broader sense, studying thermostable enzymes is important to better comprehend the evolutionary process as well as the energetics that drive protein folding, recognition and stability.

Studies comparing thermostable and mesostable enzymes have identified features associated with enhanced thermostability. Some features seen in thermophilic enzymes contribute to the stability of protein folding, such as the improved quality of packing [3, 4], improved electrostatic interactions [5–7] and increased hydrophobicity in the protein core [8, 9]. Other features diminish destabilizing forces such as decreased conformational flexibility [10–12] or entropy of unfolding [13]. The successes of protein engineering efforts based on these features support they are indeed mechanisms that can enhance thermostability [14, 15]. Often, however, these features are difficult to translate into actionable information that can direct protein-engineering efforts.

Proteins are stabilized by a network of cooperative interactions [16]. Altering one residue alters the local environment for neighboring residues, thus residues that are beneficial in one context can be deleterious in another. Xiao and Honig find electrostatic interactions are more favorable in hyperthermophilic proteins [17]. These ionic interactions are not identified by amino acid composition but instead depend on the location of the ionizable groups. Padgornaia et al. evaluate the limits of natural sequence variation by exhaustively mapping a small, defined region of the PhoQ-PhoP interface, finding functionally active sequence combinations that are not permissible individually [18]. Understanding how interacting residues adapt and evolve to achieve enhanced thermostability is an important step towards capturing beneficial epistatic substitutions. Targeted research aimed at improving our understanding of epistasis could aid the development of new rational protein engineering approaches.

Here we use structural bioinformatics to compare clusters of interacting residues from homologous thermophilic and mesophilic enzymes, allowing the comparison of interacting substitutions in structurally equivalent environments. Our approach is to identify what we call "motifs", or groups of residues that are adjacent in space and thus interacting, breaking the problem down to smaller context dependent units. By comparing motifs from homologous thermophilic and mesophilic enzymes we can better see sequence evolution within the local chemical environment. Using carefully matched and characterized proteins from different enzyme families we can observe specific trends within or across families. We place emphasis on cellulose enzymes as the absence of certain cellulase family members in thermophilic organisms raises the question of whether some protein folds are not well suited for thermostability [19]. Further, by comparing structurally equivalent clusters we are able to evaluate backbone alterations resulting from the evolutionarily selected mutations.

Results

We compare paired thermostable and mesostable homologues to investigate changes in local environment for groups of interacting residues. We select homologous enzymes covering structurally and catalytically distinct families (Table 1). Each family of enzymes contains a thermophilic or hyperthermophilic member as well as less thermostable homologues. The Carbohydrate-Active EnZymes database organizes enzymes into families based on structural similarities (CAZy, www.cazy.org) [20]. We select industrially relevant glycoside hydrolase (GH) families for investigation given the high interest research focused on the global need for alternative liquid fuel sources [21]. We add additional enzyme families from the published literature [8, 22–24].

Optimum organism growth temperature is a good indicator of enzyme activity temperature. However, the conformational flexibility needed for a given mechanism can also be a strong determinant of optimal activity temperature [25]. The questions investigated here rely on the accurate assignment of relative thermostability of homologous enzymes. Therefore, selected enzymes have experimentally determined structures and experimentally determined optimum

Table 1. Characteristics of the enzyme families selected for cluster analysis, with the most thermostable member highlighted in bold.

PDB	ORGANISM	T_{OPT} (°C)	% Sequence ID	RMSD	MW (kDa)
GH5 $(\beta/\alpha)_8$					
3amc	***Thermotoga maritima***	**80**	**100**	**0.00**	**36.5**
3jug	*Bacillus sp. N16-5*	70	17	2.45	32.7
1gzj	*Thermoascus aurantiacus*	70	15	2.11	33.6
2whj	*Bacillus agaradhaerens*	60	17	2.40	34.2
GH7 (β-jelly roll)					
4csi	***Humicola grisea* var. *thermoidea***	**72.5*** **	**100**	**0.00**	**46.8**
1cel	*Trichoderma reesei*	62.5***	58	0.56	45.9
2yg1	*Heterobasidion irregulare*	45	61	0.50	46.9
GH9 $(\alpha/\alpha)_6$					
4dod	***Caldicellulosiruptor bescii***	**~75***	**100**	**0.00**	**51.0**
1ks8	*Nasutitermes takasagoensis*	67	43	0.62	47.8
3wc3	*Eisenia fetida*	40	56	0.54	48.5
GH10 $(\beta/\alpha)_8$					
1vbr	***Thermotoga maritima***	**90**	**100**	**0.00**	**38.2**
2uwf	*Bacillus halodurans*	70	30	0.90	41.0
2f8q	*Bacillus sp. NG-27*	70	31	0.97	40.8
1hiz	*Bacillus stearothermophilus*	65	29	0.86	43.4
1b30	*Penicillium simplicissimum*	65	33	0.79	32.4
1k6a	*Thermoascus aurantiacus*	63	33	0.68	32.8
1n82	*Bacillus stearothermophilus*	60	33	0.89	38.5
GH11 (β-jelly roll)					
1f5j**	***Dictyoglomus thermophilum***	**75**	**100**	**0.00**	**22.3**
3zse**	***Thermobifida fusca***	**75**	**(100)**	**(0.00)**	**20.8**
1m4w	*Thermopolyspora flexuosa*	70	48 (82)	0.70 (0.40)	21.8
2nqy	*Bacillus agaradhaerens*	70	56 (52)	0.51 (0.58)	22.7
1igo	*Bacillus subtilis B230*	60	58 (52)	0.71 (0.64)	22.5
1hix	*Streptomyces sp. S38*	60	47 (76)	0.82 (0.61)	20.1
1bk1	*Aspergillus kawachii*	60	38 (41)	0.87 (0.69)	19.7
3m4f	*Scytalidium acidophilum*	50	37 (42)	0.97 (0.71)	19.3
1enx	*Trichoderma reesei II*	45	51 (56)	1.09 (0.81)	20.7
1xnd	*Trichoderma harzianum*	45	51 (55)	0.73 (0.57)	20.7
1xnb	*Bacillus circulans*	45	43 (62)	0.72 (0.54)	20.4
1ukr	*Aspergillus niger*	40	38 (40)	0.91 (0.71)	19.7
1xyn	*Trichoderma reesei I*	40	38 (41)	0.71 (0.59)	19.0
GH13 $(\beta/\alpha)_8$					
1ciu	***Thermoanaerobacterium thermosulfurigenes***	**60**	**100**	**0.00**	**75.4**

(*Continued*)

Table 1. (*Continued*)

PDB	ORGANISM	T_{OPT} (°C)	% Sequence ID	RMSD	MW (kDa)
1cyg	*Bacillus stearothermophilus*	55	68	0.51	75.4
1cdg	*Bacillus circulans*	35	70	0.49	74.5
Lactate dehydrogenase (βαβ)					
1a5z	***Thermatoga maritima***	**85**	**100**	**0.00**	**34.2**
1ldn	*Bacillus stearothermophilus*	55	41	1.17	34.7
5ldh	pig	37	34	2.03	35.8
6ldh	dogfish	15	37	1.33	36.3
Malate dehydrogenase (βαβ)					
1bmd	***Thermus flavus***	**75**	**100**	**0.00**	**35.3**
4mdh	pig	37	54	0.77	36.3
1emd	*Escherichia coli*	37	21	3.16	32.4
Methionine aminopeptidase (pita-bread fold)					
1xgs	***Pyrococcus furiosus***	**90**	**100**	**0.00**	**32.8**
1mat	*Escherichia coli*	-	24	1.24	29.2
1y1n	*Mycobacterium tuberculosis*	50	20	1.13	30.6

* The published activity temperature for *C. bescii* GH9 (PDB 4dod) includes a CBM3 domain.

** GH11 includes two thermophilic enzymes with the same optimum activity temperature. % Sequence ID and RMSD compared against 3zse.pdb are shown in parentheses.

