PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Shape and evolution of thermostable protein structure

Ryan G. Coleman[1,2] and Kim A. Sharp[1,2]*

[1] The Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

[2] Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, Pennsylvania 19104

ABSTRACT

Organisms evolved at high temperatures must maintain their proteins' structures in the face of increased thermal disorder. This challenge results in differences in residue utilization and overall structure. Focusing on thermostable/mesostable pairs of homologous structures, we have examined these differences using novel geometric measures: specifically burial depth (distance from the molecular surface to each atom) and travel depth (distance from the convex hull to the molecular surface that avoids the protein interior). These along with common metrics like packing and Wadell Sphericity are used to gain insight into the constraints experienced by thermophiles. Mean travel depth of hyperthermostable proteins is significantly less than that of their mesostable counterparts, indicating smaller, less numerous and less deep pockets. The mean burial depth of hyperthermostable proteins is significantly higher than that of mesostable proteins indicating that they bury more atoms further from the surface. The burial depth can also be tracked on the individual residue level, adding a finer level of detail to the standard exposed surface area analysis. Hyperthermostable proteins for the first time are shown to be more spherical than their mesostable homologues, regardless of when and how they adapted to extreme temperature. Additionally, residue specific burial depth examinations reveal that charged residues stay unburied, most other residues are slightly more buried and Alanine is more significantly buried in hyperthermostable proteins.

Proteins 2010; 78:420–433.
© 2009 Wiley-Liss, Inc.

Key words: hyperthermophiles; hyperthermostable; pockets; geometry; depth; stability.

INTRODUCTION

It seems likely that hyperthermophilic archaea occupy positions near the root of the phylogenetic tree of life. However, there is still some debate as to whether life originated in hyperthermophilic conditions.[1–3] Nevertheless, life has adapted to many niche temperatures. Of these, the high temperature niche is the most puzzling to explain from a thermodynamic perspective, due to the increased thermal disorder that favors denatured or unfolded states. In addition to insights into fundamentals of protein stability, the discovery of thermostable variants of many enzymes has led to many practical applications.[4] Understanding how these variants achieve thermostability could lead to new ways to design proteins for greater thermostability, among other applications.

Inspired by recent work examining protein structures from a range of environmental temperatures from mesophiles to hyperthermophiles[5] we wanted to examine the overall shape and structural features of these proteins using recent advances in protein shape analysis. Additionally, we wanted to perform a more detailed analysis of structure at the residue level. With the ongoing determination of structural data on many homologues from various thermophiles and mesophiles, such overall structural differences can now be examined with increasing statistical resolution.

Many structural features that could lead to increased thermostability have been examined previously, and diverse factors have been found to differ between thermostable proteins and mesostable proteins with varying degrees of significance. The picture is further complicated when considering the extremity of the temperature (thermophiles vs. hyperthermophiles)[5,6] and the evolutionary background of the organism—ancient (original?) thermophiles vs. acquired themophilicity.[7] Structural factors that have been studied include increased hydrogen bonding in thermostable proteins,[7–12] an increase in the frequency of ion pairs and electrostatic contributions in thermostable proteins[5,6,9–16] and an increase in the amount of certain apolar contacts.[17] Differences in unfolding have been studied by various methods[7,18] including differences in rotamer states.[5,16] Also the differences in solvent exposed surface area have been

Grant sponsor: NIH Structural Biology Training; Grant numbers: GM008275, GM48130; Grant sponsor: National Human Genome Research Institute (Computational Genomics Training); Grant number: T32HG000046.
*Correspondence to: Kim A. Sharp; Department of Biochemistry and Biophysics, University of Pennsylvania, 37th and Hamilton Walk, Philadelphia, PA 19104-6059. E-mail: sharpk@mail.med.upenn.edu
Received 14 May 2009; Revised 17 June 2009; Accepted 6 July 2009
Published online 24 July 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22558

© 2009 WILEY-LISS, INC.

examined.[5,6,9,11,12,19–21] Van der Waals interactions, the amount of packing, and the number and size of cavities have been examined but lead to conflicting conclusions.[7,10,11,16,22] This is only a brief review of the structural features examined on multiple sets of protein pairs; Many other features have been examined, but only on single pairs of protein structures or by sequence based analysis.

Surprisingly, there has been no definitive study of overall shape and geometry differences, such as sphericity, arising from environmental temperature differences. This work addresses this by examining the overall geometric structure of thermostable proteins. In addition, we perform a finer resolution analysis of surface exposure. Previous work[6,9,19,20] examined the solvent exposed surface area changes per atom type, residue, or residue group. Some studies examined just the nonpolar exposed surface area changes[5] or the differing counts of residues that were exposed or buried according to a cutoff of solvent accessible area.[11] Only one previous study accounted for the overall shape changes by correcting the surface areas examined.[12] In this study, we not only identify residues that are buried, but we examine how deeply they are buried, using the distance to nearest point on the protein surface, or "burial depth." We also use travel depth[23] to examine the overall structure of the pockets and clefts of the proteins. Combined, these two depth measures provide complementary measures of how spherical the proteins are, if they are closer to ideal spheres or if they have more indentations, dimples and clefts.

We use a collated data set[5] which contains homologous structures from both mesophiles and several kinds of thermophiles. Both moderate thermophiles (45°C–80°C) and hyperthermophiles (above 80°C) are examined. Additionally, we break the class of hyperthermophiles into two subsets, the Ancient hyperthermophiles that have been hyperthermophiles for their entire evolutionary history,[24] and Recent hyperthermophiles like *Thermotoga maritima* that only recently became hyperthermophiles.[3,25] This follows the lead of previous work where a similar split in the class of hyperthermophiles was used.[7]

METHODS

Data collection

We use the recent data set of Greaves and Warwicker[5] that contains pairs consisting of a thermostable protein and a homologous mesostable protein. We use primarily the "67" set of data since these structures have pairs with chain lengths differing by 30 residues or less. Only such similarly sized structures are appropriate for most shape analysis. We do, however, use the larger "291" set for some additional analyses that do not depend on overall protein size. We examine both the moderate thermophiles and hyperthermophiles, and in addition we examine two subsets of the hyperthermophiles, the Ancient and Recent. The only organism in the dataset known to have recently adapted to extreme high temperatures is *Thermotoga Maritima*.[3,7,25] To be counted as recent in our analysis, the protein must be from *T. maritima*, additionally it must not be from an Archaeal lateral gene transfer.[25] Each protein from *T. maritima* in the 67 set[5] has closest relatives from other bacteria and is therefore presumably not from lateral gene transfer from an already hyperthermophilic archaea. There are 12 pairs in the Recent-mesophile set, and 18 pairs in the Ancient-mesophile set, for a total of 30 pairs in the combined hyperthermophile-mesophile set. There are 37 pairs in the moderate thermophile-mesophile set.

Files of single domains as specified[5] were downloaded from the Protein Data Bank.[26] In the case of multiple NMR structures, the structure closest to the average structure was used as representative of the set. All waters were removed from crystal structures. Hydrogens are assigned a radius of zero in our van der Waals radii set,[27] effectively ignoring them for all analyses. A 1.8 Å probe radius was used for all packing, surface construction, travel depth, burial depth analyses, as this is consistent with a water sphere. For the burial depth and travel depth analyses cavities are removed after surfaces were constructed.

Packing

Protein packing can be calculated by using the Voronoi construction.[28–33] A Voronoi cell is defined as the volume that is closer to the given atom (or point in the more general sense) than any other atom.[34,35] Defining the packing as the percentage of the volume of the Voronoi Cell filled by the van der Waals volume of the atom, packing is well-defined for completely buried or interior atoms. However, surface atoms have infinite Voronoi cells which must be restricted if a meaningful measure of their packing is to be computed. Methods of "capping" the Voronoi cells of surface atoms include using the molecular surface as the bounding volume[36] or using crystallographic waters.[29,32,33,37] Availability of sufficient crystallographic waters to cap depends on the resolution of the structure and how it was refined. Moreover, waters are entirely absent from NMR determined structures. For these reasons we decided to analyze surface and buried atoms separately. For the former we used the solvent accessible surface of each atom to generate the Voronoi cell capping. The solvent accessible surface is generated from van der Waals radius plus probe radius, so it lies a constant distance from the atom regardless of protein shape. However, this method of capping is somewhat arbitrary, as are other methods used to determine the packing of surface atoms. For this reason, in detailed comparisons

of packing, we believe it is more reliable to use just the interior atoms, which are bounded on all sides by identically defined and generated surfaces.

