

## Research Article

# Widespread Disulfide Bonding in Proteins from Thermophilic Archaea

Julien Jorda<sup>1</sup> and Todd O. Yeates<sup>1,2</sup>

<sup>1</sup>UCLA-DOE Institute for Genomics and Proteomics, 611 Charles Young Drive East, Los Angeles, CA 90095, USA

<sup>2</sup>Department of Chemistry and Biochemistry, University of California, Los Angeles, 611 Charles Young Drive East, Los Angeles, CA 90095, USA

Correspondence should be addressed to Todd O. Yeates, yeates@mbi.ucla.edu

Received 17 May 2011; Accepted 16 July 2011

Academic Editor: M. Adams

Copyright © 2011 J. Jorda and T. O. Yeates. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disulfide bonds are generally not used to stabilize proteins in the cytosolic compartments of bacteria or eukaryotic cells, owing to the chemically reducing nature of those environments. In contrast, certain thermophilic archaea use disulfide bonding as a major mechanism for protein stabilization. Here, we provide a current survey of completely sequenced genomes, applying computational methods to estimate the use of disulfide bonding across the Archaea. Microbes belonging to the Crenarchaeal branch, which are essentially all hyperthermophilic, are universally rich in disulfide bonding while lesser degrees of disulfide bonding are found among the thermophilic Euryarchaea, excluding those that are methanogenic. The results help clarify which parts of the archaeal lineage are likely to yield more examples and additional specific data on protein disulfide bonding, as increasing genomic sequencing efforts are brought to bear.

## 1. Introduction

The archaea inhabit incredibly diverse environments [1]. Many species thrive at temperatures exceeding 100°C. Growth at such high temperatures presents special challenges, among the most serious being the problem of stabilizing cellular proteins in their natively folded configurations. For many proteins, the folded configuration is only modestly favored energetically compared to the unfolded state [2], and high temperatures irreversibly unfold the vast majority of proteins derived from organisms that live at moderate temperatures. The question of how thermophilic proteins are stabilized has therefore attracted considerable attention over the years [3, 4].

Numerous studies have concluded that thermophilic proteins are stabilized by a wide array of forces and effects, which appear to present themselves to different degrees in different proteins and organisms [3, 5–8]. Increased atomic packing [9, 10], hydrophobic interactions [11], ionic interactions [9, 12–14], and shorter loops [15] have all been

noted as providing additional noncovalent stabilization in thermophilic proteins. More unexpected was the realization that disulfide bonding—a much stronger, covalent force—might play an important role in some organisms [16, 17]. A striking clue came when the structure of the enzyme adenylosuccinate lyase from the hyperthermophilic *Pyrobaculum aerophilum* revealed that the six cysteines in the protein chain pair up to form three disulfide bonds [17]. This prompted the development by Mallick et al. [16] of genomic calculations, which supported the idea that some thermophiles use disulfide bonding as a major mechanism for protein stabilization [16, 18, 19]. Subsequent proteomic experiments on *P. aerophilum* validated that claim [20], as have recently published structures [21–23] and biochemical studies [24–26] of proteins from various hyperthermophilic archaea.

The use of disulfide bonding came as a surprise, because in well-studied organisms the intracellular environment is chemically reducing, and therefore favors the thiol form of cysteines over the disulfide form (reviewed in [27]).

Though disulfide bonds are a common mechanism for stabilizing proteins that are either secreted or reside in oxidizing extracytosolic compartments, thermodynamic considerations prevent disulfide bonds from conferring protein stability under reducing conditions. This is a general rule, notwithstanding the existence of varied cytosolic proteins that form disulfide bonds transiently or reversibly, as part of cellular redox signaling mechanisms, for example [28–30]. The prevalent use of disulfides therefore brought up new questions about the intracellular environments of archaea, and the molecular mechanisms for forming protein disulfide bonds within the cytosol. Comparative genomics studies showed that a protein known as protein disulfide oxidoreductase (PDO) was present in thermophiles, and selectively in organisms predicted to be rich in intracellular disulfide bonding [18, 31]. This helped focus attention on PDO as the presumptive key player in intracellular protein disulfide bonding (reviewed in [32–34]), a role that is consistent with *in vitro* studies on PDO from multiple thermophiles [35–37].

The importance of disulfide bonding in thermophiles emerged when complete genomes were known for only about 25 unique prokaryotes, of which seven were archaea [16]. There are presently 1031 completely sequenced prokaryotic genomes with accompanying proteomes available at the National Center for Biotechnology Information web server. Though archaeal species constitute an unfortunately small fraction of this set—90 out of 1031—their growing number provides an opportunity for an updated assessment of thermophilic protein disulfide bonding in this important and diverse branch of the tree of life.

## 2. Results and Discussion

Protein sequences from 90 complete archaeal proteome sets were obtained from the UniProt web server, release 2011-4. Sequences were also retrieved for several viruses infecting *Sulfolobus* species, and viruses infecting *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, and *Thermoproteus tenax*, along with data for five moderately thermophilic eubacteria: *Thermotoga maritima*, *Aquifex aeolicus*, *Streptococcus thermophilus*, *Thermosiphon melanesiensis*, and *Thermobaculum terrenum*. In addition, metagenomic sequence data from archaeal-rich microbial communities sampled at geothermal springs in Yellowstone National Park were obtained from Inskeep et al. [38].

We applied both of the methods introduced by Mallick et al. [16] for analyzing disulfide bonding in the present study. The first is based on an enumeration of proteins having an even versus an odd number of cysteine residues. If an organism has a strong tendency for all or most of its cysteines to be paired into disulfide bonds, then one can expect to see an overrepresentation of proteins with an even number of cysteines. Indeed, such a trend is clear for several of the hyperthermophilic archaea examined (Figure 1). No such preference is seen in control calculations involving nonthermophilic bacterial genomes, as long as care is taken to filter out proteins destined for export from the cytosol; the formation of disulfide bonds in the oxidizing environment

of the bacterial periplasm has been extensively studied, and in fact an analysis of even versus odd cysteines has been used to illustrate the abundance of disulfide bonds in secreted proteins or those destined for the bacterial periplasm [39, 40]. Although the simple cysteine counting approach gave clear results in the present study for a number of hyperthermophiles, the pattern was less clear for some organisms, including some of the *Sulfolobus* species, despite earlier data indicating that disulfides should be abundant [18]. It is notable that some archaea, including *Sulfolobus*, are relatively rich in metalloproteins [41], whose metal sites are often coordinated by cysteines, potentially hampering an accurate prediction of disulfide abundance based on a simple counting of cysteine residues. This prompted an alternate analysis based on sequence-structure mapping.