*** The stability presented is T_m (°C), not optimum activity temperature.

doi:10.1371/journal.pone.0145848.t001

activity temperatures. Further, additional protein domains can significantly alter the stability of a given muti-domain enzyme [26, 27]. For this work we select enzymes with a single catalytic domain or where an optimum activity temperature has been determined for the catalytic domain alone. We make one exception and include the industrially relevant GH9 from *C. bescii*, for which a CBM3 was present in the activity temperature measurements [28].

Identifying structurally equivalent clusters of interacting residues

Residues that are close in space, or interacting, can display energetic cooperativity [29–31]. Here, clusters of spatially adjacent residues are identified using a distance cutoff. Two residues are defined as interacting if any side chain heavy atoms (C, N, O and S) are within 3 Å (Fig 1A). A structure is tiled in clusters, starting with the N-terminal residue and identifying adjacent positions and ultimately moving to the C-terminal residue.

Equivalent clusters in homologous enzymes are identified using the structural alignment algorithm, jFatCat flexible [32]. jFatCat flexible allows for a minimum number of backbone rotations to maximize the identification of structurally similar regions. The flexible structural alignment also generates an alignment of structurally matched sequence positions (Fig 1B). Thus clusters from paired proteins can be accurately identified and compared based on local context (Fig 1C).

Residue clusters display a high degree of sequence variation between thermophilic and mesophilic enzymes

Residue substitutions can display epistasis, and the effects of multiple substitutions cannot be easily predicted from knowledge of the individual substitutions. Podgornaia and Laub,

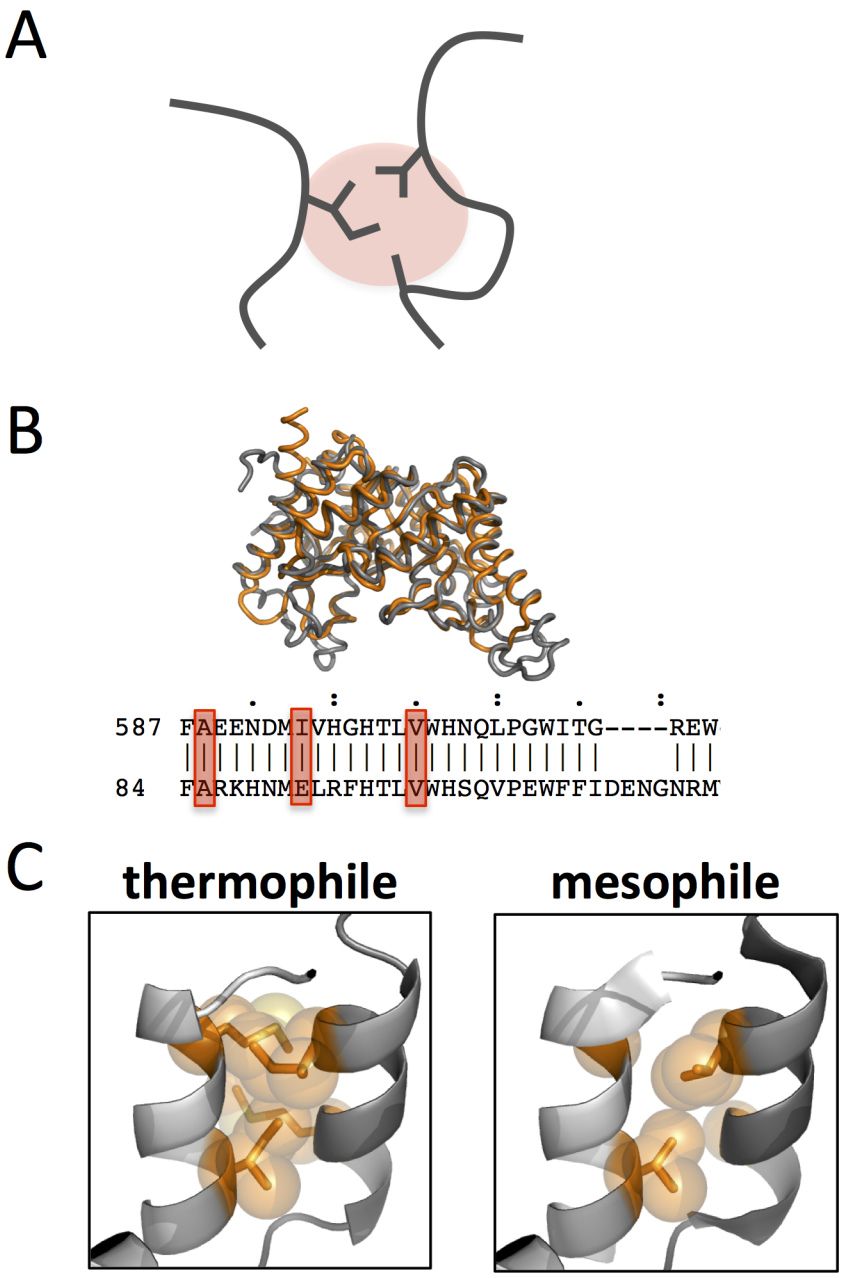

Fig 1. Identifying equivalent clusters in homologous proteins allows for direct comparison of local environments. (A) A cartoon depiction of cluster of adjacent residues is shown (red circle). (B) Structural alignment of paired enzymes is shown, with PDB 1vbr in orange and 2uwf in gray. The structurally aligned residues for the paired enzymes are shown beneath. (C) Differences in atomic packing is depicted with alternate sequences shown in stick and sphere representation on PDB 2wva.

doi:10.1371/journal.pone.0145848.g001

mapping the sequence space of the PhoQ-PhoP interface, found combinations of mutations producing functionally fit proteins in cases where the individual substitutions resulted in loss of function [18]. These results suggest evolutionary constraints and limitations of directed evolution. However, comparing homologous proteins from distantly related organisms, as is included here, can be used to investigate adaptive mechanisms taking place over long evolutionary timescales and encompassing large sampling of allowed sequence space.

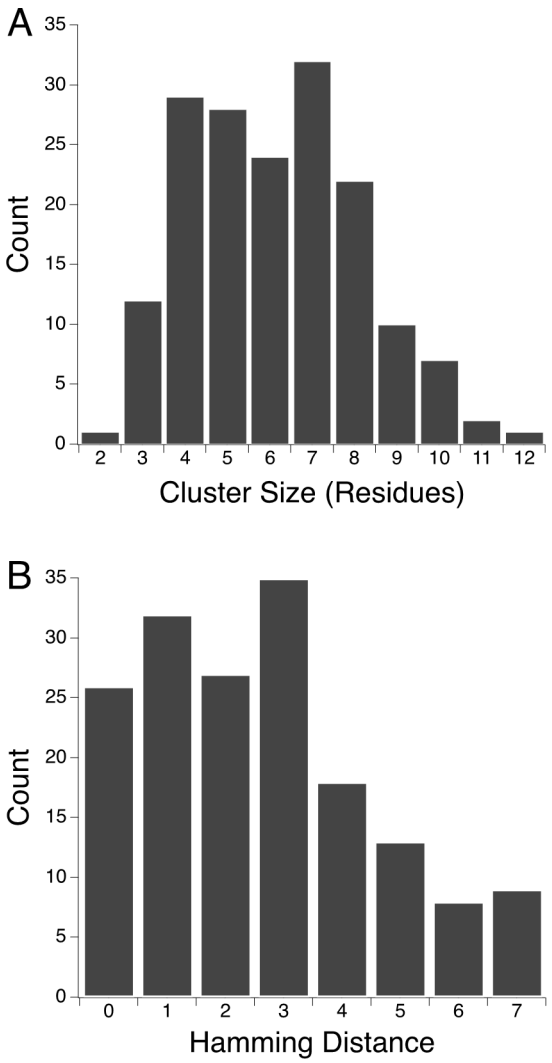

Fig 2. Evaluating the potential for epistasis. (A) The number of residues in each motif are determined for all representative thermophilic-mesophilic structure pairs and binned according to the motif size. (B) The number of residue substitutions, given as Hamming distance, in each equivalent thermophilic-mesophilic motif is determined and binned.

doi:10.1371/journal.pone.0145848.g002

Protein families in our dataset have from three to thirteen member structures (Table 1). We take a representative protein pair from each family to avoid biasing the results towards enzyme families with more members. Cluster sizes and number of amino acid substitutions are taken from pairs of structures representing the most and least thermostable members in each family. We evaluate the average size of the motifs by determining number of residues in each cluster (Fig 2A). We also evaluate the average number of amino acid substitutions, expressed as Hamming distance, between the paired clusters (Fig 2B). One hundred and fifty five motifs are identified in the representative dataset. Motifs range from two to twelve residues in size, with eighty percent of motifs having four to eight residues.