Travel depth

Previously, travel depth was established as a useful measure of depth of the molecular surface for examining pockets and ligand binding sites as well as the overall depth of the surface.[23] Travel depth is defined as the minimum distance from any surface point to the convex hull avoiding the protein interior, and is calculated using the Multiple Source Shortest Paths (MSSP) algorithm.[38] The original implementation has been improved for speed, flexibility and additional features.[39]

Burial depth

Atom burial depth has been used several times previously to analyze protein structure, although somewhat varying definitions exist in the literature, depending upon the exact implementation and desired use[40–45] see the review of Pintar et al.[46] However, no measure of burial depth has previously been applied to analyzing differences in thermostable and mesostable structures. A closely related method uses burial and counting nearby hydrophobic residues to discriminate native folds from decoys.[47] Another related concept is that of centrality or closeness of a graph connecting nearby atoms, used in various applications[48–52] Atom burial depth is defined here as the distance of the atom to the nearest point on the molecular surface. It is most efficiently calculated by starting from the molecular surface and labeling sequentially deeper points into the protein interior, using the same MSSP algorithm as for travel depth.[23,39]

Interatomic distances, Waddell sphericity, convex hull volume

We also examined several other shape metrics to see if they could discriminate between thermostable and mesostable proteins. The first was the mean interatomic distance. In principle this could be sensitive to how spherical and well-packed the folded structure is, with the advantage that it is exceptionally simple to compute, requiring only atomic coordinates, not the protein surface. The second metric was the convex hull volume, which would be larger if the protein structure is more spread out, or less "compact" at the larger, molecular scale. We note that compact is a pervasive yet ambiguous term in the literature on thermostable proteins. It has been used to refer to the efficiency of packing as measured by a Voronoi or similar analysis. It has also been used to refer to the number of contacts of a certain type, for instance hydrogen bonds, van der Waals contacts, etc. Finally, it could refer to the extent of a protein, how "splayed out" it is. Here we use compactness only in this sense, as defined by the convex hull volume. All other usages can be replaced by better terms.

The third metric, used previously to evaluate roundness of rocks and crystals, is Wadell Sphericity,[53] a dimensionless ratio of volume and surface area designed to have an upper bound of 1 (perfectly spherical), and decreasing to 0 the further from perfectly spherical the shape is. The formula for this ratio is given by Eq. (1) and it was calculated exactly from our triangulated molecular surfaces.

$$\psi = \frac{\pi^{\frac{1}{3}}(6V)^{\frac{2}{3}}}{A} \tag{1}$$

Statistical tests

Statistical significance is evaluated by permutation testing, by randomly switching (or not switching) the labels on each thermostable and mesostable pair and recomputing the difference in means of each metric across each category, and evaluating if the original difference is extremal to each permuted difference.[54] We used two individual one-tailed tests, meaning the permuted means found are checked to see if they are less than or greater than the original. In all cases the lower *P*-value is reported. Each statistical test reported was done using 1,000,000 permutations in the case of overall tests and 10,000 permutations in the case of residue-specific tests. Importantly, in residue specific tests, the overall difference in means was used as a correction factor to the difference in means when analyzing which residues become more buried or unburied. A standard threshold of 0.05 was used as a cutoff for significance.

RESULTS

Packing, mean distance, convex hull volume and Waddell sphericity

The packing analysis was performed on each atom in each structure, the results were separately accumulated over either all buried or all surface atoms in each protein and the mean of these numbers and the combined mean over all atoms was used to evaluate significance. Packing of recent and ancient hyperthermostable proteins is compared with their mesostable counterparts in Figure 1 for the buried atoms. The results are summarized in Table I for the various thermostable categories compared to their mesostable counterparts. No significant differences were found in packing of interior atoms for any thermostable set. Surface atoms, however, are significantly more tightly packed in both the ancient and combined hyperthermostable proteins.

The mean of the interatomic distance was computed across all heavy atoms, the results between the various

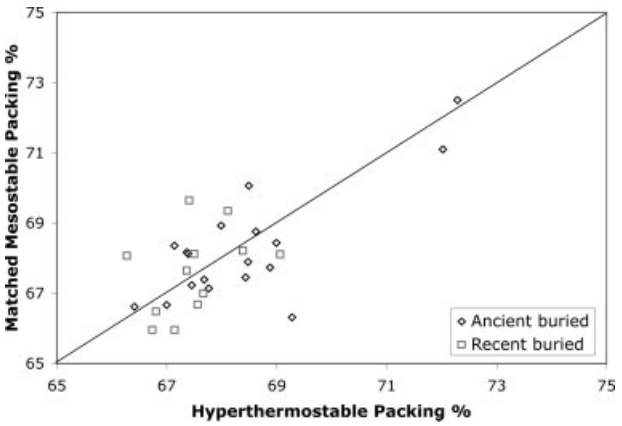

Figure 1

Packing in Hyperthermostable Proteins. Packing percentage in the buried category comparison between recent or ancient hyperthermostable vs. matched mesostable proteins for completely buried atoms.

thermostable proteins and their matched mesostable proteins are summarized in Table I and the results for the hyperthermostable categories are shown in Figure 2. No significant differences were found for any thermostable set. The convex hull volume, a metric for the overall extent of the protein also showed no significant difference in any category (Table I.)

The Wadell Sphericity[53] of each protein surface was computed by calculating the area and volume of the triangulated surfaces, and the dimensionless ratio given by Eq. (1) computed. The results are shown in Figure 3 and summarized in Table I. Hyperthermostable proteins are significantly more spherical than their mesostable counterparts. No difference is found for moderate thermostable proteins.

Wadell Sphericity is size independent so the analysis was also conducted on the "291" set.[5] The difference in mean Wadell Sphericity between the 144 pairs of hyperthermostable-mesostable structures was 0.035, with $P \ll 0.001$. These results indicate that in the larger set of hyperthermostable proteins, they too have become more spherical.

Travel depth and burial depth

The travel depth analysis was conducted on each protein structure. The mean travel depth, T_{av}, was computed by averaging over all surface points as described previously.[23] This mean was used to compare the various thermostable categories with their mesostable homologues, the results are shown in Figure 4 and summarized in Table I. Hyperthermostable proteins have significantly smaller values of T_{av}, indicating a less convoluted surface. As the maximum depth of a pocket is limited by the linear dimensions of the molecule, variation in size of proteins potentially complicates the interpretation of average travel depth. The volume of the protein is closely proportional to the number of heavy atoms, H, so $H^{1/3}$ provides a convenient measure of the average linear dimension of the molecule. Indeed Figure 5 shows that on average, travel depth increases linearly with the linear extent of the protein, for mesostable, thermostable, and hyperthermostable proteins. The scaled average travel depth, $T_{av}/H^{1/3}$ thus provides a good measure of the relative roughness of the molecule, in a fractal sense, as shown before on a larger class of small molecule binding proteins.[23] Differences in scaled average travel depth, $T_{av}/H^{1/3}$, are summarized in Table I, and are also significantly smaller for both hyperthermostable categories.

The Mean burial depth of all atoms in each protein structure was computed and the results are summarized in Table I. Atoms are significantly more deeply buried in hyperthermostable proteins, by slightly more than a tenth of an Angstrom. This is a very small difference, but it is an average over a large number of atoms, and it is statistically significant. The consistently deeper burial of atoms

Table I

Summary of Differences in Mean Values of Geometric Measures

Geometric measure[a]	Moderate thermostable	Recent hyperthermostable	Ancient hyperthermostable	All hyperthermostable
Number of pairs	37	12	18	30
Packing of buried atoms (%)	0.44 (0.06)	−0.10 (0.39)	0.16 (0.28)	0.06 (0.39)
Packing of surface atoms (%)	−0.01 (0.24)	0.28 (0.16)	**0.59** (<0.01)	**0.47** (<0.01)
Mean interatomic distance (Å)	0.08 (0.38)	−0.40 (0.27)	−0.50 (0.12)	−0.40 (0.08)
Convex hull volume (Å^3)	5 (0.5)	−1739 (0.3)	−3321 (0.08)	−2688 (0.06)
Volume (Å^3)	−430 (0.12)	1310 (0.07)	289 (0.33)	697 (0.08)
Surface area (Å^2)	−143 (0.23)	−227 (0.27)	**−533** (<0.01)	**−411** (0.02)
Wadelll sphericity	0.002 (0.42)	**0.027** (0.05)	**0.039** (<0.01)	**0.034** (<0.01)
Mean travel depth (Å)	−0.06 (0.34)	**−0.40** (0.05)	**−0.50** (<0.01)	**−0.46** (<0.01)
Mean travel depth/$H^{1/3}$ (Å)[b]	−0.001 (0.41)	**−0.044** (0.03)	**−0.047** (<0.01)	**−0.065** (<0.01)
Mean burial depth (Å)	0.12 (0.10)	**0.13** (0.01)	**0.12** (<0.01)	**0.13** (<0.01)

[a]Data is shown as the mean of the thermostable category minus the mean of the mesostable category. Statistical *P*-values for the lower of the two one-tailed tests follow in parentheses. Values with a *P*-value below the significance threshold of 0.05 are shown in bold.
[b]Mean travel depth divided by the cube root of the number of heavy atoms in the molecule, H.