A second method for disulfide bond analysis relies on the availability of known three-dimensional protein structures, not of the specific genomic proteins in question, but of homologous proteins from other organisms (Figure 2). The underlying principle is that if a given protein sequence (from a hyperthermophilic archaeon for example) has two cysteines that form a disulfide bond in the folded configuration, then when that query sequence is mapped or overlaid onto the structure of a homologous protein, the two cysteines in the query sequence should be nearby in space, as would be required if a bond between them was present [16]. The specific value,  $f$ , that we report in the present study is the fraction of all cysteine residues (among those ( $m$ ) that could be mapped onto structures) that fall within 8 Å (C-alpha to C-alpha) of some other cysteine residue in the modeled structure. We refer to this fraction,  $f$ , as the predicted disulfide abundance parameter. Whereas the cysteine counting method suffers from the oversimplified notion that all the cysteines in a protein must be disulfide bonded, the sequence-structure mapping method presents other difficulties. Only partial coverage can be expected, since many query protein sequences will not be represented by homologous structures in the PDB. In addition, in cases where a homologous structure is available, substantial evolutionary divergence between the query and the target protein can lead to unreliable alignments as well as to bona fide structural differences, both of which tend to reduce the likelihood that the cysteines will appear in close proximity as required. Nonetheless, the method provides the advantage of specificity in three dimensions.

When applied to the collection of genomic data, the sequence-structure mapping method provides a clear indication of disulfide richness for many of the thermophilic archaea (Figure 3; see Table 1-Supplementary Material available online at doi:10.1155/2011/409156). This is also the case for the *Sulfolobus* species noted above, whose analysis had been unclear by the simple cysteine counting method. In all, roughly 33 archaeal genomes (not counting closely related strains of the same genus) are judged to have significant amounts of disulfide bonding ( $f > 0.15$ ) while smaller subsets show even higher values (21 genomes with  $f > 0.25$ , and 8 with  $f > 0.35$ ) (See Supplementary Table S1). In addition, the unusual, moderately thermophilic eubacteria have a detectable but lower fraction of their cysteines in

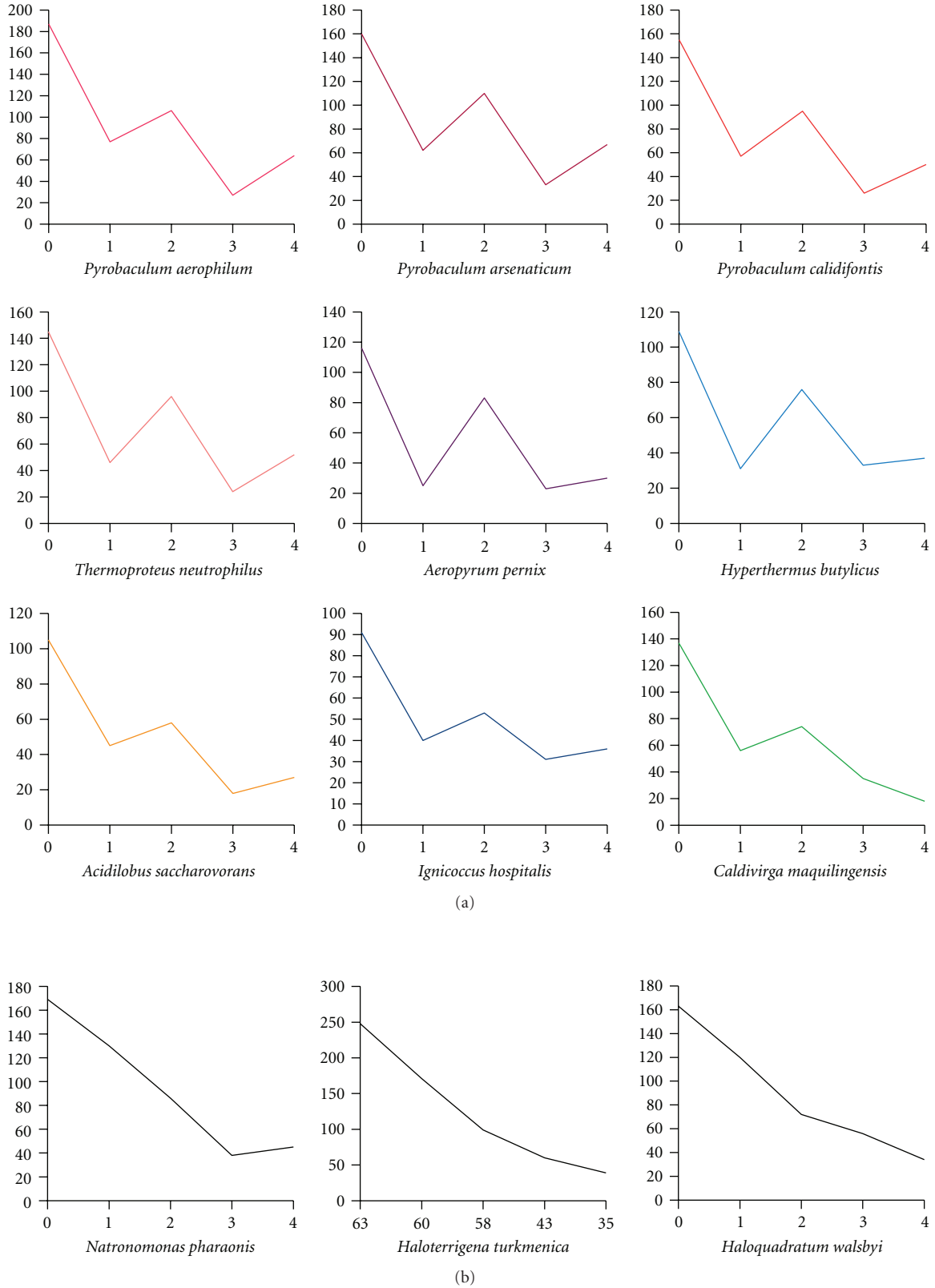


FIGURE 1: A preference for an even versus odd number of cysteines in proteins from thermophilic archaea. The dataset used for these plots consists of proteins with sizes ranging from 150 to 200 amino acids, the expected trend being more apparent for this class of proteins. (a) Numerous hyperthermophilic and thermophilic archaea show a clear propensity for even numbers of cysteine residues. This trend suggests an abundance of disulfide bonds. Nine examples are shown. (b) Selected nonthermophilic species (all halophiles) are shown as controls. In these cases, the plotted lines are nearly monotonic, indicating an absence of significant disulfide bonding in the halophiles.