The total sequence identities between paired homologous thermophilic and mesophilic enzymes in our dataset range from fifteen to eighty two percent (Table 1). Residues in the protein core are generally more conserved compared to total sequence identity, as protein function requires properly folded structure [33, 34]. Indeed, thirty five percent of the equivalent

thermophilic-meosophilic enzyme clusters from the representative dataset have zero or one residue substitution. Despite the overall high conservation within hydrophobic protein core regions, sixty five percent of the motifs have a Hamming distance equal to or greater than two, with up to seven substitutions observed. Given the clusters are composed of interacting residues, these results highlight the potential for cooperative interactions and the need for a better understanding of epistatic effects.

Residue clusters are optimized for atomic packing in thermophilic enzymes

Improved quality of side-chain packing, or absence of cavities, is often observed in more thermostable enzymes relative to mesophilic homologues [35, 36]. Here equivalent clusters are evaluated for quality of packing using the solvent accessible surface area with the van der Waals radii expanded by 1.4 Å to represent a water molecule ($SASA_{1.4}$) [37]. The $SASA_{1.4}$ is determined for all residues in a motif, comparing thermophilic enzyme clusters to equivalent clusters in paired mesophilic enzymes ($\Delta SASA_{1.4}$). Thus, a negative $\Delta SASA_{1.4}$ indicates the thermophilic enzyme cluster displays fewer or smaller cavities and thus improved atomic packing compared to the mesophilic enzyme cluster.

Again, representative thermophile-mesophile enzyme pairs with the most and least thermostable enzymes are used to prevent bias towards enzyme families with the largest representation. The majority of clusters in thermophilic enzymes display improved atomic packing relative to mesophilic enzymes, with seventy-two percent of the clusters having a negative $\Delta SASA_{1.4}$ (Fig 3A and Table 2). Proteins are dynamic molecules, often exhibiting discrete conformational substates. In this work, however, we are assessing single protein conformations. As a result $\Delta SASA_{1.4}$ values less than but close to zero could simply be the result of comparing single conformations from experimentally determined structures. We therefore consider $\Delta SASA_{1.4}$ values less than or greater than 3 Å. Of the representative structures, thirty-three percent of the clusters from thermophilic enzymes exhibit improved atomic packing of greater than 3 Å while only one percent of clusters from mesophilic enzymes exhibit superior atomic packing of greater than 3 Å (Table 2).

$\Delta SASA_{1.4}$ analysis comparing equivalent clusters in thermophilic and mesophilic enzymes generally identifies the more thermostable enzyme of a homologous pair. This approach does not, however, predict the rank-order for a family of homologous enzymes based on thermal stability. $\Delta SASA_{1.4}$ analysis was performed comparing the most thermostable enzyme in each family to every other family member. $\Delta SASA_{1.4}$ for the entire dataset shows the same trend seen with the representative structures, although some structure pairs exhibit larger differences in $\Delta SASA_{1.4}$ (Fig 3A, S1 Fig and Table 2). The maximum $SASA_{1.4}$ observed for a mesophilic enzyme motif in the representative protein set is 60.2 $Å^2$, while the maximum $SASA_{1.4}$ observed for a mesophilic enzyme motif in the entire dataset is 100.9 $Å^2$ (Table 2). The representative dataset compares the most and least thermostable enzymes from each family, yet larger differences in $SASA_{1.4}$ can be found in paired enzymes with closer optimum activity temperatures.

To verify our findings we evaluate our entire dataset with an alternative method used to measure the quality of atomic packing in protein structures. We use the VLDP web server (http://www.dsimb.inserm.fr/dsimb_tools/vldp/) to compute residue contact number for each residue cluster. VLDP uses a Laguerre Tessellation to evaluate residue volumes in protein structures.[38] The results similar to $SASA_{1.4}$, show thermophilic residue clusters are more likely to have a higher residue contact number compared to the equivalent mesophilic residue cluster (Table 2).