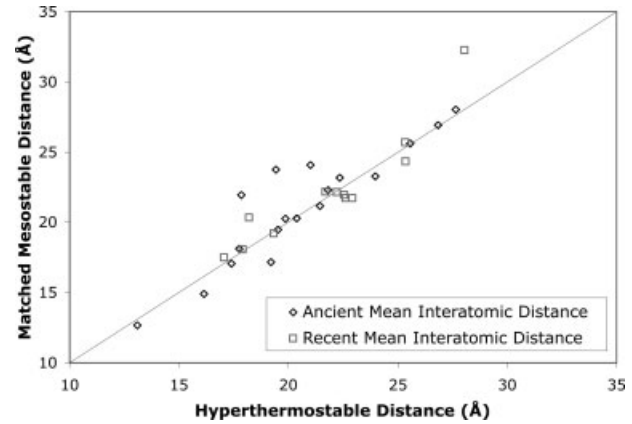

Figure 2

Interatomic distances in hyperthermostable proteins. The mean interatomic distance compared across recent or ancient hyperthermostable vs. matched mesostable proteins.

in hyperthermostable proteins is made more evident in complete histograms of burial depth accumulated over all atoms types, shown in Figure 6. The histogram for both hyperthermostable classes is consistently shifted to the right, indicating greater burial depth. This rightward shift is even clearer in the cumulative difference histogram. If just the C_β atom of each residue (C_a for glycine) is used for the burial depth analysis, very similar histograms, means, and *P*-values result. So for this kind of analysis the single atom burial is a good proxy for that of the entire residue.

Both smaller mean travel depths, and greater mean burial depths in hyperthermostable proteins indicate a more spherical shape in hyperthermostable proteins as compared to mesostable proteins. This is illustrated

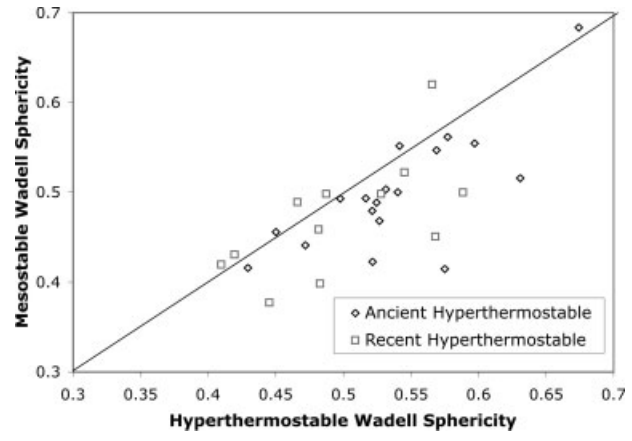

Figure 3

Wadell sphericity in hyperthermostable proteins. Wadell Sphericity compared between recent or ancient hyperthermostable vs. matched mesostable proteins.

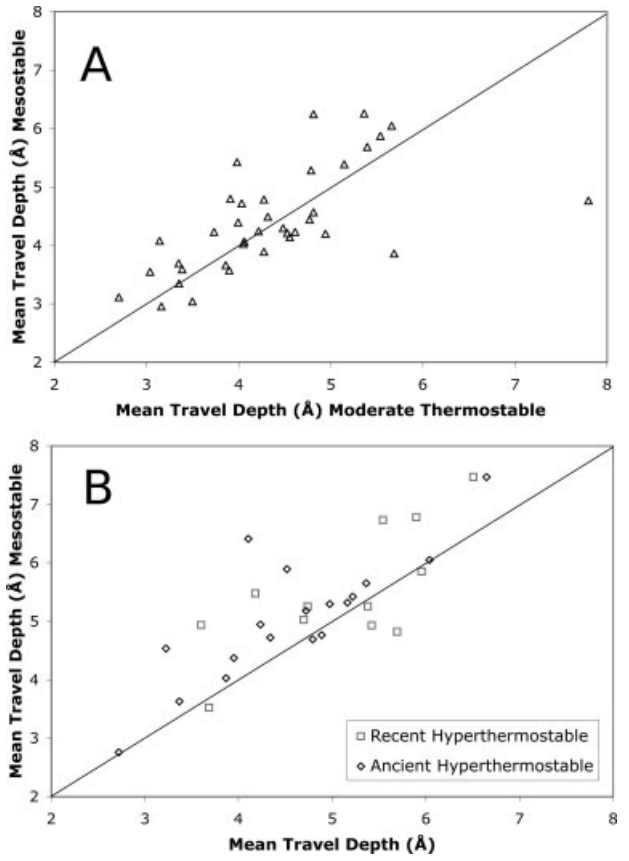

Figure 4

Travel depth in thermostable proteins. Mean travel depth compared between (**a**) Moderate thermostable, (**b**) Hyperthermostable, and the respective matched mesostable proteins.

graphically for an ancient hyperthermostable-mesostable matched pair of protein structures of phosphoserine phosphatase in Figure 7. The molecular surface on the left is colored by travel depth, while the right hand bond representation is colored by burial depth. The hyperthermostable protein (upper panels) clearly has more red colored (shallow) surface, and more red colored (deeply buried) atoms.

As the size scaled travel depth metric, $T_{av}/H^{1/3}$, largely removes the effect of protein size, one can compare mesostable and thermostable proteins that differ substantially in size, for which there are more proteins to compare. This metric was applied to the substantially larger "291" set of 144 hyperthermostable-mesostable pairs,[5] and the results are shown in Figure 8. While there is considerable scatter, the data again show that hyperthermostable proteins have significantly shallower surfaces, as can be seen in the linear trendlines. By analyzing the means of the ratios $T_{av}/H^{1/3}$, a statistical analysis was performed, resulting in a *P*-value of 5.6×10^{-5} indicating that the differences are significant.

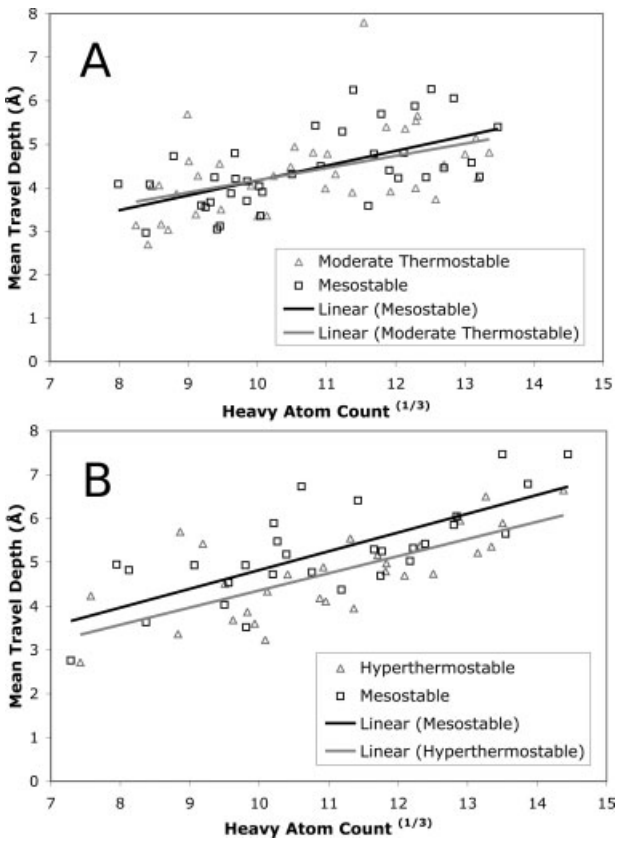

Figure 5

Size-scaled travel depth in thermostable proteins. Mean travel depth is plotted vs. the cube root of the number of heavy atoms. (**a**) Moderate thermostable and the matched mesostable proteins. (**b**) Hyperthermostable and the matched mesostable proteins. Trendlines for each set are shown on the figure.

To check the sensitivity of the burial depth and travel depth analysis to slightly different structures, we ran the analyses on a complete set of NMR structures forming one hyperthermostable-mesostable pair. The structures chosen were PDB codes [PDB:1JDQ] and [PDB:1JE3].[58] Each had 20 models. The mean burial depth had standard deviations of only 0.017 Å for both the hyperthermostable and mesostable protein. As the mean difference in burial depth from Table I is nearly ten-fold greater, this indicates that the burial depth analysis is not very sensitive to changes in which NMR structure was used. The mean travel depth had standard deviations of 0.24 Å and 0.13 Å for these two molecules, which is smaller than the difference in means (0.5 Å), also showing that travel depth is also not very sensitive to which NMR structure is used. Use of a single NMR structure out of the complete set of models seems reasonable.

The differences in burial depth of each specific residue type were also examined by comparing burials of the C_β atom (C_a for Glycine). Results are shown in Figure 9 for each of the 4 thermostable-mesostable classes. *P*-values for these individual residue burial differences were calculated by permutation, as described in the methods section. In computing the *P*-values, computed mean burial differences we first corrected by subtracting the difference in mean burial depth of C_β atoms (C_a for Glycine) over all residue types as described earlier. These corrections were 0.028, 0.13, 0.135, and 0.126 for the moderate thermostable, hyperthermostable, recent and ancient classes respectively. Residues with significant differences ($P < 0.05$) are indicated by their *P*-values on the figure.