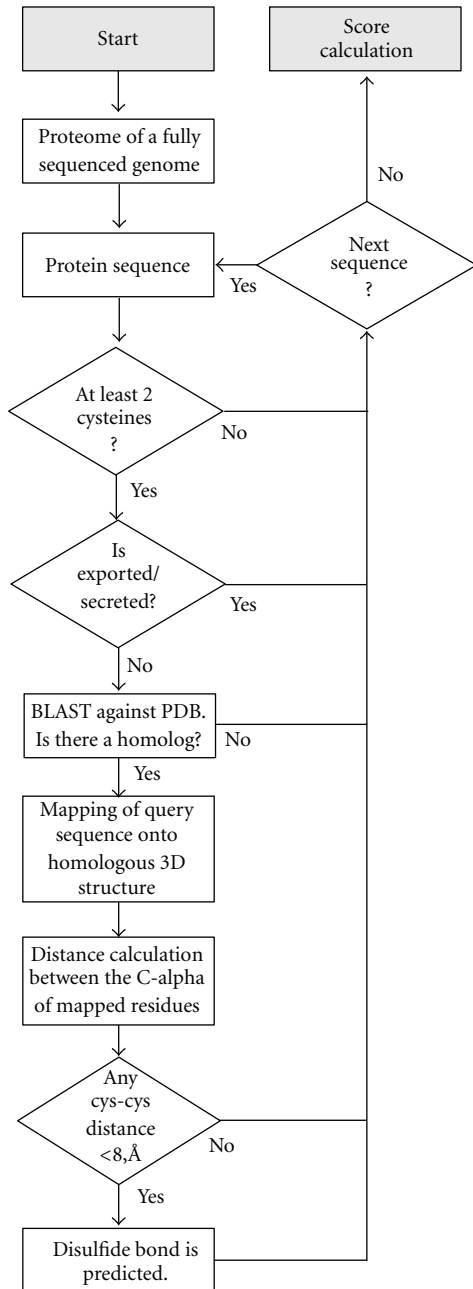


FIGURE 2: Flowchart illustrating the procedure for mapping genomic sequence data onto known three-dimensional protein structures in order to estimate disulfide abundance. Grey boxes are the starting and ending points of the pipeline, white boxes are processing steps, and diamonds are decisions made subsequent to filtering steps. The procedure loops until all the proteins from a given proteome are processed.

disulfide bonds than archaeal thermophiles, with the exception of *Aquifex aeolicus*, which stands out among this group. This result is in accordance with previous studies highlighting that *Aquifex* has a higher fraction of proteins in common with archaeal thermophiles than any other nonthermophilic eubacteria [42]. As a final source of thermophilic sequences, we performed an initial analysis, which should

be considered preliminary (see Section 4), of metagenomic sequence data derived from archaeal-rich hot springs in Yellowstone National Park [38]. Overall, these data showed strong evidence for disulfide bonding. For the cellular DNA sample from Nymph Lake (site 10, August 2009),  $f$  was 0.29 while for the sample from Crater Hills (September 2009), the value was 0.35, which is comparable to the highest values obtained for individually sequenced genomes so far.

Specific cases were investigated where disulfide bonding in a genome was evident by the sequence-structure mapping method, but was missed by the cysteine counting method. Several of the proteins from *Sulfolobus islandicus* strain M.14.25 that have exactly three cysteine residues were examined. Of the 771 sequences that could be mapped to a PDB homologue, 92 had three cysteines that could be mapped onto the structure. Among them, 53 were predicted to have a disulfide bond, and one cysteine thiol apart. Analysis of mapped structures showed that the free cysteine tended to be either involved in a putative metal-binding site or poised on the surface. The latter scenario is suggestive of potential intermolecular disulfide bonds. That hypothesis is strengthened by the observation that cases with a third, exposed cysteine occurred often in structures that were oligomeric. Two illustrative examples—one involving a likely intermolecular disulfide and one involving a likely metal binding site—are shown in Figure 4. These situations highlight a key difficulty with the cysteine counting method: proteins having an odd number of cysteines may actually represent positive cases of disulfide bonds. Furthermore, even in disulfide rich organisms where free (reduced) cysteine thiols are systematically diminished, odd numbers of cysteines may be present in proteins that form intermolecular disulfide bonds. Proteomic experiments on *P. aerophilum*, using oxidized versus reduced 2D SDS gel electrophoresis, have highlighted the abundance of intermolecular disulfide bonding in that organism [20]. Crystal structures of proteins from other thermophiles further support this point [23, 43, 44].

The predicted abundance of protein disulfides across the Archaea correlates strongly with phylogenetic divisions, but there are also notable trends within specific branches (Figure 3). Essentially all of the Crenarchaea show very high levels of protein disulfide bonding. In contrast, disulfide bonding occurs at lower levels, and more selectively, within the Euryarchaea. More specifically, the effect is practically absent from the halophiles, but present in some thermophilic Euryarchaea. The presence or absence of significant disulfide bonding within the thermophilic Euryarchaea appears to correlate most strongly with the absence of methanogenesis. That trend could reflect the incompatibility of disulfide bonding with either the redox potential required for reduction of oxidized carbon to methane, or with the oxidation-prone enzymes and cofactors that carry out methanogenesis [45]. Beyond the well-studied crenarchaeal and euryarchaea, genomes have been sequenced for microbes representing three ancient or highly divergent archaeal branches (Figure 3). Two of these, *Nanoarchaeum equitans* and *Candidatus Korarchaeum cryptofilum*, are thermophilic; of the two, *N. equitans* shows the greater degree of disulfide bonding.