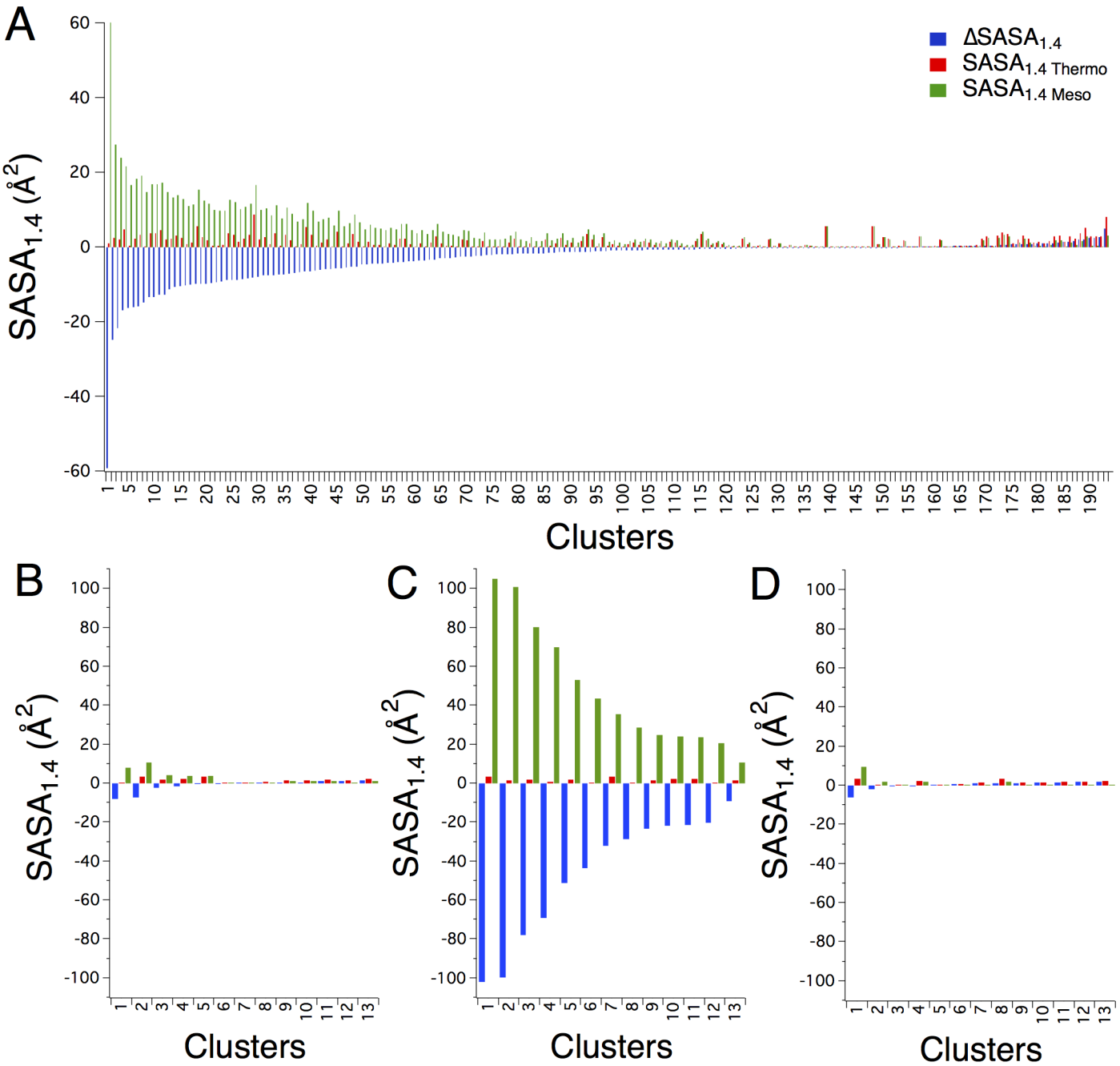

Fig 3. Thermophilic enzyme clusters display closer atomic packing compared to mesophilic enzyme clusters for most enzyme pairs evaluated. (A) $SASA_{1.4}$ values for clusters from the representative thermophilic-mesophilic structure pairs are shown, with thermophilic clusters shown in red, mesophilic clusters in green and the difference, $\Delta SASA_{1.4}$, in blue. Values are sorted by $\Delta SASA_{1.4}$. (B) $SASA_{1.4}$ values are shown comparing clusters from the thermophilic (PDB 1a5z) and mesophilic (PDB 6ldh) lactate dehydrogenase enzymes, which have a difference in optimum activity temperature of 30°C. (C) the thermophilic (PDB 1a5z) and mesophilic (PDB 5ldh) lactate dehydrogenase enzymes, with a difference in optimum activity temperature of 48°C, (D) and the thermophilic (PDB 1a5z) and psychrophilic (PDB 1ldh) lactate dehydrogenase enzymes, with a difference in optimum activity temperature of 70°C.

doi:10.1371/journal.pone.0145848.g003

Further, while only smaller void volumes are observed in thermophilic residue clusters from every family evaluated here, improved atomic packing between equivalent thermophilic and mesophilic clusters is not observed for all pairs of homologous proteins. For example,

Table 2. Comparing void volumes, as determined by $\Delta SASA_{1.4}$ and residue contact number, and percent sequence identity for paired clusters.

Representative Structures—193 clusters		
	Count	**Percentage**
$\Delta SASA_{1.4} < 0$	145	75%
$\Delta SASA_{1.4} < 1$	64	33%
$\Delta SASA_{1.4} < 2$	1	1%
	Max	**Average**
$SASA_{Thermo}$	8.7	1.5
$SASA_{Meso}$	60.2	4.8
All Structures—501 clusters		
	Count	**Percentage**
$\Delta SASA_{1.4} < 0$	388	77%
$\Delta SASA_{1.4} \leq -3$	227	45%
$\Delta SASA_{1.4} \geq 3$	6	1%
ΔContact Number ≤ 0	346	69%
	Max	**Average**
$SASA_{Thermo}$	8.7	1.6
$SASA_{Meso}$	100.9	8.4
% Sequence Identity		
	Median	**Average**
$\Delta SASA_{1.4} \leq -3$	50	49
$-3 < \Delta SASA1.4 < 3$	67	65
$\Delta SASA_{1.4} \geq 3$	44	51

doi:10.1371/journal.pone.0145848.t002

differences in atomic packing between a thermophilic lactate dehydrogenase (PDB 1a5z) and two homologous mesophilic enzymes and one psychrophilic enzyme are strikingly different. Both a mesophilic lactate dehydrogenase (PDB 1ldn) and a psychrophilic lactate dehydrogenase (PDB 6ldh) show small void volumes, very similar to the $SASA_{1.4}$ seen in the thermophilic lactate dehydrogenase (Fig 3B and 3D). Yet the differences in optimum activity temperatures compared to the thermophilic homologues are 30° and 70°C respectively. Comparing the same thermophilic lactate dehydrogenase to a different mesophilic homologue (PDB 5ldh) yields a $\Delta SASA_{1.4}$ distribution commonly seen in the protein pairs evaluated here, with significant enhancement see in atomic packing of the thermophilic residues (Fig 3C).

Similarly, the GH13 structures (PDB 1ciu) and (PDB 1cdg), with a 25°C difference in optimum activity temperatures, display $\Delta SASA_{1.4}$ values close to zero for all equivalent residue clusters (S2 Fig). The GH13 residue clusters also display high sequence identity, with the average percent sequence identity of 92, and 21 of the 33 clusters with one hundred percent sequence identity. These results indicate that while mesophilic and psychrophilic enzymes can have larger void volumes, other mechanisms also contribute to the lower thermal stability. However, the thermophilic enzymes do not appear to tolerate the destabilizing larger void volumes seen in many paired mesophilic enzymes investigated here.

Tiling clusters of interacting residues uncovers conserved "anchor" positions

The algorithm is designed to identify buried clusters of neighboring, interacting residues by searching for neighboring residues starting with N-terminal residues and cycling to C-terminal

residues. A protein structure is thus tiled in partially overlapping clusters. As a result, some buried residues with many neighbors are found in multiple clusters. These residues, described here as the anchor residues, resemble what are termed hot-spot residues when found at protein-protein interfaces. Hot spot residues at protein-protein interfaces are positions that contribute a significant amount of the stabilizing energy to drive the interaction.

Interestingly, $SASA_{1.4}$ is conserved for the anchor residues found here. For example, in Fig 4A, residues for the thermostable GH9 structure (PDB 4dod) found in six or more motifs are shown in stick representation, while the residues exhibiting the largest $\Delta SASA_{1.4}$ compared to equivalent residues in mesophile counterpart (PDB 3wc3) are highlighted in red. All residues in the GH9 thermophile are plotted according to the number of motifs in which they are found. Each sequence position is colored according to a heat map, where blue indicates the $\Delta SASA_{1.4}$ is positive, with better packing in the mesophilic motif, and red indicates negative $\Delta SASA_{1.4}$ with better packing in the thermophilic motif. The scale is set to ±10 $\Delta SASA_{1.4}$, as the lowest $\Delta SASA_{1.4}$ value for this pair of structures is -11. Blue is not observed for any sequence position (Fig 4B).

Evaluating this trend for sequence positions found in the representative structure pairs shows that in fact $\Delta SASA_{1.4}$ is conserved for all anchor residues, defined as positions found in five or more motifs (Fig 4C). The same trend holds for all sequence positions in the entire dataset, again for all anchor residues found in five or more motifs (S3 Fig). Thus the large improvements in $\Delta SASA_{1.4}$ found in thermophilic clusters compared to mesophilic clusters come from residues making fewer contacts rather than the anchor regions of the protein core. Each sequence position is colored white if the sequence identity is conserved and grey if identity is not conserved (Fig 4C). Interestingly, despite the conservation in $\Delta SASA_{1.4}$, the sequence identities for these anchor residues are not absolutely conserved. However, a higher degree of sequence conservation is seen in residue clusters with similar atomic packing, as determined by $\Delta SASA_{1.4}$ between -3 and 3 $Å^2$ (Table 2). In fact, 82% of residue clusters with 100% sequence identity are found in the residue clusters that also display conserved atomic packing.

Backbone adjustments as determined by distance differences

The networks of interacting side-chain clusters identified here tend to be large and contain multiple amino acid substitutions. Understanding how the protein backbone responds to accommodate alternate sequence combinations helps pinpoint the challenges for molecular design algorithms. Distance difference matrices are ideal for the comparison of geometric and distance similarities in enzymes and enzyme active sites as structural alignments are not necessary. Comparing distance matrices between structurally equivalent clusters is similar to comparing enzyme active sites. Distances between all Cα atoms are determined for each paired thermophilic and mesophilic cluster (Fig 5A).