The hyperthermostable-mesostable dataset has the largest amount of significant differences as shown in Figure 9(b). Alanine shows the largest change and is significantly more buried in hyperthermostable proteins as shown in detail in Figure 10. Cysteine, Tryptophan, and Valine show large trends to being more buried, but these changes are not statistically significant according to the analysis, after correcting for overall depth differences. Six

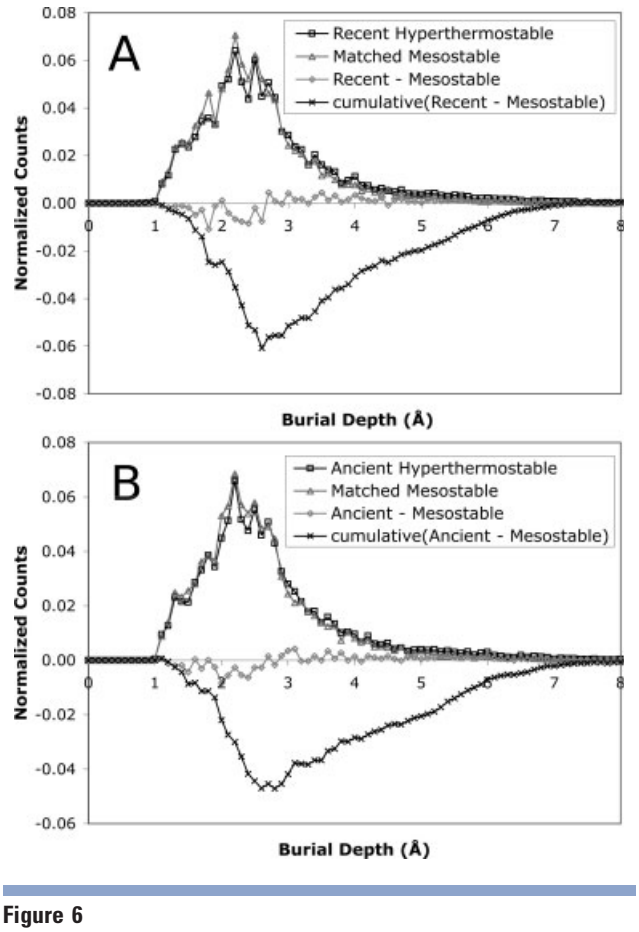

Figure 6

Burial depth in hyperthermostable proteins. The figures show normalized counts of all atoms vs. burial depth for the thermostable and mesostable proteins, the difference in frequency distributions (Thermostable-Mesostable), and the cumulative frequency difference distribution. (**a**) Ancient hyperthermostable vs. matched mesostable proteins. (**b**) Recent hyperthermostable vs. matched mesostable proteins.

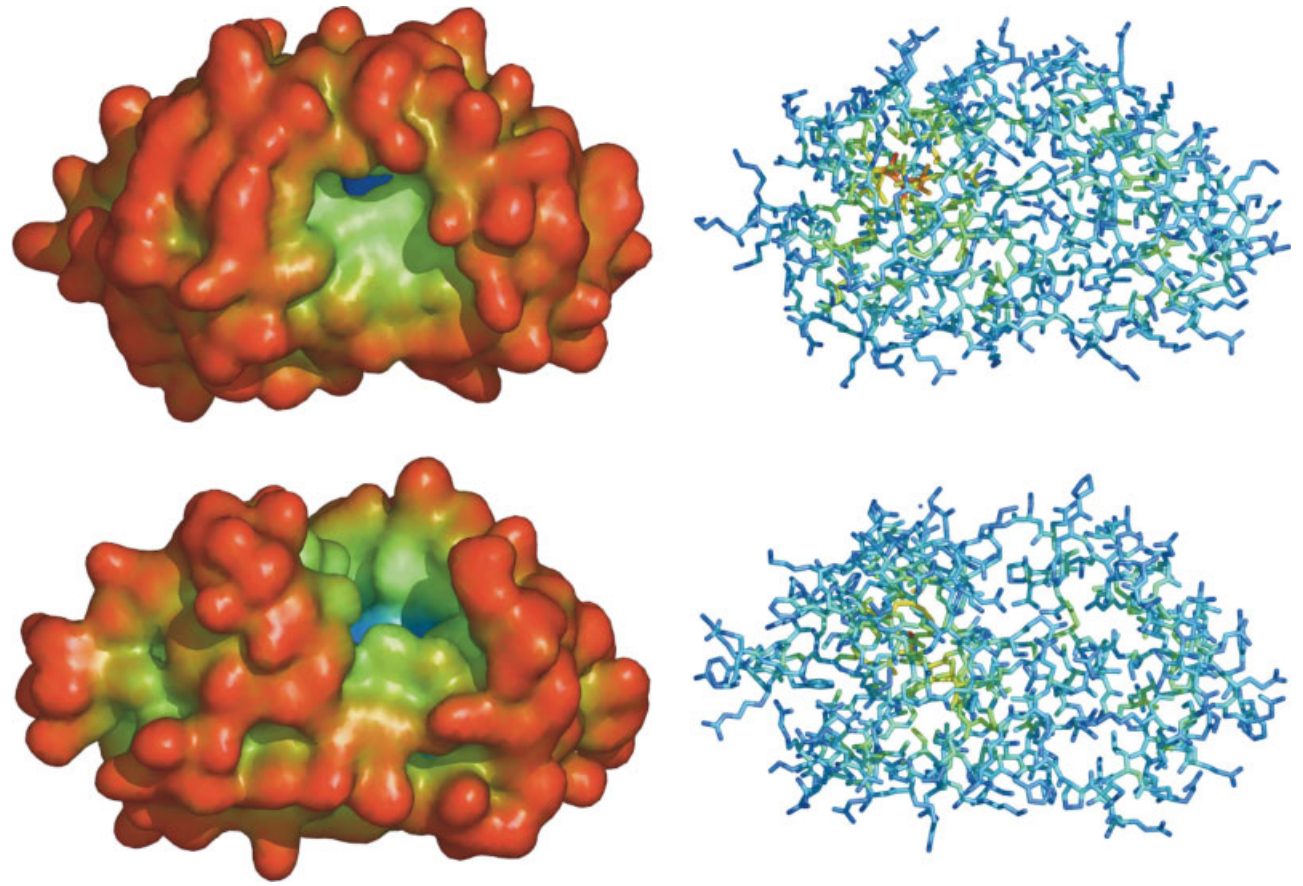

Figure 7

Example structure pair colored by travel depth and burial depth. An example matched pair of protein structures: hyperthermostable Phosphoserine Phosphatase (top, PDB code [PDB:1L7M][55]) and mesostable Phosphoserine Phosphatase (bottom, PDB code [PDB:1NNL][56]). At left, the molecular surface is colored by increasing travel depth from red to green to blue. At right the wireframe representation is colored by increasing burial depth from blue to green to red. Images were generated using a customized PyMOL.[57]

residues are less buried in the hyperthermostable proteins. Given that the correction factors are all positive (all residues on average are more buried in hyperthermostable proteins) it would be more correct to say that these six residues stay unburied in hyperthermostable proteins while all other residues get slightly more buried and alanine is much more buried. These 6 residues are the 4 charged residues (Aspartic Acid, Glutamic Acid, Lysine, and Arginine) as well as Histidine and Asparagine. Note that Histidine can also likely to be charged as the pK_a is near physiological conditions. As Asparagine is chemically labile at high temperatures[6] and may spontaneously deaminate to Aspartate, probably all the residues we find less buried in hyperthermostable proteins are charged. Only one that may be considered charged at high temperatures (Glutamine, chemically labile at high temperatures forming Glutamic Acid) is not less buried. This result is consistent with previous studies indicating increased ion pairs and ion pair networks in thermostable proteins.[5,6,9,10,13–15]

Recent and Ancient hyperthermophiles, as subsets of the entire hyperthermophile set, yield similar results as the latter [Figure 9(c,d), respectively]. Interestingly, though Alanine is more buried in the Ancient class, it is not significant according to the statistical analysis, with a p-value of 0.15. However, four of the above six "charged" amino acids are also significantly less buried in the Ancient class. In the Recent class, Alanine is significantly more buried, the difference being even more pronounced than in the combined data. Again, four of the six "charged" amino acids are significantly less buried, although it is a different four from the Ancient category. We emphasize that these are significant differences in residue burial that do not show up with just a surface/interior binary data analysis.[5,11]

The distribution of burial depth of individual residue types that are significantly more or less buried in hyperthermostable proteins was examined in more detail by comparing the complete probability distribution histogram of burial depths. Results are shown in Figure 10 for