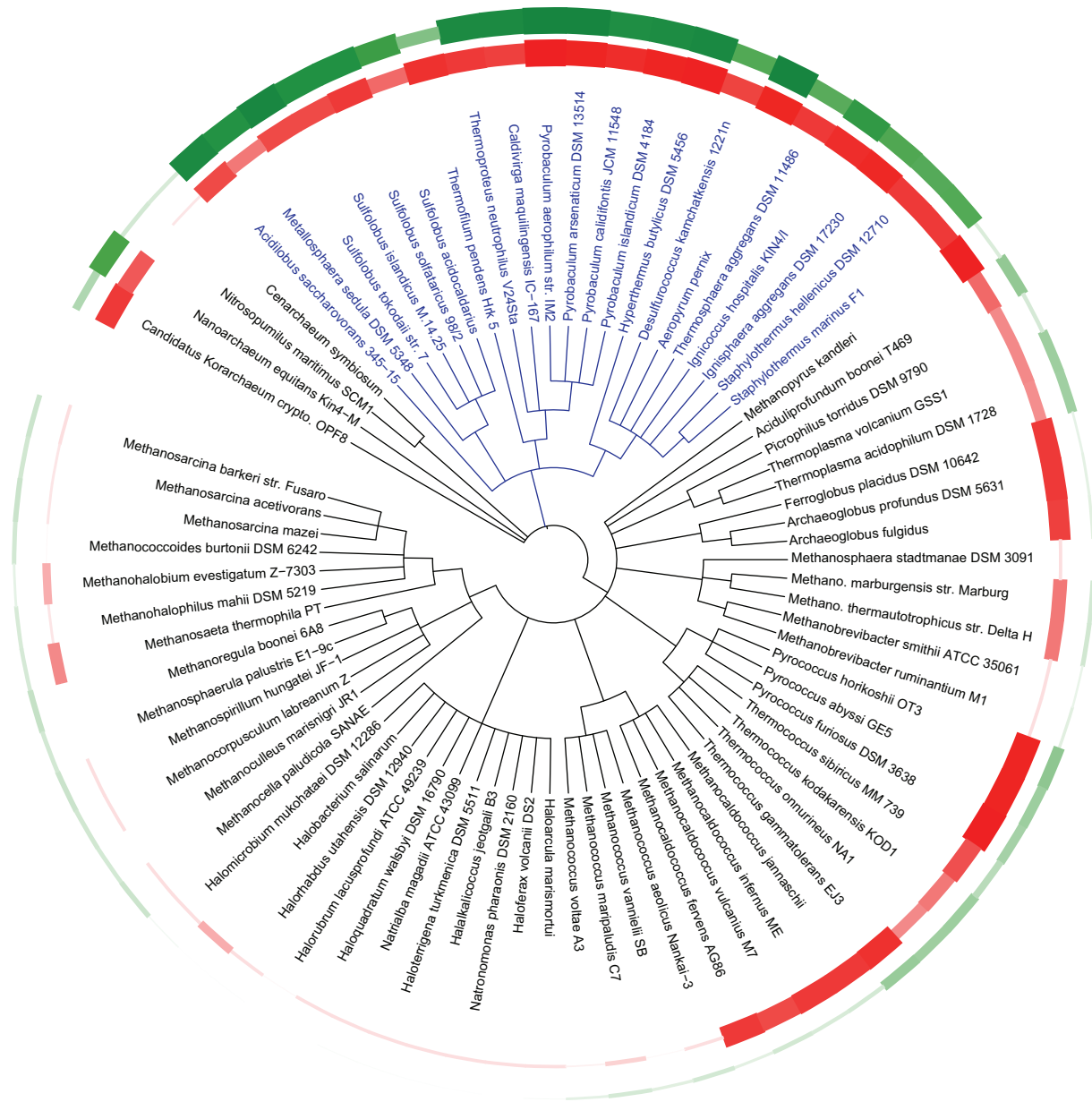


FIGURE 3: A phylogenetic tree showing the predicted abundance of disulfide bonding across the archaea. Each leaf of the tree is associated to specific values of the calculated disulfide abundance parameter ( $f$ ) and optimal growth temperature, which are depicted in respective circular plots, green (outer ring) and red (inner ring). These values are represented by the properties of the curves, which vary in thickness (and darkness) in proportion to the corresponding quantity. A general correlation is seen between disulfide richness and growth temperature; specifically, disulfide bonding is dominant in the Crenarchaea (blue) and also notable in the subset of thermophilic Euryarchaea that are nonmethanogenic.

We also investigated several viruses that infect thermophilic archaea. A nonexhaustive list of analyzed viruses is reported. The total number of cysteines that could be mapped onto structures ( $m$ ) is provided along with the estimated disulfide abundance parameter ( $f$ ): *Sulfolobus islandicus* rod-shaped virus 1 ( $m = 16$ ,  $f = 1$ ), *Sulfolobus islandicus* filamentous virus ( $m = 12$ ,  $f = 0.75$ ), *Sulfolobus* virus 1 ( $m = 18$ ,  $f = 1$ ), *Sulfolobus* virus Kamchatka ( $m = 14$ ,  $f = 0.71$ ), *Sulfolobus* turreted icosahedral virus ( $m = 14$ ,  $f = 1$ ), *Sulfolobus* virus Ragged Hills ( $m = 4$ ,

$f = 1$ ), *Pyrobaculum* spherical virus ( $m = 35$ ,  $f = 0.57$ ), *Sulfolobus* virus STSV1 ( $m = 12$ ,  $f = 0.33$ ), and *Sulfolobus* spindle-shaped virus 4 ( $m = 4$ ,  $f = 1$ ). The presence of disulfides in the *Sulfolobus* viruses has already been emphasized in reported crystal structures and by way of cysteine counting [46–48]. It is interesting that of the viruses that infect disulfide-rich thermophilic archaea, most appear to encode proteins that contain disulfide bonds. The results on individual viral genomes should be interpreted with caution, however, as most of the reported viruses have been