Taking the absolute value of each Cα distance difference allows the differences to be summed. In this way a single metric measures the degree of backbone movement between paired clusters. Further, dividing each sum by the number of residues normalizes the summed differences, resulting in a comparable metric regardless of cluster size. The resulting metric shows that, on one hand, approximately half of the clusters from the representative dataset display little backbone movement, yielding an average summed distance difference of zero to approximately one half Å (Fig 5B). The distance differences increase rapidly, however, for the remainder of motifs. Over twenty-five Å sum-of-distances is observed, and an average distance difference of up to three and a half Å. Equivalent residue clusters with an average distance difference of 3 Å have an average of 3 Å between each pair of Cα residues. Residue clusters displaying high average distance differences likely have different shapes or sizes. Thus, while some

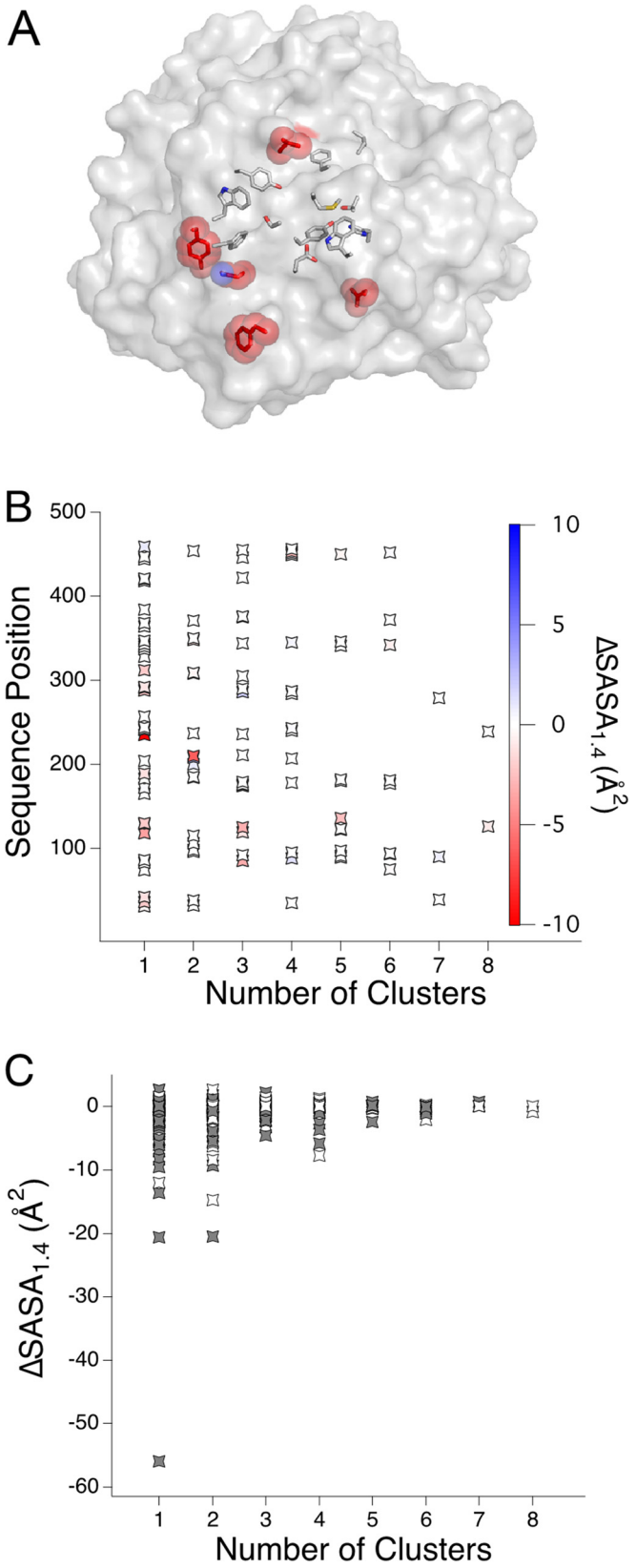
A
B
Sequence Position
500
400
300
200
100
1
2
3
4
5
6
7
8
Number of Clusters
10
5
0
-5
-10
ΔSASA1.4 (Å2)
C
ΔSASA1.4 (Å2)
0
-10
-20
-30
-40
-50
-60
1
2
3
4
5
6
7
8
Number of Clusters

Fig 4. Anchor residues are conserved in atomic packing. (A) The thermostable GH9 (PDB 4dod) is shown in surface representation, with anchor residues that are seen in a larger number of clusters shown in stick representation. Residues exhibiting the largest $\Delta SASA_{1.4}$, which are never anchor residues, are colored red. (B) Sequence positions from 4dod are binned by the number of clusters in which they are found. The heat scale indicates $\Delta SASA_{1.4}$. Importantly, blue is not observed as there are no mesophilic clusters with significantly better atomic packing relative to the matched thermophilic cluster. (C) Sequence positions from the representative set of structures are binned by the number of motifs in which they are found (x-axis), with $\Delta SASA_{1.4}$ shown for each paired position (y-axis). A white symbol indicates sequence conservation, and gray indicates the sequence differs at that position.

doi:10.1371/journal.pone.0145848.g004

paired motifs exhibit structural conservation, many display significant backbone movement. These results highlight the complexities of predicting potentially epistatic groupings of residues, even in a relatively small and defined protein region.

Discussion

A key challenge in protein engineering is accurately modeling energetic changes from mutations close in space. Protein double mutant free energy cycles show that non-additivity is a common phenomenon, especially when residue pairs are close in space [29–31]. Evolutionary studies also indicate sequence alterations can be cooperative [39, 40]. Evaluating sequence changes is more informative in context of the internal structural and chemical environment in which the substitutions are found. Ultimately, accurately predicting epistatic effects from multiple mutations requires a better understanding of how local environment affects amino acid substitutions. Here we examine how proteins evolve to function at high temperatures by evaluating local regions, applying structural informatics to investigate evolutionary patterns.

Optimized atomic-packing in protein core regions is important for enhanced thermostability. Cavities found in protein core regions diminish thermostability [36], and void volumes sized to accommodate water molecules have been observed more often in psychrophilic enzymes compared to homologous mesophilic counterparts [35]. Poorly packed protein core regions can lead to loss of conformational stability [41, 42]. Cavity filling mutations, conversely, can increase the hydrophobicity of the protein core [36]. As a protein engineering approach, Chen et al. demonstrated decreasing cavities in a protein hydrophobic core could enhance thermostability by transposing hydrophobic core regions from three thermostable enzymes into a mesophilic homologue [4]. Here, interacting residues in thermophilic enzymes display optimized van der Waals interactions, as seen by minimized cavities, compared to their mesophilic counterparts.

The results also support the role of alternative mechanisms leading to large changes in enzyme thermal stability. In addition to the examples discussed above (Fig 3B and 3D), Arimori et al., comparing two GH 9 enzymes with a 27°C difference in optimum activity temperature (1ks8 and 3nc3 in Table 1), identify an excess of negatively charged amino acids on the surface as the destabilizing mechanism for the psychrophilic homologue [43]. Kalimeri et al. report similar findings, comparing thermophilic and mesophilic malate dehydrogenase orthologues. They find that atomic volume is the same for both othologues, and instead oligomerization leads to enhanced thermal stability [44]. Importantly, regardless of the imperfections contributing to the moderate stability of mesophilic enzymes with ideal atomic packing, residue clusters from thermophilic enzymes appear to always display ideal atomic packing. Thus, as a designable element, these residue clusters represent evolutionarily optimized motifs.