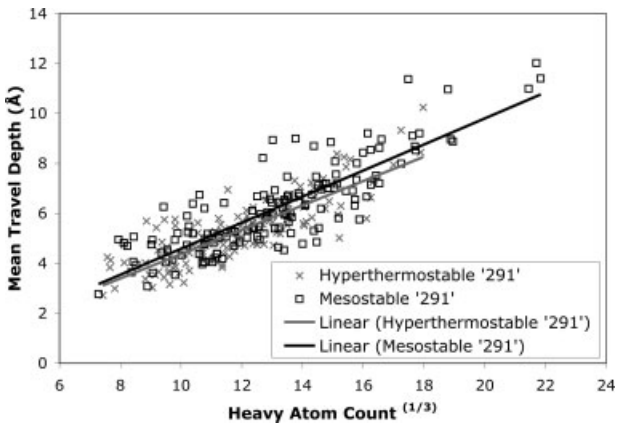

Figure 8

Size-scaled travel depth for the larger hyperthermostable set. Along the *x*-axis is the cube root of the heavy atom count, along the *y*-axis is the mean travel depth. Data for the "291" set of 144 hyperthermostable proteins and matched mesostable proteins is shown along with trendlines. The mean of the ratios for the hyperthermostable proteins is 0.445 and for mesostable proteins the mean is 0.472, the *P*-value of this difference is 5.6×10^{-5}.

just for one especially interesting case, alanine. Alanine was found to be significantly depleted overall in hyperthermostable proteins, while at the same time less exposed.[5] It has been suggested that this relative enrichment of buried alanine is due to the zero side chain entropy cost.[5] Our results also show that alanine is depleted near the surface of hyperthermostable proteins compared to mesostable proteins, and moreover that it is enriched right into the protein core (i.e., at burial depths down 6 Å).

DISCUSSION

We examined 10 different metrics of protein "shape," using three sets of thermophile-mesophile pairs: Moderate thermophiles, ancient hyperthermophiles and recent hyperthermophiles, 67 pairs of proteins in all. Each thermostable protein was matched to the mesostable homologue of similar size, in order to increase the statistical resolution of the comparisons. For the moderate thermophile-mesophile set none of the metrics were significantly different. For the hyperthermophiles, significant differences in several metrics were found, including packing of surface atoms, surface area, Wadell Sphericity, travel depth and burial depth. Of these, only Wadell Sphericity, travel depth and burial depth were significantly different in both recent and ancient hyperthermostable proteins.

Hyperthermostable proteins on average have a higher Wadell sphericity, have fewer and or less deep pockets on their surface, and their residues are on average more deeply buried than in their mesostable counterparts. Taken together, these three metrics provide the first quantitative evidence that hyperthermostable proteins are more spherical. The fact that moderate thermostable proteins show no significant differences in overall shape while hyperthermostable proteins do is in line with previous reports that moderate thermophiles and hyperthermophiles have achieved their necessary thermostability by different mechanisms.[5] We cannot rule out the possibility that some other shape metric would reveal differences between moderate thermostable proteins and their mesostable counterparts. However, given the fact that several of the shape metrics do reveal differences for hyperthermostable proteins, we conclude that adaptation to moderately elevated temperatures requires changes at the individual residue level that need produce little change in gross physical aspects of the proteins to achieve the necessary moderate increase in stability.

We consider now in more detail what the individual metrics reveal. Regarding packing efficiency, the most reliable metric is that for buried atoms, and this shows no significant difference between hyperthermostable and mesostable proteins. The similarity in interior packing is consistent with other analyses.[22] It is interesting to note that the observed increase in hydrogen bonding at higher temperatures does not correlate with increased packing in the interior.[7–10] The packing of surface atoms is greater on average in one class, ancient hyperthermostable proteins and the superclass of all hyperthermostable proteins. This is consistent with some evidence that surface residues have more contacts in thermostable proteins than mesostable proteins.[11] However, the conclusion that in one class surface atoms are better packed must be qualified. The definition of packing of surface atoms is not agreed on, and another definition may lead to different results. This ambiguity is illustrated by considering the absolute values of the packing efficiency. Moreover, surface atom packing is confounded by curvature effects in some definitions. In our method, for example, exposed atoms near convex surfaces will have lower packing than exposed atoms near flat or concave surfaces.

Hyperthermostable proteins in all categories (Ancient, Recent, and Combined) have significantly smaller mean travel depth and mean size-scaled travel depth than their mesostable counterparts, indicating fewer and shallower surface pockets. Using the size-scaled travel depth metric, we also find that hyperthermostable proteins have shallower and fewer surface pockets in a larger set of 144 protein pairs. This is one piece of evidence that hyperthermostable proteins are more spherical. The second piece of evidence for increased sphericity is that all hyperthermostable categories have higher mean burial depth than mesostable proteins, indicating that they bury more atoms overall. This agrees with increased hydrogen bonding,[7–10] increased apolar contact area[17] and increased van der Waals contacts.[7,10,22] This increased contact area that manifests across several types of interactions (hydrogen

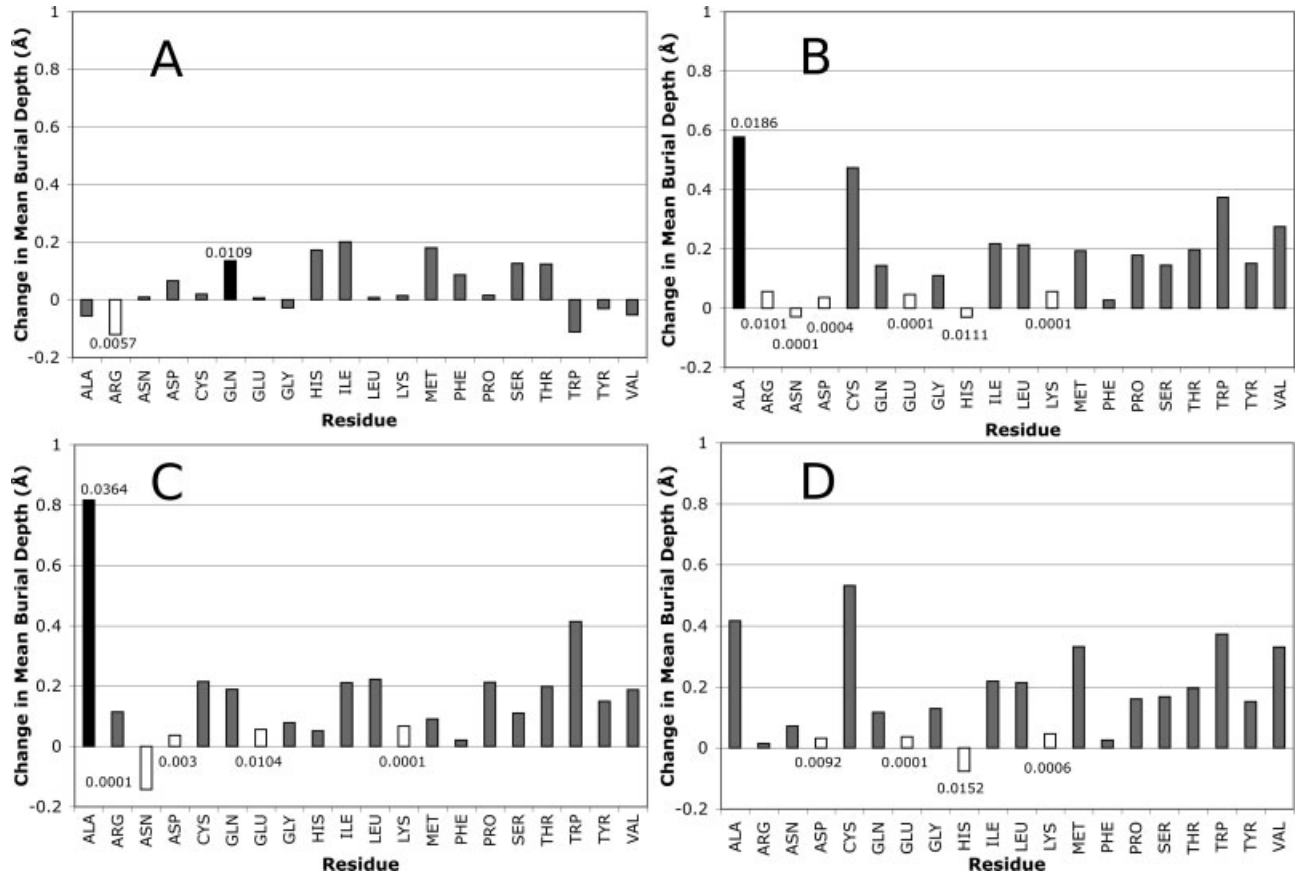

Figure 9

Residue-specific burial depth in thermostable proteins. The difference in burial depth for the C_β atom of each residue type (C_a for Glycine), expressed as Thermostable minus the matched mesostable proteins. Significantly more buried residues are shown in black, significantly less buried residues are shown in white, *P*-values for significant differences are shown above or below each bar. These *P*-values were corrected for the overall burial depth differences seen between each thermostable-mesostable set. (**a**) Moderate thermostable. (**b**) All hyperthermostable. (**c**) Recent hyperthermostable. (**d**) Ancient hyperthermostable.

bonding, apolar, van der Waals) is reflected in the overall increase in burial depth.