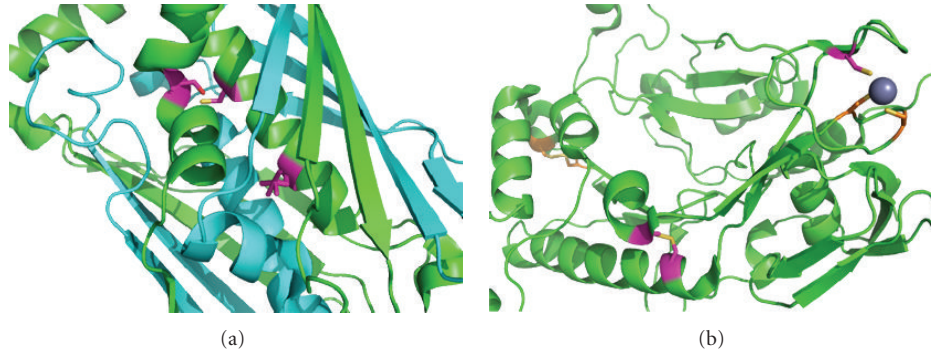


FIGURE 4: Examples of known thermophilic archaeal protein structures containing disulfide bonds, but having an odd number of total cysteines. Two proteins from *Sulfolobus islandicus* M.14.25 are shown. Mapped cysteines positions are represented in pink while cysteine residues involved in a putative metal-binding site are in orange, Zn in grey. (a) Mapping of an OsmC family protein (UniProtKB:C3MY18) onto the PDB structure 2OPL. One disulfide bond is predicted; the third cysteine is located at the surface and could participate in an intermolecular disulfide bond. (b) Mapping of a probable DNA primase small subunit (UniProtKB:C3MYF5) onto the PDB structure 1ZT2. Two mapped cysteines are poised to form a disulfide bond while the third is close to a Zn-binding site, suggesting an interaction with the metal ion. Many cases of proteins with an odd number of cysteines probably reflect the presence of intermolecular disulfide bonds or participation in metal-binding sites.

characterized from the *Sulfolobus* species, and the total number of proteins encoded by a single virus is small, making statistically significant conclusions difficult.

For reasons already noted, the computational methods discussed give only rough measures of the true disulfide abundance in the proteins from a given genome. The errors and deviations arising during sequence-structure comparison and the challenge in interpreting intermolecular disulfide bonds both tend to cause an underestimation of the actual disulfide abundance; even if all the cysteines in the proteins from some genome were disulfide bonded, only a fraction would be successfully identified by the sequence-structure mapping approach. Fortunately, a few thermophilic and hyperthermophilic microbes have been popular targets for structural studies. Cases where many protein structures have been reported from a single organism present an opportunity to evaluate the disulfide abundance directly, assuming that the structures reported are reasonably representative of the genome as a whole. Examining the deposited protein structures from several selected organisms, we evaluated directly what fraction of the cysteines were involved in disulfide bonds. We denote this fraction as  $f'$ , as it parallels the meaning of  $f$  from the sequence-structure mapping method, but is based on counting only actual reported structures. By this method, *Pyrobaculum aerophilum* stands on top with  $f' = 0.7$ ; this value is about twice the value of  $f$  predicted from sequence-structure mapping. *Sulfolobus solfataricus*, *Pyrococcus abyssi*, *Pyrococcus furiosus*, and the bacteria *Thermotoga Maritima* give values for  $f'$  of 0.37, 0.23, 0.23, and 0.091, respectively. Aside from *Pyrobaculum*, where the relatively low number of unique reported structures (25) could explain the anomalously high value of  $f'$ , the values of  $f'$  are well correlated with the estimates,  $f$ , obtained by the genomic sequence-structure mapping approach (Figure 5). This high correlation provides additional support for using the value of  $f$  as an indicator of disulfide richness in sequenced genomes.

### 3. Conclusions

Computational analysis of the growing database of sequenced genomes shows convincingly that protein disulfides are ubiquitous among the thermophilic archaea. This is especially true of the Crenarchaea, which are practically all hyperthermophilic, and which appear to be universally rich in disulfides. Disulfide bonding is less abundant and more variable across the Euryarchaea, appearing mainly in the nonmethanogenic thermophiles. Based on the single genome sequence presently available, the ancient Nanoarchaeal branch also appears to be hyperthermophilic and disulfide rich.

In multiple studies, experiments on disulfide bonded proteins from thermophilic microbes have confirmed, as expected, that those disulfide bonds play a major role in stabilizing the folded structure against unfolding and aggregation [22–26]. Proteins and enzymes derived from thermophiles have already been recognized for their utility in industrial applications [8, 49]. Specific proteins or specific homologues that have one or multiple predicted disulfide bonds could make especially attractive choices for such applications, especially since ambient conditions are generally oxidizing. The growing list of disulfide-rich organisms will continue to increase the availability of homologous enzymes for this purpose.

Finally, accelerating the acquisition of genomic data within the archaea, particularly along the Crenarchaeal branch, could have a major impact on the long-standing problem of predicting three-dimensional protein structures from sequence data. Except for a rather narrow target group—for example, very small, mainly alpha helical proteins—accurate *de novo* protein structure prediction remains unreliable [50]. However, additional information in the form of even a few spatial constraints could push structure prediction algorithms over the current barrier. Crenarchaeal proteins represent cases where such spatial