The term hot-spot residues describes key positions that contribute a majority of the binding energy to protein-protein interfaces [45]. Similar to hot-spot residues, the approach applied here uncovers anchor residues that make many contacts and are thus found in many clusters. These anchor residues are conserved with regards to atomic packing in both thermophilic and mesophilic homologues. The observed improvements in atomic packing for thermophilic

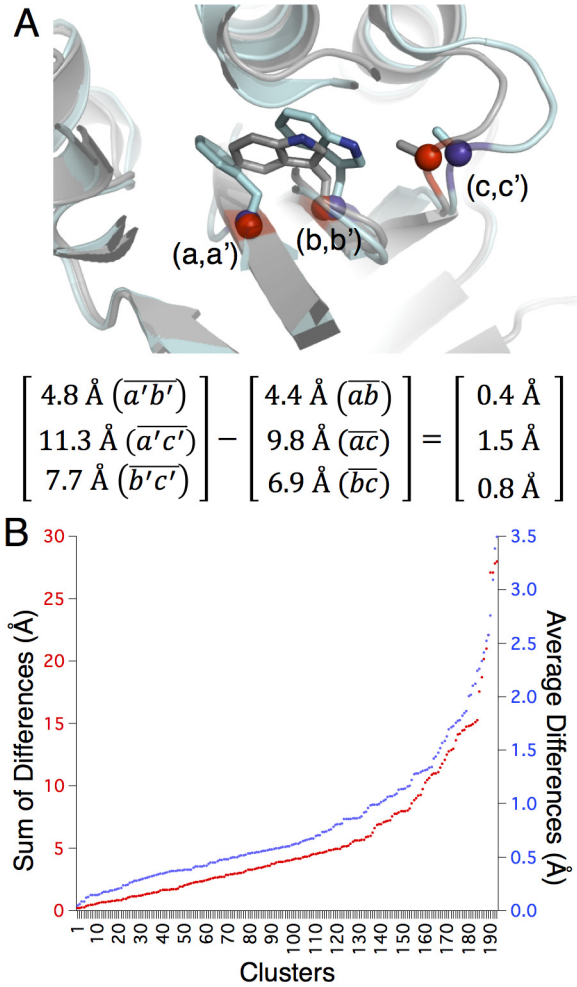

Fig 5. The backbone can move significantly in the structurally equivalent clusters. (A) Three Cα atoms from a paired cluster are shown in red spheres (thermophilic enzyme) and purple spheres (mesophilic enzyme). The atoms are labeled a, b and c for the thermophilic enzyme and a', b' and c' for the mesophilic enzyme. The Euclidian distances between Cα atoms are shown for each enzyme, with the distance differences at right. (B) The sum of the absolute values for the distance differences (red), and the average distance differences (blue) for each representative cluster are shown, sorted by summed or averaged distances.

doi:10.1371/journal.pone.0145848.g005

enzymes are thus found in residues that are peripheral to core anchor residues. These peripheral sequence positions might, therefore, serve to expand optimized protein core regions such as the residue clusters encompasing anchor residues.

The anchor residues, while well conserved in atomic packing, are not absolutely conserved in sequence. Putting other energetic contributions aside, the absence of cavities, which are known to be energetically deleterious, appears to be important for thermostability. Further, energetic contributions may be met without having to hold key sequence positions to absolute conservation. These results also support the importance of considering buried residues in structural context, as SASA for a given residue is not determined simply by that residue but by that residue and it's neighbors.

While anchor residues are not absolutely conserved in sequence, higher sequence conservation is seen in the residue clusters that display similar atomic packing between thermophilic and mesophilic enzymes. Conversely, residue clusters that differ in atomic packing also show

higher divergence in sequence. Since optimized atomic packing is seen disproportionately in the thermophilic homologues, the results indicate these protein regions have evolved to confer additional stabilization in the thermophilic homologues.

Obtaining the desired physicochemical properties for some protein targets may not always be achievable by combinations of single sequence substitutions. Yet evaluating all possible sequence space even in a small, defined region results in a combinatorial explosion that renders the approach intractable. The challenge, based on measured conformational changes, is backbone movement often seen when comparing equivalent residue clusters. Modeling, or predicting, such conformational changes with no *a priori* knowledge of optimized target sequences is not trivial and explains the challenges in predicting beneficial sequence combination *in silico*.

The method presented here identifies evolutionary optimized residue clusters with ideal sequence combinations and side-chain packing patterns. Importantly, these results suggest that while mesophilic and psychrophilic enzymes can accommodate cavities in the protein core, thermophilic enzymes cannot. As such, all residue clusters from the core of thermophilic enzymes can be viewed as potential transposable motifs to evaluate successful sequence combinations on complementary backbone structures.

Experimental Procedures

Protein Dataset

Thermophilic and mesophilic glycoside hydrolase enzymes were identified from the CAZy database (Carbohydrate-Active enZYmes), which categorizes enzymes based on structural similarity [20]. Enzymes were evaluated for the presence of additional domains using Pfam [46]. Enzyme optimum activity temperatures were found in the following publications: GH5 [26], GH 7 [47, 48], GH 9 [43, 49], GH10, GH 11, GH13 and lactate dehydrogenase and malate dehydrogenase [23], and methionine aminopeptidase [50, 51].

Sequences were aligned and analyzed using the MacVector software (MacVector, Inc., Cary, NC) [52]. Sequence alignments were performed using the GONNET substitution matrix [53], with a gap opening penalty of 10 and a gap extension penalty of 0.05. The molecular weight for each protein was computed based on the amino acid sequence using the ExPASy ProtParam tool [54]. Root-mean-square deviations (RMSD) for paired thermophilic and mesophilic enzymes were computed using PyMol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.)

Identification of Residue Clusters

Interacting residue clusters were identified using a distance cutoff of 3 Å between side chain heavy atoms (C, N, O and S) using the protein design software, Rosetta [55, 56]. Structurally equivalent residue clusters in homologous mesophilic enzymes were identified using the structural alignment algorithm, jFatCat flexible [32]. Residue clusters were filtered based on degree of solvent accessibility, selecting only clusters where each residue displayed less than 3 $Å^2$ of SASA as determined using Naccess [57]. Residue clusters were excluded if an equivalent residue from a thermophilic cluster was not found in the mesophilic cluster.

Distance difference matrices were calculated for paired thermophilic and mesophilic clusters. Accurate structural alignment of paired residues was verified manually if the sum of the distance differences for a cluster exceeded 20 Å.

Comparing Structurally Equivalent Residue Clusters

The residue accessible surface areas were computed using the program Naccess [57]. Naccess rolls a probe of a given radius over the van der Waals surface of a molecule to trace the

accessible surface. A probe of radius 1.4 Å was used here to reflect the radius of water and thus the solvent accessible surface area.

Histograms for cluster size and Hamming distance were created using StatPlus:mac [58]. Graphs were generated using IGOR Pro (WaveMetrics Inc., Lake Oswego, OR).

Supporting Information

S1 Fig. $SASA_{1.4}$ values for clusters from all remaining thermophilic-mesophilic structure pairs not shown in Fig 3A are shown, with thermophilic clusters shown in red, mesophilic clusters in green and the difference, $\Delta SASA_{1.4}$, in blue. Clusters are sorted by $\Delta SASA_{1.4}$.
(DOCX)

S2 Fig. $SASA_{1.4}$ values are shown comparing clusters from the thermophilic (PDB 1ciu) and mesophilic (PDB 1cdg) GH13 structures, which have a difference in optimum activity temperature of 25° C yet small differences in $SASA_{1.4}$ between clusters.
(DOCX)

S3 Fig. Sequence positions from all paired structures are binned by the number of motifs in which they are found (x-axis), with $\Delta SASA_{1.4}$ shown for each paired position (y-axis). A white symbol indicates sequence conservation, and gray indicates the sequence differs at that position.
(DOCX)

Acknowledgments

This work was funded by the US Department of Energy's Bioenergy Technologies Office (DOE-BETO), Contract No. DE-AC36-08GO28308 with the National Renewable Energy Laboratory and the National Institutes of Health, NIH R01GM103843 (HY), F30CA180249 (NK).

Author Contributions

Conceived and designed the experiments: DWS MEH HY MFC YJB. Performed the experiments: DWS NK. Analyzed the data: DWS NK. Contributed reagents/materials/analysis tools: DWS NK. Wrote the paper: DWS NK MEH HY MFC YJB.