The third metric related to overall sphericity of the protein is the Wadell Sphericity measure,[53] which is simply a dimensionless ratio of volume to area, scaled so that its upper bound value of 1 indicates a perfect sphere. This measure is significantly increased for all three hyperthermophile categories: recent, ancient, and combined. We note that this ratio is considerably more sensitive than changes in volume or surface area alone. There is no significant difference in volume for any thermostable category, while a significant difference in area is only seen in ancient and combined hyperthermostable proteins.

The significant differences in these three metrics lead to the conclusion that hyperthermostable proteins are more spherical than their mesostable homologues. This difference is consistent across both ancient and recent hyperthermostable proteins, despite the different evolutionary paths those organisms have used to achieve thermostability.[7]

While mean travel depth, mean burial depth, and Wadell Sphericity all indicate increased sphericity in hyperthermostable proteins, they are by no means synonymous since they each reveal different, complementary, aspects of protein shape. Each is useful in analyzing shapes as complex as those adopted by proteins since each can distinguish some feature that the other cannot. This is illustrated schematically in Figure 11, which depicts two idealized structures with identical volume and surface area (and hence Wadell sphericity), but with different mean travel depth and burial depth. The structure with one large pocket has less deeply buried atoms, and greater mean travel depth than the structure with four smaller pockets. In this case, depth measures are more discriminating than Wadell Sphericity. On the other hand, the mean travel depth of any convex shape is zero, while the Wadell Sphericity (and mean burial depth)

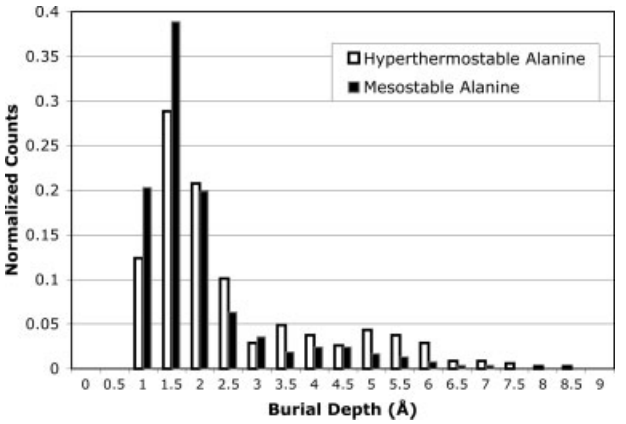

Figure 10

Histogram of burial depth of alanine. Normalized frequency histograms of burial depth of the C_β atom for alanine for mesostable and all hyperthermostable proteins.

vary depending upon the shape, so the latter two would be more discriminating. Burial depth has the additional bonus of being able to examine changes in specific residues, whereas Wadell Sphericity only measures total changes in volume and surface area. Combining information from these complementary metrics can reveal other aspects of shape. Returning to Figure 11, we see that at constant Wadell sphericity the structure with a smaller mean travel depth has a more convoluted, one might say, "rougher" surface. A straightforward measure of the roughness of the protein surface is not possible, however, when both Wadell sphericity and travel depth are different, as in the hyperthermostable-mesostable comparison. Here, as Figure 7 illustrates, the hyperthermostable protein has a smaller mean travel depth, and increased Wadell sphericity, and visually at least, has a less "rough" surface.

The burial depth and travel depth analyses while fast, do require computing and working with a molecular surface and therefore take some significant calculation time. We attempted to come up with a faster measure that would also capture the different in 'spherical property' between mesostable and hyperthermostable proteins. To this end, we calculated the interatomic distances between all pairs of heavy atoms in the proteins and examined the differences in the means. We found no significant differences as shown in Figure 2 and Table I. In retrospect, the failure of the mean interatomic distance to detect differences makes sense as it reflects to a great extent the packing, which is not significantly different. In summary, the failure of interior packing and interatomic distances to differentiate thermostable from mesostable proteins shows that the travel depth and burial depth analyses are necessary to measure the spherical property.

The shapes proteins adopt have profound energetic effects on both charge-solvent and charge-charge interactions, and both travel depth and burial depth report on this. Greater burial depth indicates that more atoms are buried further from solvent. Although many of the deeply buried residues will have apolar side chains, the backbone of each residue is still polar. Burying the backbone partial charges further from solvent lessens their favorable long range electrostatic with the higher dielectric solvent, increasing the desolvation penalty—these charges are less stable. Similarly, a charged group at the bottom of a deep pocket, as measured by travel depth, will on average has less high dielectric solvent near it, and more low dielectric protein than a charge at the bot-

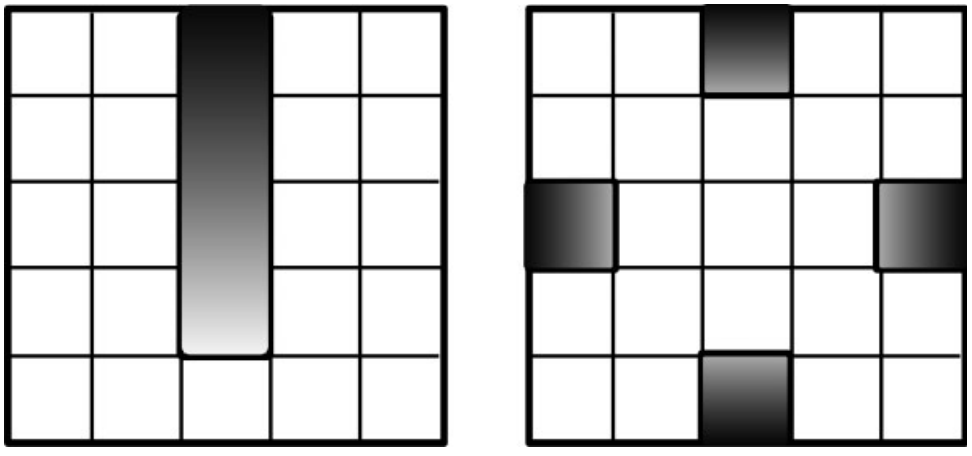

Figure 11

Equal Wadell Sphericity, different travel depth and burial depth example. 2D Schematic indicating two shapes with equal volume and surface area, and hence equal Wadell Sphericity, but different mean travel depths and different mean burial depths. The "U-shaped" volume on the left has a greater mean and maximum travel depth than the "X-shaped" volume on the right. The "X" has a higher mean burial depth, evident simply by observing the center square is not adjacent to the surface whereas all squares in the "U" are adjacent.

tom of a shallow pocket, even though their solvent accessible surface areas are the same. Charges at the bottom of a deep pocket have a greater desolvation penalty and therefore are less stable. This is a manifestation of electrostatic focusing.[56] Considering now a favorable charge–charge interaction, increased burial depth or increased travel depth will strengthen it, thereby having the opposite effect on the protein's stability. The reason is the same. There is less effective solvent interaction with the deeper charges, so less dielectric screening. Fine tuning and balancing these competing effects is one possible way for stability to be controlled in the transition between mesophile and thermophile, and a reason why travel depth and burial depth show significant differences.

Examination of the burial depth of specific residues (see Figure 9) adds another dimension to the previous surface area change analyses,[5,6,9,19,20] which measured changes in solvent accessible surface area. Surface accessibility analysis on the same dataset used here led to the conclusion that Alanine and Proline have less surface nonpolar area exposed in both moderate thermostable and hyperthermostable proteins, and that phenylalanine, methionine, tyrosine and tryptophan have more exposed nonpolar surface area in the higher temperature classes.[5] In contrast, we find here that Alanine is buried significantly more deeply in the hyperthermostable proteins, while there is no appreciable change in the moderate thermostable proteins. Proline is more buried in hyperthermostable proteins, but not significantly more buried after correcting for overall burial, and again, no appreciable change is seen in the moderate thermostable proteins. Our disagreement with the result of the four residues having more exposed nonpolar surface area could be due to two factors: The previous analysis was only of the nonpolar surface area and changes in surface area may not be directly comparable to counting residues at each burial depth. For instance a residue on the surface in our analysis will have the same burial depth no matter the local environment, however a concave region could result in less surface area where a convex region could result in more area.

Our examination of charged versus polar surface area shows that there are indeed differences between the thermostable and mesostable proteins, as first indicated by Cambillau et al.[19] Charged residues are less buried in hyperthermostable proteins and half the polar residues (Serine and Threonine) are more buried, consistent with observed changes in surface area. However, the other polar residues (Asparagine and Glutamine) are significantly less buried, which disagrees with the surface area results. Since surface area changes are analyzed as percentages, difference overall surface areas between hyperthermostable and mesostable proteins does not account for this. These unburied residues, even though they are higher in number, may in fact expose less surface area. Regardless of the disagreement on polar residues, charged residues show large changes consistent across both surface area[19,5] and burial depth analyses.

In the moderate thermostable proteins, shown in Figure 9(a), the changes are the least pronounced of any of the four categories examined. However, two residues still show up as significantly different: Glutamine is more buried and arginine is less buried, something that could not be detected in previous analysis of nonpolar surface area.