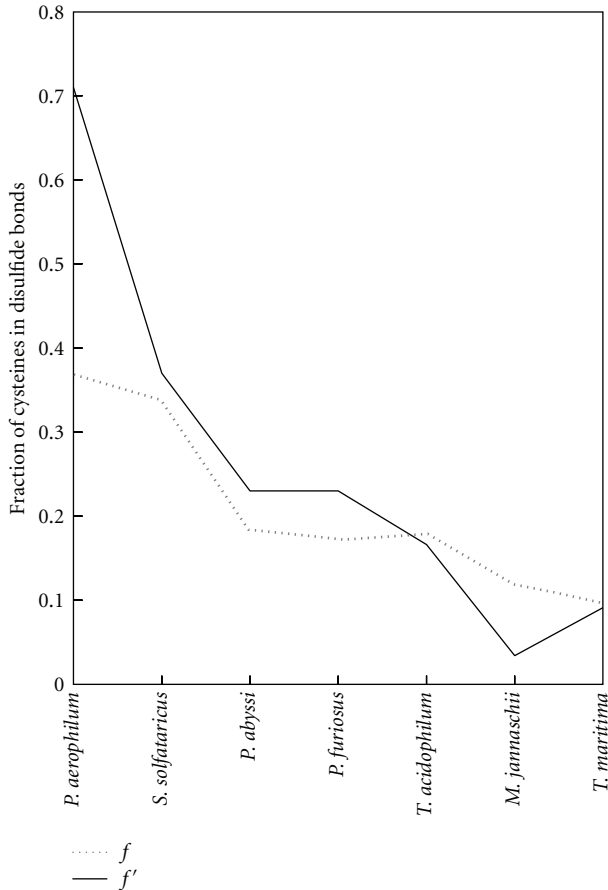


FIGURE 5: Correlation of the predicted disulfide abundance parameter,  $f$ , with the corresponding value  $f'$  determined from the protein structure databank for thermophiles with significant coverage. Selected thermophiles were analyzed if they had sufficient representation in the PDB to obtain a reasonable estimate of the disulfide abundance. Metalloproteins were excluded. See Section 4.

constraints might reasonably be inferred from sequence data. For a protein with homologues among the Crenarchaea, a correctly predicted structure will tend to place cysteine residues in proximity for disulfide bonding, whereas no such tendency would be expected for an incorrect structure prediction. Research along this line is presently underway in our laboratory.

## 4. Materials and Methods

**4.1. Proteome Datasets.** The complete proteomes used in this analysis were extracted from UniProtKB release 2011.04. A query with the keyword “complete proteome” returned 90 archaeal proteomes gathering 213232 protein sequences and stocked in FASTA format to constitute the dataset. For metagenomic data from hot springs in Yellowstone National Park, sequences of cellular DNA were included from the most recent samples at two locations, Crater Hills (1152 protein sequences from sample ID CH0909) and Nymph Lake site 10 (1762 protein sequences from sample ID NL10.0908), along

with viral DNA sequences (total of 161 protein sequences from four sites: CH, NL10, NL17, and NL18). Contigs were assembled with Newbler gsAssembler v 2.3 (98% nucleotide identity and 50 bp overlap) and translated on-the-fly in the structure-mapping procedure by using Blastx. Sources for the DNA sequences are described in [38].

**4.2. Filtering Extracytoplasmic and Metalloproteins.** As a first step, protein sequences that did not contain at least two cysteines were excluded from the analysis. Moreover, to substantially avoid a biased counting due to cysteines that could be involved in metal binding sites, cysteines falling within 5 residues of each other were not considered, based on the observation that metal binding sites are often (though not always) formed by residues closely spaced in sequence. We also filtered to remove secreted or periplasmic proteins as these are outside the scope of our study; it is already recognized that proteins in these compartments are often rich in disulfides. The PREDISI program was used to perform this filtering step, employing default parameters [51].

**4.3. Sequence-Structure Mapping.** Each protein from the dataset was processed to match it to a homologous structure in the PDB [52]. Pairwise alignment between a UniProtKB sequence and a PDB sequence was performed by BLAST. If a hit with an  $E$ -value  $< 0.0001$  was found, the sequence was mapped onto the structure. Numerous discrepancies were noted between the PDB sequences appearing in the BLAST database and the sequences reported in the corresponding PDB entry. Therefore, in order to obtain a correct mapping, the positions of the mapped residues were recalculated via a full dynamic programming alignment between the two sequences using the Needleman and Wunsch algorithm [53].

**4.4. Calculation of the Disulfide-Bond Richness Parameter,  $f$ .** A pair of cysteine residues is judged to represent a probable disulfide bond if their C-alpha atoms are spatially closer than 8 Å when they are mapped onto a homologous structure. In this case, each of the participating cysteines will be considered as a hit; if a given cysteine is within the cutoff distance of more than one other cysteine, it is only counted once. For each protein, the fraction of predicted cysteines forming disulfide bonds can be calculated. Subsequently, at the proteome level,  $f$  will stand for the ratio between the total number of hits and the total number of mapped cysteines.

**4.5. Strain Filtering.** In our reporting, the redundancy due to multiple sequencing of similar strains of the same species has been removed. For a given species, multiple strain variations were removed if they showed similar results (e.g., a number of mapped proteins within  $\pm 10\%$  of the parent strain).

**4.6. Control Datasets.** Protein structures of selected species were extracted from the Protein Data Bank (PDB). Cysteine residues involved in disulfide bonds were deduced from their presence in the SSBOND record in the PDB entry while those found in the LINK record were not taken into account,

as these typically represent cysteines involved in binding metals or other ligands. Proteins with Zn or Fe bound were excluded from the analysis. Based on the extracted PDB files for a specific organisms, the fraction of cysteines involved in disulfide bonds,  $f'$ , was calculated based on the same principle as that used to calculate  $f$  in the sequence-structure mapping approach.

**4.7. Phylogenetic Analysis.** The phylogenetic tree (Figure 3) was built with the web-based tool ITOL [54], after modifying the raw postscript output to illustrate the temperatures and computed disulfide parameters as lines of variable thickness. Optimal growth temperatures used for the circular plot were retrieved from the German Resource Centre for Biological Material website (<http://www.dsmz.de/>) and from the literature.

## Conflict of Interests

The authors declare no competing financial interests.

## Acknowledgments

The authors thank Mark Young and Ben Bolduc for their help in performing a preliminary analysis of the Yellowstone metagenomic data. This paper was supported by the BER program of the DOE Office of Science.