References

1. Takai K, Nakamura K, Toki T, Tsunogai U, Miyazaki M, Miyazaki J, et al. Cell proliferation at 122 degrees C and isotopically heavy CH4 production by a hyperthermophilic methanogen under high-pressure cultivation. Proc Natl Acad Sci U S A. 2008; 105(31):10949–54. Epub 2008/07/31. doi: 10.1073/pnas.0712334105 PMID: 18664583; PubMed Central PMCID: PMC2490668.
2. Antranikian G, Vorgias CE, Bertoldo C. Extreme environments as a resource for microorganisms and novel biocatalysts. Adv Biochem Eng Biotechnol. 2005; 96:219–62. Epub 2006/03/29. PMID: 16566093.
3. Chen J, Stites WE. Replacement of staphylococcal nuclease hydrophobic core residues with those from thermophilic homologues indicates packing is improved in some thermostable proteins. J Mol Biol. 2004; 344(1):271–80. Epub 2004/10/27. doi: 10.1016/j.jmb.2004.09.008 PMID: 15504416.
4. Chen J, Lu Z, Sakon J, Stites WE. Increasing the thermostability of staphylococcal nuclease: implications for the origin of protein thermostability. J Mol Biol. 2000; 303(2):125–30. Epub 2000/10/12. doi: 10.1006/jmbi.2000.4140 PMID: 11023780.
5. Vogt G, Woell S, Argos P. Protein thermal stability, hydrogen bonds, and ion pairs. J Mol Biol. 1997; 269(4):631–43. Epub 1997/06/20. doi: 10.1006/jmbi.1997.1042 PMID: 9217266.
6. Robinson-Rechavi M, Alibes A, Godzik A. Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: lessons from structural genomics of Thermotoga maritima. J Mol Biol. 2006; 356(2):547–57. Epub 2005/12/27. doi: 10.1016/j.jmb.2005.11.065 PMID: 16375925.

7. Greaves RB, Warwicker J. Mechanisms for stabilisation and the maintenance of solubility in proteins from thermophiles. BMC Struct Biol. 2007; 7:18. Epub 2007/03/31. doi: 10.1186/1472-6807-7-18 PMID: 17394655; PubMed Central PMCID: PMC1851960.
8. Gromiha MM, Pathak MC, Saraboji K, Ortlund EA, Gaucher EA. Hydrophobic environment is a key factor for the stability of thermophilic proteins. Proteins. 2013; 81(4):715–21. Epub 2013/01/16. doi: 10.1002/prot.24232 PMID: 23319168.
9. Dong H, Mukaiyama A, Tadokoro T, Koga Y, Takano K, Kanaya S. Hydrophobic effect on the stability and folding of a hyperthermophilic protein. J Mol Biol. 2008; 378(1):264–72. Epub 2008/03/21. doi: 10.1016/j.jmb.2008.02.039 PMID: 18353366.
10. Perl D, Welker C, Schindler T, Schroder K, Marahiel MA, Jaenicke R, et al. Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. Nat Struct Biol. 1998; 5(3):229–35. Epub 1998/03/21. PMID: 9501917.
11. Watanabe K, Suzuki Y. Protein thermostabilization by proline substitutions. J Mol Catal B-Enzym. 1998; 4(4):167–80. doi: 10.1016/S1381-1177(97)00031-3 PMID: ISI:000074643300001.
12. Richardson JS, Richardson DC. Amino acid preferences for specific locations at the ends of alpha helices. Science. 1988; 240(4859):1648–52. Epub 1988/06/17. PMID: 3381086.
13. Matthews BW, Nicholson H, Becktel WJ. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. Proc Natl Acad Sci U S A. 1987; 84(19):6663–7. Epub 1987/10/01. PMID: 3477797; PubMed Central PMCID: PMC299143.
14. Borgo B, Havranek JJ. Automated selection of stabilizing mutations in designed and natural proteins. Proc Natl Acad Sci U S A. 2012; 109(5):1494–9. Epub 2012/02/07. doi: 10.1073/pnas.1115172109 PMID: 22307603; PubMed Central PMCID: PMC3277135.
15. Miklos AE, Kluwe C, Der BS, Pai S, Sircar A, Hughes RA, et al. Structure-based design of supercharged, highly thermoresistant antibodies. Chem Biol. 2012; 19(4):449–55. Epub 2012/04/24. doi: 10.1016/j.chembiol.2012.01.018 PMID: 22520751.
16. Shakhnovich BE, Deeds E, Delisi C, Shakhnovich E. Protein structure and evolutionary history determine sequence space topology. Genome research. 2005; 15(3):385–92. doi: 10.1101/gr.3133605 PMID: 15741509; PubMed Central PMCID: PMC551565.
17. Xiao L, Honig B. Electrostatic contributions to the stability of hyperthermophilic proteins. J Mol Biol. 1999; 289(5):1435–44. doi: 10.1006/jmbi.1999.2810 PMID: 10373377.
18. Podgornaia AI, Laub MT. Protein evolution. Pervasive degeneracy and epistasis in a protein-protein interface. Science. 2015; 347(6222):673–7. Epub 2015/02/07. doi: 10.1126/science.1257360 PMID: 25657251.
19. Himmel ME, Karplus PA, Sakon J, Adney WS, Baker JO, Thomas SR. Polysaccharide hydrolase folds diversity of structure and convergence of function. Appl Biochem Biotechnol. 1997; 63–65:315–25. doi: 10.1007/BF02920433 PMID: 18576090.
20. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 2009; 37(Database issue):D233–8. Epub 2008/10/08. doi: 10.1093/nar/gkn663 PMID: 18838391; PubMed Central PMCID: PMC2686590.
21. Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, et al. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science. 2007; 315(5813):804–7. Epub 2007/02/10. doi: 10.1126/science.1137016 PMID: 17289988.
22. Sadeghi M, Naderi-Manesh H, Zarrabi M, Ranjbar B. Effective factors in thermostability of thermophilic proteins. Biophys Chem. 2006; 119(3):256–70. Epub 2005/10/29. doi: 10.1016/j.bpc.2005.09.018 PMID: 16253416.
23. Karshikoff A, Ladenstein R. Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. Protein Eng. 1998; 11(10):867–72. Epub 1998/12/23. PMID: 9862205.
24. Olaleye O, Raghunand TR, Bhat S, He J, Tyagi S, Lamichhane G, et al. Methionine aminopeptidases from Mycobacterium tuberculosis as novel antimycobacterial targets. Chem Biol. 2010; 17(1):86–97. doi: 10.1016/j.chembiol.2009.12.014 PMID: 20142044; PubMed Central PMCID: PMC3165048.
25. Daniel RM. The upper limits of enzyme thermal stability. Enzyme Microb Technol. 1996; 19(1):74–9. doi: 10.1016/0141-0229(95)00174-3 PMID: ISI:A1996UR28700013.
26. dos Santos CR, Paiva JH, Meza AN, Cota J, Alvarez TM, Ruller R, et al. Molecular insights into substrate specificity and thermal stability of a bacterial GH5-CBM27 endo-1,4-beta-D-mannanase. J Struct Biol. 2012; 177(2):469–76. Epub 2011/12/14. doi: 10.1016/j.jsb.2011.11.021 PMID: 22155669.
27. Voutilainen SP, Nurmi-Rantala S, Penttila M, Koivula A. Engineering chimeric thermostable GH7 cellobiohydrolases in Saccharomyces cerevisiae. Applied microbiology and biotechnology. 2014; 98(7):2991–3001. doi: 10.1007/s00253-013-5177-2 PMID: 23974371.

28. Yi Z, Su X, Revindran V, Mackie RI, Cann I. Molecular and biochemical analyses of CbCel9A/Cel48A, a highly secreted multi-modular cellulase by Caldicellulosiruptor bescii during growth on crystalline cellulose. PLoS One. 2013; 8(12):e84172. Epub 2013/12/21. doi: 10.1371/journal.pone.0084172 PMID: 24358340; PubMed Central PMCID: PMC3865294.