Our residue-specific results can most easily be compared to results that used an interior/exterior definition based on surface area and then counted residues in these classes to see how they differed between thermophiles and mesophiles.[11] In that study, the interior counts showed very little changes, while the exterior showed many changes, though the significances were not analyzed statistically. Lysine, arginine, and glutaminic acid had increased exterior numbers in thermostable proteins, which our analysis would agree with as those residues remain unburied or become less buried in hyperthermostable proteins. Alanine, asparagine, aspartic acid, glutamine, threonine, serine, and histidine all have less exterior residues in thermostable proteins. We agree on alanine's decreased exterior presence (and increased burial), but disagree on the other residues found less frequently on the exterior. This could be due to the differing data sets, use of a full spectrum of burial depths vs. a binary cutoff, or statistical variation.

A possible application for the burial depth preferences in hyperthermostable proteins is thermostable protein design. The amino acid burial preferences could be incorporated into models and design strategies. Additionally, using mesostable protein structures as a starting point, a design pipeline could incorporate the travel depth and burial depth analyses of the spherical property to find structures that bury more atoms and have fewer/smaller pockets. Obviously, a good protein design strategy is required as a starting point as it is not just the spherical property that ensures a protein is highly thermostable. A suggested pipeline would be to find many backbones,[60] repack the native mesostable sidechains and mutations chosen from residue-specific burial depth changes,[61] then evaluate the many possibilities to find proteins that are spherical but preserve the active site or desired function. Then traditional protein design tools, such as,[62] could be used to find further mutations that enhance the stability of the new backbone, and could be modified to include residue burial depth preferences of hyperthermostable proteins. This method is obviously an addition to already existing approaches for thermostable protein design.[63–66]

When interpreting differences between thermostable and mesostable proteins, organism sources should be considered, as temperature is not the only difference between these organisms. In the hyperthermophilic set of

30 structures, 9 organisms are represented. At least one is a piezophile, but many are not, so it is doubtful that we are seeing results from changes due to adaptation to high pressure. However, all these hyperthermophiles are unicellular and many of the mesophile homologues come from multicellular organisms. At this point there is no evidence that this systematic difference is reflected at the level of protein structure, but it is a caveat nevertheless.

Another important caveat of our analysis (indeed of any type of analysis of the PDB database) is the experimental temperature at which the structures were determined. Protein crystal structures are now almost always solved at extremely low temperatures (ca. 130 K). Even older structures or typical NMR structures are solved at room temperature, far from the environmental temperatures of hyperthermophiles. This could have several effects. Obviously at higher temperatures, the configurational entropy of the side chains will be higher and will explore more states. This could have some influence on hyperthermostability, as demonstrated by simulations showing that the charged residues in a hyperthermostable protein are able to interact cooperatively during the conformational fluctuations.[67] Our results that hyperthermostable proteins are more spherical could indicate a preference for finding conformations that have reduced flexibility, since the higher number of interactions will limit the number of states available, consistent with reduced rotameric states in hyperthermostable proteins.[5] These results are also consistent with results showing that mutations that support hyperthermostability are distributed throughout the protein and cause subtle changes in dynamics and distributed changes in stability.[68,69] However, without structures solved at the ambient temperatures for mesophiles and hyperthermophiles, it is difficult to say how our results would be affected. The effect of experimental conditions on structures is a caveat of any research based on PDB structures.[26]

Finally, in any discussion of thermostable proteins, it is important to note that the language used throughout almost all the literature (and in this work!) implies that previously mesophilic organisms have adapted to higher temperatures resulting in the hyperthermostable proteins. This is probably not the case[1,2] except in specific cases like *T. maritima*.[3,25] If the 'hot' origin of life theory is correct, then a common ancestor organism for these proteins was a hyperthermophile, though some organisms (and therefore their proteins) adapted to mesophilic conditions and then re-adapted to hyperthermophilic conditions.[3,25] While this does not affect the observed differences and their statistical significance, this "meso-centric" view does shade their evolutionary interpretation. Most mesostable proteins whose temperature dependence of stability has been examined in detail show a maximum in stability not too far from their working temperature with a substantial decrease in stability with increasing temperature above 45°C. This tends frame the question in terms of how this stability profile is changed in thermostable proteins to ensure stability at high temperatures. This could be achieved by (a) shifting the maximum of stability to a higher temperature, (b) being more stable at all temperatures, (c) reducing the rate at which stability decreases with temperature (or some combination of these effects). Viewed, however, from the perspective of the thermophile as the precursor, cases (b) and (c) present no problem in adaptation to mesophilic temperatures, since at these lower temperatures the thermostable protein is already stable. In this scenario, there would be no selective pressure, and one would not expect to see pervasive stability related structure changes of the type observed here. In case (a) however, presumably the stability of a thermostable protein at mesophilic temperatures would be low enough so that there would be selective pressure to adapt to lower temperatures, leading to significant stability correlated structure changes. Of course proteins need enough flexibility and dynamics to function, and cases (b) and (c) may result in too much stability at mesophilic temperature for optimal function, in which case again there would be selective pressure. In a recent review of available experimental evidence, hyperthermostable proteins used case (b) most often, often combined with case (a), whereas moderate thermostable proteins used case (b) often combined with case (c), however there is still not a lot of data available.[70] Considerably more data on the temperature stability profiles of matched mesostable-thermostable pairs is needed to distinguish these cases.

CONCLUSIONS

Protein structures from homologous mesophiles and hyperthermophiles have diverged due to evolution. Regardless of the age of the adaptation to hyperthermophilic conditions, the proteins adopt a more spherical structure, namely they have greater burial depth, lesser travel depth, and higher Wadell sphericity than their mesostable counterparts. The interiors of these hyperthermostable proteins are not more tightly packed, probably because mesostable proteins are already packed to near crystalline tightness. Rather these proteins have residue side-chain replacements and structural rearrangements that produce more spherical proteins. These changes are not detectable by other properties like mean interatomic distance or convex hull volume. The new metrics of travel depth and burial depth analyses are necessary to quantify the spherical property and complement Wadell Sphericity. All three metrics are applied here to proteins for the first time. In contrast to hyperthermostable proteins, moderate thermostable proteins do not show any significant differences in sphericity metric from their mesostable homologues. Moderate thermostable proteins adaptations to stability clearly do not drive

them to more spherical structures. In this way, our results support the hypothesis that moderate thermophiles and hyperthermophiles achieve the enhanced stability of their proteins by different mechanisms.

Additionally, by adding a new dimension to specific residue analysis, distance of burial instead of the binary buried/exposed metric, key observations about hyperthermostable proteins can be made, specifically that charged residues stay unburied, alanine is considerably more buried and the rest of the amino acids become slightly more buried.

ACKNOWLEDGMENTS

The authors thank the anonymous reviewers for their useful suggestions.