## References

- [1] K. O. Stetter, "Hyperthermophiles in the history of life," *Philosophical Transactions of the Royal Society B*, vol. 361, no. 1474, pp. 1837–1842, 2006.
- [2] T. E. Creighton, *Proteins: Structures and Molecular Properties*, W. H. Freeman, 1992.
- [3] R. Jaenicke and G. Böhm, "The stability of proteins in extreme environments," *Current Opinion in Structural Biology*, vol. 8, no. 6, pp. 738–748, 1998.
- [4] D. C. Rees and M. W. W. Adams, "Hyperthermophiles: taking the heat and loving it," *Structure*, vol. 3, no. 3, pp. 251–254, 1995.
- [5] S. Chakravarty and R. Varadarajan, "Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study," *Biochemistry*, vol. 41, no. 25, pp. 8152–8161, 2002.
- [6] S. Kumar and R. Nussinov, "How do thermophilic proteins deal with heat?" *Cellular and Molecular Life Sciences*, vol. 58, no. 9, pp. 1216–1233, 2001.
- [7] G. A. Petsko, "Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat?" *Methods in Enzymology*, vol. 334, pp. 469–478, 2001.
- [8] C. Vieille and G. J. Zeikus, "Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability," *Microbiology and Molecular Biology Reviews*, vol. 65, no. 1, pp. 1–43, 2001.
- [9] M. K. Chan, S. Mukund, A. Kletzin, M. W. W. Adams, and D. C. Rees, "Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase," *Science*, vol. 267, no. 5203, pp. 1463–1469, 1995.
- [10] R. Jaenicke, "Protein stability and molecular adaptation to extreme conditions," *European Journal of Biochemistry*, vol. 202, no. 3, pp. 715–728, 1991.
- [11] R. Lieph, F. A. Veloso, and D. S. Holmes, "Thermophiles like hot T," *Trends in Microbiology*, vol. 14, no. 10, pp. 423–426, 2006.
- [12] K. S. P. Yip, K. L. Britton, T. J. Stillman et al., "Insights into the molecular basis of thermal stability from the analysis of ion-pair networks in the glutamate dehydrogenase family," *European Journal of Biochemistry*, vol. 255, no. 2, pp. 336–346, 1998.
- [13] H. Hashimoto, T. Inoue, M. Nishioka et al., "Hyperthermostable protein structure maintained by infra and inter-helix ion-pairs in archaeal O6-methylguanine-DNA methyltransferase," *Journal of Molecular Biology*, vol. 292, no. 3, pp. 707–716, 1999.
- [14] A. Karshikoff and R. Ladenstein, "Ion pairs and the thermotolerance of proteins from hyperthermophiles: a 'traffic rule' for hot roads," *Trends in Biochemical Sciences*, vol. 26, no. 9, pp. 550–556, 2001.
- [15] M. J. Thompson and D. Eisenberg, "Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability," *Journal of Molecular Biology*, vol. 290, no. 2, pp. 595–604, 1999.
- [16] P. Mallick, D. R. Boutz, D. Eisenberg, and T. O. Yeates, "Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 15, pp. 9679–9684, 2002.
- [17] E. A. Toth, C. Worby, J. E. Dixon, E. R. Goedken, S. Marqusee, and T. O. Yeates, "The crystal structure of adenylosuccinate lyase from *Pyrobaculum aerophilum* reveals an intracellular protein with three disulfide bonds," *Journal of Molecular Biology*, vol. 301, no. 2, pp. 433–450, 2000.
- [18] M. Beeby, B. D. O'Connor, C. Ryttersgaard, D. R. Boutz, L. J. Perry, and T. O. Yeates, "The genomics of disulfide bonding and protein stabilization in thermophiles," *PLoS Biology*, vol. 3, no. 9, article e309, 2005.
- [19] R. Ladenstein and B. Ren, "Reconsideration of an early dogma, saying "there is no evidence for disulfide bonds in proteins from archaea,"" *Extremophiles*, vol. 12, no. 1, pp. 29–38, 2008.
- [20] D. R. Boutz, D. Cascio, J. Whitelegge, L. J. Perry, and T. O. Yeates, "Discovery of a thermophilic protein complex stabilized by topologically interlinked chains," *Journal of Molecular Biology*, vol. 368, no. 5, pp. 1332–1344, 2007.
- [21] J. A. Littlechild, J. E. Guy, and M. N. Isupov, "Hyperthermophilic dehydrogenase enzymes," *Biochemical Society Transactions*, vol. 32, no. 2, pp. 255–258, 2004.
- [22] M. Karlström, R. Stokke, I. Helene Steen, N. K. Birkeland, and R. Ladenstein, "Isocitrate dehydrogenase from the hyperthermophile *Aeropyrum pernix*: X-ray structure analysis of a ternary enzyme-substrate complex and thermal stability," *Journal of Molecular Biology*, vol. 345, no. 3, pp. 559–577, 2005.
- [23] A. Guelorget, M. Roovers, V. Guérineau, C. Barbey, X. Li, and B. Golinelli-Pimpaneau, "Insights into the hyperthermostability and unusual region-specificity of archaeal *Pyrococcus abyssi* tRNA m1A57/58 methyltransferase," *Nucleic Acids Research*, vol. 38, no. 18, Article ID gkq381, pp. 6206–6218, 2010.
- [24] T. Kaper, B. Talik, T. J. Ettema, H. Bos, M. J. E. C. Van Der Maarel, and L. Dijkhuizen, "Amylomaltase of *Pyrobaculum aerophilum* IM2 produces thermoreversible starch gels,"