29. Baik SH, Michel F, Aghajari N, Haser R, Harayama S. Cooperative effect of two surface amino acid mutations (Q252L and E170K) in glucose dehydrogenase from Bacillus megaterium IWG3 on stabilization of its oligomeric state. Appl Environ Microbiol. 2005; 71(6):3285–93. Epub 2005/06/04. doi: 10.1128/AEM.71.6.3285–3293.2005 PMID: 15933031; PubMed Central PMCID: PMC1151818.

30. Wells JA. Additivity of mutational effects in proteins. Biochemistry. 1990; 29(37):8509–17. Epub 1990/09/18. PMID: 2271534.

31. Schreiber G, Fersht AR. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. J Mol Biol. 1995; 248(2):478–86. Epub 1995/04/28. PMID: 7739054.

32. Ye Y, Godzik A. Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics. 2003; 19 Suppl 2:ii246–55. Epub 2003/10/10. PMID: 14534198.

33. Chang CM, Huang YW, Shih CH, Hwang JK. On the relationship between the sequence conservation and the packing density profiles of the protein complexes. Proteins-Structure Function and Bioinformatics. 2013; 81(7):1192–9. doi: 10.1002/Prot.24268 PMID: ISI:000320474100009.

34. Liao H, Yeh W, Chiang D, Jernigan RL, Lustig B. Protein sequence entropy is closely related to packing density and hydrophobicity. Protein Eng Des Sel. 2005; 18(2):59–64. Epub 2005/03/25. doi: 10.1093/protein/gzi009 PMID: 15788422; PubMed Central PMCID: PMC2553042.

35. Paredes DI, Watters K, Pitman DJ, Bystroff C, Dordick JS. Comparative void-volume analysis of psychrophilic and mesophilic enzymes: Structural bioinformatics of psychrophilic enzymes reveals sources of core flexibility. BMC Struct Biol. 2011; 11:42. Epub 2011/10/22. doi: 10.1186/1472-6807-11-42 PMID: 22013889; PubMed Central PMCID: PMC3224250.

36. Ohmura T, Ueda T, Ootsuka K, Saito M, Imoto T. Stabilization of hen egg white lysozyme by a cavity-filling mutation. Protein science: a publication of the Protein Society. 2001; 10(2):313–20. Epub 2001/03/27. doi: 10.1110/ps.37401 PMID: 11266617; PubMed Central PMCID: PMC2373952.

37. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol. 1971; 55(3):379–400. Epub 1971/02/14. PMID: 5551392.

38. Esque J, Oguey C, de Brevern AG. A novel evaluation of residue and protein volumes by means of Laguerre tessellation. J Chem Inf Model. 2010; 50(5):947–60. doi: 10.1021/ci9004892 PMID: 20392096.

39. Wang ZO, Pollock DD. Context dependence and coevolution among amino acid residues in proteins. Methods Enzymol. 2005; 395:779–90. Epub 2005/05/04. doi: 10.1016/S0076-6879(05)95040-4 PMID: 15865995; PubMed Central PMCID: PMC2943952.

40. Wang ZO, Pollock DD. Coevolutionary patterns in cytochrome c oxidase subunit I depend on structural and functional context. J Mol Evol. 2007; 65(5):485–95. Epub 2007/10/24. doi: 10.1007/s00239-007-9018-8 PMID: 17955155.

41. Dill KA, Shortle D. Denatured states of proteins. Annu Rev Biochem. 1991; 60:795–825. doi: 10.1146/annurev.bi.60.070191.004051 PMID: 1883209.

42. Vamvaca K, Jelesarov I, Hilvert D. Kinetics and thermodynamics of ligand binding to a molten globular enzyme and its native counterpart. J Mol Biol. 2008; 382(4):971–7. doi: 10.1016/j.jmb.2008.07.049 PMID: 18680748.

43. Arimori T, Ito A, Nakazawa M, Ueda M, Tamada T. Crystal structure of endo-1,4-beta-glucanase from Eisenia fetida. J Synchrotron Radiat. 2013; 20(Pt 6):884–9. Epub 2013/10/15. doi: 10.1107/S0909049513021110 PMID: 24121333; PubMed Central PMCID: PMC3795549.

44. Kalimeri M, Girard E, Madern D, Sterpone F. Interface matters: the stiffness route to stability of a thermophilic tetrameric malate dehydrogenase. PLoS One. 2014; 9(12):e113895. doi: 10.1371/journal.pone.0113895 PMID: 25437494; PubMed Central PMCID: PMCPMC4250060.

45. Cunningham BC, Wells JA. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science. 1989; 244(4908):1081–5. Epub 1989/06/02. PMID: 2471267.

46. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014; 42(Database issue):D222–30. Epub 2013/11/30. doi: 10.1093/nar/gkt1223 PMID: 24288371; PubMed Central PMCID: PMC3965110.

47. Momeni MH, Payne CM, Hansson H, Mikkelsen NE, Svedberg J, Engstrom A, et al. Structural, biochemical, and computational characterization of the glycoside hydrolase family 7 cellobiohydrolase of the tree-killing fungus Heterobasidion irregulare. J Biol Chem. 2013; 288(8):5861–72. doi: 10.1074/jbc.M112.440891 PMID: 23303184; PubMed Central PMCID: PMC3581431.

48. Momeni MH, Goedegebuur F, Hansson H, Karkehabadi S, Askarieh G, Mitchinson C, et al. Expression, crystal structure and cellulase activity of the thermostable cellobiohydrolase Cel7A from the fungus Humicola grisea var. thermoidea. Acta crystallographica Section D, Biological crystallography. 2014; 70(Pt 9):2356–66. doi: 10.1107/S1399004714013844 PMID: 25195749; PubMed Central PMCID: PMC4157447.

49. Ostendorp R, Auerbach G, Jaenicke R. Extremely thermostable L(+)-lactate dehydrogenase from Thermotoga maritima: cloning, characterization, and crystallization of the recombinant enzyme in its tetrameric and octameric state. Protein science: a publication of the Protein Society. 1996; 5(5):862–73. Epub 1996/05/01. doi: 10.1002/pro.5560050508 PMID: 8732758; PubMed Central PMCID: PMC2143418.

50. Tsunasawa S, Izu Y, Miyagi M, Kato I. Methionine aminopeptidase from the hyperthermophilic Archaeon Pyrococcus furiosus: molecular cloning and overexpression in Escherichia coli of the gene, and characteristics of the enzyme. J Biochem. 1997; 122(4):843–50. Epub 1997/12/17. PMID: 9399590.

51. Addlagatta A, Quillin ML, Omotoso O, Liu JO, Matthews BW. Identification of an SH3-binding motif in a new class of methionine aminopeptidases from Mycobacterium tuberculosis suggests a mode of interaction with the ribosome. Biochemistry. 2005; 44(19):7166–74. Epub 2005/05/11. doi: 10.1021/bi0501176 PMID: 15882055.

52. Rastogi PA. MacVector. Integrated sequence analysis for the Macintosh. Methods Mol Biol. 2000; 132:47–69. Epub 1999/11/05. PMID: 10547831.

53. Gonnet GH, Cohen MA, Benner SA. Exhaustive matching of the entire protein sequence database. Science. 1992; 256(5062):1443–5. Epub 1992/06/05. PMID: 1604319.

54. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999; 112:531–52. Epub 1999/02/23. PMID: 10027275.

55. Kaufmann KW, Lemmon GH, DeLuca SL, Sheehan JH, Meiler J. Practically Useful: What the ROSETTA Protein Modeling Suite Can Do for You. Biochemistry. 2010; 49(14):2987–98. doi: 10.1021/Bi902153g PMID: ISI:000276258800003.

56. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci U S A. 2000; 97(19):10383–8. Epub 2000/09/14. PMID: 10984534; PubMed Central PMCID: PMC27033.

57. Hubbard SJ, Thornton JM. 'NACCESS', Computer Program. University of College London, London; 1993.

58. Shaw M. AnalystSoft's StatPlus:mac. DoubleClick. 2008:20–1.