REFERENCES

1. Stetter KO. Hyperthermophiles in the history of life. Philos Trans R Soc B 2006;361:1837–1843.
2. Nisbet EG, Sleep NH. The habitat and nature of early life. Nature 2001;409:1083–1091.
3. Puigbòa P, Pasamontesa A, Garcia-Vallve S. Gaining and losing the thermophilic adaptation in prokaryotes. Trends Genet 2008;24:10–14.
4. Atomi H. Recent progress towards the application of hyperthermophiles and their enzymes Curr Opin Chem Biol 2005;9:166–173.
5. Greaves RB, Warwicker J. Mechanisms for stabilisation and the maintenance of solubility in proteins from thermophiles. BMC Struct Biol 2007;7:1–23.
6. Yano JK, Poulos TL. New understandings of thermostable and peizostable enzymes. Curr Opin Biotech 2003;14:360–365.
7. Berezovsky IN, Shakhnovich EI. Physics and evolution of thermophilic adaptation. Proc Natl Acad Sci USA 2005;102:12742–12747.
8. Vogt G, Argos P. Protein thermal stability: hydogren bonds or internal packing? Fold Des 1997;2:S40–S46.
9. Vogt G, Woell S, Argos P. Protein thermal stability, hydrogen bonds, and ion pairs. J Mol Biol 1997;269:631–643.
10. Szilágyi A, Závodszky P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. Structure 2000;8:493–504.
11. Glyakina AV, Garbuzynskiy SO, Lobanov MY, Galzitskaya OV. Different packing of external residues can explain differences in the thermostability of proteins from thermophilic and mesophilic organisms. Bioinformatics 2007;23:2231–2238.
12. Robinson-Rechavia M, Alibésb A, Godzik A. Contribution of electrostatic interactions. Compactness and quaternary structure to protein thermostability: lessons from structural genomics of *Thermotoga maritime*. J Mol Biol 2006;356:547–557.
13. Karshikoff A, Ladenstein R. Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads. Trends Biochem Sci 2001;26:550–556.
14. Xiao L, Honig B. Electrostatic contributions to the stability of hyperthermophilic proteins. J Mol Biol 1999;289:1435–1444.
15. Alsop E, Silver M, Livesay DR. Optimized electrostatic surfaces parallel increased thermostability. Prot Eng 2003;16:871–874.
16. Capistran-Licea VM, Millan-Pacheco C, Pastor N. Thermal adaptation strategies used by TBP [abstract]. Biophys J 2009;96:331a.
17. Paiardini A, Sali R, Bossa F, Pascarella S. "Hot cores" in proteins: comparative analysis of the apolar contact area in structures from hyper/thermophilic and mesophilic organisms BMC Struct Biol 2008;8:14–27.
18. Rader AJ. Thermostabilization due to rigidity: a case study of rubredoxin [abstract]. Biophys J 2009;96:330a.
19. Cambillau C, Claverie J-M. Structural and genomics correlates of hyperthermostability. J Biol Chem 2000;275:32383–32386.
20. Fukuchi S, Nishikawa K. Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. J Mol Biol 2001;309:835–843.
21. Lin Y-S. Using a strategy based on the concept of convergent evolution to identify residue substitutions responsible for thermal adaptation. Proteins 2008;73:53–62.
22. Karshikoff A, Ladenstein R. Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. Protein Eng 1998;11:867–872.
23. Coleman RG, Sharp KA. Travel depth, a new shape descriptor for macromolecules: application to ligand binding. J Mol Biol 2006;362:441–458.
24. Ogata Y, Imai E-I, Honda H, Hatori K, Matsuno K. Hydrothermal circulation of seawater through hot vents and contribution of interface chemistry to prebiotic synthesis. Orig Life Evol Biosph 2000;30:527–537.
25. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM. Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature 1999;399:323–329.
26. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucl Acid Res 2000;28:235–242.
27. Bondi A. van der Waals volumes and radii. J Phys Chem 1964;68:441–451.
28. Richards FM. The interpretation of protein structures: Total volume, group volume distributions and packing density. J Mol Biol 1974;82:1–14.
29. Tsai J, Taylor R, Chothia C, Gerstein M. The packing density in proteins: standard radii and volumes. J Mol Biol 1999;290:253–266.
30. Poupon A. Voronoi and Voronoi-related tessellations in studies of protein structure and interaction. Curr Opin Struct Biol 2004;14:233–241.
31. Levitt M, Gerstein M, Huang ES, Subbiah S, Tsai J. Protein folding: the endgame. Ann Rev Biochem 1997;66:549–579.
32. Gerstein M, Lynden-Bell RM. What is the natural boundary of a protein in solution? J Mol Biol 1993;230:641–650.
33. Gerstein M, Tsai J, Levitt M. The volume of atoms on the protein surface: calculated from simulation, using Voronoi Polyhedra. J Mol Biol 1995;249:955–966.
34. Voronoi GF. Nouveles applications des paramétres continus á la théorie de formes quadratiques. J Reine Angew Math 1908;134:198–287.
35. de Berg M, van Kreveld M, Overmars M, Schwarskopf O. Computational geometry: algorithms and applications. Berlin: Springer; 2000.
36. Liang J, Dill KA. Are proteins well-packed? Biophys J 2001;86:751–766.
37. Gerstein M, Chothia C. Packing at the protein-water interface. Proc Natl Acad Sci USA 1996;93:10167–10172.
38. Dijkstra EW. A note on two problems in connexion with graphs. Numer Math 1959;1:269–271.
39. Coleman RG, Sharp KA. Finding and characterizing tunnels in macromolecules with application to ion channels and pores. Biophys J 2009;96:632–645.
40. Pintar A, Carugo O, Pongor S. DPX: for the analysis of the protein core. Bioinformatics 2003;19:313–314.
41. Pintar A, Carugo O, Pongor S. Atom depth as a descriptor of the protein interior. Biophys J 2003;84:2553–2561.

42. Varrazzo D, Bernini A, Spiga O, Ciutti A, Chiellini S, Venditti V, Bracci L, Niccolai N. Three-dimensional computation of atom depth in complex molecular structures. Bioinformatics 2005;21: 2856–2860.
43. Chakravarty S, Varadarajan R. Residue depth: a novel parameter for the analysis of protein structure and stability. Structure 1999;7:723–732.
44. Yuan Z, Wang Z-X. Quantifying the relationship of protein burying depth and sequence Proteins 2007;70:509–516.
45. Liu S, Zhang C, Liang S, Zhou Y. Fold recognition by concurrent use of solvent accessibility and residue depth. Proteins 2007;68:636–645.
46. Pintar A, Carugo O, Pongor S. Atom depth in protein structure and function. Trends Biochem Sci 2003;28:593–597.
47. Huang ES, Subbiah S, Levitt M. Recognizing native folds by the arrangement of hydrophobic and polar residues. J Mol Biol 1995; 252:709–720.
48. Pereira de Araújo AF, Gomes ALC, Bursztyn AA, Shakhnovich EI. Native atomic burials, supplemented by physically motivated hydrogen bond constraints, contain sufficient information to determine the tertiary structure of small globular proteins. Proteins 2007; 70:971–983.
49. Ben-Shimon A, Eisenstein M. Looking at enzymes from the inside out: the proximity of catalytic residues to the molecular centroid can be used for detection of active sites and enzyme-ligand interfaces. J Mol Biol 2005;351:309–326.
50. del Sol A, Fujihashi H, Amoros D, Nussinov R. Residue centrality, functionally important residues, and active site shape: analysis of enzyme and non-enzyme families. Protein Sci 2006;15:2120–2128.
51. Amitai G, Shemesh A, Sitbon E, Shklar M, Netanely D, Venger I, Pietrokovski S. Network analysis of protein structure identifies functional residues. J Mol Biol 2004;344:1135–1146.
52. Brinda KV, Vishveshwara S. A network representation of protein structures: implications for protein stability. Biophys J 2005;89: 4159–4170.
53. Wadell H. Volume, shape and roundness of quartz particles. J Geol 1935;43:250–280.
54. Fisher RA. "The Coefficient of Racial Likeness" and the future of craniometry. J R Anthropol Inst 1936;66:57–63.
55. Wang W, Cho HS, Kim R, Jancarik J, Yokota H, Nguyen HH, Grigoriev IV, Wemmer DE, Kim SH. Structural characterization of the reaction pathway in phosphoserine phosphatase: crystallographic "snapshots" of intermediate states. J Mol Biol 2002;319:421–431.
56. Peeraer Y, Rabijns A, Verboven C, Collet JF, Van Schaftingen E, De Ranter C. High-resolution structure of human phosphoserine phosphatase in open conformation. Acta Crystallogr D Biol Crystallogr 2003;59:971–977.
57. DeLano WL. The PyMOL molecular graphics system. San Carlos, California: DeLano Scientific; 2002.
58. Yee A, Chang X, Pineda-Lucena A, Wu B, Semesi A, Le B, Ramelot T, Lee GM, Bhattacharyya S, Gutierrez P, Denisov A, Lee CH, Cort JR, Kozlov G, Liao J, Finak G, Chen L, Wishart D, Lee W, McIntosh LP, Gehring K, Kennedy MA, Edwards AM, Arrowsmith CH. An NMR approach to structural proteomics. Proc Natl Acad Sci USA 2002;99:1825–1830.
59. Klapper I, Hagstrom R, Fine R, Sharp K, Honig B. Focusing of electrical fields in the active site of Cu-Zn superoxide dismutase: effects of ionic strength and amino-acid modifications. Proteins 1986;1:47–59.
60. Yang Q, Sharp KA. Building alternate protein structures using the elastic network model. Protein Sci 2009;74:682–700.
61. Canutescu AA, Shelenkov AA, Dunbrack RL, Jr, Canutescu AA,Shelenkov Dunbrack RL, Jr. A graph theory algorithm for protein side-chain prediction. Protein Sci 2003;12:2001–2014.
62. Kono H, Saven JG. Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure. J Mol Biol 2001;306:607–628.
63. Shah PS, Hom GK, Ross SA, Lassila JK, Crowhurst KA, Mayo SL. Full-sequence Computational design and solution structure of a thermostable protein variant. J Mol Biol 2007;372:1–6.
64. Lehmann M, Wyss M. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution Curr Opin Biotech 2001;12:371–375.
65. van der Burg B, Eijsink VGH. Selection of mutations for increased protein stability Curr Opin Biotech 2002;13:333–337.
66. DiTursi MK, Kwon S-J, Reeder PJ, Dordick JS. Bioinformatics-driven, rational engineering of protein thermostability. Protein Eng Des Sel 2006;19:517–524.
67. Danciulescu C, Ladenstein R, Nilsson L. Dynamic arrangement of ion pairs and individual contributions to the thermal stability of the cofactor-binding domain of glutamate dehydrogenase from *Thermotoga maritima*. Biochem 2007;46:8537–8549.
68. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature 2007;450:913–916.
69. Hollien J, Marqusee S. Structural distribution of stability in a thermophilic enzyme. Proc Natl Acad Sci USA 1999;96:13674–13678.
70. Luke KA, Higgins CL, Wittung-Stafshede P. Thermodynamic stability and folding of proteins from hyperthermophilic organisms. FEBS 2007;274:4023–4033.