- Applied and Environmental Microbiology*, vol. 71, no. 9, pp. 5098–5106, 2005.
- [25] G. Cacciapuoti, S. Forte, M. A. Moretti, A. Brio, V. Zappia, and M. Porcelli, “A novel hyperthermostable 5′-deoxy-5′-methylthioadenosine phosphorylase from the archaeon *Sulfolobus solfataricus*,” *FEBS Journal*, vol. 272, no. 8, pp. 1886–1899, 2005.
- [26] G. Cacciapuoti, M. A. Moretti, S. Forte et al., “Methylthioadenosine phosphorylase from the archaeon *Pyrococcus furiosus*: mechanism of the reaction and assignment of disulfide bonds,” *European Journal of Biochemistry*, vol. 271, no. 23–24, pp. 4834–4844, 2004.
- [27] H. F. Gilbert, “Molecular and cellular aspects of thiol-disulfide exchange,” *Advances in enzymology and related areas of molecular biology*, vol. 63, pp. 69–172, 1990.
- [28] U. Jakob, W. Muse, M. Eser, and J. C. A. Bardwell, “Chaperone activity with a redox switch,” *Cell*, vol. 96, no. 3, pp. 341–352, 1999.
- [29] H. J. Choi, S. J. Kim, P. Mukhopadhyay et al., “Structural basis of the redox switch in the OxyR transcription factor,” *Cell*, vol. 105, no. 1, pp. 103–113, 2001.
- [30] M. A. Wouters, S. W. Fan, and N. L. Haworth, “Disulfides as redox switches: from molecular mechanisms to functional significance,” *Antioxidants and Redox Signaling*, vol. 12, no. 1, pp. 53–91, 2010.
- [31] E. Pedone, B. Ren, R. Ladenstein, M. Rossi, and S. Bartolucci, “Functional properties of the protein disulfide oxidoreductase from the archaeon *Pyrococcus furiosus*: a member of a novel protein family related to protein disulfide-isomerase,” *European Journal of Biochemistry*, vol. 271, no. 16, pp. 3437–3448, 2004.
- [32] E. Pedone, D. Limauro, and S. Bartolucci, “The machinery for oxidative protein folding in thermophiles,” *Antioxidants and Redox Signaling*, vol. 10, no. 1, pp. 157–169, 2008.
- [33] A. Becerra, L. Delaye, A. Lazcano, and L. E. Orgel, “Protein disulfide oxidoreductases and the evolution of thermophily: was the last common ancestor a heat-loving microbe?” *Journal of Molecular Evolution*, vol. 65, no. 3, pp. 296–303, 2007.
- [34] R. Ladenstein and B. Ren, “Protein disulfides and protein disulfide oxidoreductases in hyperthermophiles,” *FEBS Journal*, vol. 273, no. 18, pp. 4170–4185, 2006.
- [35] E. Pedone, K. D’Ambrosio, G. De Simone, M. Rossi, C. Pedone, and S. Bartolucci, “Insights on a new PDI-like family: structural and functional analysis of a protein disulfide oxidoreductase from the bacterium *Aquifex aeolicus*,” *Journal of Molecular Biology*, vol. 356, no. 1, pp. 155–164, 2006.
- [36] S. Bartolucci, D. De Pascale, and M. Rossi, “Protein disulfide oxidoreductase from *Pyrococcus furiosus*: biochemical properties,” *Methods in Enzymology*, vol. 334, pp. 62–73, 2001.
- [37] K. D’Ambrosio, E. Pedone, E. Langella et al., “A novel member of the protein disulfide oxidoreductase family from *Aeropyrum pernix* K1: structure, function and electrostatics,” *Journal of Molecular Biology*, vol. 362, no. 4, pp. 743–752, 2006.
- [38] W. P. Inskeep, D. B. Rusch, Z. J. Jay et al., “Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function,” *PLoS one*, vol. 5, no. 3, article e9773, 2010.
- [39] R. J. Dutton, D. Boyd, M. Berkmen, and J. Beckwith, “Bacterial species exhibit diversity in their mechanisms and capacity for protein disulfide bond formation,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 33, pp. 11933–11938, 2008.
- [40] R. Daniels, P. Mellroth, A. Bernsel et al., “Disulfide bond formation and cysteine exclusion in gram-positive bacteria,” *Journal of Biological Chemistry*, vol. 285, no. 5, pp. 3300–3309, 2010.
- [41] T. Iwasaki, “Iron-sulfur world in aerobic and hyperthermoacidophilic archaea *Sulfolobus*,” *Archaea*, vol. 2010, Article ID 842639, 14 pages, 2010.
- [42] L. Aravind, R. L. Tatusov, Y. I. Wolf, D. R. Walker, and E. V. Koonin, “Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles,” *Trends in Genetics*, vol. 14, no. 11, pp. 442–444, 1998.
- [43] F. G. Pearce, M. A. Perugini, H. J. McKerchar, and J. A. Gerrard, “Dihydrodipicolinate synthase from *Thermotoga maritima*,” *Biochemical Journal*, vol. 400, no. 2, pp. 359–366, 2006.
- [44] T. Toyooka, T. Awai, T. Kanai, T. Imanaka, and H. Hori, “Stabilization of tRNA (m1G37) methyltransferase [TrmD] from *Aquifex aeolicus* by an intersubunit disulfide bond formation,” *Genes to Cells*, vol. 13, no. 8, pp. 807–816, 2008.
- [45] K. F. Jarrell, “Extreme oxygen sensitivity in methanogenic archaeobacteria,” *Bioscience*, vol. 35, pp. 298–302, 1985.
- [46] E. T. Larson, B. Eilers, S. Menon et al., “A winged-helix protein from *Sulfolobus turreted* icosahedral virus points toward stabilizing disulfide bonds in the intracellular proteins of a hyperthermophilic virus,” *Virology*, vol. 368, no. 2, pp. 249–261, 2007.
- [47] E. T. Larson, B. J. Eilers, D. Reiter, A. C. Ortmann, M. J. Young, and C. M. Lawrence, “A new DNA binding protein highly conserved in diverse crenarchaeal viruses,” *Virology*, vol. 363, no. 2, pp. 387–396, 2007.
- [48] S. K. Menon, W. S. Maaty, G. J. Corn et al., “Cysteine usage in *Sulfolobus* spindle-shaped virus 1 and extension to hyperthermophilic viruses in general,” *Virology*, vol. 376, no. 2, pp. 270–278, 2008.
- [49] J. A. Littlechild, “Thermophilic archaeal enzymes and applications in biocatalysis,” *Biochemical Society Transactions*, vol. 39, no. 1, pp. 155–158, 2011.
- [50] J. Moul, K. Fidelis, A. Kryshtafovych, B. Rost, and A. Tramontano, “Critical assessment of methods of protein structure prediction—Round VIII,” *Proteins: Structure, Function and Bioinformatics*, vol. 77, no. 9, pp. 1–4, 2009.
- [51] K. Hiller, A. Grote, M. Scheer, R. Münch, and D. Jahn, “PrediSi: prediction of signal peptides and their cleavage positions,” *Nucleic Acids Research*, vol. 32, pp. W375–W379, 2004.
- [52] P. W. Rose, B. Beran, C. Bi et al., “The RCSB Protein Data Bank: redesigned web site and web services,” *Nucleic Acids Research*, vol. 39, supplement 1, pp. D392–D401, 2011.
- [53] S. B. Needleman and C. D. Wunsch, “A general method applicable to the search for similarities in the amino acid sequence of two proteins,” *Journal of Molecular Biology*, vol. 48, no. 3, pp. 443–453, 1970.
- [54] I. Letunic and P. Bork, “Interactive Tree of Life v2: online annotation and display of phylogenetic trees made easy,” *Nucleic Acids Research*, vol. 39, supplement 2, pp. W475–W478, 2011.